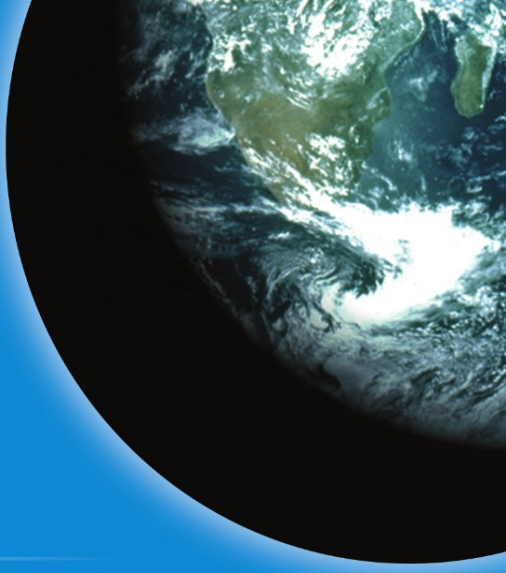
Space Technology Library



Marco D'Errico *Editor*

Distributed Space Missions for Earth System Monitoring

Space Technology Library

IAA · INTERNATIONAL ACADEMY OF ASTRONAUTICS

Springer

Distributed Space Missions for Earth System Monitoring

SPACE TECHNOLOGY LIBRARY

Published jointly by Microcosm Press and Springer

For further volumes:
http://www.springer.com/series/6575

Marco D'Errico
Editor

Distributed Space Missions for Earth System Monitoring

Published jointly by
Microcosm Press
El Segundo, California

Editor
Marco D'Errico
Dipartimento di Ingegneria Aerospaziale
Seconda Universita di Napoli
Aversa, Italy

ISBN 978-1-4614-4540-1 ISBN 978-1-4614-4541-8 (eBook)
DOI 10.1007/978-1-4614-4541-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012945401

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Milly and Pino, my roots from the past
To Loredana, my ever present companion
To Matteo and Sveva, my dear eyes into the future

Marco D'Errico

Preface

Distributed space mission concepts had first been introduced to astronomical and planetary applications, but later such approaches had also been proposed for Earth observation. This trend has led to a flurry of related researches in many fields, such as dynamics, telecommunications, remote sensing, architecture, and many others. While first pioneering studies and proposals date back to the 1980s, and further systematic works, mainly on dynamics, were published in the 1990s, it is in this last decade that a huge worldwide effort to further develop such a concept has been demonstrated. Some missions relaying on two cooperative satellites have already been successfully accomplished: GRACE (aimed at gravimetry), PRISMA (a technology demonstrator), and Tandem-X (an Earth observation mission embarking a Synthetic Aperture Radar interferometer) represent great achievements in space system engineering. They will likely be the precursors of enhanced future missions.

Thanks to these researches, it is widely recognized that upcoming space systems will make use of collaborating platforms to replace current monolithic systems and to implement missions otherwise impossible (e.g., those requiring very large sensor apertures). Such evolution calls for a revolutionary change of mentality in design, realization, and operation at various levels. At payload level, one has to assess the capability to integrate the mission payload using different elementary payloads onboard various satellites. In addition, new concepts (e.g., modularity, autonomy, standardization, plug & play components) must be explored to attain an efficient bus implementation and new subsystems (e.g., relative trajectory design, relative navigation and control, satellite interlink) are to be implemented to enable the required new functions.

Although the approach to distributed space missions is thus inherently multidisciplinary, researches have often been focused on specific topics. For instance, if on the one hand relative trajectory dynamics has been thoroughly studied and many steps forward have been made in the last decade, on the other hand such issues have often been considered independently from the payload. In addition, interesting results have been published on the distributed sensor concept, but have rarely considered the issues raised by orbital dynamics. Finally, the technological challenges (measurements, coordination, communications) behind a multi-satellite system have only been

partially tackled and exploited. This book, *Distributed Systems for Earth System Monitoring*, is an effort to present in a unique environment its state of the art and demonstrate through different experiences the variety of themes which characterize distributed space systems with the aim to better highlight their strong connections.

The book is organized in four parts and 23 chapters. Part I is devoted to the analysis of distributed Synthetic Aperture Radars: bistatic and multistatic radars are presented in two chapters. Following this, relative dynamics and guidance, navigation, and control are the subjects in Part II, which consists of five chapters, wherein the different topics are developed with reference to relative trajectory: (a) design, (b) establishment, maintenance and control, and (c) measurements. This latter topic is tackled by considering three different approaches for navigation: Global Positioning System, radio frequency, and vision-based metrologies. Technological challenges connected to distributed space missions are the topic in Part III, where the impacts of the distributed approaches are discussed with reference to autonomy, navigation, and communications (space to space and space to ground). Finally, Part IV reports on a number of missions and studies which have been performed in recent years to analyze how and to what extent the distributed concept has been studied and applied. This part covers radar missions (e.g., Tandem-X), gravimetry missions (e.g., Grace), and technology demonstration missions (e.g., PRISMA). Chapter 23, the last concluding chapter, reflects the authors' opinions on the future trends, the potentials and the risks of the distributed space mission approach.

The book, written within the activity of a Study Group of the International Academy of Astronautics,[1] is coauthored by 52 authors from seven countries and from many institutions (universities, research centers, industries, and agencies). A large group of authors with a variety of professional figures have been required to cover the wide range of topics discussed and to reflect the different opinions and approaches available. I hope that we have been successful in our efforts to offer a common but articulated synthesis.

The coordination of such a group was not an easy task. I want to thank all the authors for their contributions, for the patience in responding to my requests, and for their efforts to keep up with schedule. I am very grateful to and thank Springer staff and, in particular, Maury Solomon and Megan Ernst, who, frequently solicited by my questions, doubts, and requests, have always found a way forward.

Finally, I would like to say thank you to Antonio Moccia, who introduced me to the space field since the times of my M.Sc. dissertation. His silent and coherent example is a continuous clear guide to all those young people walking on the bridge between educational and professional lives. Last, but not least, I am grateful to Rainer Sandau, who introduced me to the International Academy of Astronautics and its international context. Without his support, advice, and encouragement, I would never have dared to propose and coordinate this book.

Aversa, Italy Marco D'Errico

[1] This is currently in the process of approval by the Board of Trustees of the International Academy of Astronautics.

Authors

Domenico Accardo He is Assistant Professor of Avionics, AIAA senior member, and International Member in the AIAA Sensor Systems Technical Committee. His research is devoted to the development of advanced star trackers, unmanned aerial systems, integrated navigation systems, airborne tracking systems for collision avoidance, and advanced tools for air traffic management.

Department of Aerospace Engineering, University of Naples "Federico II", Naples, Italy

Jean Paul Aguttes He has more than 30 years experience in space mission design and particularly in the Radio Frequency domain. His main researches and publications are related to the design of spaceborne SAR satellite and the way to reduce or circumvent the constraints of the antenna. He organized in 2002 the first symposium "Formation Flying Missions and Technologies" now recurrently held by the space agencies. Until 2006, Aguttes was responsible for CNES Advanced Projects; he is currently head of the Radio-Frequency Sub Directorate in CNES Toulouse Space Center.

CNES, Toulouse Space Center, Toulouse, France

Giovanni Alberti He received a degree "cum laude" in electronic engineering (telecommunication specialization) from the University "Federico II" of Naples, in 1989. His main research activities deals with the analysis and design of microwave sensors (Synthetic Aperture Radar, Radar Altimeter, Ground Penetrating Radar) and image processing in the field of remote sensing (SAR interferometry, in particular). He is responsible of research projects sponsored by ASI, ESA, and NASA. He has been involved in both Mars (MARSIS and SHARAD) and Saturn (CASSINI) missions. He is a member of Cassini Radar and SHARAD Science Teams. He is Italian PI deputy for SHARAD radar. Currently he is working in the framework of EXOMARS project in the field of radar altimeter test facility. He lectured on "Aerospace Remote Sensing Systems" at Seconda Università di Napoli.

Consortium for the Research on Advanced Remote Sensing Systems (CORISTA), Naples, Italy

Kyle T. Alfriend Dr. Kyle T. Alfriend is the TEES Research Professor of Aerospace Engineering at Texas A&M University. He is a member of the National Academy of Engineering, the European Academy of Sciences, and is a Fellow of the AIAA and AAS. He is the recipient of the AIAA Mechanics and Control of Flight Award and the AAS Dirk Brouwer Award.

Department of Aerospace Engineering, Texas A&M university, College Station, TX, USA

Markus Bachmann He received the Dipl.-Ing. degree in electrical engineering from the Technical University of Karlsruhe, Germany, in 2005. In 2005, he joined the Microwaves and Radar Institute at the German Aerospace Center, where he is working as system engineer in SAR calibration. In the TerraSAR-X/TanDEM-X project, Bachmann was responsible for the antenna model calibration. In TanDEM-X he also was in duty for the planning of the commissioning phase. Since 2012, he leads the mission engineering team.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Per Bodin He has worked at OHB Sweden (formerly Space Systems Division of SSC) for more than 15 years. He has worked as AOCS lead engineer and subsystem manager on SMART-1 and PRISMA and is currently head of the AOCS & Software Department at OHB Sweden.

OHB Sweden AB, Solna, Sweden

Benjamin Bräutigam He received the Dipl.-Ing. degree in electrical engineering from the Technical University Karlsruhe, Germany, in 2003. In the same year, he was a Visiting Scientist at NOAA Environmental Technology Laboratory, Boulder, Colorado, USA. In 2004, he joined the Microwave and Radar Institute at the German Aerospace Center (DLR). He was responsible for internal instrument calibration of spaceborne SAR missions like TerraSAR-X. Since 2008, he has been head of the system performance group working on TerraSAR-X and TanDEM-X.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Helko Breit Helko Breit received his diploma degree in electrical and telecommunication science from the Technische Universität München, Munich, Germany, in 1990. He joined the German Aerospace Center (DLR) in 1990 and is currently supervising the SAR processor development team at DLR's Remote Sensing Technology Institute. He was responsible for the development of the TerraSAR-X Multi-Mode SAR Processor (TMSP) and bistatic processing kernel of the Integrated TanDEM-X Processor (ITP).

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Stefano Cesare Stefano Cesare has more than 25 years experience in space missions design. In particular, he was the performance manager of the GOCE mission.

He managed several projects, among which are preparatory studies and technology developments for the Next Generation Gravity Mission. Currently he is the head of the Italian Scientific Programs Unit.

Thales Alenia Space Italia, Torino, Italy

Simone D'Amico He leads the Autonomous Formation Flying team at DLR/GSOC and is chairman of the 5th SFFMT Conf. His research focuses on space flight dynamics, autonomous navigation systems, and orbit control with key contributions to TanDEM-X and PRISMA.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Michel Delpech He is Senior Expert of CNES and leads the Formation Flying–related activities. After a long involvement in robotics for planetary exploration, his research activities focuses on relative metrology systems, autonomous navigation, guidance, and control, and system design for future formation flying concepts. He was technical manager of the French contribution to PRISMA.

CNES, Toulouse Space Center, Toulouse, France

Marco D'Errico He is associate professor of Space Systems, IAA corresponding member, and secretary of the IAA Permanent Committee on Small Satellites. His research is devoted to mission analysis of Earth observation systems, formation flying, and synthetic aperture radars.

Department of Aerospace and Mechanical Engineering, Second University of Naples, Aversa (CE), Italy

Giancarmine Fasano He is assistant professor of Aerospace Systems and member of the IAA Program Committee for the 16th Symposium on Small Satellites. His main research interests are formation flying, mission analysis for multi-satellite SAR architectures, and unmanned aircraft systems.

Department of Aerospace Engineering, University of Naples "Federico II", Naples, Italy

Hauke Fiedler He has received his Ph.D. in Astronomy and Astrophysics from the Ludwig-Maximilians-University, Munich, in 2000 and his Masters of Space S. E. from the TU Delft, the Netherlands, in 2009. From 2001 to 2009, Dr. Fiedler was a scientific assistant with the DLR Microwaves and Radar Institute working on TanDEM-X and multistatic SAR systems. He spent 2 years from 2009 at the DLR Headquarters working on Institute Strategic Development for the Executive Board Division. He is now with the German Space Operations Center responsible for Space Situational Awareness.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Thomas Fritz He received his diploma degree in physics from the University of Münster, Münster, Germany in 1996, and the Dr. rer. nat. degree from the University of Bonn, Bonn, Germany in 2000. He then had a postdoctoral position in the Radioastronomical Institute of Bonn University. In 2003, he joined the Remote

Sensing Technology Institute of the German Aerospace Center (DLR) and is currently supervising a SAR application development team. He is responsible for development and system engineering of the Integrated TanDEM-X Processor (ITP).

German Aerospace Center (DLR), Remote Sensing Technology Institute, Wessling, Germany

Eberhard Gill He is Professor and chairholder of Space Systems Engineering, corresponding member of IAA and Senior Scientist of DLR. His research interests are spacecraft autonomy, formation flying, GNSS technology, and combined communication and navigation applications. He has been co-investigator and principal investigator on various space science and technology missions.

Space Systems Engineering, Delft University of Technology, Delft, The Netherlands

Jaime Hueso González He received his M.S. degree in telecommunication engineering from the Polytechnic University of Valencia, Spain, in 2003. From 2004 to 2005, he was a Microwave Engineer with the European Space Agency, Noordwijk, the Netherlands, where he researched on microwave filter design and multipactor effects on spaceborne microwave waveguides. Since 2006, he has been with the German Aerospace Center (DLR), Wessling, Germany. He has been working as a Calibration Engineer for projects like TerraSAR-X, TanDEM-X, and GMES Sentinel-1. His major research interests include the development of innovative calibration concepts for digital elevation models and satellite baseline determination.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Michele Grassi He is associate professor of Space Systems at the University of Naples "Federico II." His research focuses on guidance, navigation, and control of aerospace systems. He has been principal investigator/co-investigator of projects on the topic.

Department of Aerospace Engineering, University of Naples "Federico II", Naples, Italy

Maria Daniela Graziano Since 2011, she is researcher for CO.RI.S.T.A., Consortium for Research on Advanced Remote Sensing System. Her research is mainly devoted to innovative methods for constellation design to enable new arising applications and formation flying mission studies.

Department of Aerospace and Mechanical Engineering, Second University of Naples, Aversa (CE), Italy

Thomas Grelier He is a radionavigation engineer at CNES. His research activities focus on GNSS signal processing and design, and development of GNSS space receivers. He was the technician responsible for the FFRF equipment on the PRISMA satellites.

CNES, Toulouse Space Center, Toulouse, France

Jian Guo He is assistant professor of Space Systems Engineering, He has coauthored more than 50 papers. His research interests include space systems engineering, distributed space systems, and miniaturized components.

Space Systems Engineering, Delft University of Technology, Delft, The Netherlands

Irena Hajnsek She is a Professor of the Earth Observation and Microwaves Remote Sensing chair at the Institute of Environmental Engineering at the Swiss Federal Institute of Technology Zürich (ETH Zürich) and is a head of the Polarimetric SAR Interferometry research group at the German Aerospace Center. She was part from the very beginning, back in 2003, of the core team proposing TanDEM-X and is today the science coordinator of the TanDEM-X mission.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Nadjim Horri He is a research fellow and lecturer at SSC, where he teaches spacecraft guidance, navigation, and control. He specializes in the attitude and orbit control of spacecraft and the control for satellite.

Surrey Space Centre, University of Surrey, Surrey, United Kingdom

Claudio Iacopino Claudio Iacopino is currently a Ph.D. student at the Surrey Space Centre, University of Surrey. He is researching on planning and scheduling systems for distributed missions using natural inspired approaches. He works in close collaboration with Surrey Satellite Technology Ltd and the European Space Operation Centre.

Surrey Space Centre, University of Surrey, Surrey, United Kingdom

Ralph Kahle He received a diploma degree in aerospace engineering from the Technical University of Dresden in 2001, and a Ph.D. degree from the Technical University of Berlin in 2005. He joined DLR's German Space Operations Center (GSOC) as a Flight Dynamics (FD) engineer in 2004. Ralph Kahle is responsible for the TerraSAR-X/TanDEM-X FD system engineering and operation. His field of work also comprises space mission analysis, FD software development, and conduction of LEOP, and routine FD operations.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Michael Kirschner He was born in 1954 in Rhineland-Palatinate/Germany. University diploma in Space Technology (Institute of Space Propulsion, Stuttgart 1982), dissertation in Space Technology (Institute of Astronautics, Munich 1990). Senior engineer at flight dynamics group of German Space Operations Center (GSOC, DLR). Since 1990, Dr. Kirschner was involved in several satellite mission operations.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Rolf Konig He received a diploma degree in Geodesy from the University of Stuttgart, Germany, in 1980; an assessor degree in legal surveying from the state survey of Baden-Wuerttemberg, Germany, in 1982; and a Dr.-Ing. degree from the University of the Federal Armed Forces in Munich, Germany, in 1988. He joined the German Geodetic Research Institute (DGFI) in 1988 and moved to the German Research Centre for Geosciences (GFZ) in 1992. Since 1988, Dr. Koenig managed various work packages connected to satellite missions mainly in the frame of satellite geodesy. He is engaged in the international science community concerning space-geodetic techniques and reference systems. Currently he is heading GFZ's section topic "Earth System Parameters".

GFZ German Research Center for Geoscience – Dept 1 c/o DLR, Oberpfaffenhofen, Wessling, Germany

Gerhard Krieger He received the Dipl.-Ing. (M.S.) and Dr.-Ing. (Ph.D.) degrees in electrical and communication engineering from the Technical University of Munich, Germany, in 1992 and 1999, respectively. In 1999, he joined the Microwaves and Radar Institute (HR) of the German Aerospace Center (DLR). From 2001 to 2007, he led the New SAR Missions group that conceived and developed advanced bistatic and multistatic radar systems and missions like TanDEM-X. After approval of TanDEM-X for implementation, he has been appointed as the mission's primary Systems Engineer. Since 2008, Gerhard Krieger has been head of the Radar Concepts Department.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Paco López-Dekker He received the Ingeniero degree in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC), in 1997; the M.S. degree in electrical and computer engineering from the University of California, Irvine, in 1998; and the Ph.D. degree from the University of Massachusetts, Amherst, in 2003. During 2003, he worked at Starlab, focusing on the development of GNSS-R sensors. From 2004 to 2006, he was a Visiting Professor at the Universitat Autonoma de Barcelona (UAB). In 2006, he joined the Remote Sensing Laboratory at UPC to conduct research on bistatic SAR under a 5-year Ramon y Cajal Grant. Since November 2009, he leads the SAR Missions group at the Microwaves and Radar Institute (HR) of the German Aerospace Center (DLR).

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Daan Maessen He is a Ph.D. candidate in the field of relative navigation for formation flying satellites at the chair of Space Systems Engineering of Delft University of Technology. He received a Master of Science in Aerospace Engineering at the same university and has contributed to more than 20 scientific publications in his field of work.

Space Systems Engineering, Delft University of Technology, Delft, The Netherlands

Franz-Heinrich Massmann He was born in 1954 in Lower Saxony/Germany. University diploma in Geodesy (1979), Senior scientist with focus on orbit and gravity determination at the University of Stuttgart, DGFI Munich, and since 1992 at the German Research Center for Geosciences (GFZ). For the last 10 years he is also involved in satellite operations management (CHAMP, GRACE).

German Research Centre for Geosciences, Geodesy and Remote Sensing, Potsdam, Germany

Didier Massonnet He is head of the PHARAO ultra-accurate space atomic clock project at CNES. His research activities comprise SAR image processing, notably in differential interferometry, as well as SAR system design, such as the Interferometric Cartwheel concept.

CNES, Toulouse Space Center, Toulouse, France

Antonio Moccia He is full professor of aerospace systems and chairman of the Ph.D. School in Industrial Engineering. His research interests include high-resolution aerospace remote sensing systems, guidance, navigation and control of aerospace systems, space systems mission analysis, space systems dynamics and control, and remote sensing data numerical analysis for operative applications. He has been principal investigator in these fields in several national and international research programs.

Department of Aerospace Engineering, University of Naples "Federico II", Naples, Italy

Oliver Montenbruck He is head of the GNSS Technology and Navigation Group at DLR/GSOC and lecturer at Technische Universität München. His research activities comprise spaceborne GNSS receiver technology, autonomous navigation systems, spacecraft formation flying, and precise orbit determination.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Alberto Moreira He is the Director of the Microwaves and Radar Institute at German Aerospace Center (DLR) and a Professor at the Karlsruhe Institute of Technology (KIT), Germany, in the field of microwave remote sensing. The mission TanDEM-X, led by his Institute, has successfully started the operational phase in December 2010. Prof. Moreira is the initiator and Principal Investigator for this mission. His professional interests and research areas encompass radar end-to-end system design and analysis, innovative microwave techniques and system concepts, signal processing, and remote sensing applications.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Ron Noteborn He is senior AOCS specialist in the AOCS and Software Department at OHB Sweden. His involvements include GNC lead engineer, GNC operations engineer, and PI of the Autonomous Rendezvous experiment on the Prisma formation flying mission.

OHB Sweden AB, Solna, Sweden

Phil Palmer Phil Palmer is currently a Reader at the Surrey Space Centre, University of Surrey. He heads a research group in Astrodynamics and works closely with Surrey Satellite Technology Ltd (SSTL) to provide constellation management and mission planning systems for commercial applications.

Surrey Space Centre, University of Surrey, Surrey, United Kingdom

Alfredo Renga He is a post-doc researcher. His research activities deal with monostatic and bistatic synthetic aperture radar for remote sensing and surveillance and relative navigation of spaceborne, airborne and ground-moving platforms by carrier-based differential GPS.

Department of Aerospace Engineering, University of Naples "Federico II", Naples, Italy

Rainer Sandau Rainer Sandau has over 30 years of experience in airborne and spaceborne remote-sensing activities. He was involved in the instrumentation of space missions to Venus, Mars, and Earth, and in numerous concepts for instruments and small satellites with different countries/space agencies. He is a member of national and international associations, for instance Technical Director of Satellites and Space Applications in the IAA and chairman of the International Policy Advisory Committee of ISPRS. He has authored/coauthored over 250 publications, holds over 25 patents, and is coauthor or editor/coeditor of 22 books and proceedings.

German Aerospace Center (DLR), Berlin, Germany

Birgit Schättler She received the Dipl.-Math. degree in mathematics from the Bayerische Julius-Maximilians-Universität of Würzburg in 1986. In 1987, she joined the German Remote Sensing Data Center of the German Aerospace Center (DLR) and is currently working as system engineer in the Remote Sensing Technology Institute. She was responsible for integration and acceptance testing of the TanDEM-X Ground Segment.

German Aerospace Center (DLR), Remote Sensing Technology Institute, Wessling, Germany

Klaus Schilling Prof. Dr. Schilling's focus was in space industry on the design of interplanetary satellites (inluding Huygens, Roetta, Mars Rovers), before he became Ordinarius Informatics VII: Robotics and Telematics at University Würzburg. In parallel he is president of the company "Zentrum für Telematik e.V.". He was Consulting Professor at Stanford University and recipient of the Walter-Reis-Award in Service Robotics 2008. In 2005, his team realized the first German pico-satellite UWE-1 (University Wuerzburg's Experimental satellite).

Julius Maximilians University, Würzburg, Germany

Marco Schmidt Marco Schmidt is the head of space activities at computer science department 7 of the University of Wuerzburg. He finished his doctoral degree in the field of ground station networks. His research interests include distributed systems, ground station networks and small satellites.

Julius Maximilians University, Würzburg, Germany

Daniel Schulze He received a diploma degree in astronautics and space technology from the Technical University of Berlin, Germany, in 2002. He joined DLR's Microwaves and Radar Institute in 2004 and started as System Engineer for the TerraSAR-X mission in the Microwaves and Radar Institute. Since 2008, he is group leader and project manager of the TerraSAR-X/TanDEM-X System Engineering and Calibration Segment.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Gianfranco Sechi He is senior AOCS engineer with specific experiences in autonomous attitude determination systems, drag-free systems, and formation flight. In particular, he was in charge of drag-free and attitude control system algorithms design for the GOCE mission. He designed the NGGM control system in the TAS-I studies.

Thales Alenia Space Italia, Torino, Italy

Tony Sephton Has spent over 20 years in the space industry, in applications, ground segment, and EO future missions. As Head of the Awareness Activities section within the Integrated Applications Department at ESA, he now has a specific interest in promoting the development of satellite services for a new range of users.

Department of Integrated Applications, European Space Agency, Noordwijk, The Netherlands

Michael Steinhoff He was born in 1951 in Frankfurt am Main, Germany. He is a graduate mathematician (Johann Wolfgang Goethe University, Frankfurt 1984) and GRACE Mission Operations Project Manager. Since 1985, he is involved in several satellite mission operations at German Space Operations Center (GSOC, DLR).

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Dieter Ulrich He received a diploma degree in Electrical Engineering from the Technical University of Karlsruhe in 1983. He joined Astrium GmbH (former Dornier System GmbH) in Friedrichshafen and worked in several programs and projects as an attitude control system engineer and project manager. From 2001, he was leading the AOCS and Operations System Engineering Group of the TerraSAR-X and TanDEM-X project. He is currently managing the PAZ platform project at Astrium GmbH.

Astrium GmbH – AED71, Friedrichshafen, Germany

Srinivas R. Vadali He is a Professor of Aerospace Engineering. He also serves as an Associate Director of the Texas Space Grant Consortium and an Associate Editor for the *International Journal of Aerospace Engineering*. He is a Fellow of AAS, Associate Fellow of AIAA and a coauthor of the book *Spacecraft Formation Flying: Dynamics, Control, and Navigation*, Elsevier, 2010.

Department of Aerospace Engineering, Texas A&M university, College Station, TX USA

Martin Wermuth He graduated in 2001 from Technische Universität München (TUM) with a diploma degree in Geodesy. Following a research fellowship at GRGS/CNES, Toulouse, he worked on global gravity field determination from the CHAMP and GOCE mission at the TUM Institute for Astronomical and Physical Geodesy (IAPG) and received a Ph.D. in 2008. Since May 2008, he is working as a flight dynamics engineer at DLR/GSOC with focus on precise orbit determination of LEO satellites.

German Aerospace Center (DLR), German Space Operations Center, Wessling, Germany

Birgit Wessel She received a diploma degree in Geodesy from the University of Hanover, Germany in 2000 and a Dr.-Ing. degree from the Technische Universitaet Muenchen (TUM), Germany in 2006. From 2000 to 2005, she was scientific collaborator at the Institute for Photogrammetry and Cartography at the TUM. Since 2005, she is with the German Remote Sensing Data Center at the German Aerospace Center (DLR), where she currently is leading the DEM calibration and mosaicking developments for the German interferometric TanDEM-X satellite mission. Her research focuses on SAR remote sensing including image analysis, characterization of objects, and geocoding.

German Aerospace Center (DLR), German Remote Sensing Data Center, Wessling, Germany

Alex Wishart Has worked at Astrium Ltd for more than 25 years mainly in the area of digital signal processing, with the last 5 years as a Systems Engineer within the Earth Observation, Navigation and Science Division. Interested in onboard signal processing techniques and technologies with application to heterodyne RF and submillimeter instrumentation.

Future Programmes Department, Astrium Ltd, Stevenage, UK

Manfred Zink He received a diploma degree in Physics from the Technical University of Graz, Austria in 1987 and a Dr.-Ing. degree from the University of Stuttgart, Germany, in 1993. He joined DLR's Microwaves and Radar Institute in 1988 and was the Lead X-SAR Calibration Engineer for both SIR-C/X-SAR missions in 1994 and for the SRTM mission in 2000. From 2000 to 2005, Manfred Zink worked as a systems engineer for the European Space Agency's (ESA) ENVISAT/ASAR and TerraSAR-L programs. After returning to the Microwaves and Radar Institute, he has been managing the TanDEM-X Ground Segment development. He is currently heading the Satellite SAR Systems Department.

German Aerospace Center (DLR), Microwaves and Radar Institute, Wessling, Germany

Contents

Part I
Distributed Radar Sensors

Chapter 1
Bistatic Synthetic Aperture Radar

Antonio Moccia and Alfredo Renga

Abstract Bistatic Synthetic Aperture Radar represents an active research and development area in radar technology. In addition, Bistatic and Multistatic SAR concepts are tightly related to formation flying and distributed space missions that also represent the new space-based remote sensing and surveillance frontiers. This chapter introduces Bistatic SAR, in particular by comparing its peculiarities, operation and performance with respect to conventional monostatic SAR. Some basic concepts of bistatic SAR image formation and the main elements of bistatic SAR geometry are preliminary presented. Performance parameters are then analyzed, including geometric resolution, radiometric resolution and bistatic radar equation. Special emphasis is placed on analytical methods to evaluate the effects of bistatic SAR geometry on image resolution. Further implementation issues, such as footprint, time and phase synchronization are also pointed out. The analysis of past bistatic radar and bistatic SAR experiments and proposed spaceborne bistatic SAR missions supplies essential information to understand how these issues have been faced and can be potentially solved in ongoing and future operational systems. Finally, several scientific applications of bistatic SAR are outlined taking advantages of different techniques and methods.

1.1 Introduction

Bistatic Synthetic Aperture Radar (SAR) is a distributed observation system based on transmission and reception of frequency modulated microwave signals. As the name suggests an exhaustive definition of Bistatic SAR involves the presentation of two distinct concepts, namely *bistatic radar* and *synthetic aperture*.

A. Moccia (✉) • A. Renga
Department of Aerospace Engineering, University of Naples "Federico II", P.le Tecchio 80, 80125 Naples, Italy
e-mail: antonio.moccia@unina.it; alfredo.renga@unina.it

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_1,

A bistatic radar is a radar with separated transmitting and receiving antennas. Even if the name bistatic radar was coined only in 1952 by K. M. Siegel and R. E. Machol [1], the early experimental radars, developed before World War II, were actually of bistatic type, with the transmitter and the receiver separated by a distance comparable to the radar-target one. Moreover, since most of the early radar experiments were conducted utilizing the existing communication technology, continuous wave (CW) transmissions and very low frequencies (25–80 MHz) were adopted, with ground-based configurations devoted to aircraft detection. Successively, the invention of the duplexer [1] allowed the radar system to use pulses of finite time duration with a common transmit and receive antenna thus leading to the *monostatic* configuration. The monostatic radar became in a short time a standard for radar implementations especially for use by moving platforms, such as aircrafts, ships and mobile ground units. However, bistatic radar retains several unique features making it very interesting both for civil and defense applications. First, the receiver, in principle, can be simple, compact and light-weight not needing radar transmitting chain. Then, bistatic receivers are able to work without revealing their position so they can become undetectable when operating in tactical scenarios. Moreover, while it is relatively affordable to develop stealth capabilities against a monostatic illuminator, echoes reflected in other directions cannot be easily reduced and therefore they can be collected by proper bistatic receivers. Finally both monostatic and bistatic data reflected by common covered areas can be collected to improve, quantitatively and qualitatively, the characterization of the observed target or scene. This last feature highlights how in bistatic radar two types of transmitting systems, or illuminators, can be envisaged:

- *Cooperating*, i.e. the transmitter and the receiver are designed and developed jointly for carrying out common and specific tasks such as the acquisition of monostatic data together with bistatic ones. However, even if monostatic operation is not enabled, cooperating systems are implemented when the illuminator provides the bistatic receiver with additional timing and navigation information, such as transmitted oscillator state for coherent processing or position and velocity of transmitting antenna phase center.
- *Non-cooperating*, i.e. the receiver exploits the electromagnetic signal radiated by an independent and uncoordinated source, a so-called *illuminator of opportunity*, such as GPS or broadcasting satellites and TV or radio towers [2]. In this framework the terms parasitic receivers or hitchhiking radar have been also utilized [1, 2].

As far as synthetic aperture is concerned, the concept of synthetic aperture radar as mapping sensor was first introduced by Carl Wiley in 1951 [3]. Basically Synthetic Aperture Radar (SAR) is a chirp radar that exploits amplitude and phase of the received echoes [4]. It is therefore able to produce an image by measuring backscattered pulses as a function of time delay and Doppler shift, thus determining the radar complex reflectivity for each range-Doppler terrain cell. Values from different cells are combined to form a two dimensional image. As in traditional real aperture radar, the synthetic radar beam is directed to the side

of the platform trajectory. The antenna is designed to have sufficiently wide beam width in the along-track direction so that a given target or area is illuminated by a number of pulses. The number of illuminating pulses is typically on the order of thousand for a spaceborne SAR and the ensemble of locations of the antenna during the illumination of a given target constitutes the *synthetic aperture* of the system. The length of the synthetic aperture is typically 3–7 km for spacecraft and 60–100 m for aircraft [4], hence impossible to be carried on a single aerospace system. Monostatic SAR, with a single antenna, high-power transmitter and high-sensitive receiver onboard the same air or space platform represents today a mature remote sensing and surveillance system and it is routinely utilized in a wide range of commercial and scientific applications.

On the basis of these preliminary remarks it is possible to state that Bistatic Synthetic Aperture Radar is a SAR with separated transmitting and receiving antennas. Similarly to multistatic SAR, it is based on multiple transmitting and receiving antenna positions. The realization of air or space bi-/multistatic SAR involves the coordinate use of multiple systems, possibly operating in separate locations or platforms, with accurate time synchronization and antenna pointing between transmitters and receivers and with accurate measurement and control of relative positions, i.e. implementing formation flying strategies, even when illuminators of opportunity are considered. Apparently bistatic/multistatic SAR is more demanding than a monostatic one, due to the necessity of forming synthetic antennas synchronizing more than one system. However the above statements on SAR, which are reported more or less similar in every classic textbook when introducing the synthetic aperture concept (see for example [5, 6]), can be rephrased with reference to formation flying and antenna synchronization as showed in Fig. 1.1, i.e. with reference to bi-/multistatic SAR.

Hence, Bistatic/Multistatic Synthetic Aperture Radar is a concept as old as Synthetic Aperture Radar, although there was not a clear awareness of that the very beginning. Actually, even a monostatic SAR is basically a distributed (i.e. multistatic) system, at least from the image formation and data processing point of view.

In addition, this way of interpreting SAR behavior highlights how Bistatic and Multistatic SAR concepts are tightly related to the new spaceborne remote sensing and surveillance frontier to which the present book is devoted: formation flying. Also in formation flying [7], indeed, groupings of duplicate or similar satellites have sensors in common or are complementary related, talk to each other, share payloads or mission functions, and are in one or more ways synchronized. In summary, it is possible to state that dealing with Bistatic/Multistatic SAR means dealing with the future of spaceborne (and airborne) remote sensing systems.

Hitherto, the related implementation challenges have contributed to restrain the development of bistatic SAR systems. Limited bistatic experiences have been therefore reported in the literature [8] and the first operative bistatic SAR mission is TanDEM-X [9, 10], launched only in 2010, and realizing a spaceborne bistatic SAR interferometer based on two LEO satellites flying in close formation. Research in airborne and spaceborne bistatic SAR, however, has been recently intensified and

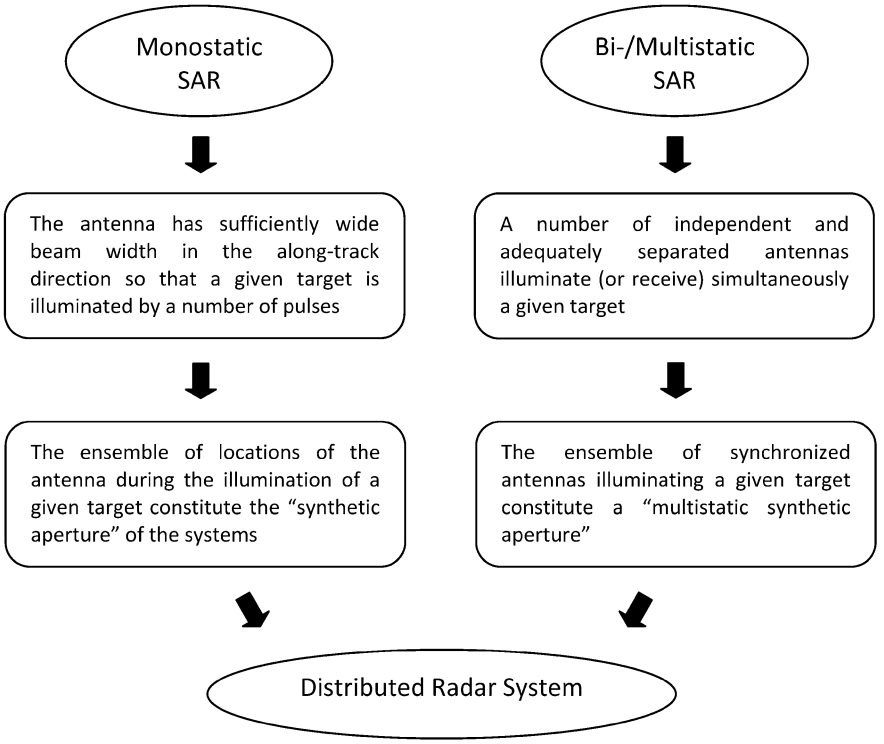


Fig. 1.1 Interpretation of SAR as distributed radar system

it is very likely that bistatic SAR will go beyond the experimental level in the near future, pressed by the potential vast amount of scientific and practical tasks bistatic SAR can fulfill.

This chapter is organized as follows. Section 1.2 introduces the basic concepts of bistatic SAR image formation and defines the main elements of bistatic SAR geometry. The concept of critical baseline is also presented to stress the different information content between coherent and incoherent combinations of bistatic data. Performance parameters are introduced in Sect. 1.3. In detail, Sect. 1.3.1 shows some analytical methods to evaluate the effects of bistatic SAR geometry on the geometric resolution, whereas radiometric resolution and bistatic radar equation are analyzed in Sect. 1.3.2. Further implementation issues, such as footprint, time and phase synchronization are presented in Sect. 1.3.3. The analysis of past bistatic radar (Sect. 1.4.1) and SAR (Sect. 1.4.2) experiments and proposed spaceborne SAR missions (Sect. 1.4.3) supplies essential information to highlight how these issues have been faced and can be potentially solved in ongoing and future operational systems. Finally, several scientific applications of bistatic SAR are outlined in Sect. 1.5 taking advantages of different techniques and methods.

1.2 Bistatic SAR Geometry and Configurations

Three elements form the basic bistatic SAR configuration: the transmitter or illuminator, the receiver and the target or observed scene. The illuminator transmits a sequence of frequency modulated pulses towards the target area while following its trajectory (see Fig. 1.2). Those pulses are reflected from the scene thus generating a sequence of bistatic echoes that are collected by the receiver. In general the receiver is a moving platform too. For this configuration the start-and-stop approximation [11] can be assumed, that is transmitter and receiver antennas are considered to be at a standstill during ranging signal propagation, and they move on the next position only after the echo is received. According to this hypothesis the signal can be modeled as a function of two independent variables: the *fast time* and the *slow time*. The fast time is the time scale which scans the transmission/reception of the pulses at the velocity of light, whereas the slow time is the time scale which scans successive positions of transmitter and receiver along the synthetic aperture, at a velocity which is negligible with respect to the velocity of light. A raw bistatic SAR image is then generated by the sequence of transmitted/received pulses as a function of the fast and slow time. Pulse compression techniques [4–6] are utilized exploiting the frequency modulation of the transmitter pulses to enhance the fast time or range resolution. As far as the slow time is concerned the Doppler frequency modulation, or Doppler history, due to the relative motion among transmitter, receiver and target can be tracked and hence compressed as done for the fast time signals. In this context, as noted in the introduction, bistatic SAR image formation and focusing resemble those of conventional monostatic SAR: coherent processing is utilized even if different relations and models have to be utilized to perform Doppler compression and to relate fast and slow time with conventional geometric and cartographic image coordinates. Indeed, in conventional monostatic SAR two directions can be individuated immediately, namely the along-track or azimuth direction, parallel to the motion of platform relative to observed scene, and across-track direction orthogonal to the previous one. The fast time or range resolution is relevant to across-track direction and the slow-time or Doppler resolution to

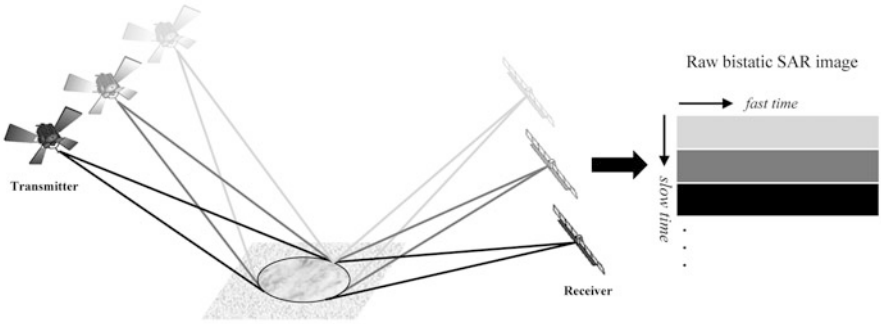


Fig. 1.2 Illustration of bistatic SAR raw image formation

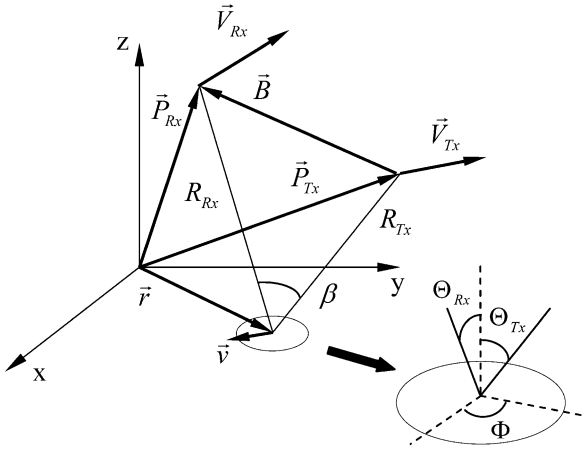


Fig. 1.3 General bistatic SAR geometry

along-track direction. A monostatic SAR image is therefore naturally defined in along-track and across-track components. In bistatic SAR, instead, the relative positions and velocities between transmitter and receiver turn out to be completely arbitrary. In such general conditions neither across-track nor along-track directions strictly exist, and therefore new definitions and models are needed as Sect. 1.3 will show. Moreover it is important to remark that a pulsed working mode has been presented for simplicity, but bistatic operation also allows for CW transmissions and receptions to be used. Possible impacts of such CW schemes on bistatic SAR image formations will be discussed in Sect. 1.3.3.

The basic bistatic SAR configuration can operate in a wide range of bistatic geometries. For the sake of individuating the most notable ones a set of geometric parameters can be introduced. The bistatic synthetic aperture, indeed, is determined by relative geometry and motion among transmitter, receiver and target area, i.e. transmitter and receiver positions, $\vec{P}_{Tx}$ and $\vec{P}_{Rx}$, transmitter and receiver velocities, $\vec{V}_{Tx}$ and $\vec{V}_{Rx}$, target position, $\vec{r}$, target velocity $\vec{v}$, and transmitter and receiver slant ranges, R_{Tx} and R_{Rx}. Classic bistatic radar literature [1] utilizes two parameters to define the bistatic geometry: the baseline, $\vec{B}$, that is the vector from the transmitter to the receiver, and the bistatic angle, β, that is the transmitter-target-receiver angle. However, as far as imaging radar applications are concerned, it is worth introducing transmitter and receiver incidence angles, Θ_{Tx} and Θ_{Rx} respectively, and transmitter out-of-plane angle, Φ, too.

Figure 1.3 represents a general bistatic SAR geometry and illustrates main relevant parameters defined with respect to an arbitrary Earth-fixed Cartesian reference frame. According to those parameters various bistatic SAR geometries can be defined:

- *In-plane same-side*, with the transmitter within the range elevation plane of the receiving antenna and at the same side with respect to the target, i.e. $\Phi = 0$.

- *In-plane opposite-side*, with the transmitter within the range elevation plane of the receiving antenna but at the opposite side with respect to the target, i.e. $\Phi = 180°$.
- *Out-of-plane*, $\Phi \neq 0$ and $\Phi \neq 180$.
- *Nadir-looking* transmitter, $\Theta_{Tx} = 0$, or receiver, $\Theta_{Rx} = 0$.

Furthermore, the receiver can perform different observations depending on its looking direction. Specifically *side-looking*, *forward-looking* and *squinted-mode* receivers are implemented when the receiving antenna is pointed, respectively to the side of the platform trajectory, along platform trajectory directed forwards or backwards, or otherwise.

The basic bistatic SAR configuration can be enriched in different ways. First of all, monostatic operation can be exploited together with bistatic ones [12], namely the transmitter also receives the echoes backscattered from the target area and therefore a pair of monostatic-bistatic images of the same region can be generated. In such a monostatic-bistatic configuration additional time requirements are set: CW transmissions cannot be used since they would hinder monostatic data from being received. Moreover, even though pulsed transmission schemes are utilized, both monostatic and bistatic Doppler bandwidths have to be unambiguously sampled for monostatic and bistatic slow time compression. This is not problematic if relatively close monostatic-bistatic configurations are utilized (e.g. for LEO satellites flying in formations), in which monostatic and bistatic Doppler histories are alike, but complications could be caused in general geometries when the motion characteristic of the transmitter are very different from those of the receiver (e.g. for transmitter and receiver flying at different altitudes or working in different looking modes).

More demanding bistatic SAR configurations involve the utilization of multiple receivers and/or multiple transmitters [13]. In this context, a single transmitter can be source of illumination for several bistatic SAR receivers covering a given target area but also MIMO (Multi-Input Multi-Output) SAR architectures can be envisaged with more, monostatic or non monostatic, transmitters.

From the data exploitation point of view when a bistatic SAR configuration is implemented, able to generate multiple bistatic (or monostatic-bistatic) images of the same area, it is necessary to compare and combine in some way these images, thus improving the achievable characterization level of the observed scene. As a result of the coherent SAR focusing the output of bistatic SAR processors is a complex image. Two or more bistatic images can be therefore compared both in amplitude and in phase. When both amplitude and phase are used a *coherent combination* is formed, whereas in an *incoherent combination* only amplitude differences among the bistatic data are taken into account.

Evidently, a coherent combination can be properly exploited only if there is an adequate phase correlation among the considered data. In bistatic SAR configurations, the degree of phase correlation is mainly determined by the separation among the receivers. In this concern, the concept of *critical baseline* can be introduced: the critical value for the baseline is the one causing complete loss of

correlation between focused SAR images [14, 15]. This means that a coherent combination can be derived only if the separation between the receivers is shorter than the critical baseline. The actual value of the critical baseline strictly depends on the intended application of the considered coherent combination, e.g. along-track interferometry or cross-track interferometry (see Sect. 1.5.2.1 and Chap. 2 for the details). The critical baseline for cross-track interferometry can be assumed as figure of merit to derive the order of magnitude of critical separation. For this application, the critical value corresponds to the baseline for which the variation in phase difference between adjacent resolution elements is 2π. According to the models developed in Refs. [14, 15] and extended to bistatic SAR operations [9, 16] the critical baseline is

$$B_{\perp} = \frac{\lambda R_{Rx}}{\Delta r_g \cos \Theta_{Rx}} \tag{1.1}$$

where λ is the wavelength and Δr_g is the bistatic ground range resolution. The symbol $\perp$ indicates that the critical baseline computed by (1.1) is not the actual separation among the receivers, but it is the baseline component along the direction orthogonal to the line of sight.

For simplicity only the orthogonal critical baseline will be considered in the remainder of this section, even if the actual baseline could be derived accordingly once the bistatic configuration is defined. Furthermore, it is important to remark that (1.1) returns the critical baseline for all introduced bistatic SAR configurations, specifically (1.1) can be utilized to calculate the critical baseline for:

- Monostatic-bistatic SAR configurations, in this case the baseline is the separation between the monostatic (i.e. transmitting/receiving) sensor and the bistatic (receiving-only) one;
- Multistatic SAR configurations, in this case the baseline is the separation among all the sensors that receive bistatic echoes from the same transmitter.

Figure 1.4 reports the critical baseline for different bistatic SAR formations, namely Low Earth Orbit (LEO), Medium Earth Orbit (MEO), Geosynchronous Earth Orbit (GEO) formations, and a bistatic SAR characterized by Moon-based stations. All of these formations have been proposed for Earth remote sensing by cross-track SAR interferometry (see for example [13, 17, 18]). From Fig. 1.4 it is immediate to conclude that the critical baseline is always a small fraction of the slant range distance.

In order to preserve phase correlation, the working baseline must be a fraction of the critical one (a value from 5% to 10% has been selected in Refs. [9, 10] for TanDEM-X mission), this means that very close formations (with baseline lower than 1–2 km) have to be considered for LEO bistatic SAR. The baselines increase from hundreds to thousands of kilometers from MEO or geosynchronous to Moon-based bistatic systems. These huge baselines, however, have to be interpreted also considering the relevant enormous slant ranges. Indeed, the same bistatic angle,

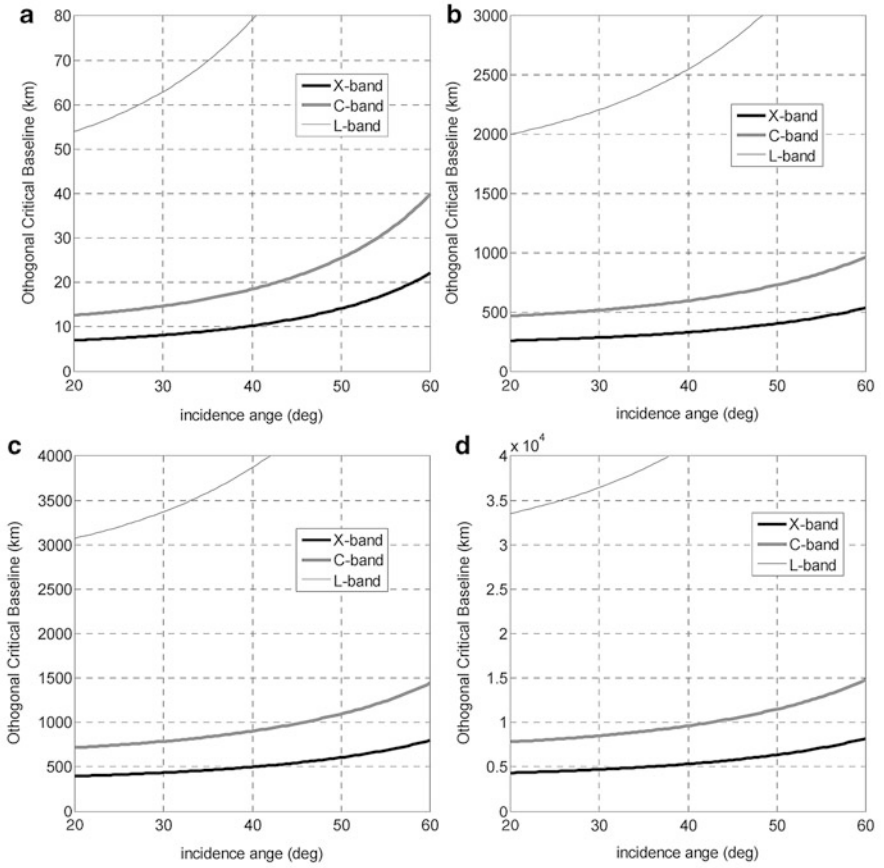


Fig. 1.4 Critical baseline for different bistatic SAR formations and wavelength (3 m ground range resolution, X-band $\lambda = 3.1$ cm, C-band $\lambda = 5.6$ cm, L-band $\lambda = 24$ cm): (**a**) LEO, (**b**) MEO, (**c**) GEO, (**d**) Moon-based

reported in Fig. 1.5, is attained for all introduced bistatic formations: as a matter of fact the bistatic angle is very small in any case.

As noted above, the critical baseline changes if coherent combinations are used for a different application, such as moving target indication (MTI) by along-track interferometry. In this case the baseline sets the maximum and the minimum target velocities the bistatic SAR can detect. In general both velocities depend on the ratio between satellite platform velocity and baseline [19]. For LEO orbits useful baselines are limited to hundreds of meters [9, 10, 20], and even shorted separations would be of interests for higher altitudes.

In conclusion, it is possible to state that in coherent bistatic combinations very limited variations of the observation geometry has to be expected among the different receivers forming the bi-/multistatic SAR system.

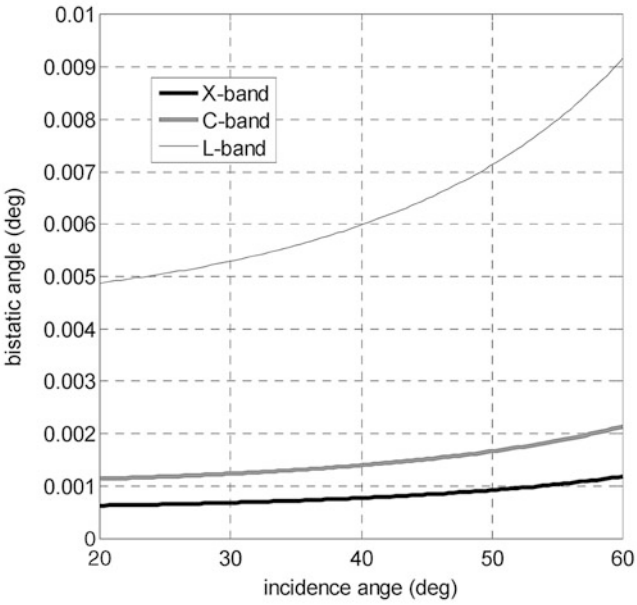


Fig. 1.5 Bistatic angles corresponding to the critical baselines reported in Fig. 1.4 (3 m ground range resolution assumed)

1.3 System Performance

Parameters determining bistatic SAR performance strictly depends on the intended application. For instance in coherent combinations the capability to reduce phase noise and biases is essential. However, as for all imaging sensors, some common parameters can be individuated to define system performance, the most notable example being the *resolution*.

Resolution is defined as the capability to distinguish close targets on the ground. Specifically, *geometric resolution* deals with the separation in one or more dimensions of two targets with the same amplitude, i.e. the same brightness in the resulting image. In synthetic aperture radar such dimensions are (slant or ground) range and Doppler [4–6]. In addition *radiometric resolution* can be introduced and it is in general defined as the capability of the sensor to distinguish between targets characterized by different brightness in SAR images, depending on their different scattering properties along with instrument sensitivity. The effects of bistatic system parameters, such as acquisition geometry, bandwidth and integration time on geometric resolution is presented in Sect. 1.3.1. As in monostatic SAR radiometric resolution is essentially influenced by Signal-to-Noise Ratio (SNR) and additional speckle-generated noise. However bistatic SAR geometry and bistatic scattering properties, as reported in Sect. 1.3.2, play an important role too.

Additional performance parameters concern timing, pointing and synchronization issues. Again, part of these problems are conceptually similar to those of monostatic SAR, such as pulse repetition frequency selection, whereas other ones are peculiar of bistatic system, e.g. low-frequency phase errors are generated by the use of two different oscillators, the first in transmission, and the second in reception. Basic concepts and relevant effects on SAR imaging are reported in Sect. 1.3.3.

1.3.1 Geometric Resolution

Synthetic aperture radar utilizes frequency modulated signals to achieve high fast time resolution, whereas the Doppler frequency modulation of the received signal, due to relative motion among transmitter, receiver and target is exploited to enhance slow time resolution. In conventional side-looking monostatic SAR fast time resolution corresponds to the capability to distinguish two close targets with different slant or ground range coordinates. Moreover, slow time resolution derives from platform velocity and leads to the well known concept of azimuth, or along-track resolution. If a bistatic SAR is considered, in which the transmitter and the receiver fly along parallel tracks with same altitude and velocity, attaining in-plane observation with no-squint angle, a similar principle holds. Indeed, under above hypotheses a straightforward computation of bistatic SAR resolution is possible because targets are focused in the range elevation plane common to the transmitter and the receiver [8], so range and azimuth directions are well defined and form a right angle. In this case ground range and azimuth resolutions are

$$\Delta r_{g|Bist} = \frac{c}{W(\sin\Theta_{Tx} + \sin\Theta_{Rx})} \tag{1.2}$$

$$\Delta a_{|Bist} = l_{Tx}\frac{R_{Rx}}{R_{Rx} + R_{Tx}} \tag{1.3}$$

where c is the velocity of light, W is the chirp bandwidth, l_{Tx} is the transmitting antenna length, and the receiver incidence angle must be considered negative when in opposite side geometry. According to these relations, bistatic resolutions can be easily expressed as a function of the corresponding monostatic ones

$$\Delta r_{g|Bist} = \frac{2\sin\Theta_{Tx}}{\sin\Theta_{Tx} + \sin\Theta_{Rx}}\Delta r_{g|Mono} \tag{1.4}$$

$$\Delta a_{|Bist} = \frac{2R_{Rx}}{R_{Rx} + R_{Tx}}\Delta a_{|Mono} \tag{1.5}$$

Figures 1.6 and 1.7 report the ratios between monostatic and bistatic resolutions according to (1.4) and (1.5). Specifically, the computation of azimuth resolution

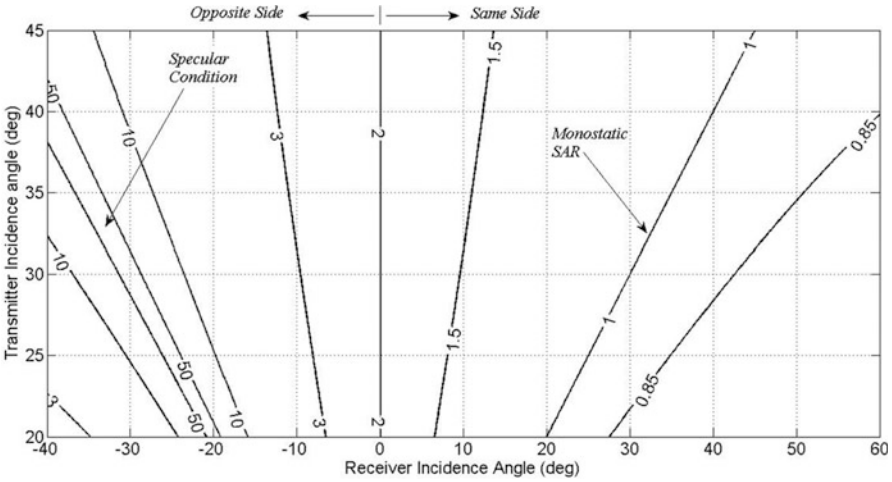


Fig. 1.6 Ratio of bistatic and monostatic ground range resolution for in-plane, parallel-track, same altitude and same velocity configuration

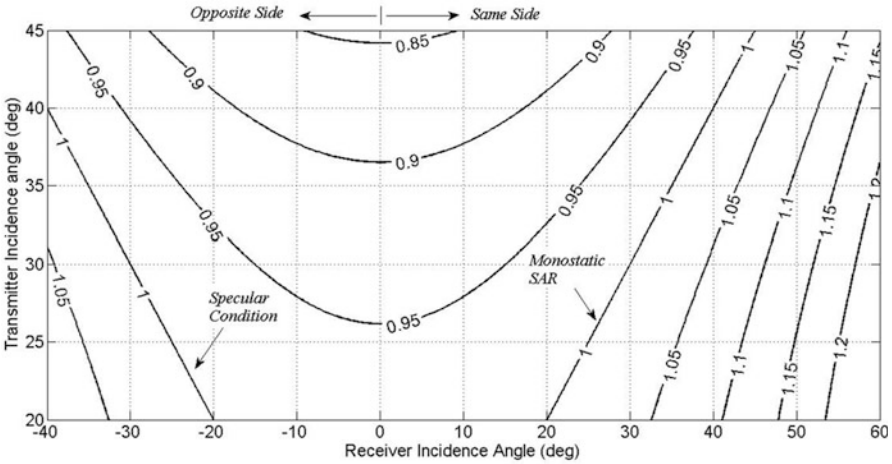


Fig. 1.7 Ratio of bistatic and monostatic azimuth resolution for in-plane, parallel-track, same altitude and same velocity configuration

ratios involves the utilization of sensor altitude to compute transmitter and receiver slant ranges: a value of 450 km is considered herein, even though the variation of sensor altitude has a negligible effect on azimuth resolution ratios when dealing with LEO formations.

On the basis of the plotted ratios, some important considerations can be stated:

- In-plane same side geometries with transmitter closer to the target ($\Theta_{Rx} > \Theta_{Tx}$) improve bistatic ground range resolution with respect to monostatic one;

- If the receiver is closer to the target ($\Theta_{Rx} < \Theta_{Tx}$), or opposite side geometry is considered, degraded bistatic ground range resolution is achieved;
- If the specular observation condition is met there is slant range ambiguity and bistatic ground range resolution diverges;
- The variation of the bistatic azimuth resolution is generally more limited and better performance then monostatic one is obtained if the receiver is closer than transmitter to the target, both in same-side and opposite-side geometry.

For bistatic SARs characterized by arbitrary geometric configurations, however, the way fast and slow time resolutions translate in different detected ground areas is not trivial. The resolution in each possible direction can be determined, in general, by computing the width of impulse response (or point spread function) along this particular direction in the focused image. This computation is complex and time-consuming and this approach does not allow for general and direct equations to be derived regulating the behavior of image resolutions in any observation geometry. Approximate solutions for the width of the point spread function along some remarkable directions can be derived by the so-called gradient method [21–23]. The method starts from the following two considerations:

- In bistatic SAR neither range nor azimuth direction strictly exists;
- Image pixels can be interpreted as generated by intersecting "iso-time delay" and "iso-Doppler" lines.

On this basis the method proposes to keep the terms range and Doppler resolutions as the minimum ground distance associated to the change of round trip delay and Doppler frequency, respectively. The direction of the gradient of time delay represents the direction which maximizes the change in time delay so it individuates the direction of best fast time resolution. The same applies to the direction of the gradient of Doppler frequency identifying the direction of the best Doppler resolution. Since the capability to distinguish close targets in the fast time is in inverse relation with transmitted signal bandwidth, the ground range resolution can be expressed as

$$\overrightarrow{\Delta r_g} = \frac{1}{W\left|\overrightarrow{\nabla t_g}\right|}\vec{i}_{t_g} \tag{1.6}$$

where $\overrightarrow{\nabla t}_g$ is the ground projection of the gradient of time delay, and $\vec{i}_{t_g}$ is the unit vector in the direction of $\overrightarrow{\nabla t}_g$. Moreover, the capability to distinguish close targets in the slow time is in inverse relation with the coherent integration time, T, that is synthetic aperture duration, so the gradient method estimates Doppler resolution as

$$\overrightarrow{\Delta a} = \frac{1}{T\left|\overrightarrow{\nabla f_g}\right|}\vec{i}_{f_g} \tag{1.7}$$

where $\overrightarrow{\nabla f_g}$ is the ground projection of the gradient of Doppler frequency and $\vec{i}_{fg}$ is its unit vector. According to the definition of gradient the unit vectors $\vec{i}_{t_g}$ and $\vec{i}_{f_g}$ are locally perpendicular to iso range and iso-Doppler lines, respectively. If the following hypotheses are assumed:

- Both transmitter and receiver fly linear path at constant velocity within the coherent integration time
- The Earth is flat within the image area

with reference to the notation introduced in Fig. 1.3, the gradients of time delay and Doppler frequencies can be expressed as

$$\overrightarrow{\nabla t} = \frac{1}{c}\left(\frac{\vec{R}_{Tx}}{R_{Tx}} + \frac{\vec{R}_{Rx}}{R_{Rx}}\right) = \frac{1}{c}\left(\vec{i}_{Tx} + \vec{i}_{Rx}\right) \tag{1.8}$$

$$\overrightarrow{\nabla f} = \frac{1}{\lambda}\left(\frac{1}{R_{Tx}}\left(\vec{V}_{Tx} - \left(\vec{V}_{Tx}\cdot\vec{i}_{Tx}\right)\vec{i}_{Tx}\right) + \frac{1}{R_{Rx}}\left(\vec{V}_{Rx} - \left(\vec{V}_{Rx}\cdot\vec{i}_{Rx}\right)\vec{i}_{Rx}\right)\right) \tag{1.9}$$

with

$$\vec{R}_{Tx} = \vec{P}_{Tx} - \vec{r} \qquad \vec{R}_{Rx} = \vec{P}_{Rx} - \vec{r}$$

From (1.6) and (1.8) and according to Fig. 1.3, the ground range resolution can be rewritten as [22]

$$\overrightarrow{\Delta r_g} = \frac{c}{W\sqrt{\sin^2\Theta_{Tx} + \sin^2\Theta_{Rx} + 2\sin\Theta_{Tx}\sin\Theta_{Rx}\cos\Phi}}\vec{i}_{t_g} \tag{1.10}$$

with

$$\vec{i}_{t_g} = \begin{bmatrix} \dfrac{\sin\Theta_{Tx}\cos\Phi + \sin\Theta_{Rx}}{\sqrt{\sin^2\Theta_{Tx} + \sin^2\Theta_{Rx} + 2\sin\Theta_{Tx}\sin\Theta_{Rx}\cos\Phi}} \\ \dfrac{\sin\Theta_{Tx}\sin\Phi}{\sqrt{\sin^2\Theta_{Tx} + \sin^2\Theta_{Rx} + 2\sin\Theta_{Tx}\sin\Theta_{Rx}\cos\Phi}} \end{bmatrix} \tag{1.11}$$

which show that the gradient of time delay points towards the bisector of the bistatic angle, so the ground range direction is always along the ground projection of this bisector. Again, as expected, if the receiver is along the specular direction of the transmitted signal ($\Phi = 180°$, $\Theta_{Tx} = \Theta_{Rx}$) ground range resolution diverges. In this case the bisector of the bistatic angle is normal to the ground plane and this means that the gradient of time delay has no component onto the local horizontal plane. As far as the Doppler direction $\vec{i}_{f_g}$ is concerned, it is determined by the relative position, velocity and geometry among transmitter, target and receiver, so

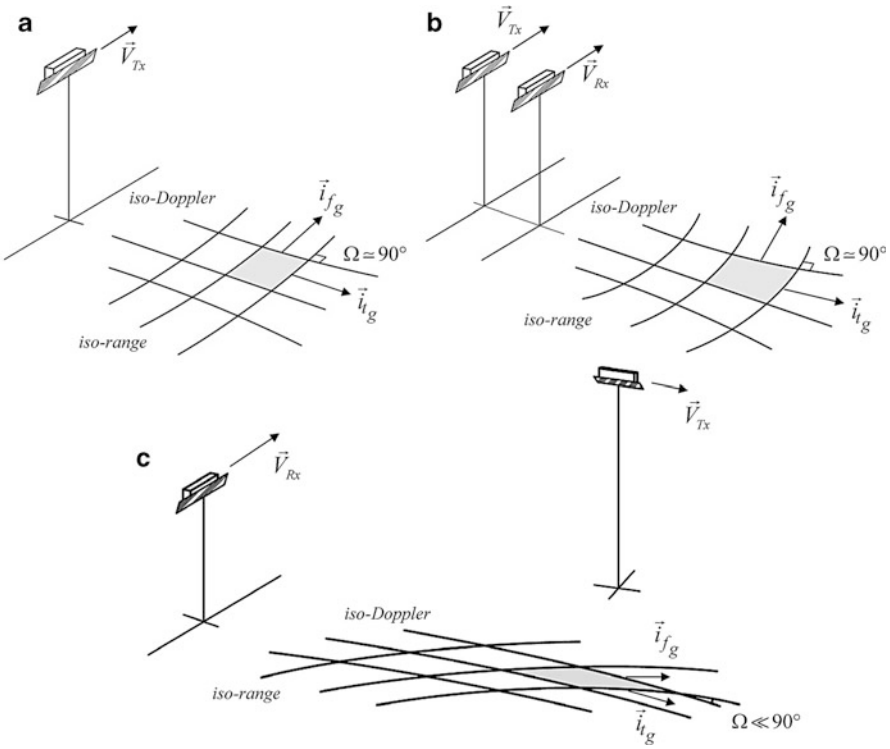


Fig. 1.8 Illustration of pixel skewing effect in monostatic (**a**) and bistatic (**b**)–(**c**) SAR

only for particular bistatic SAR configurations the direction of $\vec{i}_{f_g}$ can be easily related to system dynamics. For example, in the particular case of parallel-track, same-velocity, in plane observation Doppler direction is parallel to the velocity vectors and therefore orthogonal to the ground-range direction; further notable configurations exist and some of them will be analyzed in the following subsection.

According to the gradient method, both ground range and Doppler resolutions are completely defined in vectorial form and, in general, ground-range and Doppler directions are not perpendicular. Pixel skewing therefore arises from bistatic observation (see Fig. 1.8). Minimum pixel area is achieved if ground-range and Doppler directions are orthogonal, so the degradation with respect to this condition can be accounted for by defining the Ω angle that is the angle between iso-range and iso-Doppler lines

$$\Omega = \cos^{-1}\left(\vec{i}_{t_g} \cdot \vec{i}_{f_g}\right) \tag{1.12}$$

Equation 1.12 states that bistatic geometries exist in which ground range and Doppler resolutions are good but imaging capabilities can be very poor. In fact, if Ω

approaches zero or π the lines of constant time delay are locally parallel to the lines of constant Doppler. Hence, the Ω-angle assumes the form of an essential parameter that must be used, together with ground range and Doppler resolution, to characterize geometric imaging performance of bistatic SAR.

The above considerations put in evidence that the gradient method is characterized by the capability to cover in a simple and direct way the general bistatic geometry. However, for completeness it is important to remark that, in order to derive the gradients in (1.6) and (1.7), a definite time instant has to be individuated within the coherent integration time. Transmitter and receiver positions and velocities at this time must be considered to compute the gradients. This means that the method implicitly neglects the variations of the gradient within the synthetic aperture. However, as shown in Ref. [22], this variation is negligible, unless very long synthetic antennae are generated, leading to significant variations in target aspect angles within the synthetic aperture. In that case different approaches must be utilized [24].

The following subsections reports a series of examples to point out how different relative positions and velocities contribute to geometric resolutions. The constant gradient hypothesis is valid for those examples and therefore the gradient method is utilized with positions and velocities computed halfway through integration time [22].

1.3.1.1 Example: LEO-LEO Bistatic SAR

A bistatic SAR is considered herein in which LEO orbits are selected for both transmitter and receiver. In this framework, it is reasonable to state that formation flying satellites are utilized. The simplest example of satellite formation is the one in which transmitter and receiver fly along the same quasi-polar orbit with a slight difference in true anomaly, thus establishing a leader-follower formation in our Earth-fixed reference frame. From a theoretical point of view, this case is very interesting, since it can avoid pixel skewing effects even if it is an out-of-plane configuration ($\Phi \neq 0$), as it will be shown in the following. Actually, according to Fig. 1.9a, a leader-follower formation can be considered in which both transmitter and receiver implement a slightly squinted observation, pointing their antennas towards the ground area that is seen at boresight by an equivalent monostatic sensor placed in the midpoint of the baseline.

Indeed, in this case, the bisector of the bistatic vector and the ground projection of fast time gradient are directed along the x-axis. Moreover, as far as the Doppler gradient is concerned, the following relations hold

$$\vec{V}_{Tx} = \vec{V}_{Tx} = \vec{V} \qquad R_{Tx} = R_{Rx} = R$$
$$\left(\vec{V} \cdot \vec{i}_{Tx}\right) = -\left(\vec{V} \cdot \vec{i}_{Rx}\right)$$

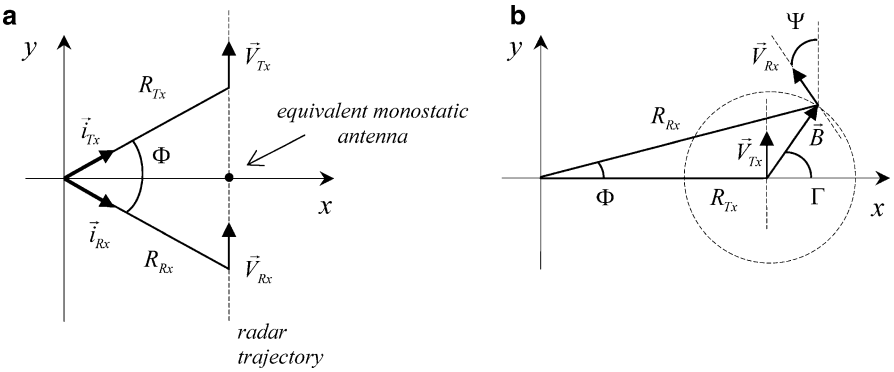


Fig. 1.9 Observation geometry for LEO-LEO bistatic SAR: (**a**) leader-follower configuration, (**b**) generic. Transmitter and receiver operate at the same non-zero z coordinate

that is

$$\overrightarrow{\nabla f} = \frac{1}{\lambda}\left(2\frac{\vec{V}}{R} - \frac{\vec{V}\cdot\vec{i}_{Tx}}{R}\left(\vec{i}_{Tx} - \vec{i}_{Rx}\right)\right) \tag{1.13}$$

since $\vec{i}_{Tx} - \vec{i}_{Rx}$ is in the plane yz it results

$$\vec{i}_{t_g} \perp \vec{i}_{f_g} \quad \Rightarrow \quad \Omega = 90^\circ$$

Ground range and Doppler resolutions, actually, differs from the monostatic case, but their variation is limited to a few percentage points depending on the along-track separation (see Fig. 1.10).

A part from this example of leader-follower formation, in general LEO-LEO bistatic SAR, both transmitter and receiver velocities affect the gradient of Doppler frequency and consequently Doppler resolution and Ω-angle. In order to point out this effect the case illustrated in Fig. 1.9b can be taken as a reference, in which two LEO satellites at the same orbital altitude are separated both in along-track and cross-track coordinates. In addition, the transmitter is assumed to be a side-looking system (actually also a monostatic sensor adopted as a reference), whereas the bistatic receiver is able to gather data from arbitrary directions. As illustrated in Fig. 1.9b, the angle Γ is introduced to simulate different along-track and cross-track baseline components. Specifically, varying Γ from zero to 2π the receiver position is on a circle around the transmitter with radius given by the baseline modulus. In addition, the angle Ψ is introduced to account for differences in flight direction between the transmitter and the receiver. It is worth noting that these geometries require more complicated orbit selections, and, in general, offer more limited bistatic coverage, with respect to the leader-follower scheme.

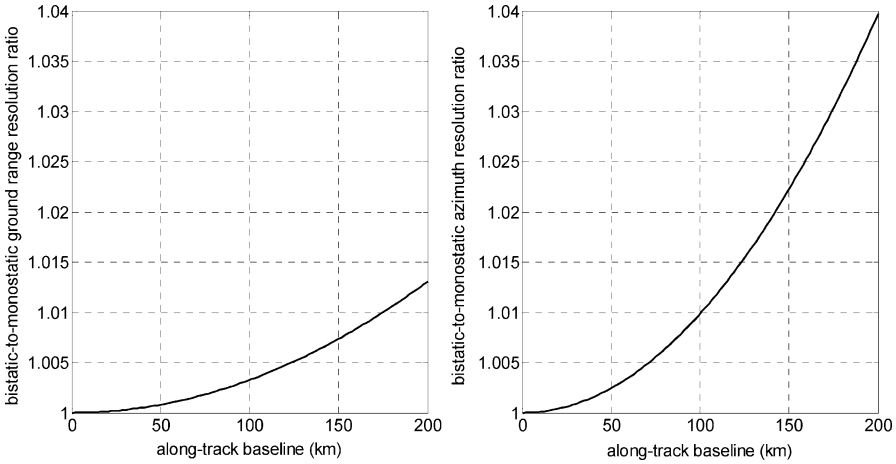


Fig. 1.10 Ratio of bistatic and monostatic ground range (*left*) and (*right*) azimuth resolutions as a function of the baseline for leader follower LEO formations (450 km altitude and 45° incidence angle for the equivalent monostatic sensor)

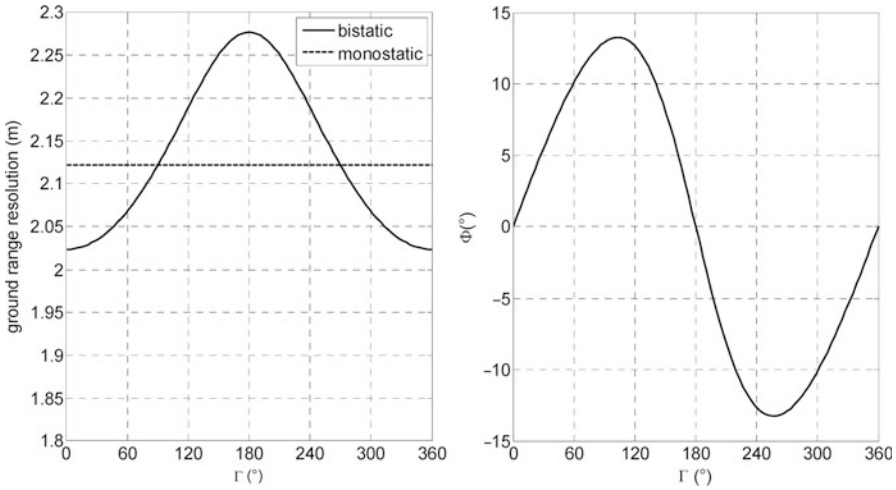


Fig. 1.11 Ground range resolution (*left*) and out-of-plane angle (*right*) for the considered LEO-LEO bistatic SAR (450 km altitude, 45° transmitter incidence angle, 100 km baseline length, 100 MHz chirp bandwidth). Monostatic resolution refers to the observation geometry of the transmitter

Figures 1.11 and 1.12 show examples of ground range and Doppler resolutions, along with the attained values of Ω-angle. Figure 1.11 also shows the corresponding out-of-plane angles. Since baseline is smaller than transmitter slant

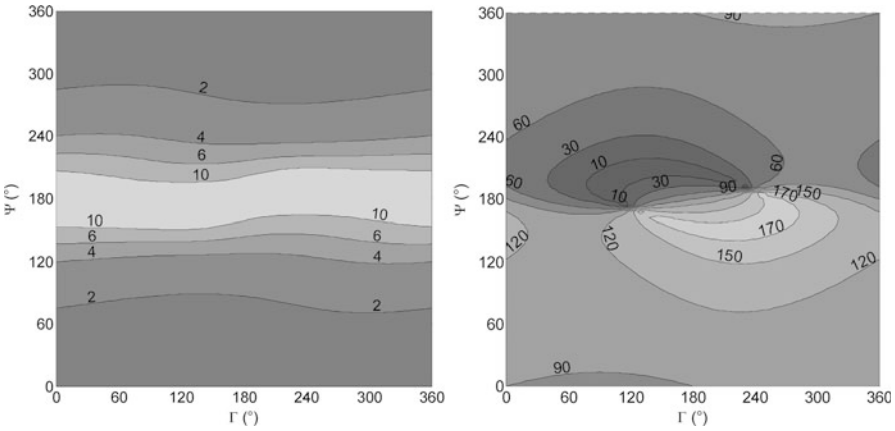


Fig. 1.12 Doppler resolution in meters (*left*) and Ω-angle in degrees (*right*) for the considered LEO-LEO bistatic SAR (450 km altitude, 45° transmitter incidence angle, 100 km baseline length, 1 s coherent integration time)

range, the variation of the out-of-plane angle with respect to Γ is limited, so the bistatic ground range resolution is close to the one achievable by the transmitter when working in monostatic mode.

The behavior of Doppler resolution and Ω-angle, instead, deserves more attention. Independently from the values of Γ, when Ψ approaches 180° Doppler resolution is degraded. Actually in this case receiver velocity is opposite to that of the transmitter and so the available Doppler bandwidth for slow time compression is very limited. Doppler resolution in Fig. 1.12 quickly diverges in the region within the contour lines at 10 m. Similar considerations can be applied to the Ω-angle: critical conditions are met for Ψ close to 180° and Γ close to 90° and 270°, i.e. when large out-of-plane angles (up to 10–15°) are established thus rotating the bisector of the bistatic angle out of the transmitter range elevation plane and closer to Doppler gradient direction.

Finally Figs. 1.13 and 1.14 report examples of iso-range and iso-Doppler contours for bad and good observation geometries, respectively. The reported contours are in fine agreement with the results derived in Figs. 1.11 and 1.12. In particular, Fig. 1.13 definitely shows the importance of adequate bistatic geometry even in presence of excellent resolutions: when iso-range and iso-Doppler lines are almost parallel any increase in chirp bandwidth and coherent integration time has no practical effect on bistatic performance. Moreover, it is worth noting the 1 s integration time, already quite large with respect to conventional spaceborne SAR, is possible assuming steerable antennas. On the other hand, Fig. 1.14 demonstrates the possibility of achieving satisfactory results even for non parallel trajectories, thus showing how formation flying design is strictly related to bistatic SAR.

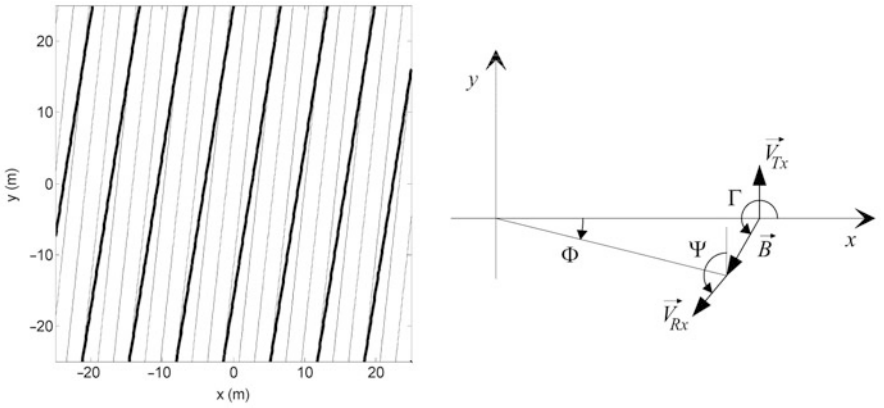


Fig. 1.13 Iso-range (*thin*) and iso-Doppler (*thick*) contours for $\Gamma = 240°$, $\Psi = 150°$ and $\Phi \cong -13°$. In this case $\Omega \cong 177°$ and ground range and Doppler resolution are 2.2 and 6.5 m, respectively

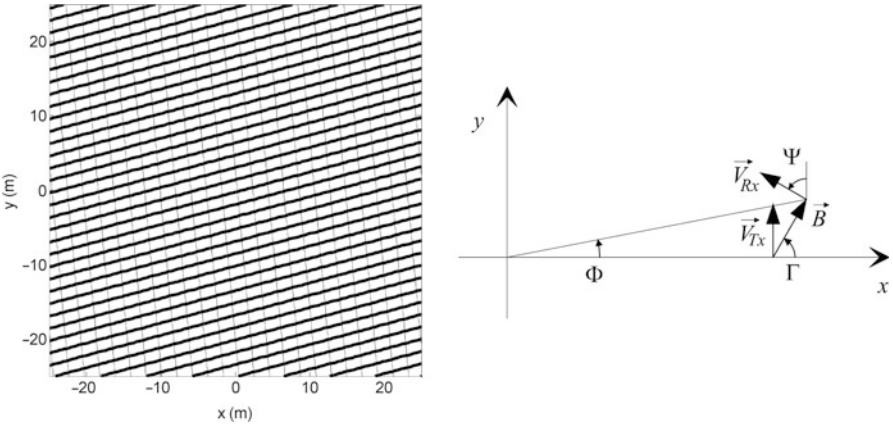


Fig. 1.14 Iso-range (*thin*) and iso-Doppler (*thick*) contours for $\Gamma = 60°$ and $\Psi = 60°$ and $\Phi \cong 10°$. In this case $\Omega \cong 100°$ and ground range and Doppler resolution are 2.1 and 1.6 m, respectively

1.3.1.2 Example: LEO Receiver and Higher Altitude Transmitter

High altitude transmitters in conjunction with a constellation of LEO receivers have been proposed to achieve a systematic reduction of revisit time [13], especially over regional areas. Indeed, high altitude satellites, such as MEO or GEO, can access wide areas easily; at the same time LEO receivers have the twofold advantage of

- Cutting down the required power in transmission, thanks to the reduction of echo propagation path with respect to the case of high altitude monostatic or bistatic SAR;

- Allowing the system to exploit observation geometries different from the conventional monostatic side-looking one, such as the forward- or backward-looking ones [25, 26].

If transmitter altitude is higher than 20,000 km and LEO receivers are considered, it results

$$\frac{\vec{V}_{Rx}}{R_{Rx}} >> \frac{\vec{V}_{Tx}}{R_{Tx}} \tag{1.14}$$

namely the velocity-to-slant-range ratio of the receiver is two orders of magnitude larger than that of the transmitter. In this case transmitter contribution to Doppler gradient can be neglected and geometric resolutions can be derived assuming a quasi-stationary transmitter ($\vec{V}_{Tx} \cong 0$). In this case, gradient method equations can be simplified notably. According to the analysis reported in Ref. [22] the following relations can be utilized for side-looking and forward-looking receivers

$$\Delta a = \begin{cases} \dfrac{\lambda R_{Rx}}{V_{Rx}} \dfrac{1}{T} & side-looking \\ \dfrac{\lambda R_{Rx}}{V_{Rx} \cos^2 \Theta_{Rx}} \dfrac{1}{T} & forward-looking \end{cases} \tag{1.15}$$

$$\Omega = \begin{cases} \pi - \left| \tan^{-1} \left(- \dfrac{\sin \Theta_{Tx} \cos \Phi + \sin \Theta_{Rx}}{\sin \Theta_{Tx} \sin \Phi} \right) \right| & side-looking \\ \pi - \left| \tan^{-1} \left(\dfrac{\sin \Theta_{Tx} \sin \Phi}{\sin \Theta_{Tx} \cos \Phi + \sin \Theta_{Rx}} \right) \right| & forward-looking \end{cases} \tag{1.16}$$

where four-quadrant inverse tangent within values $[-\pi, \pi]$ must be used to compute Ω-angle. Equation 1.15 states that Doppler resolution is degraded in forward-looking mode, with respect to side-looking mode, by a factor $1/\cos^2 \Theta_{Rx}$, that is a rise from 1 to 4 for receiver incidence angles ranging from 0° to 60°.

For the sake of a quantitative analysis of image resolutions and Ω-angle, the orbital geometry proposed in Ref. [13] can be taken as a reference (see Fig. 1.15). It relies on a GEO transmitter covering European latitudes, so 50° as transmitter incidence angle can be assumed. In addition, a constellation of LEO receivers can be designed to collect bistatic data from whatever out-of-plane configuration (i.e. $\Phi \in [0, 360°]$) and incidence angle (i.e. $\Theta_{Rx} \in [0, 60°]$), both in side-looking and in forward-looking mode.

Figure 1.16 shows an example of ground range and Doppler resolutions, while Fig. 1.17 reports the relevant Ω-angle for both side and forward-looking receivers. Due to the symmetry of the problem, the analysis is limited to $\Phi \in [0, 180°]$. As expected ground range resolution (that is the same for either side- or forward-looking receivers) is greatly affected by the out-of-plane angle, with the worst value

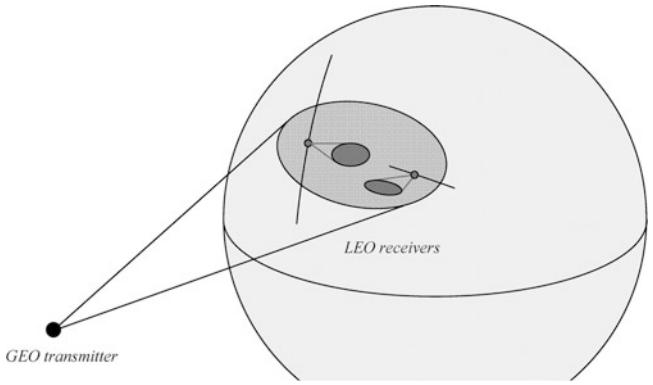


Fig. 1.15 Illustration of GEO-LEO bistatic SAR

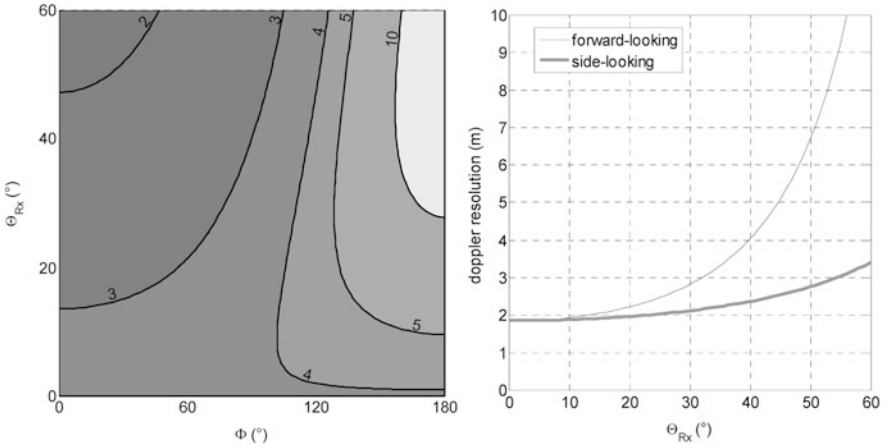


Fig. 1.16 Ground range contours in meters (*left*) and Doppler resolution (*right*) for the considered GEO-LEO bistatic SAR (450 km receiver altitude, 100 MHz chirp bandwidth and 1 s coherent integration time)

attained for opposite-side geometry. With reference to Doppler resolution, assuming a constant integration time, an increase of off-nadir angle determines a decrease of Doppler bandwidth, more relevant in forward-looking geometry.

As far as the Ω-angle is concerned, for forward-looking receivers in-plane observation is not enabled because in this case iso-range and iso-Doppler lines are locally parallel. Indeed, for forward-looking receiver with stationary transmitter the gradient of Doppler frequency is aligned to receiver velocity, but the latter is also parallel to the gradient of time delay on the ground, because receiver velocity is in the common range elevation plane. Better performance is obtained for out-of-plane observations, in which the direction of the gradient of time delay on the ground is out of the receiver range elevation plane. Opposite observations apply to

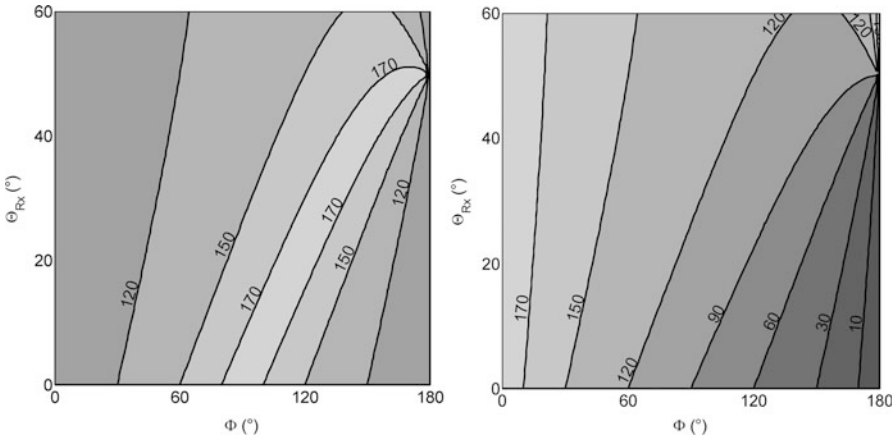


Fig. 1.17 Ω-angle in degrees for the considered GEO-LEO bistatic SAR with side-looking (*left*) and forward-looking (*right*) receivers

side-looking receiver. In this case better values of Ω are, in general, obtained, with the sole exception of strong out-of-plane geometry (Φ close to 90°).

The outlined behavior for a forward-looking receiver is a direct consequence of intrinsically ambiguous geometry. Actually, left/right ambiguity arises from monostatic forward-looking geometry [26] and therefore the ambiguity can be removed in the bistatic case only if the transmitter is able to rotate iso-range and/or iso-Doppler lines [22, 25, 26]. On the contrary, side-looking SAR geometry is unambiguous thus justifying the limited level of resulting bad values for the Ω-angle in Fig. 1.17 (left).

1.3.2 Radiometric Resolution and Signal-to-Noise Ratio

Radiometric resolution can be defined as the minimum brightness contrast necessary for discrimination of two targets. This definition, anyway, cannot be expressed in quantitative form as a function of the main SAR system parameters unless specific SAR imaging models have been introduced.

The received radar signal is typically represented as the coherent addition of all the contributions of the individual scatterers, so each pixel in SAR images can be interpreted as the phasor resulting from the focused returns of the individual scatterers within the resolution cell [4]. This coherent addition from many scatterers gives rise to a phenomenon called speckle [5] which is typical of radar or laser scattering and is caused by surface roughness when comparable to wavelength.

The main effect of speckle phenomenon on SAR imaging is that radar response exhibits a statistical behavior and therefore the amplitude and phase of the focused SAR images have to be interpreted as random variables. This means that, in general,

if homogeneous areas are observed by SAR, the statistical distribution of amplitude and phase will result in intensity image variations, indicated as speckle-noise over the observed area. As a consequence, there is no possibility to characterize radiometric aspects of a SAR target by measurements of one pixel only [27]. According to this statistical approach, radiometric resolution has to account for the possibility to consider two areas, with two different speckles, as different [27]; radiometric resolution can be therefore quantified comparing the mean intensity value (the echo power per unit area), μ, and the closest mean value that can be recognized, in statistical sense, as different, namely $\mu+s$, where s is the standard deviation of the image intensity. As dB scale is generally adopted, radiometric resolution can be finally expresses as

$$\gamma = 10\log\left(1+\frac{s}{\mu}\right) \tag{1.17}$$

From(1.17) it is immediate to understand why multi-look techniques [6] can be utilized to improve radiometric resolution. Actually, if N_L looks are utilized to generate SAR image, more samples are collected thus reducing the standard deviation by a factor $\sqrt{N_L}$, without affecting significantly the mean.

The standard deviation s includes two independent contributions, namely the variation due to speckle and that due to thermal noise [5, 6]

$$\gamma = 10\log\left(1+\frac{s_{speckle}}{\mu}+\frac{s_{noise}}{\mu}\right) \tag{1.18}$$

For the sake of quantifying those ratios, a statistical distribution for pixel amplitude and phase must be introduced. As common in monostatic SAR, for uniformly distributed phase and Rayleigh distributed amplitudes, an exponential distribution [5] can be derived for image intensity, so that (1.17) becomes

$$\gamma = 10\log\left(1+\frac{1}{\sqrt{N_L}}\left(1+\frac{1}{SNR}\right)\right) \tag{1.19}$$

that is the widespread relation utilized to compute radiometric resolution in SAR imaging. As expected, if both SNR and N_L tend to infinity (noiseless data), the radiometric resolution tends to 0 dB. Typical radiometric resolution requirements span from 0.5 to 4 dB depending on the application. Figure 1.18 reports an example of radiometric resolution versus the scattering coefficient for different numbers of looks. If SNR is 0 dB and one look is considered the expected radiometric resolution is 4.77 dB, that means a difference in mean echo intensity of this order is required to discriminate between extended areas. Multi-looking represents a valuable solution to improve radiometric resolution, however a trade-off must be found because with the increasing number of looks geometric resolution degrades accordingly [5, 6].

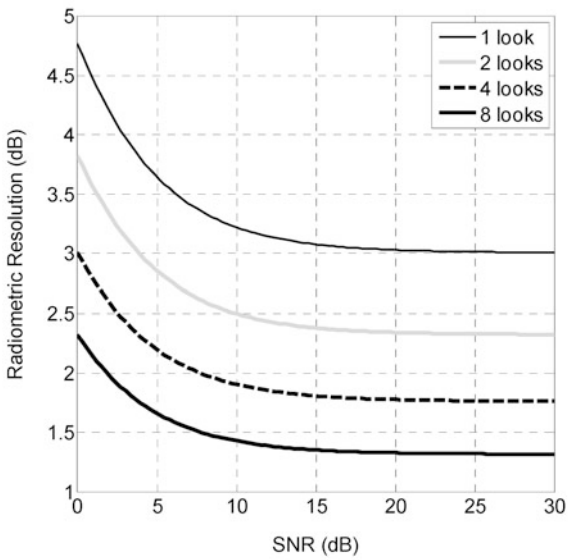


Fig. 1.18 Variation of radiometric resolution with scattering coefficient and number of looks

The actual evaluation of γ involves the determination of the signal-to-noise ratio and therefore the introduction of bistatic SAR equation. To this end different expressions have been proposed [5, 28, 29] to put in evidence more conveniently contributions of various parameters. Slight modifications to the formulations proposed in Refs. [28, 29] lead to the following

$$SNR = \frac{POW_{Tx}\lambda^2 G_{Tx} G_{Rx} \sigma_b^0 T d_c \Delta r_g \Delta a \cos\Phi}{(4\pi)^3 R_{Tx}^2 R_{Rx}^2 k_B T_{noise} F_{noise}} \tag{1.20}$$

with POW_{Tx} transmitted power, G_{Tx} transmitting antenna gain, G_{Rx} receiving antenna gain, σ_b^0 bistatic scattering coefficient, d_c duty cycle, k_B Boltzmann's constant, T_{noise} receiver noise temperature, F_{noise} receiver noise figure.

Most of the introduced parameters are commonly utilized to define monostatic SAR performance. In bistatic SAR, however, some remarks have to be made.

- The behavior of the scattering coefficient in bistatic SAR geometry is more complex than in the monostatic case. An established database for bistatic scattering of both natural scenes and artificial targets is not available at the present day. The main database for ground and sea scattering concerns X-band microwave measurements [30], and was acquired during the 1960s. According to that work, the main findings are that the backscattering coefficient approaches a minimum for strong out-of-plane geometries (i.e. Φ close to 90° or 270°) with values 10–20 dB below the monostatic value. In addition, out-of-plane values are not significantly different from in-plane ones for out-of-plane angles within 10° from the opposite side direction and within 40° from the same-side one, that is angles close to in-plane observation.

- Duty cycle is defined as the ratio between the length of the transmitter pulse and the pulse repetition interval. In monostatic SAR, with co-located transmitters and receivers, quite limited duty cycles (lower than 0.2) must be usually adopted to avoid ambiguities. In bistatic SAR, instead, no limitation is, in principle, set to the duty cycle: large values can be therefore exploited, including the case of continuous transmission and reception ($d_c = 1$) [28].
- A few parameters in (1.20) strictly depend on the transmitter, specifically POW_{Tx}, R_{Tx}, and G_{Tx}; their contribution in radar equation represents the power flux density near Earth surface

$$\rho_{POW} = \frac{POW_{Tx} G_{Tx}}{(4\pi)^3 R_{Tx}^2} \tag{1.21}$$

  and it is an important issue particularly when illuminators of opportunity are considered, as in Ref. [28]. Indeed, ρ_{POW} can be used to make a comparison among different available illuminators.

1.3.3 Timing, Pointing and Synchronization

The availability of accurate time reference is essential in SAR imaging. When monostatic observation is implemented ultra-stable local oscillators are used to define radar time unit. This unit sets references for pulse duration, modulation and demodulation, and controls the alternation of transmitting and receiving modes.

In bistatic SAR two separated systems are utilized for transmission and reception so some kind of synchronization must be used. In general, synchronization problem can be tackled at different levels because different types of synchronization may be necessary depending on the considered bistatic SAR geometry and configuration. The problem can be schematized in three different tasks: *footprint* matching, *windowing* design and *time* and phase (or frequency) *referencing*.

1.3.3.1 Footprint Matching

Footprint matching deals with the problem of illuminating the target area by the transmitter and of pointing the main lobe of the receiving antenna toward the same ground area. The pointing accuracy requirements and the relevant performance strictly depend on the intended implementation. Specifically, if LEO-LEO bistatic SAR is considered, the problem can be more challenging because transmitter and receiver orbit with similar velocity and also footprint dimensions with similar order of magnitude are expected. A strategy is presented in Ref. [31] to guarantee footprint matching for LEO pendulum formations. The strategy involves the

coordinate use of electronically steerable SAR antennas and attitude maneuvers to superimpose transmitter and receiver footprints at different observed latitudes along satellites orbit. In order to avoid residual gaps depending on attitude errors receiver footprint was assumed slightly larger than the illuminated area.

In the case of LEO receivers and higher altitude transmitters, instead, the footprint matching problem can take advantage of the altitude difference between transmitter and receiver. Transmitter is slower and generally designed to cover large areas, receiver pointing requirements therefore resemble the monostatic case.

1.3.3.2 Windowing Design

Windowing design is in charge of assuring that the opening of the receiving channel follows the pulse repetition frequency (PRF) of the transmitter. PRF selection in bistatic SAR is conceptually similar to the monostatic case, that is the optimal PRF has to be computed accounting for proper sampling of Doppler bandwidth and avoiding SAR ambiguities [4–6]. Within the framework of the start-and-stop approximation, PRF selection is primarily dictated by the need to avoid range and frequency aliasing in the range-Doppler image. Range aliasing occurs when the pulse repetition frequency is so high that several pulses illuminate different parts of the scene at the same time: this condition establish the maximum value of PRF. Moreover, according to the Nyquist's sampling criterion, a lower bound on the pulse repetition frequency is established by the Doppler bandwidth of the illuminated area. In bistatic SAR the Doppler bandwidth, to a first order, is

$$B_D(\vec{r}) = \frac{1}{\lambda}\left[\vec{V}_{Tx}\cdot\left(\frac{\vec{R}_{Tx}\left(\frac{T}{2}\right)}{\left|\vec{R}_{Tx}\left(\frac{T}{2}\right)\right|}-\frac{\vec{R}_{Tx}\left(-\frac{T}{2}\right)}{\left|\vec{R}_{Tx}\left(-\frac{T}{2}\right)\right|}\right)+\vec{V}_{Rx}\cdot\left(\frac{\vec{R}_{Rx}\left(\frac{T}{2}\right)}{\left|\vec{R}_{Rx}\left(\frac{T}{2}\right)\right|}-\frac{\vec{R}_{Rx}\left(-\frac{T}{2}\right)}{\left|\vec{R}_{Rx}\left(-\frac{T}{2}\right)\right|}\right)\right] \tag{1.21}$$

that is the Doppler bandwidth is determined by the first and last positions from which the target is seen by both radar antennas. Bistatic SAR scenario, however, features two main differences with respect to the monostatic counterpart:

- Transmitter and receiver are physically separated, so, if convenient, the system can operate in CW mode, that is the transmitted frequency modulated signal is continuous and the receiver samples the scattered signal continuously. In such a mode the concept of PRF is lost. However, range and Doppler focusing have to be performed for SAR imaging. As range focusing is the cross-correlation of the scattered echo with a copy of the transmitted signal, the CW bistatic SAR can be assimilated to a usual pulsed SAR system by introducing a stop length [28, 32]. This length is the time duration of the window used for range cross-correlation. In such way, even if CW mode is considered, the equivalent PRF can be introduced as the inverse of the stop length. Range compressed data from subsequent time windows can be then focused in azimuth by conventional

SAR processing. The time window has to be set: (a) allowing the system to be considered working in start stop fashion; (b) achieving a non-ambiguous equivalent PRF (i.e. equivalent PRF higher than processed Doppler bandwidth). Finally it is worth noting that a bistatic SAR can also operate in a partially CW mode, that is the transmitter is a pulsed radar and the receiver performs continuous data acquisition. Obviously, in this case, transmitting PRF has to be tuned to the Doppler bandwidth of bistatic data.

- SAR image ambiguities are conventionally interpreted in range and azimuth domains [6], that is ghosts in SAR images are generated in range or azimuth coordinates when, respectively, echoes from proceeding and succeeding pulses arrives back at the antenna simultaneously with the desired one or Doppler spectrum is contaminated by ambiguous signals from antenna azimuth sidelobes. In bistatic SAR range and azimuth ambiguities survive both in fast and slow time domains, and the location of the relevant ghost targets in the final image depends on the relative geometry among transmitter, target, and receiver, as observed in Sect. 1.3.1 when directions of the gradient of time delay and Doppler frequency were introduced.

1.3.3.3 Time Referencing

The definition of a common time reference between transmitter and receiver affects bistatic SAR performance at different levels, although the most stringent requirements on time synchronization are set by the need for accurate slant range measurements. A timing accuracy, Δt, of the order of a fraction of the compressed pulse width is a typical requirement [30]. If this constraint cannot be satisfied by the ultra-stable clock embarked on the receiver, a reference signal that periodically matches receiver clock with transmitter one must be provided. The reference signal can be generated either by the transmitter or by external sources (such as GPS) depending on the required accuracy. For this type of time synchronization the receiver clock stability, q, can be expressed as [30]

$$q = \frac{\Delta t}{2T_{up}} \tag{1.23}$$

where T_{up} is the adopted clock update interval.

Phase (or frequency) synchronization issues set additional, and typically tighter, constrains on time referencing accuracy and oscillator stability. In bistatic SAR, indeed, low frequency phase errors do not cancel out when the received signal is demodulated because the signal had been previously modulated by the transmitter with a different oscillator characterized, in general, by different low frequency phase errors [24, 33]. It is worth noting that those errors do not affect monostatic SAR that uses the same oscillator for either modulation or demodulation.

Low frequency phase errors can have strong impacts on bistatic SAR imaging, reducing image quality notably. Indeed, as showed in Refs. [24, 33]:

- A constant oscillator frequency offset (linear phase errors) between transmitter and receiver leads to a displacement both in range and in azimuth of the peak of the impulse response;
- A linear oscillator frequency drift (quadratic phase errors) generates a widening and a degradation of impulse response in Doppler direction. If a reference signal is exploited by receiver oscillator, phase stability is required over the coherent integration time, T, and the required phase error, $\Delta\phi$, has to be lower than 90° [30], thus leading to the following required stability for receiver oscillator

$$q = \frac{\lambda \Delta\phi}{2\pi c T} \tag{1.24}$$

- Random oscillator phase noise degrades all the image parameters, especially integrated side-lobe ratio (ISLR).

Special conditions exist in which phase synchronization requirements can be relaxed. Indeed, as experiences in Ref. [34] if the duration of the bistatic acquisition is short (few seconds at most) and high quality crystal oscillators are used, only a constant offset of carrier in carrier frequency has to be expected. The offset can be estimated from bistatic raw data and then removed.

1.4 From Ideas to Proof of Concepts

Non-systematic bistatic measurements from spacecraft have been conducted so far, in particular considering synthetic aperture radar. Actually bistatic observation has been mainly applied by using one or both Earth-fixed antennae and with reference to targets of limited dimensions. This situation can be explained considering difficulties in bistatic SAR previously outlined. Furthermore, even if very promising applications of bistatic SAR are known from a long time, only the basic principles of such applications have been investigated and achievable performance have been characterized only qualitatively. As a result it has been difficult to assess possible roles of bistatic SAR within the framework of satellite remote sensing, and to define and develop operative bistatic systems. The transition from those *ideas* to *proof of concepts*, anyway, is being currently carried out.

This section illustrates some stages in the development of bistatic systems, from the beginning theoretical speculations and the first experimentations to the most recent and important results, with more emphasis on space based investigations and scientific applications.

1.4.1 Bistatic Radar

Atmospheric studies by Earth-based bistatic radar measurements represented one of the first area of research [35] together with oceanographic applications [36]. Specifically, the observability of clear air winds by Bragg bistatic scattering from electromagnetic refractive turbulence index at short wavelength was put in evidence in Ref. [37], whereas the peculiarities of bistatic radar equation for detecting meteorological targets, such as raindrops and refractivity perturbations, were analyzed in Ref. [38]. In addition, the results of Doppler frequency maps from Loran-A transmissions and a ground-based receiver have been presented in Ref. [36], showing the possibility of tracking anisotropies of ocean-wave spectra and therefore identifying sea wave lengths and directions; the experiment also represented one of the first examples of adoption of illuminators of opportunity. More recently, networks of bistatic dual- and multiple Doppler weather radars have been proposed for obtaining vector winds [39].

An example of fully spaceborne bistatic radar experiment was presented in Ref. [40]. The L-band telecommunication link of Mir orbital station was used as transmitter and the reflected signals from the Earth were received from a geosynchronous satellite. This experiment represented the first example of bistatic radiolocation of the Earth from space and the collected data were also exploited for atmospheric investigations, such as the estimation of absorption coefficients due to atmospheric oxygen at L-band.

Planetary studies have also taken advantage of bistatic radar measurements. Different configurations were implemented, e.g. satellite-based transmitter and Earth-based receiver [41–44] or planet-based transmitter and satellite-based receiver [45]. Proposed and partially experimented planetary applications include:

- Topography, reflectivity, surface scattering, dielectric constant, rms surface slopes, surface texture and density on Moon, Mars and Venus;
- Remote probing in regions and under conditions not obtainable with Earth-based systems.

Renewed impulse to the bistatic research and the growth of bistatic techniques was given around year 2000 by the diffuse utilization of illuminators of opportunity [2] and specifically by the application of this concept to Global Navigation Satellite System (GNSS) constellations. Actually, thanks to the intrinsic characteristics of GNSSs some well-known limitations of illuminators of opportunity can be overcome. Specifically, continuous (both in space and time) transmissions can be utilized, with well-known and stable characteristics, accurate knowledge of illuminator position and velocity, and signal modulation and coding adequate for range-Doppler processing. Available signal power, non-selectable transmitter carrier frequency, bandwidth and polarization, however, still remain important issues that can have an impact on resolution, signal to noise ratio and signal processing, thus affecting potential applications. Investigated and experimented applications deal with the measurement of GPS reflected signals for estimation of sea-state and

wind speed [46–48], for rough surface characterization [49, 50], and moving target detection, such as aircrafts approaching a runway [51]. Most of the experimentations relied on ground-based receivers, mounted, for instance, on the edge of a bridge [48] or on instrumented towers [50]. However, since the first demonstration of the possibility of gathering Earth-reflected GPS signals from LEO orbits [52], the interest on spaceborne bistatic LEO receivers for large scale observation has been increasing. A few years later, indeed, the first experimental payload flew onboard UK-DMS satellite [53], showing the possibility of processing C/A codes of GPS reflected signals from oceans to obtain delay Doppler distributions. Finally, a bistatic orbital demonstrator was developed by the European Space Agency (ESA) [54], able to use the full bandwidth of GNSS reflected signals for high resolution ocean altimetry as primary objective, but potentially also for scatterometry, ice altimetry, soil moisture and biomass evaluation.

1.4.2 Bistatic Synthetic Aperture Radar

The research conducted by Stanford University and Scripps Institution of Oceanography in the mid-1970s led to the first documented experiments of synthetic aperture radar in bistatic geometry [55, 56]. Two different bistatic configurations were implemented. The first experiment [55] utilized LORAN-A 1.95 MHz navigation facility as transmitting source, whereas the receiver was mounted in a step van which was driven at uniform velocity along available aircraft taxiways, ramps and roads. Radar signals backscattered from sea surface were coherently processed. Due to the extremely long wavelength of LORAN-A, very long synthetic aperture were realized, ranging from 0.8 to 2.7 km, to achieve satisfactory angular resolution. In addition sequential observations on nearly orthogonal headings, separated only a few minutes in time, were used to provide multidirectional high resolution measures of the ocean wave spectrum. The second experimental campaign, instead, was carried out using a ship-borne bistatic radar [56]. The transmitting antenna was suspended from the ship's mast with the receiving one mounted near the bow about 25 m from the transmitting antenna. Both antennas pointed ocean surface. The motion of the ship was used to synthesize an approximately 350 m long aperture and measurements were made at four frequencies ranging from 4 to 21 MHz.

About 10 years later the first airborne bistatic SAR surveys were conducted by Environmental Research Institute of Michigan [57]. Two airborne X-band SARs flying with programmed separations along parallel trajectories were used to implement bistatic observations at three bistatic angles (2°, 40°, and 80°). In particular the collected data showed a reduction of about 10–20 dB in the dynamic range of the bistatic images with respect to the monostatic ones, due to reduced retroreflector effects in build-up areas. These experiments clearly put in evidence some basic issues of bistatic SAR implementations, such as time and phase synchronization or antenna footprint matching, but also bistatic SAR focusing approaches and performance.

Table 1.1 Main bistatic SAR experiments conducted from year 2000

Configuration	Institution	Country
Airborne illuminator/ground-based receiver	FGAN [60]	Germany
	DRDC [61]	Canada
Airborne illuminator/airborne receiver	QinetiQ [62]	Great Britain
	DLR/ONERA [63]	Germany
	FGAN [64]	Germany
Spaceborne illuminator/ground-based receiver	Politecnico di Milano [32]	Italy
	Univ. of Cataluña [65]	Spain
	Univ. of Birmingham [66]	UK
	Univ. College London [67, 68]	UK
LEO illuminator/airborne receiver	DLR [34]	Germany
	FGAN [25, 69]	Germany
Moon-orbiting illuminator and receiver	NASA/ISRO [70–72]	USA/India

In the first 1990s a more challenging bistatic SAR configuration was successfully implemented by NASA JPL [58]. Bistatic SAR images were generated using C-Band ERS-1 satellite as illuminator and NASA DC-8 aircraft as receiver. Due to the extreme differences in transmitter and receiver velocities, bistatic data collection occurred in a brief burst of signals as the spacecraft overflew the aircraft. Even if only 3 s data acquisition was performed, it was observed that this time span was too long for the two frequency standards to remain coherent, and therefore the chirps directly received from the satellite, recorded along with their echoes, were used as coherent reference for bistatic SAR processing. Similar experiments were also conducted using SIR-C as transmitter [59].

During the last decade research in bistatic SAR has been intensified and important experimental campaigns were conducted in several countries by means of different realizations, namely the transmitter was spaceborne or airborne and the receiver airborne or ground-based. Table 1.1 lists the main bistatic SAR experiments carried-out from year 2000.

With reference to configurations with airborne illuminator and fixed receiver, two experiments can be mentioned [60, 61]. In May 2004, a bistatic field test was performed by FGAN (Forschungsgesellschaft für Angewandte Naturwissenschaften, Research Establishment for Applied Science) [60]: MEMPHIS radar, operating in mm-wave range and installed on a Transall C-160, was utilized as transmitter, and the receiver unit was mounted on the roof-deck of a building nearby the scene. Same side geometry was implemented, with both side-looking and squinted transmitter (0–40° squint angle). The scene prepared for the field test included different types of targets: a fence, four spheres, one corner reflector and two military transport vehicles. Thanks to the 200 MHz bandwidth transmitted signal, 75 cm range resolution was achieved. Furthermore, a decrease of the cross-range resolution caused by the decreasing effective synthetic aperture was also verified.

A second experiment with airborne illuminator and fixed receiver was conducted by DRDC (Defense Research and Development Canada) in spring 2006 [61].

The research activity aimed at demonstrating the feasibility of a long-range (>100 km) airborne experimental bistatic SAR for moderate-to-high resolutions (<1 m). The experimentation utilized the CONVAIR 580 C/X-SAR system as transmitter and an ad-hoc developed compact two-channel, wide-band X/C band module as receiver. Rather than using a direct link between transmitter and receiver to establish phase coherence, DRDC used highly stable "free-running" crystal oscillators, in order to keep the system simple and to reduce the risk of interception. DRDC devised an experimental setup to discipline the oscillators to a three-axis set of accelerometers. The output of the accelerometers was utilized to apply an accurate correction voltage to the oscillator inputs, which are inversely proportional to the frequency offset induced by the aircraft motion. Results have shown the possibility to improve phase noise of more than 20 dB.

A fully airborne, synchronized bistatic SAR demonstration was carried out by QuinetiQ in September 2002 [62]. The experiment made use of a high transmit power radar, onboard a BAC1-11, together with a passive X-band receiver mounted on a helicopter. Both platforms operated in spotlight mode, imaging an urban environment surrounded by natural clutter regions. Synchronization was achieved using a pair of cesium atomic clocks. Data were recorded at a number of different imaging geometries, with the aim of investigating the effect of varying bistatic angle, grazing angle and platform velocity on bistatic imagery. Both mono and bistatic data were acquired thus highlighting differences in the scattering mechanisms in the two images. For instance, acquisitions at 70° bistatic angle showed that retro-reflector effects in man-made environment is more relevant in monostatic images. Moreover different building structures appear bright in the two images: in the bistatic case the sides of the buildings are more visible, whereas in the monostatic one roofs appear brighter.

DLR (Deutschland für Luft- und Raumfahrt, German Aerospace Center) and ONERA (Office National d'Etudes et de Recherches Aéronautiques, French National Aerospace Research Center) organized some joint bistatic airborne experiments [63] involving their radar systems E-SAR and RAMSES, respectively. These experiments took place between October 2002 and February 2003. Two main geometrical configurations were flown. In the first geometrical configuration, an along-track baseline of 30 m was established to investigate phase synchronization issues and single pass interferometry. The second geometrical configuration was designed to acquire images with large bistatic angles. The two planes flew on parallel tracks around 2 km apart, at about the same altitude, establishing bistatic angles up to 20°. No communication link was used for synchronization. A two steps solution was, however, adopted to guarantee accurate timing. First, before take-off, the two oscillators were connected and their frequencies matched, then just before each acquisition, the two PRFs were synchronized to the PPS (Pulse-Per-Second) signal coming from the GPS. This ensured window synchronization. Finally, residual discrepancies between the frequencies of the two oscillators were corrected in the data processing chain. In all acquisitions, both systems were used also in reception so that a monostatic image was always acquired for each bistatic image. Moreover several data takes were acquired for each geometrical configuration, with

varying system parameters, e.g. bandwidth of 50 MHz or 100 MHz, RAMSES transmitting or E-SAR transmitting. Images acquired with large bistatic angles put in evidence the possibility to use bistatic images to characterize natural surfaces and discriminate between different types of land cover.

In November 2003, also FGAN undertook a fully airborne bistatic flight campaign. The transmitting sensor, AER-II, was placed on a Dornier Do-228 and the receiving sensor, PAMIR, on a Transall C-160 [64]. Both SAR sensors operate at X-band with 300 MHz bandwidth. Parallel track bistatic geometries were implemented with tight overlap control of antenna footprints. Several acquisitions were carried out with bistatic angles ranging from 13° to 76° to explore the influence of the bistatic angle on the collected data. No synchronization link was utilized. The finest ground range resolution was approximately 0.74 m and the Doppler resolution on the ground was approximately 0.35 m.

As far as bistatic configurations with spaceborne transmitters and ground-based receivers are concerned, a series of bistatic SAR images were generated by Politecnico di Milano using a single digital TV channel, 20 MHz bandwidth, of the Hotbird-4 geostationary satellite and a fixed ground-based receiver mounted on a small tower [32]. The synthetic aperture was obtained by exploiting the small daily motion of the satellite. Range focusing was carried out using the direct signal received by a commercial parabolic antenna as reference. Achieved range and azimuth resolutions were of the order of 10 m on an area of about 200 $\times$ 200m.

A bistatic SAR receiver prototype was developed by University of Cataluña and tested from September 2005 to January 2006 using Envisat and ERS-2 satellites as non cooperative transmitters [65]. The receiver was located on the roof of the Signal Theory and Communications Department of the University of Cataluña. The experimentation was intended to deploy and synchronize several bistatic fixed receiver in order to perform single-pass interferometry. The prototype utilized the direct signal to carry-out PRF extraction and chirp and Doppler rates estimation. The first bistatic images were produced in December 2005 by sub-aperture processing with integrated topography compensations.

University of Birmingham has been carrying out active research in the area of Space-Surface Bistatic SAR (SS-BSAR) since 2003. The main aim of this research is to experimentally demonstrate the feasibility and performance of SS-BSAR, utilizing GNSS as illuminator of opportunity [66]. The experimental setup includes a moving platform (on a 27 m long track) with a heterodyne antenna pointed towards the satellite and a radar antenna pointed towards the observed ground area. The heterodyne antenna collects the direct GNSS signal that is used for reflected signal synchronization during bistatic SAR image generation. Experimental bistatic SAR images were generated using different satellites from GLONASS constellation (5 MHz signal bandwidth) and different bistatic angles. Range resolution of 30 m was therefore achieved along with 1.5 m cross-range resolution for integration times up to 45 s. In addition, a synchronization scheme based on GNSS broadcasted data has been validated.

Also University College of London (UCL) has been conducting bistatic radar and SAR experiments with spaceborne transmitters and ground-based receivers

since 2002 [67, 68]. Specifically, several images were generated using Envisat as illuminator of opportunity and a dual-channel receiver. Further experiments are planned with multi-channel receivers, allowing for the implementation Displaced Phase Center Antenna and Space-Time Adaptive Processing techniques to detect moving targets.

A bistatic X-band experiment was performed by DLR in November 2007. TerraSAR-X was used as transmitter and DLR's F-SAR was used as bistatic receiver [34]. Trajectories of both platforms were designed to be nearly parallel to maximize footprint overlapping and to avoid poor range resolution configurations. Transmitting antenna was steered in azimuth, performing a spotlight illumination of the scene, while F-SAR was receiving in regular stripmap mode. Continuous bistatic data acquisition was preferred to overcome echo window synchronization problems. The nominal mid-beam incidence angle was about 55° for the transmitter and 45° for the receiver. Selected transmitted chirp bandwidth was 100 MHz. A high PRF was chosen to guarantee high along-track oversampling rate of the bistatic data. The ground range dimension of the imaged scene was about 2,500 m whereas along-track size of the scene varied between 600 and 900 m for near and far range, respectively. Three X-band transponders were used as reference targets on ground to quantify focused image quality and to precisely synchronize bistatic data. Footprint overlapping time was limited by standard duration of TerraSAR-X high-resolution spotlight acquisitions: bistatic data therefore were recorded during approximately 3.5 s. The achieved ground range resolution varied between 2.2 and 1.7 m for near and far range, respectively, whereas Doppler resolution on the ground yielded values between 0.25 and 0.39 m for 2.77 s integration time. Peculiar effects of bistatic acquisitions were observed such as different brightness distribution between monostatic and bistatic images and skewed point target response from the transponders, with different skewing angles due to the different equivalent squint angles.

Various spaceborne-airborne bistatic SAR experiments were conducted by FGAN using TerraSAR-X satellite as transmitter and the airborne PAMIR system as receiver [69]. The first experiment was conducted in July 2008. The implemented bistatic configuration was similar to the previous DLR experiment, i.e. parallel-tracks with TerraSAR-X working in spotlight mode and PAMIR operating in stripmap mode. Some differences between those experiments, anyway, must be put in evidence. First of all, a different geometry was exploited with 47° transmitter incidence angle and 70° receiver incidence angle thus attaining about 23° bistatic angle at the scene centre. In addition, the 150 MHz TerraSAR-X bandwidth was utilized. Finally, an auxiliary antenna was mounted on the top of aircraft fuselage to gather the direct signal from the transmitter. The direct signal was used to set the pulse repetition frequency of the receiver, but also to determine the residual unknown frequency offset between transmitter and receiver oscillators. A second experiment was carried out in the beginning of year 2009. Again parallel tracks were selected but in this case also receiver antenna beam was steered during synthetic aperture formation and the so-called "double sliding spotlight mode" was implemented to widen azimuth scene extent. The TerraSAR-X 300 MHz transmitting

bandwidth was also utilized. From the focused images a good agreement between the theoretical geometric resolutions and the achieved ones was verified. Finally, during the bistatic campaigns also monostatic data were collected by either TerraSar-X and PAMIR. From the comparison between focused monostatic and bistatic images noticeable differences were observed in terms of scattering behavior but also of geometric effects such as shadowing, foreshortening and layover.

In the framework of the same research activity an additional spaceborne-airborne experiment was conducted by FGAN in November 2009 [25]. Specifically, spotlight TerraSAR-X transmission was exploited and a backward-looking antenna, mounted on the loading ramp of PAMIR aircraft, was used to gather bistatic echoes. As a matter of fact imaging conditions in forward- or backward-looking directions are the same, so the experiment also demonstrated the feasibility of bistatic SAR with forward-looking receivers. Specifically, the experiment clearly put in evidence the effects of pixel skewing, illustrated in Sect. 1.3.1, due to the peculiar observation geometry in which the platforms did not move along parallel tracks.

In August 2009, NASA's Lunar Reconnaissance Orbiter (LRO) and Indian Space Research Organization's Chandrayaan-1 spacecraft attempted a bistatic SAR experiment to demonstrate the presence of ice deposits in permanently shadowed craters of the Lunar South Pole [70]. Both Chandrayaan-1 and LRO embarked a compact S-band SAR. The systems were designed to operate in bistatic mode with either Chandrayaan-1 or LRO transmitting and receiving. The main goal of the experiment was to measure Circular Polarization Ratio (CPR) of lunar polar regions from about 5° bistatic angle. Indeed, high CPR signatures had been observed in Earth-based Arecibo monostatic measurements, indicating the possible presence of water ice. Different phenomena, however, can explain the observed high CPR in monostatic geometry, such as increased blockiness of lunar regolith [71]. Bistatic SAR measurements could be used to separate the effects of surface roughness from those of ice. Unfortunately, the experiment encountered technical difficulties [72] and the Chandrayaan-1 mission ended before another attempt could be initiated.

1.4.3 Proposed Spaceborne Missions

During last decades several examples of spaceborne bistatic mission have been proposed. Most of them characterized by spaceborne illuminator and airborne or ground-based receivers but some examples of fully spaceborne missions also exist.

In the mid of 1980s a feasibility study was presented [73], describing a bistatic system for surveillance, essentially devoted to tactical scenarios. The proposed solution relied on a constellation of L-band MEO satellites (about 10,000 km altitude) coupled with airborne and ground-based receivers. The constellation was developed to achieve continuous global coverage whereas bistatic receivers had to perform conventional search and track radar functions, using SAR imaging and

MTI techniques too. Some peculiarities of bistatic geometry were further put in evidence in Refs. [74, 75] where a constellation of X-band MEO illuminators (about 7,000 km altitude) was proposed for worldwide surveillance by airborne bistatic receivers. The orbital altitude was selected to match, approximately, the velocity-to-range ratio of the transmitter to that of the receiver. In this way a proper control over Doppler spectrum spreading due to ground clutter returns could be achieved, thus lowering, notably, the minimum detectable velocity with respect to monostatic systems, when working in MTI mode. Different solutions exploiting geosynchronous transmitters [76–78] were also proposed to reduce the number of required illuminators in surveillance and moving target detection applications. In addition, the massive exploitation of non-cooperative LEO communication satellites together with Earth-based receivers for vessel and target detection [79] can be interpreted as a different approach to reduce the impact of the space segment on the complexities and costs of bistatic systems.

More recently, new fully spaceborne bistatic mission concepts have been proposed relying on both radar and SAR system.

As far as scientific bistatic radar missions are concerned, fully spaceborne systems have been proposed for oceanography by bistatic radar altimeters [80–82]. Specifically, a LEO constellation was presented in Ref. [80] able to reduce the revisit time with respect to a constellation of monostatic altimeters. In Ref. [82] similar ideas have been presented, where the reduction of the revisit time is achieved collecting both monostatic and bistatic measurements by transmitting and receiving systems equipped with multi-beam antennas.

Finally, several scientific mission concepts have been developed in the last decade to realize bistatic SAR observations based on different LEO satellite formations. In this framework it is worth mentioning BISSAT/SABRINA mission study [20, 83], Cartwheel concept [84], Radarsat 2/3 program [85], and TanDEM-X mission [9, 10]. Similarities and differences exist between these missions. As part of mission objectives all the mission examples aimed at gathering bistatic SAR images from close inter-satellite separations to apply along-track or cross-track interferometric techniques (see Sect. 1.5.2.1). Cartwheel and Radarsat 2/3 were based on dedicated satellite systems, whereas BISSAT/SABRINA and TanDEM-X were conceived to complement a preexisting monostatic SAR mission by an additional satellite to realize a spaceborne monostatic/bistatic SAR system. Mission phases with very large baselines, up to hundreds of kilometers, had been envisaged in the framework of BISSAT/SABRINA mission study. These acquisitions, able to produce couples of monostatic/bistatic images characterized by large bistatic angles (up to 45°), were proposed to experiment innovative bistatic SAR techniques other than interferometric ones (see Sect. 1.5.2.2). Large baseline bistatic SAR acquisitions, even though with reduced baselines, are currently planned in the last part of TanDEM-X mission. Further details about the most remarkable examples of bistatic SAR spaceborne mission studies and concepts can be found in Part 4.

1.5 Bistatic SAR Techniques and Applications

Possible applications of bistatic SAR depend on the number and type of data simultaneously collected from a given target area. The following subsections deal with the cases of images generated from a single receiving platform and from multiple receiving ones, and illustrate the main applications. For each case an example of promising technique is selected and the relevant applications are analyzed in detail.

1.5.1 Images from a Single Receiving Platform

In the simplest bistatic SAR configuration only images from a single receiving platform can be produced. Nevertheless interesting scientific and practical applications can be accomplished. Indeed, bistatic SAR scattering mechanisms [86, 87] show many aspects that can be utilized to enhance imaging performance. For instance, thanks to the reduced retro-reflector effects in bistatic SAR geometry, a reduction of about 10–20 dB in image dynamic range has been experienced in airborne X-band surveys at three different bistatic angles [57]. Since signals of strong retro-reflectors are reduced, weak signals can be made more prominent by an appropriate dynamic range at the receiver. Consequently, more details can be detected in a bistatic SAR image than in a monostatic one: this peculiarity can be used in bistatic observations aimed at the assessment of the extent or the monitoring of the growth of urban areas, allowing for better biomass evaluation [8]. Furthermore, bistatic SAR mapping potentiality for several land applications can be emphasized by using polarization modes. In this context, target detection in a natural media, such as the maritime environment, can be improved by using matched polarization states for which the clutter return is low [87].

The reported two examples of bistatic SAR applications have been validated by analysis and experiments. However, as already noted, there are no assessed databases for bistatic scattering. This means that the number and type of bistatic SAR techniques and applications could potentially be wider. In this context a bistatic system, working in the single bistatic receiver configuration can be designed to collect data from different observation geometries and polarizations thus measuring and characterizing different, and innovative, aspects of bistatic SAR scattering mechanisms.

Oceanographic applications also take advantage of bistatic SAR implementations. Bistatic data, as conventional monostatic SAR images, suffer from important limitations when sea surface is observed. Specifically, Doppler resolution is usually degraded in monostatic SAR by the sea wave motion during the formation of the synthetic aperture. In addition, sea wave motion causes a shift between scattering element real position and imaged one. Bistatic SAR features, able to reduce those limitations, are reported in Sect. 1.5.1.1.

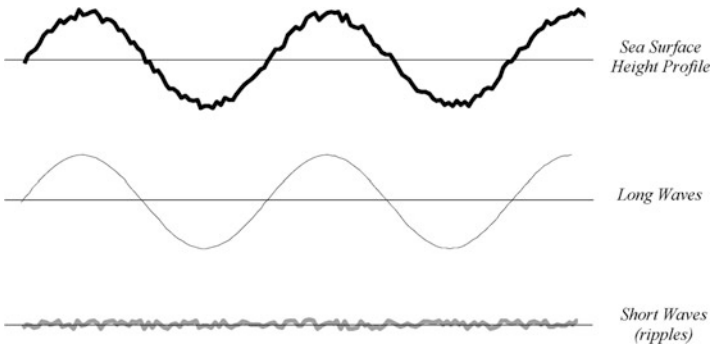


Fig. 1.19 Illustration of a two scale model for surface height profile [91]

1.5.1.1 Example: Bistatic SAR Oceanography

Advantages of bistatic SAR observation in oceanographic applications can be put in evidence by characterizing well-known SAR imaging models of ocean waves to a bistatic geometry [88]. SAR observation of the sea is modeled as the scattering from a rough moving surface [89, 90]. The height profile of sea surface in stationary conditions can be expressed as a sum of (infinite) harmonic components, each of them characterized by its amplitude, phase, angular frequency and wavenumber or wavelength. Such a height profile generates different scales of roughness that affect radar observation differently. In radar scattering theory, surface roughness is represented in terms of its standard deviation and correlation length. Even if from a theoretical point of view a continuous roughness distribution should be considered (i.e. both standard deviation and correlation length ranging from zero to infinity), discrete models are successfully adopted in practical applications and numerical studies. Specifically a two scale model is considered in this section [91]. The model represents sea wave as a sum of two different waves (see Fig. 1.19):

- Short waves, i.e. ripples;
- Long waves.

Both long and short waves have significant effects on SAR images. Ripples or capillarity waves comprise the sea wave components that match the Bragg resonance condition for radar signal wavelengths utilized in microwave remote sensing [89]. In this condition strong scattered signals are generated by sea surface and therefore brighter pixels are derived in SAR images because the Bragg resonant waves produces local echoes adding in phase. In monostatic SAR geometry the following relation among radar signal wavelength, incidence angle and cross-track wavelength component of capillarity wave, Λ, holds

$$\Lambda_m = \frac{\lambda}{2 \sin \Theta_{Tx}} \tag{1.25}$$

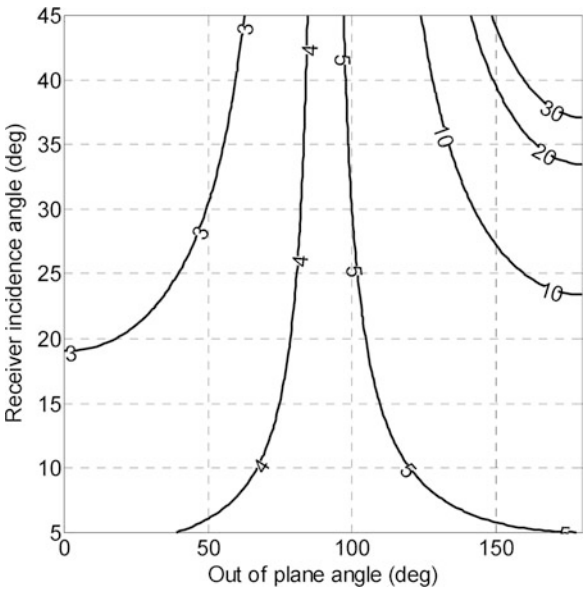


Fig. 1.20 Wavelength of Bragg-resonant sea wave components (cm) for X-band bistatic SAR as a function of receiver pointing angles at 45° transmitter incidence angle

whereas in bistatic geometry the receiver incidence angle and the out-plane angle must be taken into account too [88]

$$\Lambda_b = \frac{\lambda}{\sin\Theta_{Tx} + \sin\Theta_{Rx}\cos\Phi} \tag{1.26}$$

A monostatic X-band SAR, working with 45° incidence angle, meets the Bragg condition when observing sea short waves with about 2 cm wavelength. Figure 1.20 reports the Bragg-resonant sea wave components for X-band bistatic SAR. It is immediate to note that additional information on sea state can be collected by bistatic SAR since different sea wave components contribute to SAR scattering in bistatic geometry.

As far as the scattering from long wave is concerned, three main effects arise:

- Tilt effect, it is a purely geometric effect. The long waves generate a local variability of the observation angles, that is different Bragg-resonant waves are imaged by the radar depending on their location along the long wave.
- Hydrodynamic effect, cross-section modulation due to non uniform distribution of short waves with respect to long waves and due to their motion.
- Orbital motion effects.

The orbital motion is the back-and-forth and up-and-down motion of sea surface generated by the motion of water particles within the waves. Those particles

describe a circular orbit with a diameter equal to the height of the wave. This motion affects SAR images when [92]

$$\frac{\omega T}{2} \leq 1 \tag{1.27}$$

where ω is the frequency of the wave. In such a case the orbital motion is responsible for two important effects on SAR images: *velocity bunching* and *acceleration smearing*. The velocity bunching is a SAR peculiar artifact generated by SAR focusing algorithms. When a moving target is imaged, SAR focusing algorithm, that is developed for stationary scenes, mistakes its Doppler history for that of a stationary target placed in a different azimuth position. This results in a shift of moving target in the Doppler direction by an amount depending on target velocity. When observing sea surface the velocity distribution produced by the orbital motion is not uniform and so the Doppler displacements vary within the imaged area: there are regions where image elements are grouped (i.e. bunched) but others where they are emptied out. Velocity bunching can be in principle removed from SAR image. In general non-linear techniques are needed and numerical iterations have to be implemented [93]. Particular conditions, however, exist in which velocity bunching has a linear effect on SAR images, that is it can be modeled, and hence removed too, by a linear modulation transfer function [88]

$$MTF = -\frac{C}{H} \tag{1.28}$$

where C is a parameter depending on the observation geometry and H indicates the height profile of the sea surface. The linear method can be effectively utilized when $C<<1$. For the sake of concreteness, a maximum values of 0.3 can be utilized as a boundary to limit the region of validity of the linear method. Bistatic SAR geometry can be exploited to achieve linearity regions wider than in the monostatic case. Indeed with reference to parallel track, in-plane, same side bistatic SAR configurations the following relations has been derived in Ref. [88] for the ratio between the parameters C in bistatic and monostatic case

$$\frac{C_b}{C_m} = \frac{R_{Rx}}{R_{Tx} + R_{Rx}} \left(1 + \frac{G(\Theta_{Rx}, \Xi)}{G(\Theta_{Tx}, \Xi)}\right) \tag{1.29}$$

where

$$G(\Theta, \Xi) = \sin^2\Theta \sin^2\Xi + \cos^2\Theta \tag{1.30}$$

and Ξ represents the angle between sea wave propagation direction and platforms motion. In addition C_b and C_m are in the same linear relation with the main parameter of the sea state, such as wavelength and angular frequency [88].

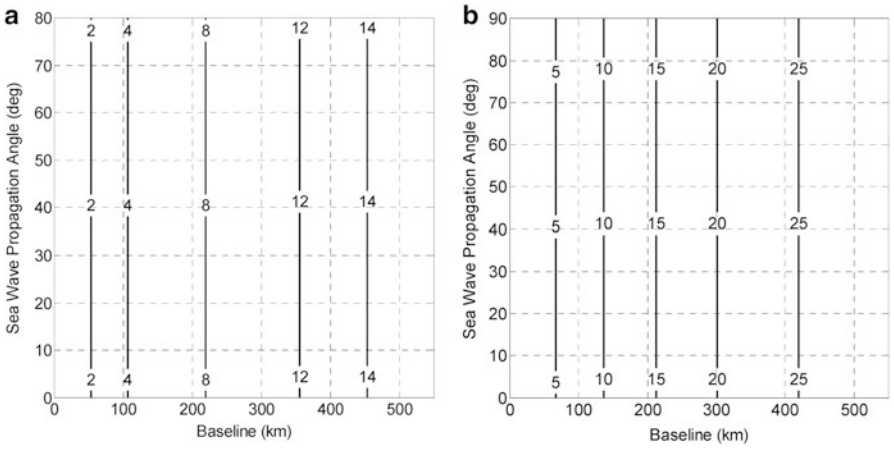


Fig. 1.21 Improvements (in percentage) of linearity region in bistatic geometry when accounting for velocity bunching (**a**), and acceleration smearing effects (**b**). Transmitter and receiver are supposed to orbit at 620 km altitude, 45° transmitter incidence angle has been assumed and the baseline is so that the bistatic receiver is closer than transmitter to the target

It is worth noting that C_b is always larger than C_m in the proposed same side geometry. Figure 1.21a reports the percentage difference between C_b and C_m as a function of the incidence angle and the sea wave propagation angle: as a matter of fact the influence of sea wave propagation angle is very limited, moreover when large separations are achieved, the bistatic configuration offers a wider range (upper to 10–15%) of sea wave spectrum in which a linear SAR MTF is applicable for removing velocity bunching effects.

Orbital motion produces also a smearing in the focused SAR images of sea surface, i.e. a distortion of the resolution cell and a degradation of the azimuth resolution, depending on the time variability (acceleration) of sea surface velocity. Since the acceleration smearing can be interpreted as the results of an instantaneous velocity variation and therefore of a time-varying velocity bunching, bistatic SAR operation is able to reduce this effect with respect to monostatic SAR. With reference to parallel track, in-plane, same side bistatic SAR configurations, the bistatic improvement [88] can be quantified as the square of the ration between C_b and C_m (see Fig. 1.21b). The obtainable improvement is larger than 20% for large baselines and is not affected by the actual sea wave propagation angle.

1.5.2 Images from Two Platforms

If the transmitter is also a monostatic SAR or a multistatic system is considered in which more than one bistatic receiver is utilized, two or more images can be derived from the target area at the same time. In such a case, as introduced in

Sect. 1.2, the separation between the receiving platforms becomes a crucial parameter. If the baseline is larger than the critical value the system can be defined as Large Baseline Bistatic (LBB) SAR [13, 20, 83], where the adjective "large" does not identify a specific baseline length, but it simply remarks that phase coherence cannot be used for data processing and interpretation. Concerning this, a LBB SAR can be interpreted as the opposite of an Interferometric Bistatic SAR or bistatic SAR interferometer (InSAR), where phase coherence is mandatory thus limiting the baseline to be shorter than the critical value.

1.5.2.1 Coherent Combinations

The exploitation of phase coherence in Bistatic InSAR is the base for a variety of well-known and widely investigated techniques and applications:

- *Along-track interferometry for moving target indication (MTI)* in which operation of an along-track interferometer is based on the acquisition of two SAR images taken under identical geometry of observation, but separated by a short time interval. When this is accomplished, any difference between the two images results from changes in the scene. In particular, the phase difference between the echoes from the same target allows its radial speed to be measured [94]. An along-track interferometer can be realized by means of a monostatic-bistatic configuration, where one antenna is used for transmitting [19, 95]. In such a case the effective along-track baseline (i.e. the separation between antenna phase centers), is half the along-track physical separation between the antennas.
- *Cross-track interferometry for digital elevation model (DEM) generation and tomography* in which operation of a cross-track interferometer is based on coherent combination of the signals from two, horizontally or vertically separated SAR antennas, to calculate the interferometric phase difference at each point of the image. The phase difference directly depends on the local relief [96]. DEM generation can greatly benefit from the simultaneous acquisition of the two images in the bistatic configuration when compared to the classic monostatic repeat-pass case. In fact, a considerable drop in the attainable performance results from the decorrelation of interferometric image pairs acquired with time separation [97], due to radar wavelength-scale changes occurred in the scene between the observations. In addition, atmospheric artifacts [9, 98] will equally affect the two acquisitions of the bistatic configuration, thanks to the simultaneity and the slight difference in viewing geometry. With regard to SAR Tomography, it is a technique to achieve focused 3-D SAR images [99], thus overcoming the limitation of conventional monostatic SAR whose images represent a projection of three-dimensional distributed scene reflectivity on the two-dimensional plane individuated by slant range and azimuth directions. The technique relies on an aperture synthesis along the direction normal to the slant range within the elevation plane. This aperture is

synthesized by properly processing several SAR images of the same area gathered with (slightly) different off-nadir/incidence angles. Bi-/multistatic SAR tomography could be properly utilized for the analysis of volume scattering mechanisms, such as when vegetated areas or semi transparent media like snow or ice are observed [99], or in the framework of 3-D imaging and monitoring of urban areas [100].

- *Resolution enhancement* thanks to the coherent combination of monostatic-bistatic or multistatic images [98, 101]. Indeed, if there is a difference in looking angle between the data, different spectral components of the ground range reflectivity spectrum can be sampled. Actually, their coherent combination results in a larger signal bandwidth and therefore in a better geometric resolution.

The quantitative analysis of coherent bi-/multistatic SAR techniques is reported in Chap. 2, whereas the remainder of this chapter deals mainly with LBB SAR or bistatic SAR with a single receiving platform.

1.5.2.2 Incoherent Combinations

The scientific applications of LBB SAR proposed in this section take advantage of the difference in observation geometry between monostatic and bistatic sensors, or among different bistatic receivers. Possible applications include:

- *Stereo-radargrammetry*: it is well-known that SAR images taken under very different viewing angles can be used to perform three-dimensional terrain measurements by making use of the stereo effect. Up to now stereo-radargrammetric applications were realized only by repeat-track monostatic data, i.e. un-simultaneously collected data. Monostatic-Bistatic or Multistatic Stereo-Radargrammetry is able to avoid time decorrelation thus achieving better accuracy (see Sect. 1.5.2.2.1 for the details).
- *Acquisition of terrain slope*: the bistatic scattering coefficient can be related to the root mean square surface slope within a resolution cell [102], and on this basis a model can be developed to derive terrain slope from bistatic data [83].
- *Measurements of surface roughness, dielectric constant and soil moisture*: the availability of both monostatic and bistatic data with a reasonably large (about 30°) difference in viewing angle would significantly help the retrieval of surface roughness and dielectric constant [89]. In fact, the difference in the scattering coefficient at two bistatic angles is small for very rough surfaces and large (several dB) for smooth surfaces. With regard to soil moisture, its effect on the amplitude of the backscattered monostatic signal has been largely investigated, and the effectiveness of its retrieval by radar operating at a low frequency (such as the L-band) has been well established, particularly in presence of moderate vegetation cover. The potential of using bistatic SAR measurements for retrieving soil moisture has been investigated in Ref. [103]. The results indicate that, particularly at the C-band, the estimation of soil moisture can strongly be

improved (namely the error standard deviation is reduced up to a factor of 3 with respect to a monostatic radar) by complementing monostatic and bistatic measurements.

- *Velocity measurements by Doppler analysis*: the relations among Doppler centroid frequencies, position and velocity vectors of the platforms and bistatic slant range can be utilized to develop a procedure to compute the slant range components of target velocity along bistatic slant range. The technique applies to both monostatic-bistatic and multistatic configurations. The most attractive applications are oceanographic, being possible to generate velocity and direction maps of sea waves [8].

1.5.2.2.1 Example: Bistatic Stereo-Radargrammetry

Four different techniques exist to derive terrain height from SAR data: clinometry, polarimetry, stereo-radargrammetry and interferometry [104]. Clinometry uses shadows and occulted areas to extract relative elevations of specific targets from a single image. After a good development during the mid-1990s, SAR clinometry remained a marginal technique, useful only where it is quite difficult to derive ground truth, such as for extraterrestrial sites. Polarimetry allows measurement of azimuthal slopes and so can provide an estimation of terrain elevation [105]. The technique is strongly affected by the quality of selected complex scattering models (especially over forested and agricultural lands) and two orthogonal passes are needed for DEM generation [106], for this reason up to now only airborne data were utilized. SAR interferometry, based on the coherent combination of two or more SAR images, is the most mature and accurate method for wide-area DEM generation, as described in Sec. 5.2.1. SAR interferometry, however, suffers from some main drawbacks. Indeed, the phase interferogram must be unwrapped to resolve the modulo 2π ambiguity before estimating topography. Different approaches exist for phase unwrapping [96], usually a time-consuming and difficult task that cannot be avoided. Moreover height estimation accuracy shows high sensitivity to phase decorrelation. In forested or mountain slope areas phase coherence can be very low, and a incorrect solution of the phase unwrapping problem might be derived, often leading to diverging errors in interferometric DEM generation [107]. Finally very accurate antenna relative positioning, up to millimeter level [9, 10], is needed to limit DEM errors. Part of those constraints can be mitigated by bistatic stereo-radargrammetry.

Classic stereo-radargrammetry refers to the processing of pairs of monostatic images, taken on the same area under different viewing angles, for computing terrain elevation [108]. It exploits the stereo-effect, as in optical photogrammetry, but also accounts for side-looking monostatic SAR geometry: points on the Earth's surface exhibit different displacements in SAR images as a function of their elevation, position in the range line, and radar pointing angles. The relief reconstruction procedure is based on the measurement of the difference (parallax) in target positions in the two images forming the stereoscopic pair and on the

equations relating the positions and heights of viewed targets. Typically, the target height is directly linked to the parallax difference between the two observations, and the parallax is expressed as a function of target slant ranges or ground ranges, which are the typical cross-track coordinates in radar images. As a matter of fact, both interferometry and stereo-radargrammetry are based on the notion of detecting relative displacements between two images, except that InSAR uses phase domain data, whereas stereo-radargrammetry exploits range/Doppler and amplitude measurements. Hence, stereo-radargrammetry is intrinsically more robust to compute terrain elevation with respect to interferometry because it is less sensitive to decorrelation and is not affected by phase-unwrapping problems. Moreover stereo-radargrammetry, unlike InSAR, does not operate on wavelength scales thus also relaxing the relative positioning requirements. On the other hand, just due to the wavelength scale operation, in high correlation areas data provided by SAR interferometry are, no doubts, more accurate and with a finer resolution [107].

Large separations between the receiving platforms are needed to generate the necessary stereo-effect for high accuracy DEM generation by stereo-radargrammetry. This is the reason why stereo-radargrammetric applications were realized only by monostatic data, which so far have been collected along separated passes, to achieve very large (from tens to hundreds of kilometers) cross-track baselines in LEO. In LBB SAR two (or more) images are acquired simultaneously, i.e. with no time dependence, so LBB SAR guarantees an accuracy improvement with respect to classic stereo-radargrammetry.

The experiences on monostatic repeat-pass data show that different acquisition geometries are exploitable [109]: same and opposite sides, steep and shallow look angles. From a geometric point of view larger stereo-intersection angles (i.e. very large baselines and opposite-side stereo pairs) provide the best results. Actually, when dealing with real-world data, large radiometric discrepancies are generated in opposite-side geometry thus limiting previous advantages, especially as terrain slope and time effects due to non-simultaneity between the images increase. Therefore, provided that a minimum intersection angle of 5–10° is guaranteed in repeat-pass same-side geometry, it is possible to state that there is not a significant correlation between DEM accuracy and acquisition geometry in monostatic radargrammetry.

In a spaceborne bistatic scenario, however, further considerations are needed. First of all it is worth noting that the bistatic angle plays the same role in bistatic radargrammetry as the stereo-intersection angle in monostatic repeat-pass one. The bistatic image quality, in terms of geometric and radiometric resolutions, is greatly affected by bistatic geometry, even before stereo reconstruction, therefore the effect of the acquisition geometry on bistatic image parameters must be accounted for when selecting the bistatic configuration for radargrammetric DEM generation. Moreover, when bistatic images are adopted to form a stereoscopic pair, new relations are needed to define target height as a function of the peculiar parameters of bistatic surveying geometry [8]. Namely, the models of classical stereo-radargrammetry can be specialized to the bistatic configuration, but also new models can be developed. In Ref. [110] various models have been proposed and

analyzed for stereo-radargrammetric relief reconstruction by spaceborne bistatic SAR. The models have been specified to LEO monostatic-bistatic observation geometry. This configuration is herein taken as a reference, even if the basic principles of those methods hold for multistatic configurations too and the relevant equations could be derived.

Monostatic-bistatic SAR stereo-radargrammetry can take advantage of in-plane same-side observation geometry and parallel tracks. In-plane configuration reduces geometric distortions and pixel skewing effects that are usually generated in out-of-plane bistatic SAR acquisitions. Moreover same-side data collection, even though limited in forming large bistatic angles, allows the system to reduce bistatic ground range resolution degradation. In addition, if the bistatic sensor is closer to the target, it can take advantages of stronger echoes, also benefiting from an improvement in Doppler resolution. Finally, LEO formation flying satellites can implement stable pendulum configurations [83], able to attain same side bistatic geometry close to in-plane conditions with almost parallel tracks, thus also reducing maintenance operations.

As noted above, there is not a unique approach for bistatic stereo radargrammetry. Even if the stereo-effect is exploited in any case, two different methodology can be individuated:

- *Parallax-based methods*, which principle is borrowed from optical photogrammetry.
- *True stereo radar methods*, which rely on the peculiarity of radar and SAR acquisition.

Parallax-based methods utilize the concept of parallax difference between the observations forming the stereoscopic pair [108]. In general, the parallax difference, $d\vec{p}$, is a two-dimensional vector because the same target, due to differences in observation geometry, can be projected in points on the two images that present both cross-track and along-track displacements. The height reconstruction by the parallax method therefore can be resumed by the following equation

$$h(\vec{r}) = \vec{F} \cdot d\vec{p} \tag{1.31}$$

where h is the altitude of the target $\vec{r}$ and the function $\vec{F}$ depends, through the selected method, on the parameters of the observation geometry, essentially transmitter and receiver positions and velocity with respect to the target [110–112]. Parallax-based height computation relies on relative measurements, i.e. a reference surface, datum, must be defined and height are computed with respect to datum level. Hence, whereas the absolute parallax is defined with reference to the target relief displacement in the radar image, parallax differences are obtained with respect to datum, which allows definition of a reference parallax that must be calculated to derive target relative height [8]. Finally, accurate knowledge of the height of a limited number of ground control points is needed to evaluate topography. In parallel-track same-side in-plane monostatic-bistatic geometry, with no squint angle, targets are focused in the common range elevation plane, so a

scalar parallax difference can be utilized, as in Ref. [110] where parallax-based reconstruction equations derived under flat-Earth hypothesis are presented.

Parallax-based methods strongly rely on image quality and on capability to extract information from images. However, the intrinsic nature of radar, and of SAR too, allows direct access to slant range and Doppler measurements. On this basis the parallax difference can be expressed as a function of SAR measurements and not obtained just as result of image processing procedures [110]. This statement is of fundamental importance to pass from optical derived approaches to true stereo radar methods. The most notable example of such methods is the rigorous stereo SAR problem [108]. The equations defining the problem in monostatic bistatic geometry are

$$R_{Tx} = \left|\vec{P}_{Tx} - \vec{r}\right| \tag{1.32}$$

$$R_{Rx} = \left|\vec{P}_{Tx} + \vec{B} - \vec{r}\right| \tag{1.33}$$

$$f_{Tx} = \frac{2\left(\vec{V}_{Tx} - \vec{v}\right) \cdot \left(\vec{P}_{Tx} - \vec{r}\right)}{\lambda\left|\vec{P}_{Tx} - \vec{r}\right|} \tag{1.34}$$

$$f_{Rx} = \frac{1}{\lambda} \frac{\left(\vec{V}_{Tx} - \vec{v}\right) \cdot \left(\vec{P}_{Tx} - \vec{r}\right)}{\left|\vec{P}_{Tx} - \vec{r}\right|} \cdot \frac{\left(\vec{V}_{Tx} + \vec{V}_B - \vec{v}\right) \cdot \left(\vec{P}_{Tx} + \vec{B} - \vec{r}\right)}{\left|\vec{P}_{Tx} + \vec{B} - \vec{r}\right|} \tag{1.35}$$

where f_{Tx} and f_{Rx} are the monostatic and bistatic Doppler centroid frequencies, respectively, and $\vec{V}_B$ indicates the relative velocity between the platforms (bistatic with respect to monostatic one). Equations 1.32 and 1.33 define monostatic and bistatic range spheres, while (1.34) and (1.35) individuate monostatic and bistatic Doppler cones. Differently from classic photogrammetry, where triangulation equations are applied on corresponding targets identified in photographic images by means of the optical stereo reconstruction obtained thanks to stereocomparators [113], monostatic and bistatic slant ranges and Doppler centroid frequencies are derived for each image point of co-registered data, as in numerical photogrammetry. Concerning monostatic radargrammetry least square methods can be used to derive target positions, and target height too, accounting for spherical or ellipsoidal zero relief surface and assuming a static scene (i.e. $\vec{v} = 0$). The equations valid in bistatic geometry are quite similar to the monostatic repeat-pass case, however an essential difference arises [110]: the introduction, in the equations of the problem, of the baseline vector and of the relative velocity, instead of receiver position and velocity, allows for a reduction of the overall error in the determination of the target position with respect to the monostatic case, because relative positioning and navigation of formation flying satellites is able to achieve better accuracy than the absolute navigation utilized in monostatic repeat-pass algorithms. Finally, it is worth noting that, thanks to its vector formulation, the rigorous stereo SAR problem can be readily applied to any bistatic and multistate geometry.

The first step of both parallax-based and true stereo SAR methods is the image matching: that is relating a pixel of the bistatic image to the corresponding one in the monostatic image. In this way it is possible to associate both monostatic and bistatic SAR measurements (range/Doppler, ground range/azimuth…) of each observed target. Concerning this, it is important to remark that, as noted in Sect. 1.3.2, speckle noise is generated in the images produced by the considered monostatic-bistatic system. For this reason proper pattern recognition and image co-registration algorithms have to be used to guarantee sub-pixel co-registration accuracy [109], that is fundamental for accurate DEM generation. A quality decay of image matching can result from angle decorrelation and SNR degradation. As shown in Ref. [110], the performance of parallax-based method are strongly affected by the quality of the image matching, whereas true stereo SAR methods are characterized by greater robustness since their performance is only partially influenced by co-registration errors. The following sub-section reports an example of error budget model for the rigorous stereo SAR problem and shows the achievable height estimation accuracy considering same side monostatic-bistatic acquisitions performed by two LEO satellites flying in formation.

Error Budget of Height Estimation Accuracy

A general approach for the development of an error budget is to perform a propagation of the height estimation error. For any given reconstruction method, the height is a function of a series of parameters, S_1, S_2, …, S_n

$$h = h(S_1, S_2, \ldots, S_n) \tag{1.36}$$

If only random errors are considered, i.e. assuming that a limited number of ground control points can be used to reduce systematic errors, the basic propagation for the variance of h is

$$\sigma_{h|TOT}^2 = \sum_i \left(\frac{\partial h}{\partial S_i}\right)^2 \sigma_{S_i}^2 + \sum_{\substack{i,j \\ i \neq j}} \left|\frac{\partial h}{\partial S_i}\frac{\partial h}{\partial S_j}\right| \sigma_{S_i S_j} \tag{1.37}$$

where σ_{S_i} is the uncertainty of the parameter S_i, $\sigma_{S_i S_j}$is the cross correlation between the parameters S_i and S_j, and the partial derivatives represent the height sensitivity to uncertainties in the parameters. For the monostatic-bistatic rigorous stereo SAR problem, the functional model of height reconstruction is

$$h = h\left(P_{Tx}, \vec{V}_{Tx}, \vec{B}, \vec{V}_B, R_{Tx}, R_{Rx}, f_{Tx}, f_{Rx}\right) \tag{1.38}$$

and all parameters can be considered uncorrelated with the exception of the monostatic and bistatic slant ranges and Doppler centroid frequencies. Equations 1.32–1.35 do not provide an explicit formulation for the height computation functional model, however, the height sensitivities to the different parameters

can be derived numerically from the equations of the problem when observation geometry is assigned.

For evaluating the overall height estimation accuracy a pendulum formation is considered, in which monostatic and bistatic satellites share semi-major axis, eccentricity, inclination and argument of perigee. The remaining parameters can be then selected to maximize the area on which bistatic acquisition can be achieved [31]. Assuming the Italian COSMOSkyMed mission as a reference for the monostatic sensor [114], and considering the bistatic sensor working with constant off-nadir angle, the covered latitudes ranges from about ±65° for 5° off-nadir angle, to ±80° for 20° off-nadir angle. In the first case the baseline ranges from 200 to 550 km with 7–27° bistatic angle, whereas for 20° off-nadir angle shorter separations are attained ranging from 100 to 400 km with 2–14° bistatic angle. Even if important out-of-plane baseline components (up to 50 km) are generated by the relative orbital motion, the considered pendulum formation establishes bistatic observation geometries quite close to the in-plane condition. In this case height sensitivities to $\vec{V}_{Tx}, \vec{V}_B, f_{Tx}, f_{Rx}$ and to the along-track baseline are very limited and the height uncertainty generated by errors on those parameters is negligible. On this basis the error budget model for height estimation accuracy can be written as

$$\sigma^2_{h|TOT} = \sigma^2_{\vec{P}_{Tx}} + \left(\frac{\partial h}{\partial B_c}\right)^2 \sigma^2_{B_c} + \left(\frac{\partial h}{\partial B_r}\right)^2 \sigma^2_{B_r} + \left(\frac{\partial h}{\partial R_{Tx}}\right)^2 \sigma^2_{R_{Tx}} + \left(\frac{\partial h}{\partial R_{Rx}}\right)^2 \sigma^2_{R_{Rx}} + 2\left|\frac{\partial h}{\partial R_{Tx}} \frac{\partial h}{\partial R_{Rx}}\right| \sigma_{R_{Tx}R_{Rx}} \tag{1.39}$$

where B_c and B_r are cross-track and radial baseline components, respectively. The uncertainty $\sigma_{R_{Tx}R_{Rx}}$ can be expressed as the product of monostatic and bistatic slant range uncertainties [110]. In (1.39) the height sensitivity to monostatic sensor position deserves more attention. Indeed, according to (1.37) and (1.38), for $\vec{P}_{Tx} = (X_{Tx}, Y_{Tx}, Z_{Tx})$, the relevant term should be

$$\sigma^2_{h|\vec{P}_{Tx}} = \left(\frac{\partial h}{\partial X_{Tx}}\right)^2 \sigma^2_{X_{Tx}} + \left(\frac{\partial h}{\partial Y_{Tx}}\right)^2 \sigma^2_{Y_{Tx}} + \left(\frac{\partial h}{\partial Z_{Tx}}\right)^2 \sigma^2_{Z_{Tx}} \tag{1.40}$$

however the following result can be verified [110]

$$\left(\frac{\partial h}{\partial X_{Tx}}\right)^2 + \left(\frac{\partial h}{\partial Y_{Tx}}\right)^2 + \left(\frac{\partial h}{\partial Z_{Tx}}\right)^2 = 1 \tag{1.41}$$

so, if the same uncertainty is assumed for each component of the monostatic sensor, it is immediate to conclude that an error in monostatic platform position merely produces an identical error on target altitude (a result shared with SAR interferometry [96]). The uncertainties of the monostatic antenna position basically depend on the accuracy of the absolute navigation system. A value of 1 m for each component

of position is compatible with present generation medium accuracy GPS-based navigation solutions. As regard to the baseline, differential GPS algorithms allow dynamic relative state with decimeter-to-centimeter accuracy to be computed, therefore a conservative value of 0.5 m can be assumed for the uncertainties of each baseline component. Monostatic slant range uncertainties are linked to the dimension, Δr_{Tx}, of monostatic slant range resolution element. A reasonable expression can be derived considering the quantization error and assuming that the actual slant range to the target has a uniform distribution within a cell [110]

$$\sigma^2_{R_{Tx}} = \frac{1}{12}\Delta r^2_{Tx} \tag{1.42}$$

COSMOSkyMed stripmap products are characterized by a final ground range resolution of 3–15 m, so values of 3 and 0.75 m can be assumed for ground range and azimuth resolution of single-look-complex data. Bistatic slant range uncertainty depends on bistatic slant range resolution, as the monostatic one, and it is also related to the error resulting from the application of co-registration procedures required to form stereo pairs. A value of 1/10 of an image pixel is compatible with areas where medium-to-high correlation is achieved. Hence, bistatic slant range uncertainty can be calculated as follows

$$\sigma^2_{R_{Rx}} = \frac{1}{12}\Delta r^2_{Rx} + \left(\frac{\Delta r_{Rx}}{10}\right)^2 + \left(\frac{\Delta a_{Rx}}{10}\right)^2 \tag{1.43}$$

where Δa_{Rx} is bistatic azimuth resolution. Bistatic resolutions can be easily computed for the selected pendulum configuration by the gradient method (see Sec. 1.3.1). Figure 1.22 reports the contributions of the considered error sources to height estimation uncertainty; the overall height estimation error is also shown (lower right plot). It is worth noting that the reported slight increase of the height uncertainty for small off-nadir angles of the bistatic satellite and low latitudes does not rely on a change of the sensitivity function, but it is due to the increase of bistatic slant range uncertainty. In fact a degradation of the bistatic slant range resolution is attained for very large bistatic angles.

In conclusion, according to the present analysis, monostatic-bistatic spaceborne stereo-radargrammetry is able to measure local target relief with a metric accuracy. The best results can be obtained when a small off-nadir angle is selected for the bistatic sensor, so that long baselines can be reached and large bistatic angles are established. Finally, it is worth noting that considering 5–10° bistatic off-nadir angles the height estimation accuracy is practically constant, i.e. it is not much affected by target latitude, hence very accurate DEMs can be generated on very large areas.

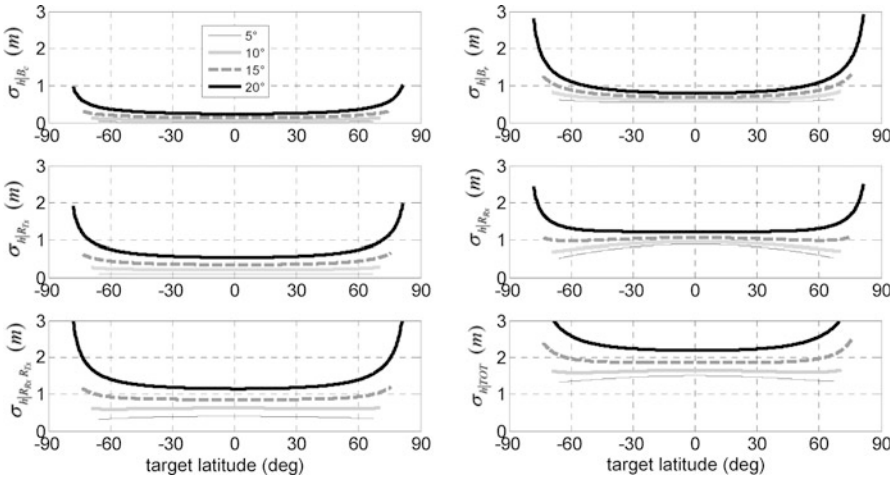


Fig. 1.22 Contributions to height uncertainty as a function of target latitude for four bistatic antenna off-nadir angles

References

1. Willis NJ (1990) Bistatic radar. In: Skolnik MJ (ed) Radar handbook. McGraw-Hill, New York
2. Howland PE, Griffiths HD, Baker CJ (2008) Passive bistatic radar systems. In: Cherniakov M (ed) Bistatic radar: emerging technology. Wiley, Chichester
3. Wiley CA (1985) Synthetic aperture radar, a paradigm for technology evolution. IEEE Trans Aerosp Electron Syst 21:440–443
4. Moccia A (2010) Synthetic aperture radar. In: Blockley R, Shyy W (eds) Encyclopedia of aerospace engineering. Wiley, Chichester
5. Ulaby FT, Moore RK, Fung AK (1982) Microwave remote sensing: active and passive. In: Radar remote sensing and surface scattering and emission theory, vol II. Addison-Wesley, Advanced Book Program, Reading, MA
6. Curlander JC, McDonough RN (1991) Synthetic aperture radar systems & signal processing. Wiley, Wiley Series in Remote Sensing, New York
7. Hobish MK (2001) Satellite formation flying. In: NPOESS remote sensing tutorial, NASA. http://rst.gsfc.nasa.gov/. Latest access on 13th Jan 2012
8. Moccia A (2008) Fundamentals of bistatic synthetic aperture radar. In: Cherniakov M (ed) Bistatic radar: emerging technology. Wiley, Chichester
9. Krieger G, Moreira A et al (2007) TanDEM-X: a satellite formation for high-resolution SAR interferometry. IEEE Trans Geosci Remote Sens 45(11):3317–3341
10. Krieger G, Hajnsek I et al (2010) Interferometric synthetic aperture radar (SAR) missions employing formation flying. Proc IEEE 98(5):816–843
11. Barber BC (1985) Theory of digital imaging from orbital synthetic aperture radar. Int J Remote Sens 6(7):1009–1057
12. Moccia A, Chiacchio N, Capone A (2000) Spaceborne bistatic synthetic aperture radar for remote sensing applications. Int J Remote Sens 21(18):3395–3414
13. Krieger G, Moreira A (2006) Spaceborne bi- and multistatic SAR: potential and challenges. IEE Proc Inst Electr Eng-Radar Sonar Navig 153(3):184–198

14. Zebker H, Villasenor J (1992) Decorrelation in interferometric radar echoes. IEEE Trans Geosci Remote Sens 30(5):950–959
15. Rodriguez E, Martin JM (1992) Theory and design of interferometric synthetic aperture radars. IEE Proc F 139(2):147–159
16. Moccia A, Fasano G (2005) Analysis of spaceborne tandem configurations for complementing COSMO with SAR interferometry. EURASIP J Appl Signal Process 20:3304–3315
17. GESS Science Definition Team (2003) GESS: Global earthquake satellite systems. A 20-year plan to enable earthquake prediction. NASA JPL, California Institute of Technology, Pasadena, CA 400–1069
18. Moccia A, Renga A (2010) Synthetic aperture radar for earth observation from a lunar base: performance and potential applications. IEEE Trans Aerosp Electron Syst 46(3):1034–1051
19. Moccia A, Rufino G (2001) Spaceborne along-track SAR interferometry performance analysis and mission scenarios. IEEE Trans Aerosp Electron Syst 37(1):199–213
20. Renga A, Moccia A, D'Errico M et al (2008) From the expected scientific applications to the functional specifications, products and performance of the SABRINA mission. In: Proceedings of the IEEE radar conference, Rome, Italy, doi:10.1109/RADAR.2008.4720935
21. Cardillo GP (1990) On the use of gradient to determine bistatic SAR resolution. Proc Antennas Propag Soc Int Symp 2:1032–1035
22. Moccia A, Renga A (2011) Spatial resolution of bistatic synthetic aperture radar: impact of acquisition geometry on imaging performance. IEEE Trans Geosci Remote Sens 49 (10):3487–3503
23. Zeng T, Cherniakov M, Long T (2005) Generalized approach to resolution analysis in BSAR. IEEE Trans Aerosp Electron Syst 41(4):461–474
24. Gierull C (2004) Bistatic synthetic aperture radar. Defence R&D Canada, Ottawa, Technical report DRDC-OTTAWA-TR-2004-190
25. Walterscheid I, Espeter T, Klare J, Brenner A (2010) Bistatic spaceborne–airborne forward-looking SAR. In: Proceedings of the 8th European conference on synthetic aperture radar, Aachen, Germany, pp 986–989
26. D'Errico M, Moccia A, Renga A et al (2009) Satellite-unmanned airborne systems cooperative approaches for the improvement of all-weather day and night operations. ESA contract 22449/09/F/MOS, Final report, Second University of Naples, Aversa, Italy
27. Keydel W (1992) Basic principles of SAR. In: Fundamentals & special problems of synthetic aperture radar, Advisory Group for Aerospace Research & Development, Neuilly-sur-Seine, France, AGARD-LS-182
28. Cherniakov M, Zeng T (2008) Passive bistatic SAR with GNSS transmitters. In: Cherniakov M (ed) Bistatic radar: emerging technology. Wiley, Chichester
29. Walterscheid I, Klare J, Brenner AR, Ender JHG, Loffeld O (2006) Challenges of a bistatic spaceborne/airborne SAR experiment. In: Proceedings of the 6th European conference on synthetic aperture radar, Dresden
30. Willis NJ (1991) Bistatic radar. Artech House, Boston
31. Moccia A, D'Errico M (2008) Bistatic SAR for earth observation. In: Cherniakov M (ed) Bistatic radar: emerging technology. Wiley, Chichester
32. Cazzani L, Colesanti C, Leva D, Nesti G, Prati C, Rocca F, Tarchi D (2000) A ground-based parasitic SAR experiment. IEEE Trans Geosci Remote Sens 38(5):2132–2141
33. Krieger G, Yuonis M (2006) Impact of oscillator noise in bistatic and multistatic SAR. IEEE Geosci Remote Sens Lett 3(3):424–428
34. Rodriguez-Cassola M, Baumgartner SV, Krieger G, Moreira A (2010) Bistatic TerraSAR-X/F-SAR spaceborne–airborne SAR experiment: description, data processing, and results. IEEE Trans Geosci Remote Sens 48(2):781–794
35. Doviak RJ, Goldhirsh J, Miller AR (1972) Bistatic radar detection of high altitude clear air atmospheric targets. Radio Sci 7:993–1003

36. Peterson AM, Teague CC, Tyler GL (1970) Bistatic-radar observation of long-period, directional ocean-wave spectra with Loran A. Science 170:158–161
37. Atlas D, Naito K (1968) Carbone RE Bistatic microwave probing of a refractively perturbed clear atmosphere. J Atmos Sci 25:257–268
38. Rogers PJ, Eccles PJ (1971) The bistatic radar equation for randomly distributed targets. Proc IEEE 59(6):1019–1021
39. Wurman J, Heckman S, Boccippio D (1993) A bistatic multiple-Doppler network. J Appl Meteorol 32:1802–1814
40. Pavelyev AG, Volkov AV, Zakharov AI, Krutikh SA, Kucherjavenkov AI (1996) Bistatic radar as a tool for earth investigation using small satellites. Acta Astronaut 39(9–12):721–730
41. Parker MN, Tyler GL (1973) Bistatic-radar estimation of surface-slope probability distributions with applications to the moon. Radio Sci 8(3):177–184
42. Simpson RA, Tyler GL (1982) Radar scattering laws for the lunar surface. IEEE Trans Antennas Propag 30(3):438–449
43. Simpson RA (1993) Spacecraft studies of planetary surfaces using bistatic radar. IEEE Trans Geosci Remote Sens 31(2):465–482
44. Tyler GL, Howard HT (1973) Dual-frequency bistatic-radar investigations of the Moon with Apollos 14 and 15. J Geophys Res 78(23):4852–4874
45. Tang CH, Boak TIS, Grossi MD (1977) Bistatic radar measurement of electrical properties of the Martian surface. J Geophys Res 82:4305–4315
46. Fung AK, Zuffada C, Hsieh CY (2001) Incoherent bistatic scattering from the sea surface at L-band. IEEE Trans Geosci Remote Sens 39(5):1006–1012
47. Zavorotny VU, Voronovich AG (2000) Scattering of GPS signals from the ocean with wind remote sensing applications. IEEE Trans Geosci Remote Sens 38(2):951–964
48. Martín-Neira M, Caparrini M, Font-Rossello J, Lannelongue S, Serra Vallmitjana C (2001) The PARIS concept, an experimental demonstration of sea surface altimetry using GPS reflected signals. IEEE Trans Geosci Remote Sens 39(1):142–150
49. Zahn D, Sarabandi K (2000) Simulation of bistatic scattering for assessing the application of existing communication satellites to remote sensing of rough surfaces. Proc IGARSS 4:1528–1530
50. Zavorotny VU, Voronovich AG (2000) Bistatic GPS signal reflections at various polarizations from rough land surface with moisture content. Proc IGARSS 7:2852–2854
51. Zavorotny VU, Voronovich AG, Katzberg SJ, Garrison JL, Komjathy A (2000) Extraction of sea state and wind speed from reflected GPS signals: modeling and aircraft measurements. Proc IGARSS 4:1507–1509
52. Lowe ST, LaBrecque JL, Zuffada C, Romans LJ, Young LE, Hajj GA (2002) First spaceborne observation of an earth-reflected GPS signal. Radio Sci. doi:10.1029/2000RS002539
53. Gleason S, Hodgart S, Sun Y, Gommenginger C, Mackin S, Adjrad M, Unwin M (2005) Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing. IEEE Trans Geosci Remote Sens 43(6):1229–1241
54. Martín-Neira M, D'Addio S, Buck C, Floury N, Prieto-Cerdeira R (2011) The PARIS ocean altimeter in-orbit demonstrator. IEEE Trans Geosci Remote Sens 49(6):2209–2237
55. Teague CC, Tyler GL, Joy JW, Stewart RH (1973) Synthetic aperture observations of directional height spectra for 7 s ocean waves. Nat Phys Sci 244:98–100
56. Teague CC, Tyler GL, Stewart RH (1977) Studies of the sea using HF radio scatter. IEEE J Oceanic Eng OE-2(1):12–19
57. Auterman JL (1984) Phase stability requirements for a bistatic SAR. In: Proceedings of the IEEE national radar conference, Atlanta, pp 48–52
58. http://trs-new.jpl.nasa.gov/dspace/handle/2014/33128. Latest access on 13th Jan 2012
59. Martinsek D, Goldstein R (1998) Bistatic radar experiment, In: Proceedings of the European conference on synthetic aperture radar, Berlin

60. Balke F (2005) Field test of bistatic forward-looking synthetic aperture radar. In: Proceedings of IEEE radar conference, Washington, DC, pp 423–429
61. Gierull C (2006) Mitigation of phase noise in bistatic SAR systems with extremely large synthetic aperture. In: Proceedings of the 6th European conference on synthetic aperture radar, Dresden
62. Yates G, Horne AM, Blake AP, Middleton R (2006) Bistatic SAR image formation. IEE Proc Radar Sonar Navig 153(3):208–213
63. Dubois-Fernandez P, Cantalloube H, Vaizan B, Krieger G, Horn R, Wendler M, Giroux V (2006) ONERA-DLR bistatic SAR campaign: planning, data acquisition, and first analysis of bistatic scattering behaviour of natural and urban targets. IEE Proc Radar Sonar Navig 153 (3):214–223
64. Walterscheid I, Brenner AR, Ender J (2004) Geometry and system aspects for a bistatic airborne SAR-experiment. In: Proceedings of the 5th European conference on synthetic aperture radar, Ulm, pp 567–570
65. Sanz-Marcos J, Mallorqui JJ, Aguasca A (2005) First steps towards single-pass interferometry based on a bistatic fixed receiver SAR system. In: Proceedings of the FRINGE, Frascati, Italy
66. Antoniou M, Saini R, Cherniakov M (2007) Results of a space-surface bistatic SAR image formation algorithm. IEEE Trans Geosci Remote Sens 45(11):3359–3371
67. Griffiths HD, Baker CJ, Baubert J, Kitchen N, Treagust M (2002) Bistatic radar using spaceborne illuminators. In: Proceedings of the IEE International Conference (RADAR 2002), Edinburgh, UK, pp 1–5
68. Griffiths HD (2008) New directions in bistatic radar. IEEE radar conference, Rome, Italy, pp 1–6
69. Walterscheid I, Espeter T, Brenner AR, Klare J, Ender JHG, Nies H, Wang R, Loffeld O (2010) Bistatic SAR experiments with PAMIR and TerraSAR-X – setup, processing, and image results. IEEE Trans Geosci Remote Sens 48(8):3268–3278
70. www.nasa.gov/mission_pages/Mini-RF/news/tandem_search.html. Latest access on 13th Jan 2012
71. Nozette S, Spudis P, Bussey B et al (2010) The lunar reconnaissance orbiter miniature radio frequency (Mini-RF) technology demonstration. Space Sci Rev 150:285–302
72. www.nasa.gov/offices/oce/appel/ask-academy/issues/volume3/AA_3-7_F_outside.html. Latest access on 13th Jan 2012
73. Hsu YS, Lorti DC (1986) Spaceborne bistatic radar – an overview. IEE Proc 133 (F7):642–648
74. Chen P, Beard JK (2000) Bistatic GMTI experiment for airborne platforms. In: The record of the IEEE international radar conference, Alexandria, VA, pp 42–46
75. Moore KL, Richards CL, Chen P (2003) Bistatic radar system using transmitters in mid-earth orbit. US Patent 6614386
76. Ogrodnik RF, Wolf WE, Schneible R, McNamara J, Clancy J, Tomlinson PG (1997) Bistatic variants of space-based radar. In: Proceedings of the IEEE aerospace conference, Snowmass at Aspen, Co, vol 2, pp 159–169
77. Guttrich GL, Sievers WE, Tomljanovich NM (1997) Wide area surveillance concepts based on geosynchronous illumination and bistatic unmanned airborne vehicles or satellite reception. In: Proceedings of the IEEE national radar conference, Syracuse, pp 126–131
78. Hartnett MP, Davis ME (2003) Operations of an airborne bistatic adjunct to space based radar. In: Proceedings of the IEEE radar conference, Huntsville, Alabama, USA, pp 133–138
79. Cherniakov M, Kubik K, Nezlin D (2000) Bistatic synthetic aperture radar with non-cooperative LEOS based transmitter. Proc IGARSS 2:861–862
80. Martín-Neira M, Mavrocordatos C, Colzi E (1998) Study of a constellation of bistatic radar altimeters for mesoscale ocean applications. IEEE Trans Geosci Remote Sens 36 (6):1898–1904

81. Picardi G, Seu R, Sorge SG, Martin-Neira M (1998) Bistatic model of ocean scattering. IEEE Trans Antennas Propag 46(10):1531–1541
82. Alberti G, Zelli C (1999) Design of bistatic altimetric mission for oceanographic applications. Space Technol 19(2):83–96
83. Moccia A, Rufino G, D'Errico M et al. (2001) BISSAT: a bistatic SAR for earth observation. Phase A study – final report, ASI research contract I/R/213/00, University of Naples, Naples, Italy
84. Massonet D (2001) Capabilities and limitations of the interferometric cartwheel. IEEE Trans Geosci Remote Sens 39(3):506–520
85. Caves R, Luscombe AP, Lee PF, James K (2002) Topographic performance evaluation of the RADARSAT-2/3 tandem mission. Proc IGARSS 2:961–963
86. Hauck B, Ulaby F, DeRoo FR (1998) Polarimetric bistatic measurement facility for point and distributed targets. IEEE Antennas Propag Mag 40:31–41
87. Airiau O, Khenchaf A (2000) A methodology for modeling and simulating target echoes with a moving polarimetric bistatic radar. Radio Sci 35(3):773–782
88. Moccia A, Rufino G, De Luca M (2003) Oceanographic applications of spaceborne bistatic SAR. In: Proceedings of the IGARSS, Toulouse
89. Ulaby FT, Moore RK, Fung AK (1986) Microwave remote sensing: active and passive – vol. III: from theory to applications. Artech House, Norwood, MA
90. Alpers W, Ross D, Rufenach C (1981) On the detectability of ocean surface waves by real and synthetic aperture radar. J Geophys Res 86(C7):6481–6498
91. Hasselmann K, Raney RK, Plant WJ, Alpers W, Shuchman RA, Lyzenga DR, Rufenach CL, Tucker MJ (1985) Theory of synthetic aperture radar ocean imaging: a MARSEN view. J Geophys Res 90(C3):4659–4686
92. Alpers W, Rufenach C (1979) The effect of orbital motions on synthetic aperture radar imagery of ocean waves. IEEE Trans Antenn Propag AP-27(5):685–690
93. Hasselmann K, Hasselmann S (1991) On the nonlinear mapping of an ocean wave spectrum into a synthetic aperture radar image spectrum and its inversion. J Geophys Res 96 (C6):10713–10729
94. Goldstein RM, Zebker HA (1987) Interferometric radar measurement of ocean surface currents. Nature 328(6132):707–709
95. Romeiser R, Runge H (2007) Theoretical evaluation of several possible along-track InSAR modes of TerraSAR-X for ocean current measurements. IEEE Trans Geosci Remote Sens 45 (1):21–35
96. Rosen PA, Hensley S, Joughin IR, Li FK, Madsen SN, Rodriguez E, Goldstein RM (2000) Synthetic aperture radar interferometry. Proc IEEE 88(3):333–382
97. Li FK, Goldstein RM (1990) Studies of multibaseline spaceborne interferometric synthetic aperture radars. IEEE Trans Geosci Remote Sens 28(1):88–97
98. Massonnet D (2001) The interferometric cartwheel: a constellation of passive satellites to produce radar images to be coherently combined. Int J Remote Sens 22(12):2413–2430
99. Reigber A, Moreira A (2000) First demonstration of SAR tomography using multibaseline L-band data. IEEE Trans Geosci Remote Sens 38(5):2142–2152
100. Fornaro G, Serafino F (2006) Imaging of single and double scatterers in urban areas via SAR tomography. IEEE Trans Geosci Remote Sens 44(12):3497–3505
101. Prati C, Rocca F (1993) Improving slant-range resolution with multiple SAR surveys. IEEE Trans Aerosp Electron Syst 29(1):135–144
102. Khenchaf A (2001) Bistatic scattering and depolarization by randomly rough surfaces: application to the natural rough surfaces in X-band. Waves Random Media 11(2):61–89
103. Pierdicca N, Pulvirenti L, Ticconi F et al (2007) Use of bistatic microwave measurements for earth observation. ESA/ESTEC contract 19173/05/NL/GLC, University of Rome, Rome, Italy
104. Toutin T, Gray AL (2000) State-of-the-art of extraction of elevation data using satellite SAR data. ISPRS J Photogramm Remote Sens 55(1):13–33

105. Schuler DL, Lee JS, De Grandi G (1996) Measurement of topography using polarimetric SAR images. IEEE Trans Geosci Remote Sens 34(5):1266–1277
106. Toutin T (2004) Radarsat-2 stereoscopy and polarimetry for 3D mapping. Can J Remote Sens 30(3):496–503
107. Gelautz M, Paillou P, Chen C, Zebker H (2003) Radar stereo- and interferometry-derived digital elevation models: comparison and combination using Radarsat and ERS-2 imagery. Int J Remote Sens 24(24):5243–5264
108. Leberl F (1990) Radargrammetric image processing. Artech House, Boston
109. Toutin T (1999) Error tracking of radargrammetric DEM from RADARSAT images. IEEE Trans Geosci Remote Sens 37(5):2227–2238
110. Renga A, Moccia A (2009) Performance of stereo radargrammetric methods applied to spaceborne monostatic-bistatic synthetic aperture radar. IEEE Trans Geosci Remote Sens 47(2):544–560
111. Renga A, Moccia A (2009) Effects of orbit and pointing geometry of a spaceborne formation for monostatic-bistatic radargrammetry on terrain elevation measurement accuracy. Sensors 9:175–195
112. Rigling D, Moses RL (2005) Three-dimensional surface reconstruction from multistatic SAR images. IEEE Trans Image Process 14:1159–1171
113. Lillesand TM, Kiefer RW (1979) Remote sensing and image interpretation. Wiley, New York, pp 283–289
114. Italian Space Agency (2007) COSMO-SkyMed system description & user guide. ASI-CSM-ENG-RS-093-A

Chapter 2
Multistatic Radar Systems

Paco López-Dekker, Gerhard Krieger, and Alberto Moreira

Abstract This chapter discusses the key elements in the design of a distributed multistatic synthetic aperture radar mission. A number of application domains are discussed, with an emphasis on single- and multi-baseline interferometric techniques, deriving lower and upper bounds to the required spacecraft separation. Several multistatic formation concepts (Cartwheel, Helix, etc.) are discussed within the general framework provided by the Clohessy-Wiltshire equations. Several canonical multistatic acquisition modes (bistatic, alternate bistatic, etc.) are introduced, and the particularities of standard SAR modes in a distributed mission scenario are discussed. Here, the drawbacks of burst-modes such as ScanSAR or TOPS in a multistatic configuration are highlighted. Relevant theory with regard to oscillator phase noise is introduced and followed by a discussion of several phase synchronization approaches. These include TanDEM-X like synchronization links, which are the preferred option for most high frequency mission concepts, and GNSS or data driven approaches, which may be adequate for less demanding or for lower frequency systems. At the end of the chapter, three novel proposed missions are discussed: the high-end Tandem-L mission; SIGNAL, a compact Ka-band mission; and PICOSAR, a C-band low cost passive add-on mission concept.

2.1 Introduction

Bistatic radars are strictly defined as systems where the transmitter and receiver are physically separated. A multistatic radar system extends this concept by using multiple physically separated receivers and/or transmitters so that at least two simultaneous radar data sets are generated. Although the term multistatic probably

P. López-Dekker (✉) • G. Krieger • A. Moreira
German Aerospace Center (DLR), Microwaves and Radar Institute, Münchner Straße 20, Wessling 82234, Germany
e-mail: Francisco.LopezDekker@dlr.de; Gerhard.Krieger@dlr.de; Alberto.Moreira@dlr.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_2,

brings to mind a system with a single transmitter and multiple receivers, a two element distributed radar system is actually sufficient to generate multistatic measurements: a full monostatic radar system serving as an illuminator for itself and for a second, receive-only, sub-system.

There are several aspects that make multistatic radar concepts appealing. From an application point of view, the multiple combinations of transmitters and receivers increase the dimensionality of the measurement space. For example, the TanDEM-X mission [1] exploits this to derive highly accurate Digital Elevation Models (DEMs) from single-pass interferometric Synthetic Aperture Radar (SAR) measurements (InSAR), thus adding a third dimension to the 2-D SAR data products. From an engineering point of view, multistatic systems are appealing because of their efficient use of resources, since a larger fraction of the scattered radar signal is collected, and because they allow breaking up a large and complex system into a collection of smaller and simpler individual subsystems.

This chapter discusses general aspects of space-borne multistatic SAR concepts. The discussion is restricted to formation-flying systems, with a strong emphasis on single- or multi-baseline interferometric applications. The goal is to provide the reader with an overview of the principal aspects to consider in the design of a multistatic system or mission. Detailed examples of several multistatic missions or mission concepts are discussed in Part 4 of this book.

2.1.1 Historical Overview

Building on the success of many spaceborne monostatic radar missions, which have become a fundamental component of the Earth Observation (EO) mix, during the last two decades, a number of multistatic radar mission concepts have been proposed. One driving concept was the idea of acquiring single-pass interferometric data to generate global DEMs. For example, in the early 1990s, NASA's Jet Propulsion Laboratory (JPL) proposed the TOPSAT mission concept. One of the two alternative implementations considered was an L-band formation flying multistatic system, which would have provided 30 m horizontal resolution DEMs with height errors between 2 and 5 m [2].

Following the success of their RADARSAT program, the Canadian Space Agency (CSA) also proposed the formation flying RADARSAT-2/3 C-band topographic mission concept [3, 4], which presents many commonalities with Germany's TanDEM-X mission. During the same period of time, France's Centre National d'Etudes Spatiales (CNES), proposed the revolutionary Interferometric Cartwheel concept, which added a formation of three compact receive-only spacecraft to an existing C- or L-band monostatic mission such as ESA's ENVISAT mission, or JAXA's ALOS [5].

The crown to this series of multistatic topographic missions is the aforementioned German TanDEM-X (TerraSAR-X add-on for Digital Elevation Measurements) mission [1], proposed in 2003 and in operation since the launch

of the TanDEM-X satellite in June 2010. Its successful operation has demonstrated the value and feasibility of multistatic radar missions, producing 12 m horizontal resolution DEMs with height accuracies better than 2 m, and successfully addressing the synchronization and relative positioning and navigation challenges associated with such a mission.

A non-interferometric mission concept was TechSat-21 (Technology Satellite of the twenty-first century) [6], studied by the US Air Force Research Laboratory as an experimental system to demonstrate a distributed, and in-space reconfigurable, radar using low-cost individual systems. This program was cancelled in 2003.

2.1.2 Definitions and Basic Principles

2.1.2.1 Monostatic and Bistatic Phase Center

One recurring term in this chapter will be that of the equivalent phase center. The concept of phase center is fundamentally monostatic in nature. For a monostatic radar the phase center is the electrical center of the antenna. The radar range to a target is defined as the distance between the target and the phase center of the radar. It is worth mentioning that for a moving radar, the physical position of the phase center changes between the transmission of the radar pulse and the reception of the corresponding radar echoes. In this case, the equivalent phase center for each pulse is located at the mid-point between those two positions. The same concept can be extended to a bistatic measurement, for which the equivalent phase center can be placed approximately in the mid-point between the position of the transmitter at the instant it transmits the pulse, and the position of the receiver when it receives the echoes.

For a synthetic aperture, the equivalent phase center depends on the position of the target and on the SAR processing. In general, it can be placed at the orbital position at which the radar sees a target with the Doppler frequency corresponding to the center of the processed Doppler band. For SAR systems using a Total Zero Doppler Steering law [7], this position will correspond to the point at which the radar sees the target in a direction orthogonal to its trajectory in Earth Centered Earth Fixed (ECEF) coordinates.

2.1.2.2 Multistatic Configurations

In this chapter three types of configurations are considered:

- *Fully-active* systems are formed by a number of in principle identical spacecraft with full radar systems [8] (Fig. 2.1, left panel). Therefore, each spacecraft can acquire monostatic images, or be used as a passive receiver recording the bistatic radar echoes of the signal transmitted by another satellite. It is also

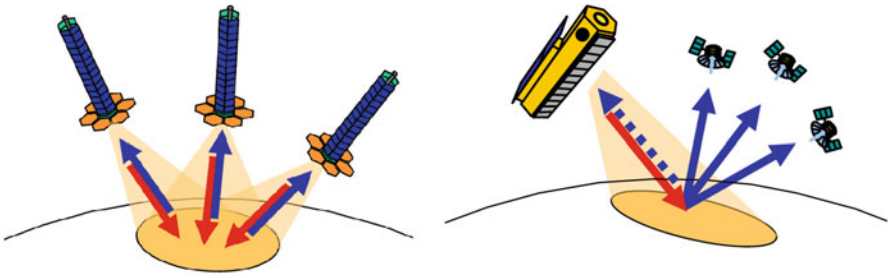


Fig. 2.1 *Left*: illustration of a fully-active multistatic system. *Right*: illustration of a semi-active system with a full SAR system as transmitter and a constellation of small receiver spacecraft

possible to interleave transmit pulses from multiple spacecraft and receive with all. This means that for N_s spacecraft, the number of non-redundant possible phase centers is

$$N_p = \frac{N_s \cdot (N_s + 1)}{2}. \tag{2.1}$$

Note that there is a quadratic relation between the number of spacecraft and the number of possible phase centers. However, fully benefiting from all the possible phase centers with more than two spacecraft would require complex and innovative techniques that are currently far beyond the state of the art. Fully-active configurations offer the maximum flexibility and redundancy, and are an attractive and cost-effective configuration for multistatic systems with a small number of spacecraft. A prime example of a fully-active system is TanDEM-X [1].

- *Semi-active* systems [8] are formed by a single active illuminator and a number of receive-only passive systems (Fig. 2.1, right panel) . Examples of semi-active concepts are the Interferometric Cartwheel [5, 9], or the BISSAT mission [10], both of which were conceived as low-cost add-ons to existing monostatic missions. Strictly speaking, this kind of semi-active add-on system yields one bistatic image for each passive spacecraft plus the original monostatic data. However, for most applications (e.g. SAR interferometry) only the bistatic ones can be combined, since the transmitter and the passive receivers are typically separated by a large along-track distance for safety reasons. Besides these opportunistic mission concepts, for future missions conceived from their beginning as multistatic, a semi-active architecture allows also the simplification of the illuminator, which could be transmit-only. In this case, continuous wave (CW) radar concepts could be developed, allowing for significantly lower peak power requirements [11].
- *Multi-monostatic* SAR systems consist of a number of monostatic SAR spacecraft operated as a distributed sensor [12]. Although these are not multistatic in nature, they have many commonalities both in terms of system design and of potential applications. In fact, a fully-active system can easily be operated in a

multi-monostatic mode, as it was done, for example, in the *pursuit monostatic* phase of the TanDEM-X mission.

Besides these three canonic configurations, it is clear that a number of hybrid combinations are possible. For example, a semi-active configuration with two illuminators would duplicate the number of bistatic phase centers.

2.1.2.3 Cooperative Versus Opportunistic Systems

Multistatic missions can also be classified as cooperative or opportunistic. Opportunistic systems, such as the Interferometric Cartwheel, are monostatic systems that use an existing system as a transmitter of opportunity. This system can be a monostatic SAR system, but also another type of source such as, for example, GNSS satellites [13]. Opportunistic systems are cost-effective, since the illuminating component is already provided. Their main technological drawback, or challenge, is the lack of explicit synchronization mechanisms.

In contrast, cooperative systems are from their onset designed to allow multistatic operations, including, for example, dedicated synchronization sub-systems.

2.1.2.4 Slow-Time and Frequency Synchronization

A key aspect of any multistatic mission is the synchronization of the system. Here we distinguish between two levels of synchronization:

- With *slow-time synchronization* it is meant the alignment of the transmit pulses and the acquisition of the echo windows by the multiple receivers. This level of synchronization requires a common time-frame. GNSS (GPS) time is accurate enough to meet the requirements, so a link between the on-board navigation system and the Instrument Control Unit (ICU) is required.
- *Phase synchronization* means the alignment in phase of the frequency sources used by the radar sub-systems, which include the phases of the Local Oscillators (LO), but also that of the sampling clocks. This level of synchronization is orders of magnitude more demanding than the slow-time alignment and will usually require dedicated sub-systems or the use of sophisticated processing techniques.

2.2 Application Examples

2.2.1 Cross-Track Interferometry for Topographic Mapping

Single-pass cross-track InSAR [14–16] is the most straightforward application of a multistatic SAR system. Since the launch of the second TerraSAR-X like satellite

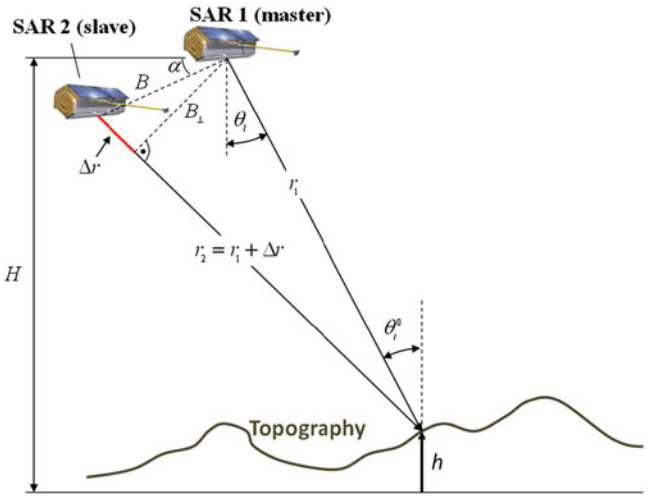


Fig. 2.2 Cross-track InSAR geometry. B and $B_\perp$ are the cross-track baseline and the normal component, respectively. θ_l and θ_i^0 are the look and nominal incidence angle, respectively. H and h are the flight altitude of the satellites and the topographic height, respectively. The slant-range distance of the two phase centers to the resolution cell is given by r_1 and r_2, and their difference by Δr

and the start of the TanDEM-X mission [1], it is also the first (and at the time of writing only) operational distributed radar application.

Cross-track interferometry (XTI) has been used successfully to derive digital elevation models (DEMs) of Earth for decades, in most cases combining data acquired during repeated passes of a single spacecraft. In these repeat-pass scenarios, the main contribution to the final error budget is the loss of coherence due to temporal decorrelation: the radar scattering coefficient is generally speaking a random process. To avoid this loss of coherence, single-pass interferometric configurations are required. This was done for the first time from space during the Shuttle Radar Topography Mission (SRTM) [17], which used a 60 m boom to provide a physical baseline between two receive antennas.

Cross-track InSAR is a powerful and well established technique to measure surface topography. The principle of cross-track interferometry is illustrated in Fig. 2.2 for the case of two spacecraft. Due to the separation between the two satellites, the distance from a point on the surface to the first satellite (r_1) and to the second one (r_2) differ by a factor Δr that can be written as

$$\Delta r = B \cdot \sin(\theta_l - \alpha) \tag{2.2}$$

where B is the cross-track separation of the two phase centers, and θ_l and α represent the look angle w.r.t. nadir and the orientation of the baseline with respect to the horizontal direction, respectively. Note that, due to the curvature of Earth, the look

angle is smaller than the nominal incidence angle θ_i^0. For a given resolution cell, this range difference is translated to an interferometric phase difference

$$\phi = n_w \cdot k_0 \cdot \Delta r \tag{2.3}$$

where k_0 is the wavenumber associated to the center frequency of the system and n_w is 2 for the multi-monostatic case and 1 when the receivers share a common transmitter. The power of interferometry results from the fact that the measurement of the interferometric phase allows estimating range differences with an accuracy of small fractions of a wavelength. This is typically several orders of magnitude better than the range resolution of the radar.

For a given resolution cell, the interferometric phase can now be related to the topographic height, h, with respect to a reference surface by

$$\phi = n_w \cdot k_0 \cdot \left(B \cdot \sin(\theta_l - \alpha) - \frac{B_\perp}{r_1 \sin \theta_i^0} h \right) \tag{2.4}$$

where $B_\perp$ is the nominal normal baseline component. The first term in (2.4) is a systematic phase, while the second term reflects the interferometric sensitivity to height, which is characterized by the vertical wavenumber

$$k_z = \frac{\partial \phi}{\partial h} = n_w \cdot k_0 \cdot \frac{B_\perp}{r_1 \cdot \sin \theta_i^0} \tag{2.5}$$

or by the so-called height of ambiguity

$$h_{amb} = \frac{\lambda \cdot r_1 \cdot \sin \theta_i^0}{n_w \cdot B_\perp}, \tag{2.6}$$

which is the change of topographic height that caused a phase difference of 2π.

These expressions show that the interferometric sensitivity to height and, therefore, the final height resolution is proportional to the cross-track separation of the two receiving phase centers. There are three factors that impose an upper limit to the cross-track interferometric baseline. Looking at a resolution cell with different incident angles introduces a so-called geometric decorrelation term, or spectral shift, between the interferometric pair [18]. This results from the fact that individual scatterers at different range positions within the resolution cell have different interferometric phases. The baseline for which this interferometric phase differences reaches a full cycle is called the critical baseline

$$B_{\perp,\, crit} = \frac{2 \cdot B_{rg} \cdot \lambda \cdot r_1 \cdot \tan(\theta_i^0 - \zeta)}{n_w c_0} \tag{2.7}$$

where B_{rg} is the transmit pulse bandwidth, and ζ is the local slope component in the ground range direction. When the baseline reaches this critical value there is total

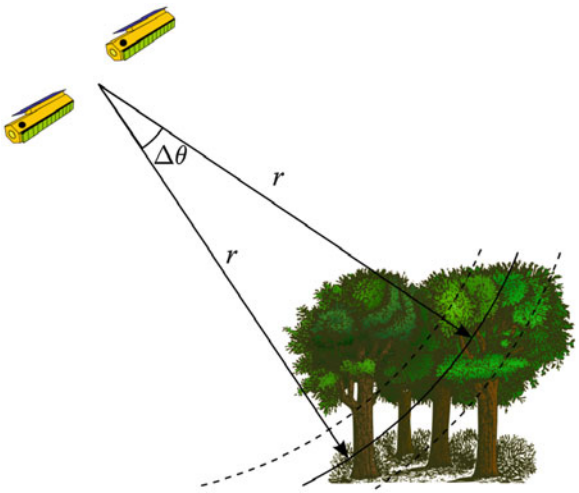


Fig. 2.3 Illustration of volume decorrelation in the case of a forest. Since the tree canopy is semi-transparent to radar waves, the received radar signal will contain contributions of scattering centers located at exactly the same slant range but at different elevations

loss of coherence. The impact of geometric decorrelation is limited by using baselines that are small compared to the critical one. This can be achieved simply by choosing a small physical cross-track separation between the interferometric acquisitions, or by increasing the pulse bandwidth, i.e. the range resolution, of the system at the cost of increased data rate and reduced system sensitivity. Interferometric SAR processing usually includes a common-band filtering step [18] that, in the case of distributed targets on a more or less constant slope surface, eliminates the decorrelating effect at the cost of a reduced range resolution.

The second limiting effect is volume decorrelation. Like in geometric decorrelation, volume decorrelation happens due to the existence of scatterers within a resolution cell with different interferometric phases. The difference in this case is that these scatterers may be located at exactly the same slant-range distance from the reference phase center but at different elevation angles due to their different vertical positions. This is illustrated in Fig. 2.3 for the typical case of a forest [19, 20].

For layered structures, the contribution of the volume decorrelation to the total interferometric complex coherence is given by

$$\gamma_{vol} = \exp(jk_z z_0) \frac{\int_0^{h_V} F(z') \exp(jk_z z') dz'}{\int_0^{h_V} F(z') dz'}, \tag{2.8}$$

where $F(z)$ represents the vertical scatterers distribution, h_v is the volume height, j is the square root of -1, and z_0 is the reference height. The first term in (2.8) represents the interferometric phase for a scattering center at the reference height. The absolute value of this volumetric coherence term is always smaller or equal

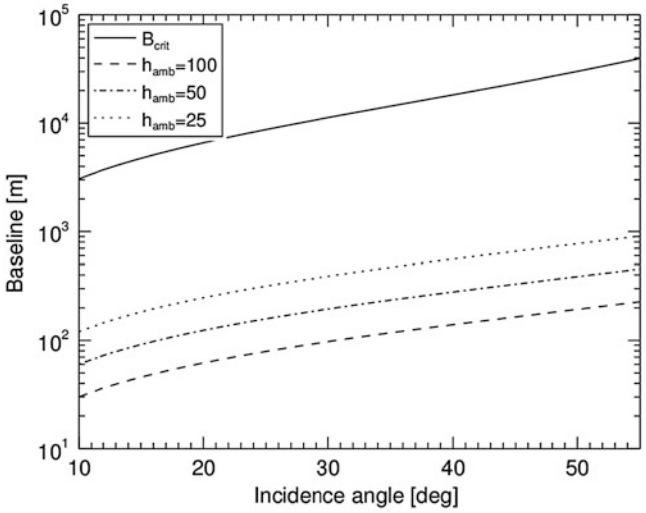


Fig. 2.4 Example of critical baseline and baselines corresponding to heights of ambiguity of 25, 50 and 100 m, respectively, as a function of the incidence angle. An X-band system with 150 MHz pulse bandwidth and an orbital height of 550 km has been considered

than one. For smooth distributions, the magnitude of γ_{vol}tends to be a decreasing function of the $k_z \cdot h_v$product, while it may show an oscillatory behavior if vertical discontinuities or discrete layers are present. In contrast to geometric decorrelation, the impact of volume decorrelation cannot be mitigated by improving the system resolution, nor can it be reduced using common-band filtering techniques. The only way to limit its impact is to have a height of ambiguity, h_{amb} that is large in comparison to the effective vertical spread of the scatterers.

The third limitation is the fact that the interferometric phase is wrapped modulo 2π, meaning that two heights that differ by a multiple of the height of ambiguity produce the same measured interferometric phase [21]. To resolve this ambiguity, a phase unwrapping step is required [22]. Phase unwrapping algorithms basically integrate phase differences between pairs of pixels. Generally speaking, since these phase differences are also ambiguous, there is no unique solution, so phase unwrapping becomes a matter of finding the most likely solution. In order to minimize phase unwrapping errors, phase differences between pairs of pixels should be smaller than half a cycle, which means that the height differences between those pixels should be smaller than half the height of ambiguity. It is worth noticing that these heights are always with respect to a reference surface, since it's a priori-known topographic phase is usually subtracted from the measured interferometric phase.

As an example, Fig. 2.4 shows the critical baseline and the normal baseline corresponding to heights of ambiguity of 25, 50 and 100 m, respectively, as a function of incidence angle for an X-band system with 150 MHz pulse bandwidth and an orbital height of 550 km. As a reference, typical heights of ambiguity used

for the TanDEM-X mission are in the range from 25 to 50 m. The reason not to choose larger baselines in this case, which would improve the topographic sensitivity, is to avoid phase unwrapping errors.

On top of the just discussed limitations on the cross-track baseline, there is also a limitation on the along-track separation of the spacecraft in order to receive interferometry-compatible echoes from a single transmitter. For any given target, the instantaneous relative Doppler shift, i.e. the difference of Doppler shifts of the radar echoes received by each spacecraft, can be approximated as [1]

$$\Delta f_D = \frac{v_{\text{sat}} \cdot B_{AT}}{\lambda \cdot r}, \tag{2.9}$$

where v_{sat} is the velocity of the spacecraft, B_{AT} is the along-track separation (or along-track baseline) and r is the mean slant range $r = (r_1 + r_2)/2$. To determine if this relative Doppler shift is significant, it must be compared to the available Doppler bandwidth. If the transmit and receive antenna have the same length (the typical monostatic case) this available Doppler bandwidth can be approximated by

$$B_{D,\,aval} \approx \frac{2 \cdot v_{ground}}{L}, \tag{2.10}$$

where v_{ground} is the ground velocity and L the antenna length. For unequal antennas, an approximation to the effective antenna length, L_{eff}, is given by

$$L_{eff}^2 = \frac{L_{tx}^2 + L_{rx}^2}{2}. \tag{2.11}$$

Now, it is clear that in order to receive any signal the footprints of the transmit and receive antenna patterns must overlap, which requires defining an antenna pointing strategy. The two most basic options are:

- Maximum gain steering: in this case the receive beams are steered such that the maxima of the transmit and receive antenna patterns point to the same point on the ground. This maximizes the power content of each individual Doppler spectrum, but reduces the available common Doppler spectrum by Δf_{D}.
- Maximum common Doppler steering: in this case the receive beams are steered in order to equalize the main-lobe of the Doppler spectrum. For a system with identical transmit and receive antenna patterns (e.g. TanDEM-X), this is accomplished by applying the same steering to both receive antennas.

The effect of an along-track separation on the Doppler spectra is illustrated in Fig. 2.5, which shows the two-way antenna pattern parameterized as a function of Doppler (this can be directly interpreted as Doppler spectra) frequency for several cases. In all plots, the blue and red solid lines correspond to the maximum gain optimized master and slave patterns, respectively. The discontinuous lines correspond to the common-Doppler optimized patterns, with the 3-dB Doppler spectrum indicated by the thicker lines.

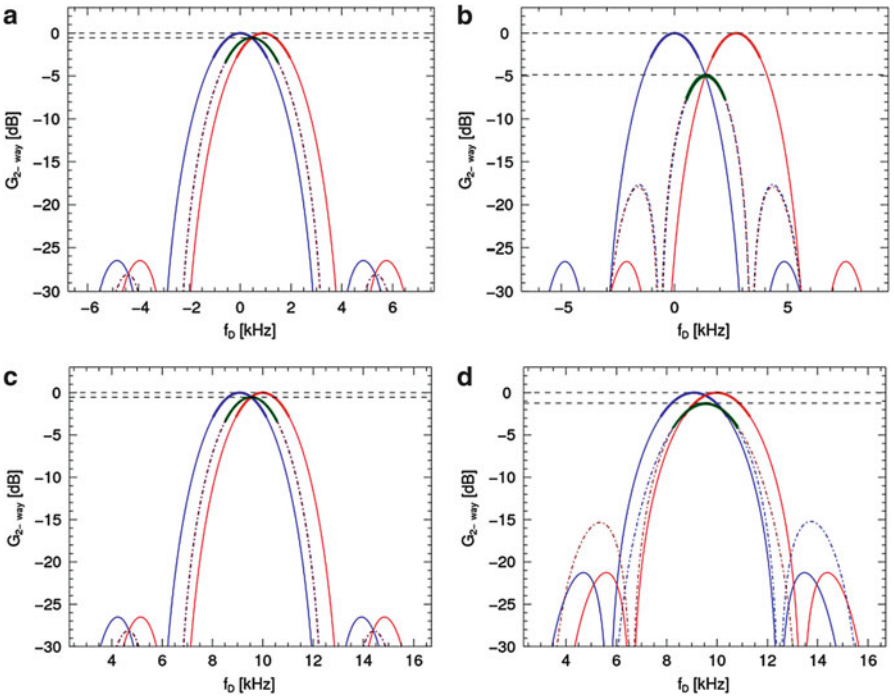


Fig. 2.5 Examples of two-way antenna patterns as a function of the Doppler frequency. In all cases, the blue and red solid lines correspond to the maximum gain optimized master and slave patterns, respectively. The discontinuous lines correspond to the common-Doppler optimized patterns, with the 3-dB Doppler spectrum indicated by the thicker lines. (**a**) A TanDEM-X like system with $B_{AT} = 2$ km. (**b**) Same as (**a**) but for $B_{AT} = 6$ km. (**c**) Same as (**a**), but assuming a separated independent transmitter 20 km ahead of the master receiver. (**d**) Same as (**c**), but reducing the receive antenna length to half its original size (2.25 m instead of 4.5 m)

The first case (a) corresponds roughly to TanDEM-X for an along-track separation of 2 km. Clearly, when both two-way patterns are optimized for gain, the result is a significant loss of common Doppler bandwidth (considering the 3-dB bandwidths). By applying an identical steering to both receive patterns, identical Doppler spectra are obtained. The cost, in this example, is a small two-way gain loss of roughly 0.5 dB. It is worth pointing out that for the case of TanDEM-X the steering applied is simply the one-way total zero Doppler steering. The result is that one of the patterns (the monostatic one) has a higher gain than the other (the bistatic one), which is a slightly sub-optimal but simple and robust solution.

The second case (b) illustrates what happens if B_{AT} becomes too large (6 km in this TanDEM-X like scenario). Now the gain optimized spectra are totally non-overlapping and, therefore, useless for interferometry. For the common-Doppler optimized patterns we can observe: a considerable gain loss (roughly 5 dB); a reduced 3-dB Doppler bandwidth; and a very large increase of side-lobe levels relative to the main-lobe level that would result in much larger azimuth ambiguity levels.

The third case (c) considers a multistatic system consisting of a common transmit spacecraft located 20 km ahead of the first of two receive-only satellites, which are themselves separated by $B_{\mathrm{AT}} = 2$ km. Otherwise the system specifications are still roughly those of TanDEM-X. The transmit system is assumed to point at its zero Doppler. Both receive systems have a large common-mode steering so that the receive- and transmit-pattern footprints overlap, resulting in large Doppler centroids. Other than that, the resulting Doppler spectra are virtually identical to those of the first example.

The last case (d) is similar to the previous one, with the important difference that now the antenna length of the receiver (L_{rx}) has been cut in half, yielding a 2.25 m long antenna instead of the original 4.5 m, which is reflected in the wider gain optimized two-way patterns. Different from all previous cases, and in order to achieve equal Doppler spectra, it is necessary to apply significantly different squint angles to each receive antenna: in this particular case this squint difference was 0.3°, which is almost as large as the one-way transmit 3-dB beamwidth. This results in degraded side-lobes and in a considerable gain loss (roughly 2 dB, which is much worse than the 0.5 dB of the reference case). Also, most of the 3-dB main-lobe of the common-Doppler optimized spectrum is below the gain-optimized spectra. This suggests that, in this particular case, the global interferometric performance would be better for gain-optimized patterns than for common-Doppler optimized ones, and that there must be some globally optimum in-between solution.

To achieve single-pass baselines in the order of hundreds of meters, as in the previous example, a formation-flying multistatic system is the only viable option. Since the required baselines scale with the wavelength, this distributed mission solution becomes even more attractive as lower frequencies are used. For example, in the Tandem-L mission proposal baselines of several km are required to achieve the mission objectives.

Besides achieving the desired large baselines, a distributed system allows modifying this baseline during the mission. For example, a smaller baseline may be used during a first phase of a topographic mission in order to obtain a first iteration of a DEM without suffering from phase unwrapping problems. In a second phase, the use of a larger baseline will then allow the refinement of this DEM. Such a strategy has been employed for TanDEM-X (cf. Chap. 13).

2.2.2 Multi-Baseline Cross-Track Interferometry

From the discussion in the previous section, it should be intuitively clear that a multi-baseline single-pass cross-track interferometric system could add significant value with respect to a single baseline configuration such as implemented in the TanDEM-X mission. While the simplest approach is to iteratively process interferograms with decreasing height of ambiguities, there are a number of approaches that directly combine all the available information. This allows the

unambiguous estimation of the topography from a set of SAR images that taken pair-wise would in all cases suffer from unsolvable ambiguities.

Some of these techniques evaluate the phase increments between adjacent pixels in multiple co-registered interferograms, thereby taking advantage of the deterministic relation between the baseline length and the phase gradient [23]. Other techniques resolve height ambiguities independently for each pixel [24], e.g. in an Earth based coordinate system by using a maximum likelihood approach [25]. This offers the opportunity to fuse multiple height estimates acquired in completely different imaging geometries, but requires also a precise knowledge of the remainder phase (more precisely: remainder topographic height) for each interferometric image pair. Still other techniques regard height determination as a direction of arrival (DOA) estimation in a sparse array and apply e.g. model based spectral estimation techniques to fuse the information from multiple SAR images in an unambiguous height estimate [26]. Interferometric acquisitions with large baselines may also use a joint multi-channel processing to minimize deteriorations from slope dependent baseline decorrelation [27].

2.2.3 Polarimetric SAR Interferometry

The combination of InSAR with SAR Polarimetry [28, 29] results in Polarimetric SAR Interferometry (PolInSAR) [30, 31]. In PolInSAR the capability of SAR Polarimetry to separate scatterers by their polarimetric signature is combined with the capability to measure height, via the direction of arrival in elevation estimation, provided by cross-track InSAR. For example, in the case of a forest, the canopy will generate a cross-polar contribution in the radar echo that is associated to volume scatterers. In contrast, the ground return will show the polarimetric signature of single-bounce scattering, associated to rough surfaces, and double bounce scattering, associated to the dihedrals formed by the trunks and the ground.

As suggested by this example, one of the primary applications of PolInSAR is the estimation of vegetation heights. It is worth noting that this inversion is, generally speaking, non-trivial. In addition, in many cases, for example forests observed at C-, X-band, or higher frequencies, the attenuation of the signal by the canopy is so large that the contribution of the ground return is below the noise floor.

In practice, PolInSAR inversion algorithms extract the desired information from the estimated multi-channel (6×6) covariance matrices, which are estimated by spatially averaging over a region that is assumed to be homogeneous. The inversion in order to obtain, for example, forest height, is done using an assumed layered model for which the expected covariance matrix can be derived analytically. The simplest (and most robust) model is the Random Volume over Ground (RVoG) [20, 32] model, which assumes that a forest can be modelled as a homogeneous layer of volume scatterers of a given thickness at some distance of a homogeneous ground.

PolInSAR inversion can greatly benefit from a multi-baseline scenario [33]. First, different baselines will be optimal for single-baseline inversion for different vegetation heights. Second, the availability of multiple baselines can allow estimating vertical profiles more complex than a simple RVoG. These profiles can be described, for example, by a truncated Legendre series and, therefore, parameterized by a small number of coefficients [34]. Three or four baselines seem sufficient to accurately retrieve the mean vertical profile.

2.2.4 SAR Tomography

A natural extension to SAR interferometry is SAR tomography [35–38]. In SAR tomography, a number of acquisitions from (slightly) different cross-track positions are treated conceptually as a sparse array, allowing the discrimination of signal contributions (radar echoes) corresponding to the same slant-range and azimuth positions by their direction of arrival in elevation. SAR tomography allows real 3-D imaging of semi-transparent volume scatterers, and can be applied to retrieve the vertical structure of, for example, forests, dry soil, or ice.

The SAR tomography concept is also useful for non-transparent scatterers, where it allows to resolve geometric distortions inherent to SAR imaging such as foreshortening and, in particular, layover [39, 40]. The problem of layover is that targets at different ground range distance from the sensor, but with different heights, may be at the same slant-range distance of the SAR system. This requires slopes larger than the incidence angle. Layover effects are ubiquitous in the case of man-made targets, as illustrated in the example in Fig. 2.6. In this case, scatterers on

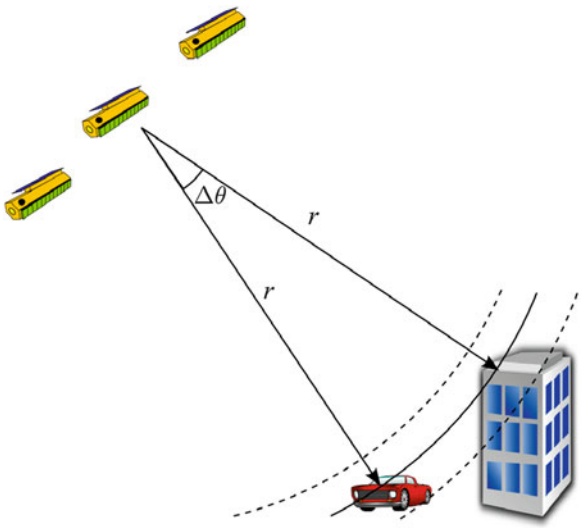


Fig. 2.6 Illustration of a layover situation: scatterers on the roof of the building will appear on the same range resolution cell in the SAR image as the car. Since both scattering centers are located at a different elevation angle w.r.t. the radar, they can be discriminated using array processing if many simultaneous acquisitions from slightly different positions are available

the roof of a building would appear in the same position as the car that is situated at ground level, but at a smaller ground range distance from the radar. The roof-top and the car would contribute to the same resolution cell (pixel) in the resulting SAR image. Since single-baseline interferometry assumes implicitly a model with a single effective scattering center to which a unique height can be attributed, it is clear that in the case of layover the phase to height conversion will yield an incorrect result. With more than two phased centers, array processing techniques (also known as beamforming techniques) can be used to separate the contribution of both targets, and to correctly position each of them.

SAR tomography from space has been demonstrated already by a number of authors [37, 41] using series of repeat-pass acquisitions. Like in interferometric applications, temporal decorrelation becomes a major challenge, in particular if multiple acquisitions are required to construct the tomographic data set. This restricts its applicability to scenes with very long coherence times, for example urban areas or natural scenarios observed at very low frequencies (e.g. at P-band).

Aside from temporal decorrelation, which can be eliminated in a single-pass multistatic configuration, the second challenge is the distribution of the phase centers. The ideal tomographic configuration is the one where the different effective phase centers are distributed uniformly in elevation, therefore regularly sampling the elevation signature of the radar signal. The optimum spacing between these phase centers can be derived from Nyquist sampling theory. For this, the signal received for echoes at a slant-range distance r, as a function of the effective normal baseline, $B_\perp$, with respect to some reference position can be written as

$$s_r(B_\perp, r) = C_0 \int \gamma(r, \theta_i) e^{-jn_w k_0 B_\perp \sin(\theta_i - \theta_i^0)} d\theta_i, \tag{2.12}$$

where $\gamma(\cdot)$ represents the complex reflectivity function that characterizes a given point in space, and C_0 groups all remaining terms that affect the amplitude of the received signal. Parameterizing the position as a function of height, using (2.5) in order to use k_z instead of $B_\perp$, and dropping the explicit dependence from R yields the following relation

$$s_r(k_z) = C_1 \int \gamma(h) e^{-jk_z h} dh, ss \tag{2.13}$$

where the new constant C_1 includes the scaling due to the variable change. This result shows a Fourier transform relationship between the received signal (as a function of the baseline/tomographic dimension) and the complex reflectivity (as a function of height). For a volume with effective thickness h_v, which can now be interpreted as a bandwidth, the normal baseline between adjacent acquisitions that meets Nyquist's sampling criterion is given by [42]

$$\Delta k_z \leq \frac{2\pi}{h_v}. \tag{2.14}$$

In terms of the minimum cross-track normal baseline, this means

$$b_{\perp,\min} \leq \frac{\lambda \cdot r \cdot \sin\theta_i^0}{n_w \cdot h_v}. \tag{2.15}$$

The vertical resolution is given by the length of the tomographic aperture. Thus, for a target resolution Δh, the maximum baseline must be such that

$$k_{z,\max} \geq \frac{2\pi}{\Delta h}, \tag{2.16}$$

or, in terms of baseline,

$$b_{\perp,\max} \geq \frac{\lambda \cdot r \cdot \sin\theta_i^0}{n_w \cdot \Delta h}. \tag{2.17}$$

Combining these results, it is easy to see that the number of effective phase centers in a canonical tomographic configuration is given by

$$N_{\text{tomo}} = \frac{h_v}{\Delta h}, \tag{2.18}$$

which is an intuitively satisfactory result stating that the number of tomographic samples should be equal to the number of independent tomographic resolution cells in the resulting volumetric image. From the system designer point of view, however, this result may easily lead to an unrealistic number of phase centers. It is worth stressing the point that the number of tomographic samples is not restricted to the number of receivers. Instead, it is given by the number of unique combinations between transmitter and receiver.

2.2.5 Sparse Tomography

The previous discussion suggests that a useful multistatic system capable of single-pass tomography requires a complex constellation with many spacecraft. It is obvious that this canonical solution has many drawbacks:

- High costs associated to such a system, even though the receiving systems could be low cost and compact, with little or no redundancy, and with unit costs reduced thanks to small series production.
- Data management. On-board data storage and data down-link capacity are already critical design constraints for single-satellite radar systems. A tomographic system with N effective phase centers will multiply this problem by the same factor N. The reader should note that this problem is not relaxed using a multiple active configuration.

- Formation design and control. A single-pass tomographic system requires a formation that provides the desired cross-track baselines with a relatively small along-track separation (to keep Doppler Centroid compatibility). While arbitrarily complex formations can be designed that fulfil these requirements, the addition of each new satellite tightens the formation control requirements and reduces safety margins.

In light of these issues, it is clear that solutions requiring less phase centers are desired. Fortunately, SAR tomography literature already offers a variety of processing approaches to handle these sparse scenarios using sparse array processing techniques [42–44]. The basic idea behind these methods is the assumption that there are only a few dominant scattering centers, which can be understood either as point-like scatterers (typical of urban environments) or as distributed volume scatterers spanning an angle smaller than the tomographic resolution of the system. Experimental studies suggest that, indeed, the number of tomographically separable scattering centers, N_{scat}, is typically below 3–5 [45]. This is in particular true for high resolution SAR systems.

If the number of scattering centers is known, then the tomographic imaging can be done using subspace methods such as the MUSIC algorithm or, in a brute force approach, finding the Non-linear Least Square (NLS) error solution (i.e. directly inverting a possibly over determined non-linear problem) [46]. For these approaches, a main challenge is to determine the number of relevant phase centers, which can be done using a priori knowledge or using information-theoretic criteria. For example several authors [35, 46] suggest using the Akaide Information Criterion (AIC) [47] to estimate the dimension (e.g. number of scatterers) of the tomographic data set. If the number of scatterers is known or can be robustly determined, then the number of required tomographic phase centers is lower bounded by

$$N_{\text{tomo}} \geq N_{scat} + 1. \tag{2.19}$$

Another promising tomographic processing approach uses the Compressive Sensing (CS) paradigm [43–45]. CS-based algorithms are capable of providing super-resolution in the tomographic dimension, i.e. a vertical resolution for a given maximum baseline that is better than the one given by (2.17), with a reduced number of measurements and without a priori knowledge of the number of scatterers. The number of required phase centers depends on the number of tomographic resolution cells, the number of scatterers, and the SNR. For 3–5 scatterers it appears that 15 phase centers ($N_{\text{tomo}} = 15$) provide robust performance.

2.2.6 *Along-Track Interferometry and Moving Target Indication*

The basic idea behind along-track interferometry [22, 48–51] (ATI) is the repeated observation of a scene under nearly identical geometry with a given temporal lag, τ_{ATI}. If a target is moving and this movement has a non-zero component in the

line-of-sight (LOS) direction, the resulting change of the slant-range will result in an interferometric phase

$$\phi_{ATI} = 2k_0 \cdot (\mathbf{v}_{\text{target}} \cdot \mathbf{r}_{LOS}) \cdot \tau_{ATI}, \tag{2.20}$$

where $\mathbf{v}_{\text{target}}$represents the target's velocity vector, and $\mathbf{r}_{\text{LOS}}$the line-of-sight vector from the center of the synthetic aperture to the target. Since the interferometric phase is wrapped modulo 2π, it is generally desirable that the ATI phase is constrained to the interval $[-\pi, \pi]$. Considering the range of velocities to be covered, this can directly yield the order of magnitude of the desired temporal lag. For very slow movements, for example those associated to terrain subsidence, these temporal scales are in the order of days to months, or even years. These scales are addressed by means of repeat-pass interferometry: repeated observations made by different passes of a single spacecraft. For large velocities, such as those of vehicles or ocean currents, these scales are typically in the order of milliseconds to tenths of seconds. This is the realm of along-track interferometry, requiring two acquisitions in a single pass by means of two along-track separated phase centers. Temporal lag and along-track separation or baseline, B_{AT}, are related by

$$\tau_{\text{ATI}} = \frac{n_w \cdot B_{\text{AT}}}{2 \cdot v_{\text{sat}}}, \tag{2.21}$$

where B_{AT} is the physical separation between the phase centers. Typically, ATI systems use a common transmitter, so that $n_{\text{w}} = 1$. The term $B_{\text{AT}}/2$ is then the effective along-track baseline between the equivalent bistatic phase centers.

Analogous to the height of ambiguity in XTI, it is useful to define an unambiguous velocity range, v_{amb}, which results from setting the ATI phase to 2π, and solving for the velocity.

Figure 2.7 shows the horizontal unambiguous velocity range as a function of along-track baseline for L-, C-, X- and Ka-band. For ground moving targets, these horizontal velocities can reach around 50 m/s or more (high speed trains, cars on highways), making v_{amb} in the order of 100 m/s a reasonable design goal. The associated baselines are relatively small, for example, around 10 m at C-band and 5 m at X-band. These short scales are in principle not well suited for a distributed system, calling instead for single platform solutions such as the Dual Receive Antenna (DRA) available in TerraSAR-X.

Instead, a multistatic ATI system is ideally suited for ocean currents and other geophysical phenomena resulting in velocities in the order of 1 m/s. For the case of ocean currents, there is an additional effect that needs to be taken into account when considering the optimal along-track baseline: temporal decorrelation of the surface. InSAR assumes that the target being observed is the same for all the acquisitions. Since the ocean surface is constantly changing, the assumption its radar scattering coefficient does not change is only valid for a short period of time. There is therefore a trade-off between the long baseline desired to maximize the ATI sensitivity, and the short baseline needed to maintain coherence between

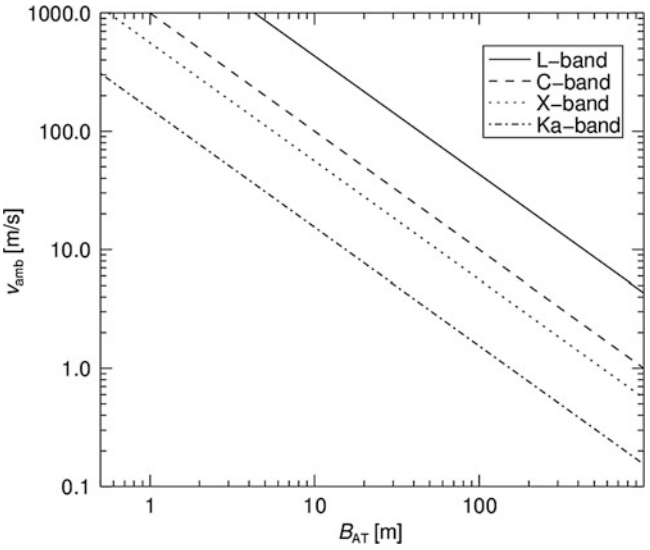


Fig. 2.7 Horizontal velocity of ambiguity as a function of the along-track baseline (B_{AT}) for L-, C-, X- and Ka-band

acquisitions. This will be discussed in some detail in Sect. 2.4.3, in the context of the PICOSAR mission concept.

2.2.7 Geometric and Radiometric Resolution Enhancement

In the previous sub-sections we have discussed applications in which multiple receivers are used to add an additional dimension to the data: for each resolution cell we now have two or more channels available. The next two sub-sections discuss the use of distributed multistatic systems to improve the quality of, or extend, the single-channel SAR images. The goal, therefore, is to use a distributed system, typically comprised of relatively small spacecraft, to synthesize the equivalent to a much larger SAR system.

One of the main effects hindering the interpretation of SAR data is the presence of multiplicative speckle *noise* [52–54]. The result is that the radar cross section measured for a resolution cell containing multiple scattering centers is essentially a random variable which, in the case of fully distributed scatterers, follows an exponential distribution (for which the standard deviation equals the mean value). To obtain reliable estimates of the true radar cross section, or the normalized radar cross section, some sort of averaging is required. For single acquisitions using a single-channel system, this averaging is done in the spatial domain, by averaging a number of neighboring pixels, or, equivalently, by dividing the SAR image in sub-bands in the Doppler frequency domain and/or in the fast-time frequency domain and then

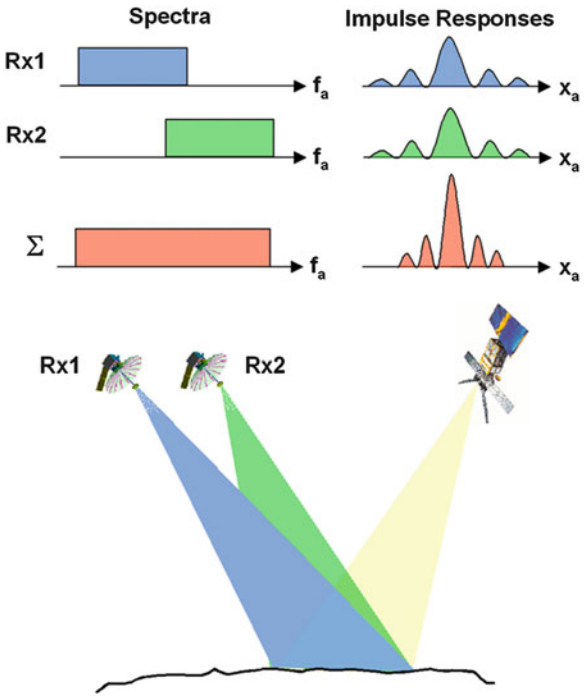


Fig. 2.8 Super-resolution with a passive microsatellite formation. A coherent combination of the spectra obtained from slightly different view angles will allow for an improved resolution in range and/or azimuth

averaging the powers of the sub-band SAR images. The result of this multi-looking process is a better estimate of the Normalized Radar Cross Section (NRCS) at the cost of a reduced product resolution. Also, multi-looking assumes implicitly that the scene can be considered homogenous over the averaged pixels. The monostatic approach to achieve high spatial and good radiometric resolution is to increase the single-look resolution of the system. One approach is to increase the transmitted pulse bandwidth, which is legally upper limited by spectrum allocation regulations, and which requires a proportional increase of the transmit power in order to maintain sensitivity. The second approach is to increase the Doppler bandwidth, which also requires an increase in the transmit power.

An alternative way to increase the resolution (and the number of available looks) is to combine acquisitions that sample different parts of the spectrum [55, 56] (Fig. 2.8). In the Doppler domain this requires acquiring images with different Doppler centroids: while the common parts of the Doppler spectra yield interferometric information, the non-overlapping parts can be used to increase the azimuth resolution. As discussed already in Sect. 2.2.1 (see Fig. 2.5), a set of passive along-track separated receivers will naturally yield a Doppler centroid difference. From a power budget point of view, this is a very appealing way to gain resolution because it does not require broadening the transmit beam, like in a monostatic system, so the additional resolution does not require additional transmit power. If the primary goal is to increase the resolution, then the different multistatic images need to be

coherent and the aggregated Doppler spectrum must be continuous. However, if the purpose is multi-looking, both of these requirements may be dropped.

Super-resolution can also be achieved in the range domain. Ground range resolution is determined by the portion of the ground spectrum (in the wavenumber domain) sampled by the radar pulse. The ground range wavenumber, k_y, is related to the slant range frequency by [18]

$$k_y = 2 \cdot k \cdot \sin(\theta_i^0 - \zeta) = \frac{4\pi f}{c_0} \sin(\theta_i^0 - \zeta), \tag{2.22}$$

where, as before, θ_i^0 is the nominal incidence angle and $\zeta$is the terrain slope. For a single SAR system, the ground range resolution is given by

$$\Delta y = \frac{2\pi}{\Delta k_y} = \frac{c_0}{2 \cdot \Delta f \cdot \sin(\theta - \zeta)}, \tag{2.23}$$

where Δf is the pulse bandwidth, and c_0 the speed of light. From (2.22) it is clear that changing the incidence angle will change the ground range wavenumber, so SAR images acquired with a different incidence angle, which results from having a cross-track baseline can be combined into a single higher resolution image. Again, a multistatic constellation with a set of cross-track separated receivers, would allow achieving this range super-resolution without increasing the transmit power. Also, as before, for the sake of multi-looking it is not necessary to coherently combine the different images, which also allows baselines well above the critical one.

It is maybe worth noting that while the combination of different portions of the Doppler spectrum is a natural extension of the synthetic aperture concept, and can be understood as an optical imaging process, super-resolution in ground range is a mixture of slant range resolution with angular resolution. This combination is less natural and more prone to problems. For example, a good knowledge of the terrain slopes is required.

2.2.8 *Unambiguous Wide-Swath Imaging*

The last application that will be discussed here is the use of a multistatic configuration to achieve High-Resolution Wide-Swath (HRWS) SAR imaging. One of the traditional limitations of single-channel SAR systems is the relation between resolution, swath width, and antenna size. The requirement for a high strip-map mode azimuth resolution implies a large Doppler bandwidth, traditionally implying a short antenna (since the best achievable strip-map resolution can be approximated by $L/2$ where L is the antenna length), which in turn requires a high PRF to avoid azimuth ambiguities. The high PRF implies a narrow unambiguous swath, which means that a wide antenna is needed in order to provide the narrow beam in elevation required to suppress range ambiguities. Aside from the resulting tradeoff

between azimuth resolution and swath width, this argumentation suggests a lower bound on the antenna area, which indeed is reported in many SAR references [57, 58]. This bound is given by

$$A = W \cdot L > \frac{4 \cdot v_{sat} \cdot \lambda \cdot r_m}{c_0} \tan \theta^0_{i,m}, \tag{2.24}$$

where A is the total antenna area, W is the antenna width, and r_m and $\theta^0_{i,m}$ are the slant range and the nominal incidence angle at the center of the swath, respectively. Although this lower bound can be relaxed if a resolution worse than $L/2$ is accepted, which allows lowering the PRF, or if a swath width smaller than the unambiguous swath is allowed [59], it is a good starting point for a system design.

For single-channel systems, given an antenna area, azimuth resolution may be traded off against swath width using burst modes like ScanSAR or TOPS (Terrain Observation by Progressive Scans) [60]. In this trade off, the amount of independent resolution cells per azimuth unit length remains, at best,[1] constant. Azimuth resolution can be improved using Spotlight modes, limiting the continuous scene length. Here again the average information rate remains constant, since azimuth SAR image segments with increased information density are followed by segments for which no information can be acquired.

While the minimum area bound cannot really be overcome without performance degradation, azimuth resolution can be decoupled from swath width by introducing multiple channels and applying digital beamforming (DBF) techniques [61–63]. From an intuitive point of view, it seems plausible that increasing the number of samples acquired by a SAR system will increase the number of independent pixels in the resulting SAR image, either by improving its resolution, or by increasing its extent. DBF SAR systems use one, or a combination, of the following two principles:

- An array of receivers in the along-track direction allows reducing the PRF required to sample a given Doppler bandwidth. From a beamforming point of view, this is accomplished by generating a number of narrow receive beams in azimuth. From Nyquist sampling theory it follows that the narrower Doppler bandwidth associated to these narrow beams allows a lower PRF. An alternative way of understanding the reconstruction process is the following: for each receiver an independent image may be formed which, due to the low PRF, will suffer from high azimuth ambiguities. Then, these images are combined using DBF methods that will place nulls at the positions of these ambiguities.
- An array of receivers in the cross-track direction allows discriminating echoes from different range directions by their direction of arrival. This allows the simultaneous acquisition of a number of strip-map sub-swaths. Here again, the key of DBF is the capacity to place zeroes at the range ambiguities.

[1] Usually there are unavoidable gaps between the modes which means that the independent pixel count goes down with respect to strip-map.

DBF SAR techniques will play an important part in future and at time of writing planned monostatic missions. But they can also be implemented in a distributed multistatic mission concept [64–67]. The idea is that instead of requiring a large and complex system, a multi-channel receiver may be distributed in a number of small spacecraft.

Aside from the obvious implementation differences, the main conceptual difference between a single spacecraft DBF system and its distributed, multistatic counterpart, is that in the first case we are dealing with an array of optimally and regularly spaced elements, while in the second we have a more or less random sparse array. For regular dense arrays DBF can be implemented using matched filter based approaches.

For sparse arrays, however, a matched filter may have large side-lobes at the locations of the ambiguities that need to be suppressed. One possible solution is to use a Maximum Likelihood (ML) beamformer, which, in the case of white additive noise, basically pseudo-inverts the received signal, placing nulls at the ambiguities [67]. Although ML-based DBF gives the best possible performance in terms of Signal-to-Clutter (ambiguities) Ratio (SCR), depending on the positions of the receivers the pseudo-inverse problem may become ill-conditioned, which in the presence of noise can result in the appearance of large imaging artifacts.

An alternative processing approach is the use of Minimum Mean Square Error (MMSE) estimator DBF. Note that although MMSE will avoid the introduction of large noise induced artifacts, it does not really solve the problem. At positions where the ML estimator would generate a large artifact, the MMSE estimator will put a value based on the a priori assumed scene statistics and not on the measurement. The real solution is to over-dimension the system by introducing extra receiving elements in order to decrease the probability of having an ill-conditioned problem.

Another problem associated to sparse arrays is that the large separation between elements makes them very sensitive to phase errors, which is a particular issue for ML or MMSE processing approaches. This means that very precise phase synchronization/calibration is required between the multiple receivers, something that is already generally complicated in a multistatic context. All together, the appeal of using a distributed system for HRWS imaging is that the system may be used at the same time for some of the previously discussed interferometric or tomographic applications.

2.3 Mission Design

2.3.1 Formation Selection

The multistatic systems and applications considered in this chapter assume a constellation of formation flying spacecraft. In this section several formation concepts will be discussed. In general, an important design goal is to have well controlled and more or less constant baselines between the spacecraft. For along-track interferometry this can be accomplished by placing the spacecraft on the same orbital plane and with the same orbital parameters but with a small

along-track separation. For this kind of configuration the only, but important problem, is dealing with random orbital drifts that, if not taken care of, could lead to a collision between the spacecraft.

However, such a simple solution is not possible in the case of cross-track interferometry since neither the vertical nor the horizontal cross-track separations remain constant for free-flying satellites on natural orbits, which are necessary to keep the fuel consumption within reasonable limits. The relative motion of free-moving satellites in close formation may be approximated by Hill's equations [68], also known as Clohessy-Wiltshire equations [69], which describe the satellite movements in a rotating reference frame [70]. This transformation allows a linearization of the differential equations characterizing the satellite dynamics. For unperturbed, Keplerian motion and a circular reference orbit with period T_0, a solution of the Clohessy-Wiltshire equations is given by [71, 72]

$$\begin{aligned}\Delta r_r(t) &= -a \cdot \delta e_i \cos\left(\frac{2\pi}{T_0}t - \alpha_i\right)\\ \Delta r_t(t) &= 2a \cdot \delta e_i \sin\left(\frac{2\pi}{T_0}t - \alpha_i\right) + a \cdot \Delta u_i,\\ \Delta r_n(t) &= a \cdot \delta i_i \sin\left(\frac{2\pi}{T_0}t - \beta_i\right)\end{aligned} \tag{2.25}$$

where a is the semi-major axis, Δu_i the relative argument of latitude of satellite i, and the magnitude δe_i together with the angle α_i and the magnitude δi_i together with the angle β_i define the so called relative inclination and relative eccentricity vectors, respectively. The variables Δr_{r}, Δr_{t}, and Δr_{n}, represent the relative position of satellite in the radial (vertical) direction, the direction of movement (along-track), and the cross-plane direction (horizontal across-track), respectively. The term $a\Delta u$ introduces a constant along-track offset. The first two equations describe an ellipse in the (nominal) orbital plane, yielding a harmonic vertical cross-track baseline oscillation and a 90° out of phase harmonic along-track baseline oscillation with twice the amplitude. The last term introduces an independent harmonic oscillation in the cross-plane direction.

2.3.1.1 Interferometric Cartwheel

One of the first and best known suggested multistatic configurations is the Interferometric Cartwheel suggested by Massonet in 1998 [5, 73]. The Cartwheel formation consists of a set of receive-only microsatellites flying in close formation on the same orbital plane along slightly elliptical orbits. In its basic configuration, all satellites have the same eccentricity and the same semi-major axis, i.e. $\delta e_{\mathrm{i}} = \delta e$ in Hill's equations. Since all satellites share the same orbital plane ($\delta i_{\mathrm{i}} = 0$), the horizontal cross-track displacement, Δr_{n}, vanishes for all orbital positions. To obtain the interferometric baselines, the arguments of perigee differ between the satellites,

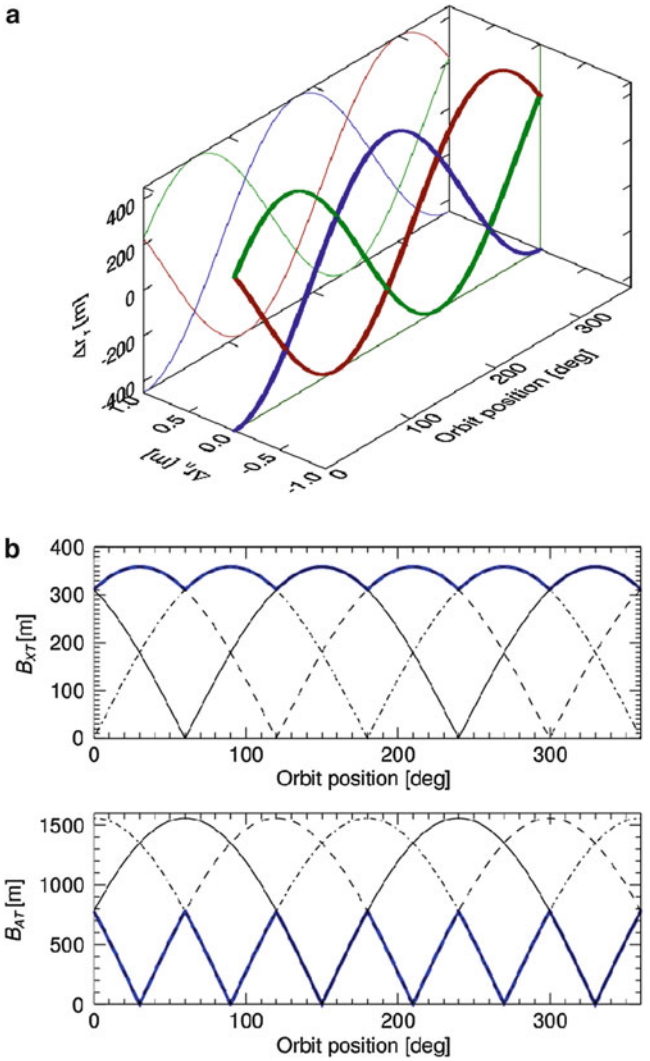


Fig. 2.9 (**a**) Relative cross-track deviations with respect to mean (nominal) orbit for a 3-spacecraft Interferometric Cartwheel as a function of the mean argument of latitude (u). There are no deviations in direction normal to the common orbital plane. (**b**) Resulting effective cross-track (*top*) and along-track baselines (*bottom*)

leading to a relative phase shift between the radial components $\Delta r_r(t)$ due to different α_i. For example, a Cartwheel with three satellites will have relative phase shifts

$$\alpha_3 - \alpha_2 = \alpha_2 - \alpha_1 = 120^\circ. \tag{2.26}$$

The upper panel in Fig. 2.9 shows an example of Δr_r (vertical axis) and Δr_n (which is always zero) for this kind of 3-spacecraft configuration for $a \cdot \delta e = 450$ m.

Note that, as long as there are no uncompensated along-track relative drifts with respect to the nominal formation, the 90° out-of-phase Δr_t ensures that the spacecraft do not collide.

The lower panel shows the norms of the effective cross-track and along-track baselines for each combination of spacecraft. It also shows (blue thick line) the vertical baseline envelope (maximum of the three baselines) and the corresponding along-track baselines. A remarkable feature of this configuration is that the maximum baseline is very stable along the orbit, with the smallest value only 13% below the largest. Another desirable feature is that these maximum cross-track baselines appear always in combination with minimum along-track baselines.

The Interferometric Cartwheel assumes a separated transmit satellite. In fact, it was proposed as a passive add-on formation to existing or planned monostatic missions that would operate without any impact on the primary system.

2.3.1.2 Cross-Track Pendulum

The cross-track Pendulum is another three passive spacecraft formation concept which can be understood as the complement (or opposite) to the Cartwheel. Instead of using slightly elliptic orbits and adjusting the argument of perigee, in the Pendulum configuration what changes is the orbital plane, either by changing the orbital inclination and/or the ascending nodes. Thus, for the Pendulum $\delta e_i = 0$, so that the along-track separation, according to (2.25), is nominally always constant. Note that this forces setting and maintaining different Δu values for each spacecraft, in order to avoid collisions. Figure 2.10 illustrates a Pendulum configuration with

$$\beta_3 - \beta_2 = \beta_2 - \beta_1 = 120^\circ. \tag{2.27}$$

To follow the previous example, the amplitude of the harmonic oscillation is set by $a \cdot \delta i = 250$ m. The cross-track baseline follows the same behavior as for the Interferometric Cartwheel, with a very constant envelope, but with the difference that in this case the baselines are horizontal. For incidence angles smaller than about 52°, a horizontal baseline gives a larger effective baseline than a vertical one, which means that, generally speaking, a Pendulum configuration would require smaller relative motions than its Cartwheel counterpart.

Since the along-track baselines are constant, the one associated to the largest cross-track baseline jumps between two values in this example where the three spacecraft were uniformly distributed with 200 m along-track separation between consecutive ones. The minimum spacing has to be large enough to guarantee formation safety even in the event that one or more spacecraft have to go into save mode.

A disadvantage of this three satellite Pendulum configuration is that it requires that the satellites have slightly different inclinations, which means that their slightly different orbital planes will have slightly different precession rates. In particular, at most one of the spacecraft can have the exact inclination required to achieve a frozen Sun-synchronous orbit, as required for most SAR missions. This requires

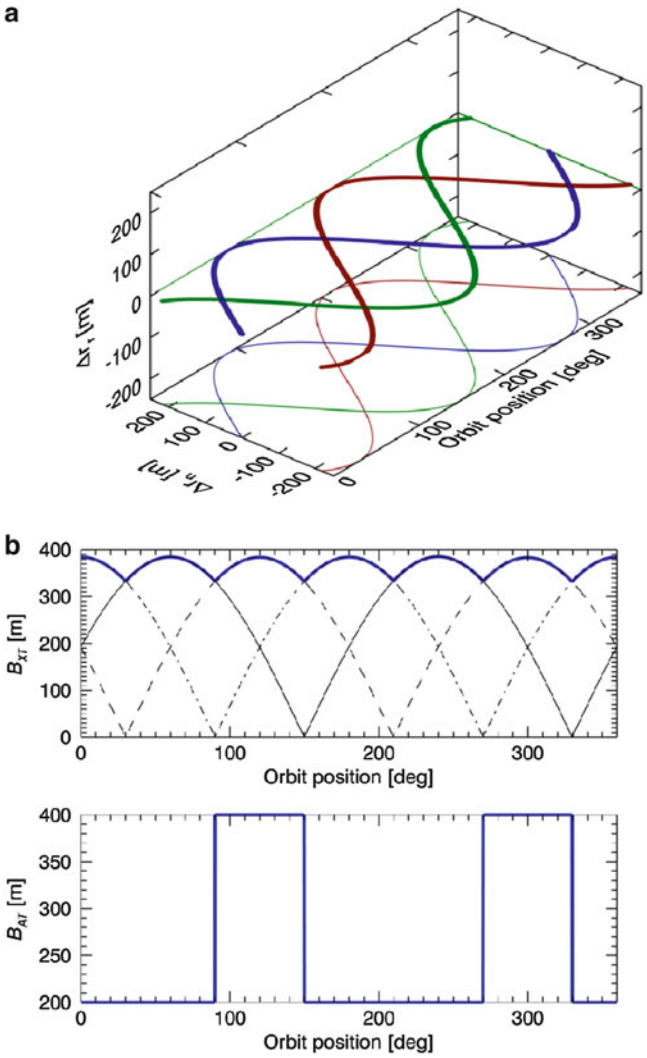


Fig. 2.10 (**a**) Relative cross-track deviations with respect to mean (nominal) orbit for a 3-spacecraft cross-track Pendulum as a function of the mean argument of latitude (u). (**b**) Resulting effective cross-track and along-track baselines

some extra fuel for formation keeping. This last problem can be avoided by using a so-called Trinodal Pendulum formation in which the three satellites have the same inclination and eccentricity but different ascending nodes. This leads to a constant baseline ratio, which is of interest for multi-baseline applications. The drawback is that all the XTI baselines tend to vanish at high latitudes.

2.3.1.3 CarPe

The Cartwheel and the Cross-track Pendulum concepts can be combined in a CarPe configuration. In this case two spacecraft can have a slightly different ascending node (but identical inclination), and the third one has a small relative eccentricity and argument of perigee that maximizes the deviation from the reference orbit when the other two spacecraft have a vanishing cross-track separation:

$$\begin{array}{l} \beta_1 = 90^\circ, \beta_2 = 270^\circ \\ \alpha_3 = 90^\circ \end{array} . \tag{2.28}$$

Figure 2.11 shows an example with $\mathrm{a}{\cdot}\delta i_1 = \mathrm{a}{\cdot}\delta i_2 = 450$ m and $\mathrm{a}{\cdot}\delta e_3 = 650$ m. These numbers have been chosen semi-arbitrarily so that the vertical and horizontal baselines result in similar effective cross-track baselines. This effective baseline, however, will depend on the incidence angle. This means that, except for a nominal incidence angle, the envelope of the effective baseline will have larger variations than the Cartwheel or the Pendulum configurations. This is illustrated in the bottom panel of the figure, for an incidence angle of 45°, which corresponds to a 40.5° look angle for an orbit height of 550 km.

The advantage of a CarPe formation with respect to a Cartwheel is that it combines a quite stable cross-track effective baseline envelope with one constant along-track baseline, which is appealing for ATI applications.

2.3.1.4 Helix

The last formation considered here is the Helix formation used in the TanDEM-X mission [1, 71, 72]. In this particular case, the formation consists of two spacecraft with

$$\begin{array}{l} \delta e_i = \delta e \neq 0 \\ \delta i_i = \delta i \neq 0 \\ \alpha_i = \beta_i = a_{i-1} + \Delta\alpha \end{array} . \tag{2.29}$$

In the particular case of TanDEM-X we have $\alpha_1 = 90^\circ$, and $\Delta\alpha = 180^\circ$. The appeal of this solution is that it is passively safe: there is always a cross-track separation between the spacecraft, so that along-track drifts can be tolerated without increasing the collision risk.

Figure 2.12 shows the relative orbits and the resulting effective cross- and along-track baselines. The cross-track projections of the relative orbits describe an ellipse and never cross. Obviously, with only two spacecraft, the effective baseline becomes the absolute value of a harmonic function, which implies that it vanishes at two points along each orbit, once during the ascending part and once during the descending one. The latitudes corresponding to these two points are different, which implies that all locations can be observed with a sufficiently large baseline at least for one semi-orbit. Furthermore, for a systematic mapping mission, the

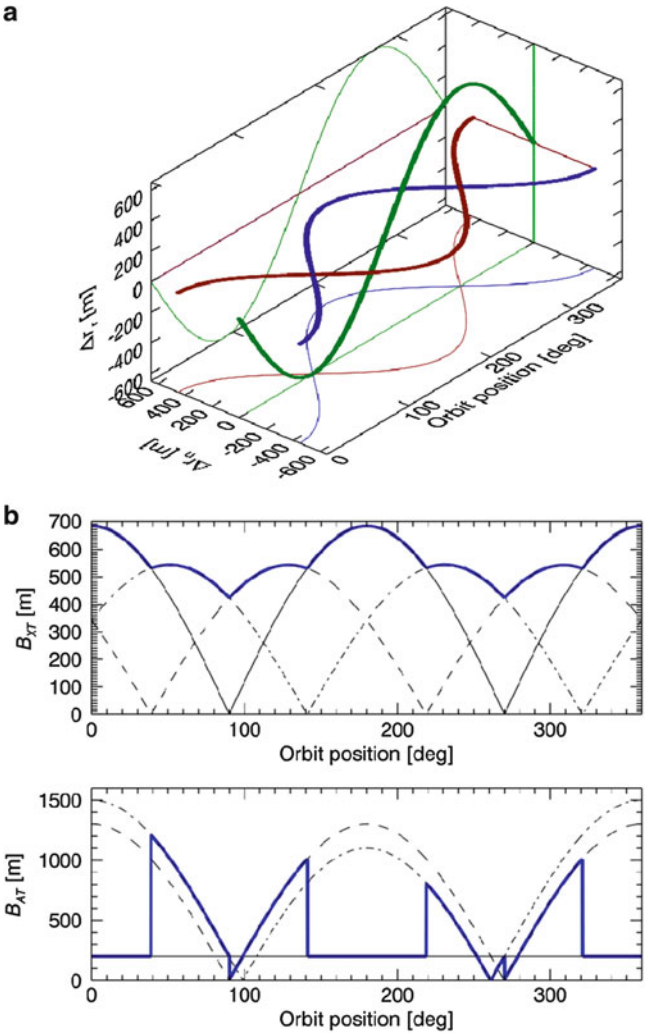


Fig. 2.11 (**a**) Relative cross-track deviations with respect to mean (nominal) orbit for a 3-spacecraft CarPe as a function of the mean argument of latitude (u). (**b**) Resulting effective cross-track and along-track baselines

formation parameters can be changed during different mission phases in order to optimize the baseline obtained at each latitude.

In principle, considering the Clohessy-Wiltshire equations, the Helix concept could be scaled to N_s spacecraft by taking $\Delta\alpha = \Delta\beta = 360^\circ/N_s$. However, this would imply a small change in the orbit inclination, which would result in an undesired relative drift of the ascending nodes. An interesting alternative could be a multi-satellite Helix with an even number of spacecraft with

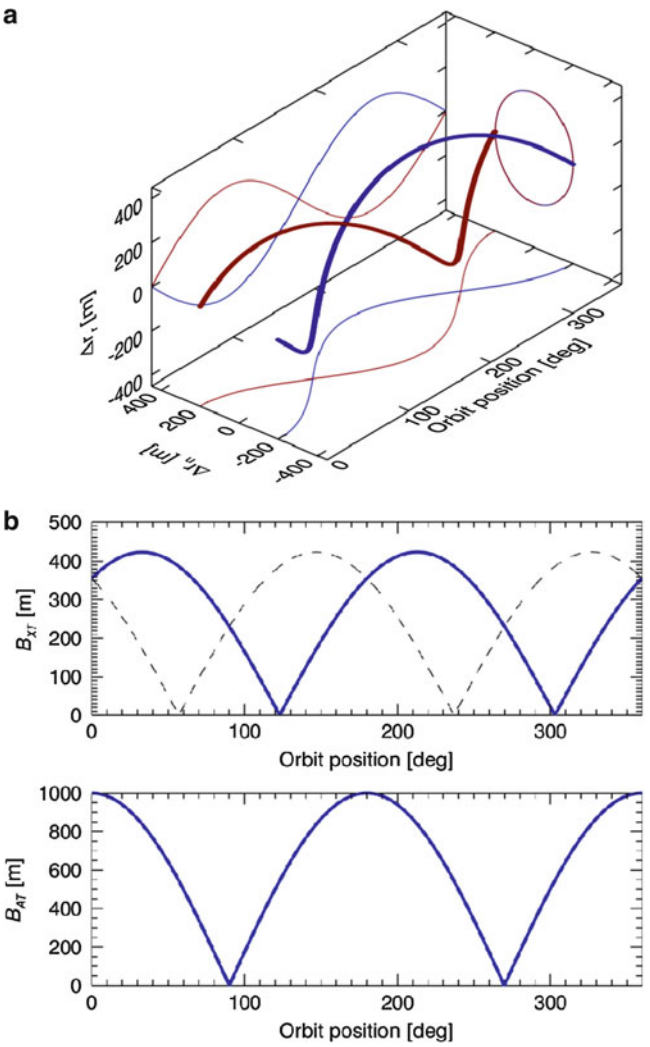


Fig. 2.12 (**a**) Relative cross-track deviations with respect to a mean (nominal) orbit for a 2-spacecraft Helix as a function of the mean argument of latitude (u). (**b**) Resulting effective cross-track and along-track baselines. The dashed cross-track baseline represents that for the complementary ascending/descending pass over a given latitude

$$\begin{aligned} \delta e_{2i} &= \delta e_{2i+1} = s_i \cdot \delta e \neq 0 \\ \delta i_{2i} &= \delta i_{2i+1} = s_i \cdot \delta i \neq 0 \end{aligned} . \tag{2.30}$$

This would result in a number of concentric Helix formations, with each formation scaled by a factor s_i

2.3.2 *Operational Modes*

This section discusses the principal operation and acquisition modes in which a multistatic system can be operated.

2.3.2.1 Multistatic Modes

With multistatic modes we refer to how a multistatic system is operated in terms of which satellite is receiving and which one is transmitting. This is closely linked to the type of multistatic configuration (see Sect. 2.1.2.2). We consider the following modes:

- *Bistatic mode*. We follow the convention set for the TanDEM-X mission and loosely denominate as bistatic mode an operating way in which all enabled receive satellites acquire the radar echoes of a common transmitter. Logically, this is the only way in which a semi-active system would be operated. For fully-active systems such as TanDEM-X or Tandem-L, this definition of the bistatic mode implies that the transmitting satellite is actually acquiring monostatic SAR data. Note for these and similar mission concepts the single transmitter bistatic mode is also the default operating mode.
- *Alternate bistatic mode*. In this mode, pulses will be transmitted sequentially by at least two (enabled) transmitting spacecraft, while all enabled receivers collect the corresponding radar echoes. For two-spacecraft systems it is also referred to as Ping-Pong mode. Alternate bistatic operations have the advantage of increasing the number of phase centers (and baselines), and may also be used for phase synchronization (see Sect. 2.3.3.5).

  The drawback of alternate bistatic operations is that, in order to adequately sample the radar echoes in azimuth (for a given desired azimuth ambiguity level), the PRF associated to each transmitter is constant. This implies that the total PRF is multiplied by the number of enabled transmitters with respect to the single transmitter bistatic mode. Generally speaking, this leads to higher range ambiguities and/or a reduced swath width.
- *Multi-monostatic mode*. This is a non-bistatic operation mode in which each spacecraft in a fully-active system acquires an independent monostatic image. Mutual interference between the satellites can be suppressed by introducing a sufficient along-track separation, in which case we can refer to this mode as *pursuit-monostatic* mode. Multi-monostatic acquisitions can also be useful for non interferometric applications for which the different spacecraft acquire completely different information content. For example, several spacecraft may acquire different portions of the Doppler spectrum, in which case mutual interference is suppressed by the different azimuth pointing of the radar beams. Likewise, by using different elevation beams, a distributed system operating in this mode may simultaneously acquire two independent swathes.
- *MIMO modes*. The idea in Multiple-Input Multiple-Output SAR modes [74, 75] is that it is possible to transmit simultaneously with multiple transmitters using

orthogonal waveforms. An often overlooked limitation is that one-dimensional orthogonal waveforms are only possible if they have non overlapping spectral content, which makes them useless for most multistatic applications (in particular for interferometric applications). However, MIMO concepts are possible using digital beamforming systems and using multi-dimensional waveform encoding [62]. The drawback of possible MIMO modes is, therefore, the increased complexity of the receivers. Despite these limitations, MIMO modes are probably the correct approach for multistatic concepts requiring a large number of phase centers.

2.3.2.2 SAR Acquisitions Modes

SAR images can be acquired in different modes. The standard operating mode is strip-map, which uses a fixed beam and for which the azimuth resolution may be approximated as half the effective two-way antenna length, L_{eff}, defined by (2.11). Due to the different along-track positions of the effective mono- or bistatic phase centers, the instantaneous Doppler frequency with which they see a given target has an offset given by

$$\Delta f_D = K_a \cdot \frac{\Delta r_t}{v_{sat}}, \tag{2.31}$$

where Δr_t is the along-track separation of the phase centers (which, for example, in the case of a semi-active system would be half the physical along-track separation between the receivers), and K_a is the Doppler rate. In the case of a strip-map acquisition, this instantaneous offset does not necessarily imply an offset in the acquired Doppler bandwidth, it just means that a given Doppler spectral component is acquired at a different time. As discussed in Sect. 2.2.1 the different antenna patterns can be optimized in several ways, causing loss of common Doppler bandwidth when the SNRs are maximized, or a loss of sensitivity if the common Doppler bandwidth is maximized. The consequence, from a mission designer point of view, is an upper limit to the along-track spread of the multiple mono- and/or bistatic phase centers.

It is also possible to operate a multistatic system in spotlight [76, 77] or sliding spotlight modes. In those modes, the beam is rotated in azimuth so that the targets remain for a longer time within the main-lobe of the two-way antenna beam pattern. This increases the available Doppler bandwidth and, therefore, the azimuth resolution of the system. In the multistatic case this makes it easier to obtain the required common Doppler bandwidth, thus allowing larger spreads between the phase centers. The drawback is that it adds the complexity of synchronizing the steering of the multiple transmit and receive patterns.

Finally, it is common to trade-off resolution for swath width by operating the SAR system in burst modes such as ScanSAR [78, 79] or TOPS [60]. In these

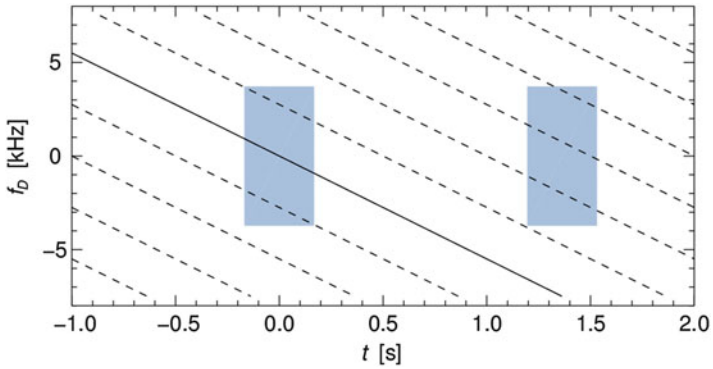


Fig. 2.13 Time-Doppler diagram for a *bistatic* ScanSAR acquisition. Each same-color line represents a target at a different azimuth position, while the *blue* and *black* color coding correspond to two different bistatic phase centers. The shaded areas represent the time-frequency acquisition window of a ScanSAR burst

modes the radar interleaves bursts of pulses with antenna patterns pointed at different elevation angles in order to acquire a number of sub-swathes. This time-multiplexing of the elevation beams implies that the length of the synthetic aperture, and thus the available Doppler bandwidth, is reduced by a factor at least as large as the number of sub-swathes.

Figure 2.13 shows the time-Doppler diagram for a two-phase center multistatic ScanSAR acquisition. The black lines correspond to the Doppler frequency history of targets located at different azimuth positions as seen by the leading bistatic phase center, while the blue lines correspond to the same targets, but seen by the trailing phase center. The vertical separation between each pair of lines is given by (2.30), for a semi-active X-band system with a 550 km orbit height and 1 km along-track separation. The slope of the lines is the Doppler rate, K_a. The rectangular shaded areas represent the time-Doppler frequency window of the ScanSAR bursts. This clearly illustrates how the range of acquired Doppler frequencies depends on the target position and on the bistatic phase center. To generate sets of multistatic images with compatible Doppler frequency components Δf_D must be small compared to the burst bandwidth

$$B_{burst} = |K_a| \cdot T_{burst}. \tag{2.32}$$

Unlike for strip-map mode, the loss of common Doppler bandwidth in multistatic ScanSAR cannot be avoided by an optimized steering. Therefore, it needs to be kept small, in relative terms, by constraining the separation between the phase centers. A similar analysis, with similar conclusions, is applicable to multistatic TOPS acquisitions. Multistatic wide-swath burst-modes are fundamentally inefficient and, therefore, unappealing.

2.3.3 Radar Synchronization

2.3.3.1 Oscillator Noise and Relative Phase Errors

One of the characteristic challenges of a multistatic system is the need to synchronize the distributed sub-systems [80–83]. This includes a temporal synchronization in order to align, for example, the sampling windows, and the much more demanding phase synchronization required to have a truly coherent system. Precise phase synchronization is required for any application that aims at coherently combining the information acquired by different receives. This includes all interferometric and tomographic applications, but also distributed imaging approaches.

For distributed systems using a single dedicated and common transmitter with a number of passive receive systems, it may be sufficient, for some applications, to limit the precise phase synchronization to the receivers.

The instantaneous phase of an oscillator can be written as

$$\theta_{osc}(t) = 2\pi \cdot f_0 \cdot t + \phi(t) + \phi_0, \tag{2.33}$$

where f_0 is the center frequency, $\phi(t)$ a time-varying phase error, and ϕ_0 a constant, arbitrary phase. It is worth mentioning explicitly that $\phi(t)$ is a non-stationary random process, which may be modeled as the sum of a zero-mean stationary term and a random walk term

$$\phi(t) = \phi_{st}(t) + \phi_{rw}(t). \tag{2.34}$$

The random walk term is essential to understand the long-term evolution of the total phase error [84–86]. For any transmitter-receiver or receiver-receiver pair, the relative phase difference will be the difference of two independent random processes, and can be expressed as

$$\Delta\phi(t) = 2\pi \cdot \Delta f_0 \cdot t + \phi_{st}(t) + \phi_{rw}(t), \tag{2.35}$$

a relative frequency term resulting in a linear phase component, and, again, the sum of a stationary process and a random walk component. Let us consider these three terms and their impact on the multistatic SAR products.

- An uncompensated (residual) linear phase error between a transmitter and the receivers will introduce an apparent Doppler for all targets, with the corresponding azimuth displacement in the processed image. After SAR processing each image would have a relative phase error proportional to the relative along-track position of the spacecraft. For ATI or GMTI applications, the error would introduce a bias in the target velocity estimation. For XTI applications, due to the inevitable along-track separation between the spacecraft,

this phase error would be interpreted as a height offset. In theory, applications requiring the combination of the signal of more than two receivers (e.g. single-pass tomography) could be totally impaired by these errors. Nevertheless, this could be casted as the typical phase calibration problem present in repeat-pass tomography, which can be solved using data-based calibration approaches.

- A residual linear term between different receivers would introduce relative phase ramps in azimuth between the SAR images. In many cases this error can be calibrated out using a priori information. For example, in most cases there would be enough a priori topographic knowledge to eliminate an erroneous azimuth ramp from a generated Digital Elevation Model.
- The stationary term is a colored random process whose single sideband (SSB) power spectrum is usually described by the superposition of (up to) five frequency components [87],

$$S_{\phi}^{SSB}(f) = \sum_{i=0}^{4} b_i f^{-i} \text{ for } f > f_c, \tag{2.36}$$

where f_c is a lower cut-off frequency that is required to keep the total power finite, so that (2.35) describes a proper stationary process. High-frequency components will have an effect similar to that of thermal noise, and will result in an increase of spurious side-lobe levels. Lower frequency terms, with time scales in the order of the aperture length, modify the apparent phase history of the received signal. This results in a defocusing of the SAR image, with a widening of the azimuth point-target response, and random azimuth and range displacements of the focused targets. Frequency components greater or equal to the inverse of the aperture length, $f > 1/T_a$, will result in an azimuth phase modulation in the focused image and, consequently, in the resulting interferograms.

- The random walk component will result in a phase error whose variance increases linearly with time [84]

$$\sigma_{\phi,rw}^2(t) = (2\pi f_0)^2 c_{rw}^2 |t|, \tag{2.37}$$

where c_{rw} is a constant that characterizes this process. This describes the expected unbounded drift of the relative phase of two oscillators. For most practical purposes, for acquisitions of a limited duration, the short-term drift can be described by the low frequency spectral components of a stationary phase error. Long-term drifts can be characterized as phase offsets that change randomly from acquisition to acquisition.

In radar systems, like in many telecommunication systems, all frequency references are always derived from a common master oscillator. Radars use Ultra Stable Oscillators (USO) as this reference. Usually these USOs operate at relatively low frequencies (values in the 5–10 MHz range are typical). The higher

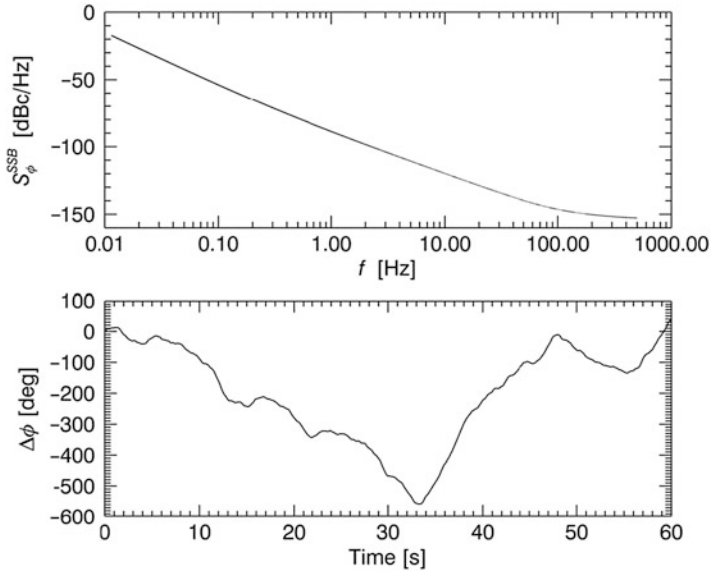


Fig. 2.14 *Top*: example of the single sideband phase-noise spectrum of a typical spaceborne 10 MHz USO. *Bottom*: difference between two random realizations of the corresponding random process, scaled to C-band. Linear trends have been removed

Table 2.1 Examples of single sideband phase noise levels, given in dBc/Hz, for space qualified USOs

	1 Hz	10 Hz	100 Hz	1 kHz	10 kHz
USO according to Ref. [82]	−89	−120	−146	−154	−155
PHARAO USO [88]	−131	−147	−156	−156	−156

frequencies required through-out the systems are derived from these master frequency source using combinations of phase-locked synthesizers, Direct Digital Synthesizers (DDSs) and/or frequency multipliers. For low frequency phase noise components all these alternatives end-up multiplying the phase noise by the ratio between the synthesized frequency and that of the USO. The consequence is that with the same USO technology, the severity of oscillator induced phase errors is directly proportional to the operating frequency of the radar. For a mission design, this may lead to different synchronization approaches depending on the operating band.

Figure 2.14 shows and example of the single sideband phase noise spectrum of a typical space qualified USO and the difference between two random realizations of the corresponding random processes. A 10 MHz reference oscillator is assumed with the phase noise levels versus frequency given in the first row of Table 2.1 [82] with a multiplication factor of 530 to reach a typical C-band operating frequency of 5.3 GHz. To get a sense of what is possible in terms of space qualified USOs, the second row in the table shows the phase noise requirements for the 5 MHz reference USO for the PHARAO system [88, 89].

2.3.3.2 Synchronization by Means of Dedicated Synchronization Links

From the previous analysis, it is clear that some strategy is required to synchronize in slow time and in phase the different subsystems in a distributed multistatic radar mission. The most direct approach is to use dedicated synchronization links in order to exchange synchronization information. The simplest conceivable scheme would be to designate one spacecraft as the master reference and let it distribute a reference synchronization signal to all other spacecraft by means of radio frequency links and use it to correct the phase of frequency references of the slave satellites. Calling the phase of the master frequency reference ϕ_1, and ϕ_i the phase of the oscillator at a sub-system i, the phase difference measured at i would be

$$\Delta\tilde{\phi}_{i1}(t) = \phi_i(t) - \phi_1(t) - \frac{2\pi f_{ref}}{c_0} \cdot B_{i1}(t), \tag{2.38}$$

where f_{ref} is the reference frequency, and $B_{i1}(t)$ is the time varying physical separation, or baseline, between the two spacecraft. Due to the phase offset resulting from the spacecraft separation, no absolute phase synchronization can be achieved with a one-way link unless this baseline is known within a small enough fraction of a wavelength. The solution to this problem is to use a two-way link, as done in the TanDEM-X mission. The addition of the return link results in an additional phase difference equation

$$\Delta\tilde{\phi}_{1i}(t) = \phi_1(t) - \phi_i(t) - \frac{2\pi f_{ref}}{c_0} \cdot B_{i1}(t). \tag{2.39}$$

Now from the addition of (2.38) and (2.39) it is possible to estimate the physical separation between the satellites, while the baseline term vanishes when both equations are subtracted. The actual phase difference between the oscillators can then be estimated from

$$\Delta\phi_{i1} = \frac{\Delta\tilde{\phi}_{i1} - \Delta\tilde{\phi}_{1i}}{2}. \tag{2.40}$$

This expression hides a subtlety: because phase differences are measured modulo 2π, this by 2 division results in an undesired π ambiguity in the resulting phase. If the signals exchanged are not pure sinusoidal but, for example, pulsed signals (as in TanDEM-X [82]), the π ambiguity may be resolved from the estimated pulse delays.[2] Since this ambiguities matter only at the operating frequency of the radar, an alternative approach to solving them is to operate the sync-link at a

[2] Note that the synchronization equations can be written in terms of time delays instead of phases, in which case there are no ambiguities. Nevertheless, the resolution provided by phase measurements is generally speaking orders of magnitude better than the estimated delays.

reference frequency that is an even fraction of the radar frequency. In this case, the frequency multiplication operation required to generate the desired radar reference frequency will eliminate any π difference. This approach, however, requires the introduction of sub-systems at new frequency which complicates the overall design.

A two-way synchronization link assumes reciprocity, so that hardware induced delays or phase shifts are identical in both directions. This assumption has to be considered in the implementation of the link and, in practice, it can become a limiting factor, since hardware components are never perfectly identical.

It is also desirable to have a near ideal radio channel, i.e. a multi-path free line-of-sight path between the spacecraft. In most formation flying concepts the relative positions of the spacecraft change along the orbit. Therefore, it is necessary to have a number of sync-antennas placed strategically around the satellites in order to cover all directions. For example, in the case of TerraSAR-X and TanDEM-X, each spacecraft has 6 synchronization horns.[3]

For a pulse-based sync-link another important consideration is the required pulse repetition frequency. In the case of TanDEM-X the synchronization pulses are interleaved with the regular radar pulses, which results in azimuth gaps in the resulting SAR raw data. In order to avoid the appearance of artifacts in the focused SAR image, the number of gaps needs to be as small as possible, which implies a low sync-pulse frequency. From a sampling theory point of view this is possible because oscillator phase noise is a low-pass random process. However, it is also necessary to estimate the difference between the reference frequencies, which can be in the order of tens or hundreds of Hertz. This can be done by including a frequency offset estimation step using a high sync-pulse frequency before the start of an acquisition.

A practical implementation of the sync-link processing is quite more complex than a direct application of (2.40) For example, it needs to take into account that the phase measurements may not be performed simultaneously, requiring the interpolation of them to an appropriate common time base. It also needs to account for Doppler effects (time varying inter-spacecraft separation) or with relativistic effects: sync-pulse measurements and calculations done in a moving reference frame tied to the spacecraft formation result in small errors if the SAR processing is done in a reference frame tied to the Earth.

2.3.3.3 GNSS Based Synchronization

A widespread approach to synchronize distributed systems is the use of the Global Navigation Satellite Systems (GNSS) [90, 91] such as the GPS system or the upcoming Galileo constellation. GPS disciplined oscillators are also a common solution for systems requiring very good long-term frequency stability. Generally

[3] In this particular case, the number 6 is also related to the overall RF design of the radar synchronization system in order to achieve an omnidirectional field of view.

speaking, GNSS frequency references have very good long-term stability. For the short term, however, the quality of the phase is limited by the low SNR of the GNSS signal, which requires long integration times in order to reduce the phase noise. Thus, a GNSS disciplined oscillator will use a high performance crystal oscillator to provide the best possible short-time stability and low phase noise, while relying on the GNSS reference for longer-term frequency and phase stability.

It is clear that, where feasible, GNSS based synchronization would be the first choice for a multistatic system. Most SAR spacecraft incorporate high performance GNSS units, which are used for positioning. For formation flying systems, GNSS will also typically be used for relative position estimation. As it is generally known, GNSS receivers need to solve the 3-D position vector and the time offset (phase) of their internal time reference w.r.t. the time reference of the GNSS system. Likewise, Differential GNSS solutions need to resolve the 3-D differential position vector and also the relative time offsets. This implies that a formation flying multistatic system using Differential GNSS processing for relative position determination is automatically also providing a common time reference. Using Carrier Differential GPS signals (CDGPS), the relative position can be estimated with a 3D RMS error in the order of 1.5 mm [92, 93], with the individual components in the 1 mm error range. This implies a long-term relative timing error in the order of $10^{-3}/3 \cdot 10^{8} \approx 3 \cdot 10^{-12}$ s, which represents about 12° of RMS error at X-band and only about 1.5° at L-band. From this brief analysis it may be concluded that GNSS-based synchronization may be good enough for low frequency (P-, L- and maybe S-band) multistatic radar missions. For higher frequencies GNSS synchronization by itself seems insufficient.

2.3.3.4 Multisquint Processing

The last phase-synchronization approach considered is to, somehow, try to estimate the phase error from the acquired radar data. There are actually several methods to calibrate (or partially calibrate) phase errors using the radar echoes.

The first such method is the Multisquint processing approach [94, 95], developed as a Motion Compensation (MoCo) technique in repeat-pass airborne SAR interferometry. In that case, baseline errors are the result of errors in the position estimation by the Inertial Navigation System (INS) of the aircraft. These errors are due to high frequency components of the flight track due to turbulence or pilot manoeuvres. At the temporal scales of the integration time corresponding to the synthetic aperture length, these errors are typically much more severe than relative phase differences of two oscillators.

Before explaining the Multisquint processing approach, it is important to remark that a time varying phase error implies a relative Doppler shift, which in turn will result in a relative azimuth shift in the focused image. This azimuth shift will introduce decorrelation due to co-registration errors in the uncompensated

interferogram. In fact, Multisquint processing was originally developed as an azimuth coregistration error estimation technique.

Multisquint processing of an interferometric pair follows the following steps:

1. A number (N_{sub}) of sub-look master and slave images is generated corresponding to different narrowband Doppler spectra. These images will have a reduced azimuth resolution. Therefore, the coherence loss due to the azimuth coregistration errors will have a smaller impact.
2. For each pair of sub-look SLC images, a corresponding interferogram is generated. This stack of sub-look interferograms contain the same topographic phase information, but the systematic error is the one corresponding to the beam center position of each look, and thus to different azimuth positions.
3. The phase difference between consecutive sub-looks is calculated, generating N_{sub}-1 double interferograms. Here, the topographic phase contribution cancels out. The resulting phase is the difference of the phase errors at the two beam center positions, and can be scaled to a derivative of the relative phase error versus azimuth position.
4. The double difference interferograms now need to be aligned such that each azimuth position corresponds to a beam center position (or to the raw data position) instead of corresponding to the same ground-azimuth (zero-Doppler) position. Once they are properly aligned, the N_{sub}-1 double interferograms can be averaged.
5. Now the double interferogram phases can be averaged in range (in the original MoCo application there is a range dependence that can be exploited to separate different relative motion components).
6. The last step is to integrate in azimuth the resulting estimated double differences, obtaining an estimate of the phase error versus azimuth. This integration step is a source of numerical errors, since zero mean random errors will result in a random walk error component in the solution. Nevertheless, since a lot of averaging is done prior to the integration, this resulting random walk component results in a very low frequency error. It should also be clear that a constant error term cannot be recovered through this process.

Multisquint processing has been proposed and successfully tested as a possible back-up solution for the TanDEM-X mission, to cover the unlikely event that the synchronization link would fail [96]. Figure 2.15 shows an example of an overlay of the interferometric phase on the intensity before (top) and after (bottom) Multisquint phase synchronization. The nominal flat Earth phase has been subtracted, so that the remaining fringes are, in theory, due to the topography. The low frequency range fringes are a residual flat earth component due to the difference of the mean height of the scene and the reference height used to flatten the interferogram. Comparing the two phase images it can be appreciated how after Multisquint processing an azimuth modulation has been removed. The estimated phase error, after removal of the linear component, is shown in Fig. 2.16.

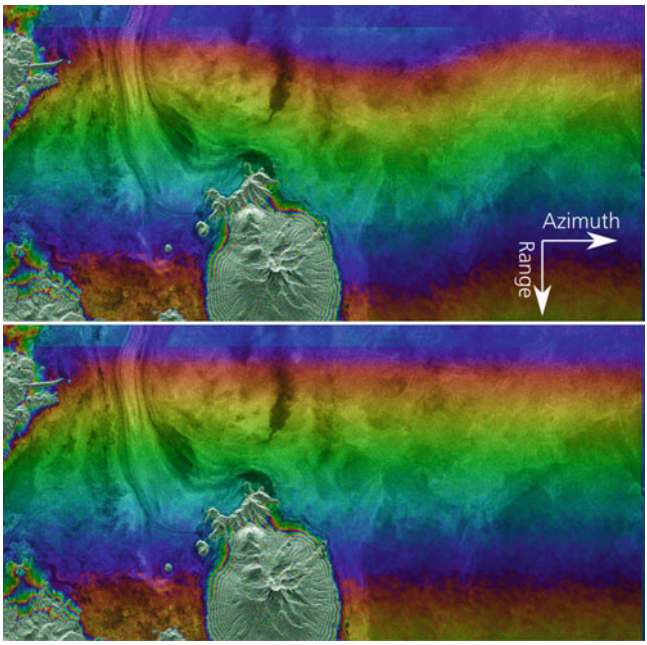


Fig. 2.15 Overlay of interferometric phase of a TanDEM-X acquisition on the corresponding intensity image. *Top*: before phase synchronization (the linear phase term due to the small frequency offset has been removed). *Bottom*: after phase correction using Multisquint processing. In both cases a nominal flat Earth term has been removed (Courtesy of Marc Rodriguez-Cassola, DLR, Microwaves and Radar Institute)

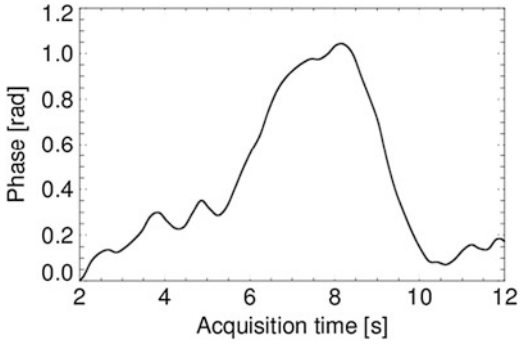


Fig. 2.16 Example of Multisquint estimated phase error corresponding to the example shown in Fig. 2.15

2.3.3.5 Alternate Bistatic Redundant Imaging

The second and last data driven calibration approach considered is applicable to distributed systems where all satellites transmit. An example would be the alternate bistatic mode of TanDEM-X, which is equivalent to the so-called

simultaneous monostatic and bistatic mode (SiMB-static) discussed for the RADARSAT-2/3 topographic mission concept [3, 4]. The concept is quite simple: due to reciprocity, the two (or more) bistatic images acquired in an alternate bistatic operation mode are, in theory, identical and, therefore, redundant. However, the bistatic images will be corrupted by the relative oscillator phase error, which will be of the same magnitude but opposite sign for the two images in a redundant pair. Thus, the phase error can be estimated directly from the phase differences between the redundant raw bistatic images. It is interesting to note that this is conceptually equivalent to a dedicated sync-link, replacing the nearly ideal LOS channel by a random time-varying channel. While it seems appealing to use this approach instead of a dedicated sync-link, it has two clear disadvantages:

- In order to obtain really redundant bistatic pairs, each of them needs to adequately sample the Doppler spectrum. This means that the total PRF of the system needs to be roughly a factor of two higher than that of a regular acquisition, which will, in general, tend to reduce the achievable swath width and reduce the achievable sensitivity.
- The data rate is also duplicated.

A possible way to relax these problems could be to not to transmit the full pulse bandwidth with both systems. Instead, each system could transmit a bit more than half the total bandwidth, with an overlapping fraction that would be used for calibration. This solves the data rate and sensitivity (radar duty cycle) problems, but, generally speaking, it does not solve the high PRF problem.

Despite the aforementioned issues, this calibration approach may be of high interest for missions considering alternate bistatic as the default mode in order to exploit the resulting multi-baseline data set.

2.3.4 Baseline Determination and Calibration

All multistatic mission concepts in which the data acquired by the different subsystems is combined coherently require very good knowledge of the relative position of the spacecraft, and, in particular, of their relative cross-track position. Although relative navigation is discussed in detail in Part 2 of this book, in this section we want to briefly discuss the general requirements and the impact of relative position knowledge errors. Here, we can distinguish between perpendicular baseline knowledge-error, i.e. the error in the relative position component perpendicular to the line-of-sight (LOS) to the target, and LOS error. As discussed in detail in Chap. 5, with the use of differential GNSS techniques, the relative position can be determined with a 3-D uncertainty in the order of a millimeter, so a main question is if this baseline knowledge is sufficient.

Although the perpendicular baseline is what determines the interferometric or tomographic sensitivity of a given configuration, small errors have a mild effect. Generally speaking, the perpendicular baseline knowledge error should be a small

fraction of the actual perpendicular baseline. This is in most cases fulfilled using the aforementioned GNSS based relative position estimates.

In contrast, while the LOS baseline is irrelevant from a sensitivity point of view, LOS baseline errors have a large impact: a LOS baseline uncertainty ΔB_{LOS} will result in a phase error

$$\Delta\phi = n_w \cdot k_0 \cdot \Delta B_{LOS} = 2\pi \cdot n_w \frac{\Delta B_{LOS}}{\lambda_0}, \tag{2.41}$$

where, as before, n_w is 1 if a single transmitter is used, and 2 for multi-monostatic data sets. From (2.41) it is clear that the LOS relative position must be known with a small fraction of the wavelength.

For single baseline cross-track interferometric applications, a LOS baseline error will introduce a height offset and a (usually small) slope in ground range in the resulting DEM. For multi-baseline applications the resulting phase errors make the array processing more difficult.

2.4 Study Examples

In this section we discuss three current multistatic mission concepts: Tandem-L, SIGNAL, and PICOSAR. The reader is referred to Part 4 of this book for much more detailed discussions of several other mission concepts.

2.4.1 Tandem-L

Tandem-L [97–99] is a German proposal for a two-spacecraft formation-flying L-band SAR system conceived to monitor the Earth, with an emphasis on dynamic processes (Fig. 2.17). The most relevant mission objectives are global inventories of forest height and above-ground biomass, large-scale measurements of Earth surface deformations due to plate tectonics, erosion and anthropogenic activities, observations of glacier movements and 3-D structure changes in land and sea ice, and the monitoring of ocean surface currents.

In order to monitor dynamic processes at continental scales implies unprecedentedly high system requirements in terms of swath width, duty cycle, and data rate. A key to satisfy these requirements is the use of advanced Digital Beamforming (DBF) techniques to achieve a very wide swath combined with a fine spatial resolution.

In order to acquire the forest height and biomass inventories, Tandem-L relies on the ability of performing single-pass fully polarimetric and interferometric (Pol-InSAR) measurements.

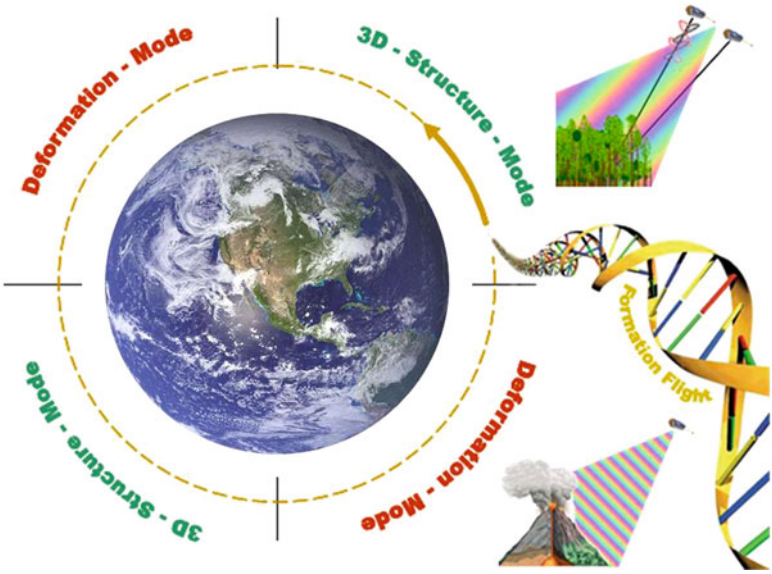


Fig. 2.17 The Tandem-L mission concept relies on a systematic data acquisition strategy using a pair of L-band SAR satellites flying in close formation. The satellite system is operated in two basic data acquisition modes: 3-D structure mode and deformation mode. New SAR imaging techniques enable frequent coverage with high geometric resolution

2.4.1.1 Science Requirements

The primary mission goals of Tandem-L can be grouped in three categories:

- Biosphere (3-D vegetation monitoring):
  - Measurement of forest height and structure
  - Global inventory of above ground forest biomass
  - Detecting vegetation disturbances and biomass changes
- Geo-/lithosphere (deformation measurements):
  - Understanding earthquake and volcano eruption cycles
  - Quantifying the magnitude of events
  - Determination and forecasting the probability of events
- Hydro- and cryosphere (structure and deformation):
  - Measurements of ice structure and its changes
  - Monitoring soil moisture and surface water changes
  - Observation of ocean currents and wave field dynamics

Table 2.2 provides a summary of some basic user requirements for the most important applications.

Table 2.2 Tandem-L requirements summary

	Science product	Coverage	Product resolution	Product accuracy
Biosphere	Forest height	All forest areas	50 m (global) 20 m (local)	~10%
	Above ground biomass			
	Vertical forest structure		100 m (global)	~20% (or 20 t/ha)
	Underlying topography		50 m (regional)	
			50 m (global) 20 m (local)	3 layers
			50 m	<4 m
Geo-/lithosphere	Plate tectonics	All risk areas	100 m (global) <20 m (fault)	1 mm/year (after 5 years)
	Volcanoes	All land volcanoes	20–50 m	5 mm/week
	Landslides	Risk areas	5–20 m	5 mm/week
	Subsidence	Urban areas	5–20 m	1 mm/year
Cryo- and hydrosphere	Glacier flow	Main glaciers	100–500 m	5–50 m/year
	Soil moisture	Selected areas	50 m	5–10%
	Water level change	Regional	50 m	10 cm
	Snow water equivalent	Local (exp.)	100–500 m	10–20%
	Ice structure changes	Local (exp.)	100 m	>1 layer
	Ocean currents	Prio. areas	~100 m	<1 m/s
All	Digital terrain and surface model	Global	~20 m (bare) ~50 m (forest)	2 m (bare) 4 m (veg.)

2.4.1.2 Mission Concept

A salient aspect of the Tandem-L mission is a systematic data acquisition plan using a two-spacecraft multistatic (fully-active) configuration. The two-satellites will be operated in two alternating modes:

- The **3-D structure mode**, during which the two spacecraft will be used only to acquire Pol-InSAR measurements of forests and other semi-transparent volume scatterers (ice, sand, etc.).
- The **deformation mode**, which relies on repeat-pass DInSAR measurements. In this mode, the two spacecraft will be used to double the acquired swath, operating in a multi-monostatic way.

For the deformation mode, the two main performance limiting factors are temporal decorrelation and atmospheric effects (the atmospheric phase screen). While temporal decorrelation can be addressed by increasing the resolution in order to be able to average more independent looks, the spatially smooth atmospheric phase errors can only be mitigated by maximizing the number of acquisitions. For that purpose, a 8-day repeat-cycle is proposed.

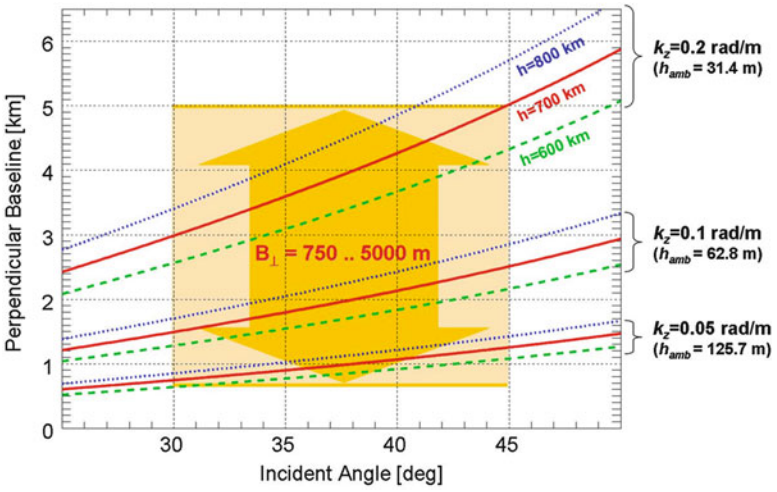


Fig. 2.18 Correspondence between vertical wavenumbers and perpendicular baselines ($B_\perp$) as a function of the incident angle and the orbital altitude

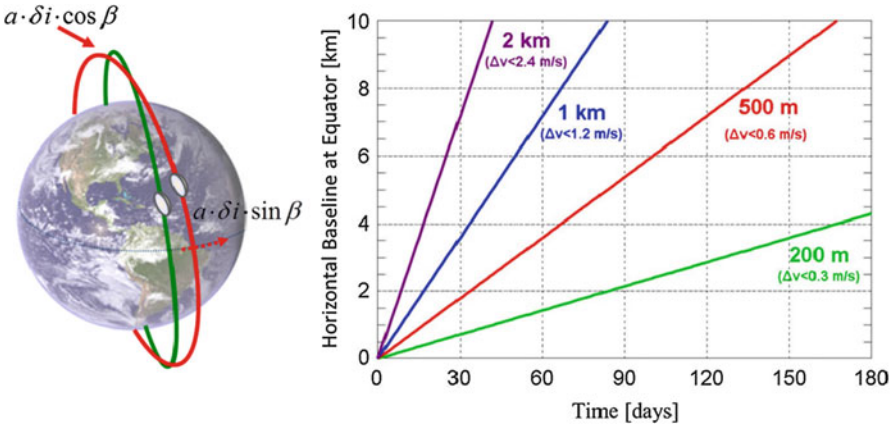


Fig. 2.19 A systematic variation of the equatorial cross-track baselines can be achieved by using orbits with slightly different inclinations. The inclination offset causes a relative drift of the ascending nodes

As discussed in Sect. 2.2, the optimal baseline for the structure mode is dependent on the structure height to be observed (see Fig. 2.18). Moreover, biomass estimation accuracies below 10% can be achieved by combining acquisitions with different baselines. To address this, the mission plan includes systematic variations of the interferometric baseline. An elegant technique to provide this wide range of cross-track baselines exploits the naturally occurring differential secular variations of the right ascension of the ascending nodes in response to slightly different inclinations. This is illustrated in Fig. 2.19 which

shows the evolution of the horizontal baselines at the equator for different inclination offsets (expressed as horizontal baselines at the northern and southern orbit turns). The figure also indicates the Δv required to introduce the inclination offset. Recall that this $\Delta v$is a measure of the impulse (normalized to unit mass) required to achieve the modified orbit, and goes directly into the fuel budget, and eventually the lifetime, of the mission. As also illustrated by this figure, Tandem-L relies mainly on the horizontal baseline, which yields smaller cross-track baselines for higher latitudes. In order to obtain the desired range of baselines over tropical and boreal regions, the Tandem-L acquisition plan will combine the use of mainly bistatic (with one of the two spacecraft acting as common illuminator for both) acquisitions in equatorial regions, and alternate bistatic or multi-monostatic acquisitions at higher latitudes.

2.4.1.3 System Concept

The Tandem-L system concept includes the use of a large reflector antenna combined a digital feed array, using SCan-On-REceive (SCORE) [100–102] in order to obtain a sufficiently wide swath despite the narrow, high-gain, receive beam. SCORE exploits the one-to-one relation between range delay and direction of arrival present in side-looking radar systems. SCORE combines multiple sub-aperture signals such that at each instant of time a narrow beam is steered towards the echo's expected direction of arrival. This allows a high antenna gain (on receive) without automatically coupling it to a narrow swath. A narrow receive beam has the further advantage of attenuating range ambiguities. By combining the SCORE technique with a broad transmit beam, a highly sensitive SAR system with wide swath coverage can be built. While all feed elements are used during transmission, allowing the illumination of a large image swath, a small number of feed elements (2–5, depending on the implementation) are activated at one time during the receive window. The active feed element positions are periodically shifted in synchrony with the systematic variation of the direction of arrival from the swath echoes. The advantages of this concept are manifold. First, the use of a large reflector antenna in connection with digital beamforming allows the reduction of the transmit power by a factor of 3–4 in comparison to the traditional SAR concept for the same imaging parameters. Second, it allows the mapping of a much wider swath (ca. 350 km) in high resolution strip-map mode with low range ambiguity levels [62, 98]. The fully polarimetric acquisition in strip-map mode with a wide swath is possible without the constraints of conventional SAR systems. This leads, however, to a large data rate and requires the implementation of advanced technologies for high data rate downlink. Table 2.3 provides a summary of preliminary system parameters.

Table 2.3 Preliminary system requirements for Tandem-L

Parameter	Value	Parameter	Value
Orbit height	760 km	Reflector size	15 m (diameter)
Repeat cycle	8 days	Focal length	10 m
Incident angle	26.3°–46.6°	Feed location	Centre
Look angle	23.3°–40.5°	Feed elements	24
Swath width	349 km	Feed length	3.43 m
Ground range	331–680 km	Tilt angle	31.9°
Tx power (avg.)	96 W	System loss	1 dB
Duty cycle	4%	Rx noisetemp.	420 K
Bandwidth	85 MHz	PRF	2,365 Hz
Polarization	Quad (linear)	Az. resolution	10 m

2.4.2 SIGNAL

SIGNAL (SAR for Ice, Glacier aNd globAL Dynamics) [103–105] is a formation flying Ka-band SAR mission concept that has as main objective to estimate accurately and repeatedly topography and topographic changes associated with mass change or other dynamic effects on glaciers, ice caps and polar ice sheets.

2.4.2.1 Science Requirements

Recent observations indicate a dramatic increase of ice mass losses from glaciers, ice caps, and the Greenland and Antarctic ice sheets [106]. If confirmed, these observations could lead to a mean sea level rise more as twice as large as the projections from the IPCC 2007 report. While current observations clearly show the sensitivity of the Cryosphere to climate change, the feedbacks to the global climate system are still not well understood. To advance this understanding it is necessary to extend and improve the available observational data. In particular, it is necessary to acquire more quantitative data regarding the global water mass balance. One key parameter that is missing to improve this understanding of the intricate dynamics of ice mass balances is a detailed and accurate knowledge of the 3-D ice surface topography and its changes at fine spatial and temporal resolutions.

The main objective of SIGNAL is to fill major observational gaps with regard to mass-balance and dynamics of global glacier ice. The mission addresses those components of the ice budget that have been subject to accelerated downwasting during the last decade and for which the knowledge of the present mass balance and temporal trends exhibits large error bars: the mountain glaciers and ice caps, and the outlet glaciers of the boundary zones of the Greenland and Antarctic ice sheets. This mass balance can currently be studied with data from gravity missions like GRACE [107] or GOCE [108], which provide data from which absolute mass changes of these components of the Cryosphere can be observed with a low spatial resolution

(in the order of 100–300 km). Radar altimeter missions, such as ICESat and CryoSAT provide highly accurate but discontinuous measurements of Ice topography. There is, however, no mission in space or in preparation that can provide the required observations of these changes with high spatial detail (50–200 m) and continuous spatial coverage.

The primary mission objectives are thus defined as:

- Reducing the uncertainty in the mass balance of glaciers and ice caps;
- Improving the knowledge on mass depletion of outlet glaciers in Greenland and Antarctica.

The secondary mission objectives are:

- Downscaling of altimetric elevation data over ice sheets;
- Mapping the motion of calving glaciers and ice streams in support of mass balance retrieval;
- Supporting the protection from natural hazards related to major mass movements;
- Assessing the new opportunities of high-frequency, high resolution interferometric SAR data for sea ice surface parameter retrieval.

2.4.2.2 Mission Concept

In order to achieve these scientific goals, SIGNAL aims at generating time series of digital elevation models with height accuracies in the order of a few decimeters. This leads to two fundamental mission aspects:

- The use of Ka-band (35 GHz) to minimize the penetration into the ice or snow cover, in order to obtain a DEM that is truly representative of the surface.
- The use of a pair of formation flying satellites. This is the only way to obtain the long baselines, roughly 100 m, required to achieve the desired height sensitivity and measurement stability, avoiding temporal decorrelation effects.

SIGNAL is designed to obtain sub-meter height accuracies ranging from 10 cm to 1 m, depending on the application and the required product resolution. For most applications, the required spatial resolution of the final product (50–200 m) is coarse compared to that of a topographic mission like TanDEM-X. This opens the opportunity of using a very large number of looks resulting from the combination of these moderate product resolution requirements with a high resolution SAR system. Note that while high system resolution is not a strong requirement, this high resolution emerges naturally from the compact system design (Fig. 2.20).

SIGNAL is a proposal for a systematic mapping mission with a lifetime of at least 5 years that will generate seasonal DEMs of the areas of interest. In addition to its mission driving interferometric capabilities, SIGNAL will also be a very capable system to estimate glacier velocities using incoherent feature tracking techniques.

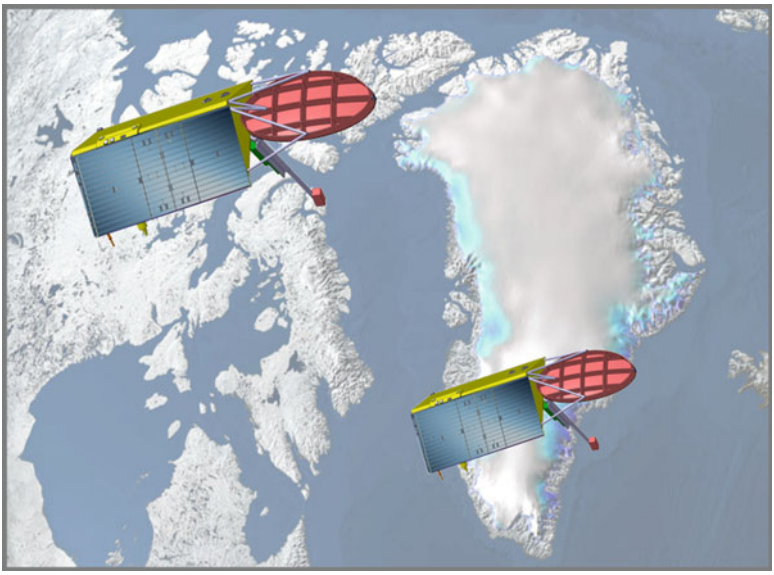


Fig. 2.20 Artist impression of SIGNAL. The preliminary design of the spacecraft is the result of a study conducted jointly by the German Aerospace Center (DLR) and Astrium GmbH

These techniques are of particular interest for fast-flowing glaciers, for which coherent techniques fail due to temporal decorrelation.

To optimally address the main topographic mission requirements but also the secondary glacier velocity measurements, the mission will be divided into three recurring scientific phases:

- A DEM acquisition phase aimed at obtaining an initial DEM. During this phase, the cross-track baseline will be set to around 40 m.
- A DEM tracking phase aimed at monitoring DEM variations with very high accuracy. During this phase the distance between the two spacecraft will be increased in order to achieve height of ambiguities in the 10–20 m range.
- A glacier velocity monitoring phase. During this, one of the two spacecraft will be rotated with respect to the other, thereby doubling the spatial coverage.

2.4.2.3 System Concept

In order to provide the required SAR performance in terms of data sensitivity, ambiguity rejection, and coverage, SIGNAL would use a reflector-based design in combination with digital beamforming techniques such as SCORE. The total coverage is further increased by simultaneously acquiring two sub-swathes separated by a gap approximately twice the width of a single sub-swath (~25 km)

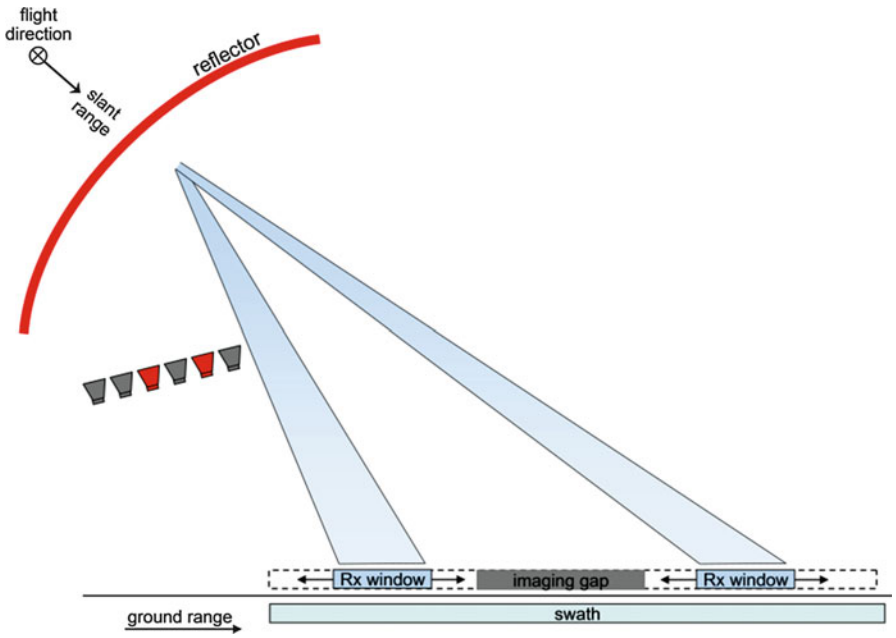


Fig. 2.21 Dual-diamond operation of SIGNAL instrument where two swathes are imaged simultaneously in strip-map mode

[105], by operating the instrument in the so-called multi-diamond SCORE mode [62, 109], as illustrated in Fig. 2.21.

On transmit two different transmit feed configurations are possible:

- **Pursuit-monostatic (multi-monostatic) operation**: in this case the single reflector is required to simultaneously illuminate both sub-swathes, without the in-between gap. The total transmit power is thus distributed on both swathes.
- **Bistatic operation**: Here each satellite can illuminate one of the two sub-swathes with the complete available power. On receive each SAR receives both sub-swathes. In this case, the transmit feed network is simplified. Compared to the monostatic operation the transmit power can be approximately halved for the same NESZ.

2.4.3 PICOSAR

PICOSAR (Passive Interferometric ocean Currents Observation Synthetic Aperture Radar), is a multistatic mission concept consisting of two small, low-cost and low power spacecraft carrying a passive, receive-only SAR payload. These two spacecraft fly at a safe distance of a Sentinel-1 (S1) [110, 111] class satellite, using it as a transmitter of opportunity and enhancing it by adding a unique along-track

interferometer dedicated to ocean surface current measurements. The passive nature of this system and the focus on a single application and single operation mode allows the implementation of PICOSAR using a very cost-effective payload design and the use of a compact and low-cost micro-satellite bus. While it has many commonalities with similar mission concepts, such as the Interferometric Cartwheel [5, 9, 112] (see Chap. 14), it is unique because of its focus on ocean currents observation using along-track interferometry, which conditions many aspects of the concept.

2.4.3.1 Science Requirements

Ocean surface currents are continuous streams of ocean water induced by global wind fields. Due to the Coriolis effect, open ocean surface currents tend to organize in large circular patterns called gyres, which rotate clockwise in the Northern Hemisphere, and counter clockwise in the Southern Hemisphere. Surface currents have a major impact on regional climates. For example, the North Atlantic gyre, which includes the Gulf Stream, carries warm water, and therefore vast amounts of heat, North along the East North-American coast and eventually, through the North Atlantic Drift, to Western Europe. Surface currents also have a significant direct economic impact due to their effect on shipping costs. Measuring and monitoring changes in surface current patterns is of high importance to the understanding of the global sea-atmosphere interaction and the global climate and climate change.

For coastal regions, the IGOS (Integrated Global Observing Strategy) coastal theme report sets a target resolution of 3–10 cm/s with horizontal resolutions under 300 m.

2.4.3.2 Mission Concept

Ocean currents are traditionally measured using drifters and, in coastal regions, using land based HF radars. Single-channel SAR systems have also been used for years to retrieve surface current information from the estimation of the Doppler Centroid anomaly [113]. However, the addition of a second channel, in an along-track interferometric (ATI) configuration, is what can turn a SAR system into a highly accurate, high-resolution surface current mapping system [15, 114–116]. ATI-SAR has been demonstrated with a number of airborne systems [117] and, more recently, from space, using TerraSAR-X. By using two phase centers with an along-track separation, an ATI system obtains two observations of the scene with exactly the same geometry, but separated by a short temporal lag. This temporal lag is proportional to the physical separation and inversely proportional to the velocity of the platform. Increasing the temporal lag increases the ATI sensitivity to the surface velocity. However, for too large ATI lags, temporal decorrelation of the sea surface degrades the final performance.

In PICOSAR, two receive-only satellites are flown in close formation, with an along-track separation in the order of 100 m. This close formation will require

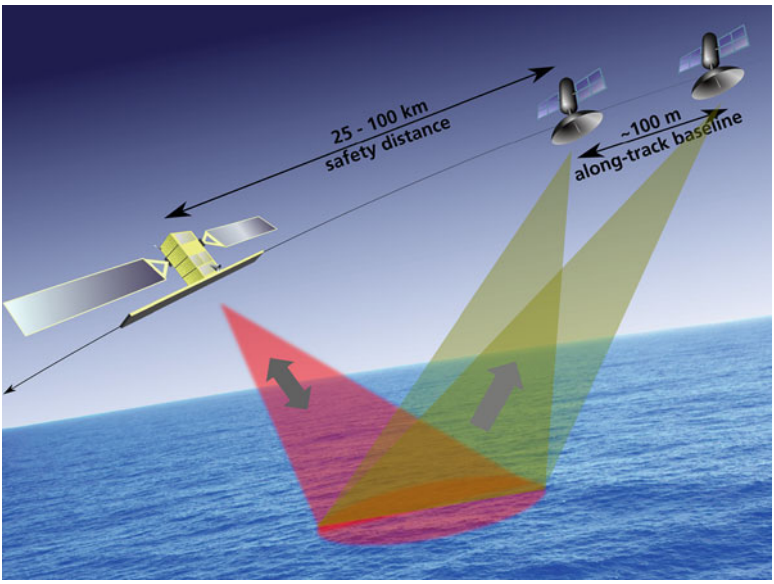


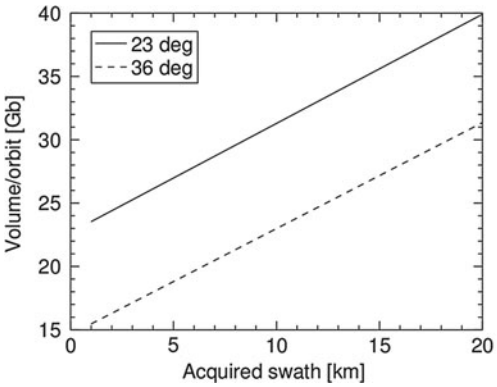
Fig. 2.22 Conceptual view of PICOSAR's multistatic configuration

autonomous formation control, although a small Helix may be added to introduce some level of passive formation safety, at the cost of introducing also a cross-track baseline component. As illustrated in Fig. 2.22, the two PICOSAR satellites would fly in a constellation with S1 at a distance of 25–100 km, determined by safety considerations.

In its default operation mode, PICOSAR will acquire data when S1 operates its Wave Mode [111]. In this mode, S1 acquires 20×20 km^2 vignettes every 100 km, with an incidence angle toggling between 23° and 36°. In its simplest incarnation, PICOSAR would acquire only a single incidence angle, which allows a single-, fixed-beam system design. Note that since S1 illuminated a swath significantly wider than 20 km, PICOSAR's swath width could be extended to about 100 km. Nevertheless, a 20 km swath width is well suited to a circular reflector antenna with a diameter in the order of 2 m. Sentinel-1 will operate in Wave Mode by default every time it flies over the oceans, thus roughly 70% of the time.

Figure 2.23 shows the data volume per orbit acquired by each receive spacecraft as a function of the acquired swath width, for both possible incidence angles. For the near range beam, S1 uses a larger pulse bandwidth in order to keep the ground range resolution fixed at approximately 5 m. The higher sampling rate is compensated by the fact that the echo window length needed to acquire the same swath is shorter in the same proportion. However, there is an overhead due to the need to sample the entire pulse length, which given the 12% transmitter duty cycle of S1 results in a larger total overhead for the near range beam. The total data volume, for 20×20 km^2 vignettes is between 30 and 40 Gigabit/orbit, which is a relaxed requirement.

Figure 2.24 provides a first-order performance analysis for PICOSAR. In particular, the analysis corresponds to a PICOSAR system flying in formation with ESA's



Beam	23°	36°
Ground range Res [m].	5	5
Pulse bandwidth [MHz]	75	49
Pulse overhead [MS/s]	18	11.6
BAQ	8:3	
Orbit duty cycle	70%	

Fig. 2.23 *Left*: Data volume per orbit as a function of desired swath width for two incidence angles. *Right*: summary of main acquisition parameters

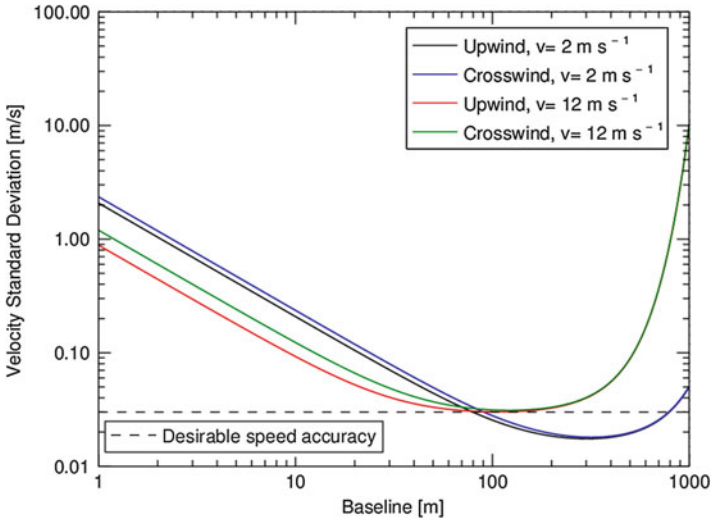


Fig. 2.24 PICOSAR surface velocity estimation uncertainty (standard deviation) as a function of along-track separation for different wind speeds and directions. A phase center separation between the two PICOSAR satellites in the order of 100 m seems desirable

Sentinel-1 mission. For the calculation of the velocity standard deviation versus the baseline, the coherence times have been calculated according to the Pierson-Moskowitz wave spectrum [118, 119], which gives 83 and 43 ms for wind speeds of 2 and 12 m/s, respectively. The number of looks considered is 256, getting a final spatial resolution of 80 m both in range and in azimuth. For the estimation of the SNR, the Normalized Radar Cross Section (NRCS) is obtained using the CMOD5 [120] model for an incidence angle of 23° for the upwind and crosswind cases and both evaluated wind velocities. On the other hand, a Noise Equivalent Sigma Zero

(NESZ) of −17 dB is considered assuming a loss of 5 dB w.r.t. Sentinel-1 due to the smaller receive antenna (a 2 m diameter reflector is assumed but a smaller antenna size would also be possible at a slightly worse performance). The results show that an optimum baseline can be selected for each wind speed considered. Effective baselines between 100 and 300 m show velocity accuracies well below the 0.1 m/s desired performance for the cases evaluated. An even better velocity accuracy is obtained for the crosswind case due to the higher NRCS obtained in this case.

2.4.3.3 System Concept

Like in the case of the Interferometric Cartwheel, the PICOSAR space segment is built around a reflector based receiver architecture with a dual polarized feed, since S1 may transmit in vertical or horizontal polarization. A double pole double throw switch between the dual polarization feed and a two channel RF front-end and digitalization stage would make the receiver redundant and would also allow experimental dual pol acquisitions. An antenna with a diameter between 1.5 and 2.5 m would suffice to meet the mission requirements.

2.4.3.4 PICOSAR Synchronization

Probably the most challenging aspect of PICOSAR and similar mission concepts is the synchronization. The along-track interferometry nature of the mission makes it extremely sensitive to systematic phase errors, which translate directly in a velocity estimate error. In contrast to cross-track interferometric mission concepts, there is little room for data based calibration of this phase. In order to meet the synchronization requirements, the use of high quality USOs in combination with an explicit synchronization link between the PICOSAR spacecraft is required.

Precise phase synchronization between S1 and both receive spacecraft is not required, since a phase offset will be common to both receivers and, therefore, cancel out in the ATI phase. However, very good relative frequency knowledge is a strong requirement, since an unaccounted frequency offset will be interpreted as a Doppler frequency offset, which will introduce a systematic bias. The required relative frequency knowledge is in the order of 1 Hz, which is within the expected short term (one orbit) frequency stability of S1. Using acquisitions over land to calibrate longer term relative frequency drifts, this implies that PICOSAR can meet its requirements without an explicit synchronization with S1.

2.5 Conclusions

There is a wide range of applications for distributed multistatic radar missions. Indeed, many possible SAR applications benefit clearly, or even require, a distributed system. The most evident examples are single-pass interferometric

applications and their multi-baseline or tomographic extensions, which require baselines of orders of magnitude that are only realizable using a distributed SAR system employing two or more formation flying satellites. Their usefulness and feasibility has been fully validated by the TanDEM-X mission that is currently in operation.

Most appealing applications require that the multistatic system is designed like a truly distributed but tightly integrated system, and not as a loose collection of cooperating independent sensors. Relatively tight and carefully designed formations are required in most cases, with strict requirements on the relative position knowledge, which has to be in the order of a small fraction of the operating radar wavelength. Also, the systems need to be carefully synchronized, usually including a synchronization link as part of the payload.

While TanDEM-X has inaugurated the era of multistatic radar missions with a fully-active system consisting of two nearly identical spacecraft, it is very likely that in the near future semi-active mission concepts will develop. In particular, due to their low marginal cost, add-on mission concepts using two or three compact receive-only spacecraft should be expected to be realized in the near future.

References

1. Krieger G, Moreira A, Fiedler H, Hajnsek I, Werner M, Younis M, Zink M (2007) TanDEM-X: a satellite formation for high-resolution SAR interferometry. IEEE Trans Geosci Remote 45(11):3317–3341
2. Zebker HA, Farr TG, Salazar RP, Dixon TH (1994) Mapping the world's topography using radar interferometry: the TOPSAT mission. Proc IEEE 82(12):1774–1786
3. Girard R, Lee PF, James K (2002) The RADARSAT-2&3 topographic mission: an overview. In: Proceedings of IEEE geoscience and remote sensing symposium (IGARSS), Toronto, Canada, 2002, vol 3, pp 1477–1479
4. Caves R, Luscombe AP, Lee PF, James K (2002) Topographic performance evaluation of the RADARSAT-2/3 tandem mission. In: Proceedings of geoscience and remote sensing symposium (IGARSS), Toronto, Canada, 2002, vol 2, pp 961–963
5. Massonnet D (2001) Capabilities and limitations of the interferometric cartwheel. IEEE Trans Geosci Remote 39(3):506–520
6. Winter JE, Anderson NC (2003) Distributed aperture implementation on the techsat 21 satellites. In: Proceedings of IEEE aerospace conference, Big Sky, Montana, USA, 2003, vol 2, pp 815–823
7. Fiedler H, Boerner E, Mittermayer J, Krieger G (2005) Total zero Doppler steering-a new method for minimizing the Doppler centroid. IEEE Geosci Remote Sens Lett 2(2):141–145
8. Krieger G, Moreira A (2006) Spaceborne bi- and multistatic SAR: potential and challenges. IEE Proc Radar Sonar Navig 153(3):184–198
9. Amiot T, Douchin F, Thouvenot E, Souyris JC, Cugny B (2002) The interferometric cartwheel: a multi-purpose formation of passive radar microsatellites. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Toronto, Canada, 2002, vol 1, pp 435–437
10. Moccia A, Rufino G, D'Errico M, Alberti G, Salzillo G (2002) BISSAT: a bistatic SAR for Earth observation. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), vol 5, pp 2628–2630

11. Krieger G, Moreira A (2003) Potential of digital beamforming in bi- and multistatic SAR. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), vol 1, pp 527–529
12. Ithapu VK, Mishra AK (2010) Cooperative multimonostatic SAR: a new SAR configuration for improved resolution. IEEE Antennas Wireless Propag Lett 9:701–704
13. Martin-Neira M, Caparrini M, Font-Rossello J, Lannelongue S, Vallmitjana CS (2001) The PARIS concept: an experimental demonstration of sea surface altimetry using GPS reflected signals. IEEE Trans Geosci Remote Sens 39(1):142–150
14. Graham LC (1974) Synthetic interferometer radar for topographic mapping. Proc IEEE 62(6):763–768
15. Goldstein RM, Zebker HA (1987) Interferometric radar measurement of ocean surface currents. Nature 328(6132):707–709
16. Rosen PA, Hensley S, Joughin IR, Li FK, Madsen SN, Rodriguez E, Goldstein RM (2000) Synthetic aperture radar interferometry. Proc IEEE 88(3):333–382
17. Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf D (2007) The shuttle radar topography mission. Rev Geophys 45:33
18. Gatelli F, Monti Guarnieri A, Parizzi F, Pasquali P, Prati C, Rocca F (1994) The wavenumber shift in SAR interferometry. IEEE Trans Geosci Remote Sens 32(4):855–865
19. Hagberg JO, Ulander LM, Askne J (1995) Repeat-pass SAR interferometry over forested terrain. IEEE Trans Geosci Remote Sens 33(2):331–340
20. Treuhaft RN, Madsen SN, Moghaddam M, van Zyl JJ (1996) Vegetation characteristics and underlying topography from interferometric radar. Rad Sci 31(6):1449–1485
21. Eineder M (2004) Problems and solutions for InSAR digital elevation model generation of mountainous terrain. FRINGE 2003 workshop, Frascati, Italy, vol 550, p 18, June 2004
22. Goldstein RM, Zebker HA, Werner CL (1988) Satellite radar interferometry: two dimensional phase unwrapping. Rad Sci 23(4):713–720
23. Ferretti A, Monti Guarnieri A, Prati C, Rocca F (1997) Multi-baseline interferometric techniques and applications. In: Proceedings presented at the ERS SAR interferometry, Florence, Italy, vol 406, p 243
24. Massonnet D, Vadon H, Rossi M (1996) Reduction of the need for phase unwrapping in radar interferometry. IEEE Trans Geosci Remote Sens 34(2):489–497
25. Eineder M, Adam N (2005) A maximum-likelihood estimator to simultaneously unwrap, geocode, and fuse SAR interferograms from different viewing geometries into one digital elevation model. IEEE Trans Geosci Remote Sens 43(1):24–36
26. Lombardini F, Griffiths HD (2001) Optimum and suboptimum estimator performance for multibaseline InSAR. Frequenz 55(3–4):114–118
27. Fornaro AM, Guarnieri AP, Tebaldini S (2005) Joint multi-baseline SAR interferometry. EURASIP J Adv Signal Process 2005(20):3194–3205
28. Durden SL, van Zyl JJ, Zebker HA (1989) Modeling and observation of the radar polarization signature of forested areas. IEEE Trans Geosci Remote Sens 27(3):290–301
29. Freeman A, Durden SL (1998) A three-component scattering model for polarimetric SAR data. IEEE Trans Geosci Remote Sens 36(3):963–973
30. Cloude SR, Papathanassiou KP (1998) Polarimetric SAR interferometry. IEEE Trans Geosci Remote Sens 36(5):1551–1565
31. Papathanassiou KP, Cloude SR (2001) Single-baseline polarimetric SAR interferometry. IEEE Trans Geosci Remote Sens 39(11):2352–2363
32. Lee JS, Cloude SR, Papathanassiou KP, Grunes MR, Woodhouse IH (2003) Speckle filtering and coherence estimation of polarimetric SAR interferometry data for forest applications. IEEE Trans Geosci Remote Sens 41(10):2254–2263
33. Papathanassiou KP, Cloude SR, Reigber A, Boerner WM (2000) Multi-baseline polarimetric SAR interferometry for vegetation parameters estimation. In: Proceedings of IEEE

international geoscience and remote sensing symposium, Honolulu, Hawaii, USA, vol 6, no 44, pp 2762–2764
34. Toraño Caicoya A, Kugler F, Papathanassiou K, Biber P, Pretzsch H (2010) Biomass estimation as a function of vertical forest structure and forest height: potential and limitations for radar remote sensing. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Aachen, Germany, pp 901–904, June 2010
35. Fornaro G, Serafino F (2006) Imaging of single and double scatterers in urban areas via SAR tomography. IEEE Trans Geosci Remote Sens 44(12):3497–3505
36. Reigber A, Moreira A (2000) First demonstration of airborne SAR tomography using multibaseline L-band data. IEEE Trans Geosci Remote Sens 38(5):2142–2152
37. Fornaro G, Reale D, Serafino F (2009) Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers. IEEE Trans Geosci Remote Sens 47(1):224–237
38. Tebaldini S (2010) Single and multipolarimetric SAR tomography of forested areas: a parametric approach. IEEE Trans Geosci Remote Sens 48(5):2375–2387
39. Lombardini F, Montanari M, Gini F (2003) Reflectivity estimation for multibaseline interferometric radar imaging of layover extended sources. IEEE Trans Signal Process 51(6):1508–1519
40. Gini F, Lombardini F, Montanari M (2002) Layover solution in multibaseline SAR interferometry. IEEE Trans Aerosp Electron Syst 38(4):1344–1356
41. Zhu XX, Bamler R (2010) Very high resolution spaceborne SAR tomography in urban environment. IEEE Trans Geosci Remote Sens 48(12):4296–4308
42. Nannini M, Scheiber R, Moreira A (2009) Estimation of the minimum number of tracks for SAR tomography. IEEE Trans Geosci Remote Sens 47(2):531–543
43. Xiao Xiang Zhu, Bamler R (2011) Sparse reconstruction techniques for SAR tomography. In: Paper presented at the 17th international conference on digital signal processing (DSP), Corfu, Greece, pp 1–8, 2011
44. Zhu XX, Bamler R (2010) Tomographic SAR inversion by -norm regularization – the compressive sensing approach. IEEE Trans Geosci Remote Sens 48(10):3839–3846
45. Budillon A, Evangelista A, Schirinzi G (2011) Three-dimensional SAR focusing from multipass signals using compressive sampling. IEEE Trans Geosci Remote Sens 49(1):488–499
46. Xiao Xiang Zhu, Adam N, Brcic R, Bamler R (2009) Space-borne high resolution SAR tomography: experiments in urban environment using TS-X Data. In: Urban remote sensing event, 2009 Joint, Shanghai, China, pp. 1–8, 2009
47. Schwarz G (1978) Estimating the dimension of a model. Ann Statist 6(2):461–464
48. Chapin E, Chen CW (2008) Along-track interferometry for ground moving target indication. IEEE Aerosp Electron Syst Mag 23(6):19–24
49. Moccia A, Rufino G (2001) Spaceborne along-track SAR interferometry: performance analysis and mission scenarios. IEEE Trans Aerosp Electron Syst 37(1):199–213
50. Gill E, Runge H (2003) Tight formation flying for an along-track SAR interferometer. In: Proceedings of 54th international astronautical federation congress, Bremen, pp 473–485
51. Lombardini F, Bordoni F, Gini F, Verrazzani L (2004) Multibaseline ATI-SAR for robust ocean surface velocity estimation. IEEE Trans Aerosp Electron Syst 40(2):417–433
52. Goodman JW (1976) Some fundamental properties of speckle. J Opt Soc Am 66(11):1145–1150
53. Lee JS (1981) Speckle analysis and smoothing of synthetic aperture radar images. Comput Graph Image Process 17(1):24–32
54. Lopez-Martinez C, Fabregas X (2003) Polarimetric SAR speckle noise model. IEEE Trans Geosci Remote Sens 41(10):2232–2242
55. Prati C, Rocca F (1993) Improving slant-range resolution with multiple SAR surveys. IEEE Trans Aerosp Electron Syst 29(1):135–143

56. Prats P, Lopez-Dekker P, De Zan F, Wollstadt S, Bachmann M, Steinbrecher U, Scheiber R, Reigber A, Krieger G (2011) Distributed imaging with TerraSAR-X and TanDEM-X. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), 2011, pp 3963–3966
57. Curlander JC, McDonough N (1991) Synthetic aperture radar: systems and signal processing. Wiley, New York
58. Ulaby FT, Moore RK, Fung AK (1986) Microwave remote sensing: active and passive, volume II: radar remote sensing and surface scattering and emission theory. Artech House, Dedham, MA
59. Freeman A, Johnson WT, Huneycutt B, Jordan R, Hensley S, Siqueira P, Curlander J (2000) The 'Myth' of the minimum SAR antenna area constraint. IEEE Trans Geosci Remote Sens 38(1):320–324
60. De Zan F, Monti Guarnieri A (2006) TOPSAR: terrain observation by progressive scans. IEEE Trans Geosci Remote Sens 44(9):2352–2360
61. Younis M, Fischer C, Wiesbeck W (2003) Digital beamforming in SAR systems. IEEE Trans Geosci Remote Sens 41(7):1735–1739
62. Krieger G, Gebert N, Moreira A (2008) Multidimensional waveform encoding: a new digital beamforming technique for synthetic aperture radar remote sensing. IEEE Trans Geosci Remote Sens 46(1):31–46
63. Gebert N, Krieger G, Moreira A (2009) Digital beamforming on receive: techniques and optimization strategies for high-resolution wide-swath SAR imaging. IEEE Trans Aerosp Electron Syst 45(2):564–592
64. Goodman NA, Stiles JM (2001) The information content of multiple receive aperture SAR systems. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), 2001, vol 4, pp 1614–1616
65. Goodman NA, Stiles JM (2002) Synthetic aperture characterization of radar satellite constellations. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Toronto, Canada, 2002, vol 1, pp 665–667
66. Goodman NA, Stiles JM (2003) Resolution and synthetic aperture characterization of sparse radar arrays. IEEE Trans Aerosp Electron Syst 39(3):921–935
67. Goodman NA, Sih Chung Lin NA, Rajakrishna D, Stiles JM (2002) Processing of multiple-receiver spaceborne arrays for wide-area SAR. IEEE Trans Geosci Remote Sens 40(4):841–852
68. Hill GW (1878) Researches in the lunar theory. Am J Math 1(1):5–26
69. Clohessy W, Wiltshire R (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 270(9):653–658
70. Prussing JE, Conway BA (1993) Orbital mechanics. Oxford University Press, USA
71. Moreira A, Krieger G, Mittermayer J (2004) Satellite configuration for interferometric and/or tomographic remote sensing by means of synthetic aperture radar (SAR). US Patent 667788413 Jan 2004
72. D'Amico S, Montenbruck O (2006) Proximity operations of formation-flying spacecraft using an eccentricity/inclination vector separation. J Guid Contr Dyn 29(3):554–563
73. Massonnet D (1998) Roue interferometrique. US Patent 236910D17306RS
74. Klare J (2008) Digital Beamforming for a 3D MIMO SAR – improvements through frequency and waveform diversity. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), vol 5
75. Ender JH, Klare J (2009) System architectures and algorithms for radar imaging by MIMO-SAR. In: Proceedings of IEEE radar conference, Pasadena, California, USA, pp 1–6
76. Carrara WC, Majewski RM, Goodman RS (1995) Spotlight synthetic aperture radar: signal processing algorithms. Artech House, Boston
77. Mittermayer J, Moreira A, Loffeld O (1999) Spotlight SAR data processing using the frequency scaling algorithm. IEEE Trans Geosci Remote Sens 37(5):2198–2214

78. Currie A, Brown MA (1992) Wide-swath SAR. IEE Proc F Radar Signal Process 139(2):122–135
79. Monti Guarnieri A, Prati C (1996) ScanSAR focusing and interferometry. IEEE Trans Geosci Remote Sens 34(4):1029–1038
80. Krieger G, Cassola MR, Younis M, Metzig R (2005) Impact of oscillator noise in bistatic and multistatic SAR. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), 2005, vol 2, pp 1043–1046
81. Weib M (2004) Synchronisation of bistatic radar systems. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Anchorage, Alaska, USA, 2004, vol 3, pp 1750–1753
82. Younis M, Metzig R, Krieger G (2006) Performance prediction of a phase synchronization link for bistatic SAR. IEEE Geosci Remote Sens Lett 3(3):429–433
83. Lopez-Dekker P, Mallorqui JJ, Serra-Morales P, Sanz-Marcos J (2008) Phase synchronization and Doppler centroid estimation in fixed receiver bistatic SAR systems. IEEE Trans Geosci Remote Sens 46(11):3459–3471
84. Vannicola V, Varshney P (1983) Spectral dispersion of modulated signals due to oscillator phase instability: white and random walk phase model. IEEE Commun 31(7):886–895
85. Demir A (2002) Phase noise and timing jitter in oscillators with colored-noise sources. IEEE Trans Circuits Syst I, Fundam Theory Appl 49(12):1782–1791
86. Demir A, Mehrotra A, Roychowdhury J (2000) Phase noise in oscillators: a unifying theory and numerical methods for characterization. IEEE Trans Circuits Syst I, Fundam Theory Appl 47(5):655–674
87. Rutman J, Walls FL (1991) Characterization of frequency stability in precision frequency sources. Proc IEEE 79(7):952–960
88. Candelier V, Canzian P, Lamboley J, Brunet M, Santarelli G (2003) Space qualified 5 MHz ultra stable oscillators. In: Proceedings of the 2003 I.E. international frequency control symposium and PDA exhibition jointly with the 17th European frequency and time forum, pp 575–582
89. Guillemot P, Dutrey J-F, Vega J-F, Chaubet M, Chebance D, Sirmain C, Santarelli G, Chambon D, Laurent P, Rousselet M, Locke C, Ivanov E, Tobar M, Potier T (2004) The PHARAO time and frequency performance verification system. In: Proceedings of the 2004 I. E. international frequency control symposium and exposition, Montreal, Canada, pp 785–789
90. Wang W-Q (2009) GPS-based time & phase synchronization processing for distributed SAR. IEEE Trans Aerosp Electron Syst 45(3):1040–1051
91. Wen-Qin Wang (2008) Baseline estimation in distributed spaceborne interferometry SAR systems. In: Proceedings of 2008 I.E. aerospace conference, Big Sky, Montana, USA, pp 1–8
92. Montenbruck O, Leung S (2005) Real-time navigation of formation-flying spacecraft using global-positioning-system measurements. J Guid Contr Dyn 28(2):226–235
93. Kroes R, Montenbruck O, Bertiger W, Visser P (2005) Precise GRACE baseline determination using GPS. GPS Solutions 9(1):21–31
94. Reigber A, Prats P, Mallorqui JJ (2006) Refined estimation of time-varying baseline errors in airborne SAR interferometry. IEEE Geosci Remote Sens Lett 3(1):145–149
95. Prats P, Scheiber R, Reigber A, Andres C, Horn R (2009) Estimation of the surface velocity field of the Aletsch Glacier using multibaseline airborne SAR interferometry. IEEE Trans Geosci Remote Sens 47(2):419–430
96. Rodriguez-Cassola M, Prats P, Lopez-Dekker P, Krieger G, Moreira A (2010) General Processing Approach for Bistatic SAR Systems: description and Performance Analysis. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), pp 998–1001, June 2010
97. Moreira A, Hajnsek I, Krieger G, Papathanassiou K, Eineder M, De Zan F, Younis M, Werner M (2009) Tandem-L: monitoring the Earth's Dynamics with InSAR and Pol-InSAR. In: Proceedings of the international workshop on applications of polarimetry and polarimetric interferometry (Pol-InSAR), Frascati, Italy, p 5, January 2009

98. Krieger G, Hajnsek I, Papathanassiou K, Eineder M, Younis M, De Zan, Prats P, Huber S, Werner M, Fiedler H, Freeman A, Rosen P, Hensley S, Johnson W, Veilleux L, Grafmueller B, Werninghaus R, Bamler R, Moreira A (2009) The tandem-L mission proposal: monitoring earth's dynamics with high resolution SAR interferometry. In: Proceedings of 2009 I.E. radar conference Pasadena, California, USA, pp 1–6
99. Moreira A, Krieger G, Younis M, Hajnsek I, Papathanassiou K, Eineder M, De Zan F (2011) Tandem-L: a mission proposal for monitoring dynamic earth processes. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Vancouver, Canada, pp 1–4, July 2011
100. Krieger G, Gebert N, Younis M, Bordoni F, Patyuchenko A, Moreira A (2008) Advanced Concepts for Ultra-Wide-Swath SAR Imaging. In: Proceedings of 2008 7th European conference on synthetic aperture radar (EUSAR), pp 1–4
101. Freeman A, Krieger G, Rosen P, Younis M, Johnson W, Huber S, Jordan R, Moreira A (2009) SweepSAR: beam-forming on receive using a reflector-phased array feed combination for spaceborne SAR. In: Proceedings of 2009 I.E. radar conference, Pasadena, California, USA, pp 1–9
102. Younis M, Huber S, Patyuchenko A, Bordoni F, Krieger G (2009) Performance comparison of reflector- and planar-antenna based digital beam-forming SAR. Int J Antennas Propag 2009:1–13
103. Villano M, Moreira A, Miller H, Rott H, Hajnsek I, Hajnsek I, Bamler R, Lopez-Dekker P, Boerner T, Zan FD, Krieger G, Papathanassiou KP (2010) SIGNAL: mission concept and performance assessment. In: Proceedings of 2010 8th European conference on synthetic aperture radar (EUSAR), Aachen, Germany, pp 1–4
104. Börner T, De Zan F, López-Dekker F, Krieger G, Hajnsek I, Papathanassiou K, Villano M, Younis M, Danklmayer A, Dierking W, Nagler T, Rott H, Lehner S, Fügen T, Moreira A (2010) Signal: SAR for ice, glacier and global dynamics. In: Proceedings of IEEE geoscience and remote sensing symposium (IGARSS) 2010, Honolulu, 2010
105. López-Dekker P, Börner T, Younis M, Krieger G (2011) SIGNAL: a Ka-band digital beam-forming SAR system concept to monitor topography variations of ice caps and glaciers. In: Proceedings of advanced RF sensors and remote sensing instruments (ARSI), Noordwijk, The Netherlands, pp 1–10, September 2011
106. Pritchard HD, Arthern RJ, Vaughan DG, Edwards LA (2009) Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. Nature 461(7266):971–975
107. Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: Mission overview and early results, Geophys Res Lett 31, L09607, doi:10.1029/2004GL019920, http://www.agu.org/pubs/crossref/2004/2004GL019920.shtml
108. Klees R, Koop R, Visser P, van den IJssel J (2000) Efficient gravity field recovery from GOCE gravity gradient observations. J Geodesy 74(7–8):561–571
109. Younis M, Bordoni F, Gebert N, Krieger G (2008) Smart multi-aperture radar techniques for spaceborne remote sensing. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Boston, Massachusetts, USA, 2008, vol 3, pp 278–281
110. Rostan F, Riegger S, Pitz W, Torre A, Torres R (2007) The C-SAR instrument for the GMES sentinel-1 mission. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Barcelona, Spain, 2007, pp 215–218
111. Snoeij P, Attema E, Davidson M, Duesmann B, Floury N, Levrini G, Rommen B, Rosich B (2010) Sentinel-1 radar mission: status and performance. IEEE Aerosp Electron Syst Mag 25(8):32–39
112. Fjortoft R, Souyris J-C, Gaudin J-M, Durand P, Massonnet D (2004) Impact of ambiguities in multistatic SAR: some specificities of an L-band interferometric cartwheel. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Anchorage, Alaska, USA, 2004, vol 3, pp 1754–1757

113. Chapron B, Collard F, Ardhuin F (2005) Direct measurements of ocean surface velocity from space: Interpretation and validation. J Geophys Res 110, C07008, doi:10.1029/2004JC002809, http://www.agu.org/pubs/crossref/2005/2004JC002809.shtml
114. Romeiser R, Thompson DR (2000) Numerical study on the along-track interferometric radar imaging mechanism of oceanic surface currents. IEEE Trans Geosci Remote Sens 38(1):446–458
115. Romeiser R, Hirsch O (2001) Possibilities and limitations of current measurements by airborne and spaceborne along-track interferometric SAR. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Sydney, Australia, 2001, vol 1, pp 575–577
116. Romeiser R, Breit H, Eineder M, Runge H (2002) Demonstration of current measurements from space by along-track SAR interferometry with SRTM data. In: Proceedings of IEEE international geoscience and remote sensing symposium (IGARSS), Toronto, Canada, 2002, vol 1, pp 158–160
117. Romeiser R (2005) Current measurements by airborne along-track InSAR: measuring technique and experimental results. IEEE J Ocean Eng 30(3):552–569
118. Pierson WJP Jr, Moskowitz L (1964) A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. J Geophys Res 69(24):5181–5519
119. Fung A, Khim Lee A (1982) A semi-empirical sea-spectrum model for scattering coefficient estimation. IEEE J Ocean Eng 7(4):166–176
120. Hersbach H, Stoffelen A, de Haan S (2007) An improved C-band scatterometer ocean geophysical model function: CMOD5. J Geophys Res 112:18

Part II
Relative Dynamics and GNC

Chapter 3
Relative Trajectory Design

Marco D'Errico and Giancarmine Fasano

Abstract An analysis of orbital relative motion models is presented with emphasis on their application to formation design. Relative motion model evolution from the first Hill's schematization (circular orbit, close satellites) is described, considering the inclusion of chief's orbit eccentricity and orbital perturbations. In particular, the inclusion of J_2-secular effects is treated in depth considering various approaches in literature. Literature is also reviewed for both small and large eccentricities. Further details are presented to model formations with small chief's eccentricity (order of 10^{-3}), which are typical of Earth observation missions, for both the case of close formations, i.e. with satellite distance of the order of tens of kilometers, and large formations, i.e. satellite distance up to hundreds of kilometers. Finally, design applications are presented, with derivation of relative trajectories from application requirements. As an example, relative orbits for SAR interferometry are derived from the requested altitude measurement uncertainty and considering different candidate geometries (pendulum, cartwheel, etc.). Relative orbits for SAR tomography and large baseline bistatic SAR applications are also analyzed.

M. D'Errico (✉)
Department of Aerospace and Mechanical Engineering, Second University of Naples,
Via Roma 29, 81031 Aversa (CE), Italy
e-mail: marco.derrico@unina2.it

G. Fasano
Department of Aerospace Engineering, University of Naples "Federico II", P.le Tecchio 80,
80125 Naples, Italy
e-mail: g.fasano@unina.it

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_3,

3.1 Introduction

Formation flying builds on a specific mission-dependent relative motion between platforms. Therefore, adequate modeling of orbital relative dynamics is the key to enable distributed missions. Hill's equations [1–3], developed in the ambit of Lunar research, are worldwide considered as the first schematization of orbital relative motion, later independently replicated by Clohessy and Wiltshire [4] and developed by many other authors in the ambit of rendezvous research (e.g. [5–7]). In this perspective, formation flying is quite an old theme because it has been studied since the very beginning of manned space programs to develop rendezvous and docking technologies.

In the last couple of decades, new mission ideas have been arising, which foresee multiple spacecraft system working for the entire lifetime in either close or loose proximity. It is then mission duration that makes modern formation flying different from rendezvous, leading to completely different requirements, including relative motion schematization which needs further terms with respect to former models [8]. Thus, the description of relative motion on near circular orbits through linearized differential equations based on the assumptions of circular reference orbit, small relative distances, and spherical Earth [1–4] has been gradually improved. In particular, one of the major error source is the Keplerian dynamics approximation, which renders the solution useless on long timescales. J_2 potential is the main perturbing effect in LEO, thus different approaches have been followed to incorporate it. For example, Vadali et al. [9] and Wiesel [10] developed linearized time-varying coefficients differential equations, while Schweighart and Sedwick [11] found for a circular reference orbit a set of constant coefficient equations similar in form to Hill's equations, but able to capture Earth oblateness effects. It is worth noting that only the latter model exhibits an analytical solution. More recently, relative dynamics for near circular perturbed LEO orbits has been treated by Halsall and Palmer [12]. A significant research effort has also been performed in the case of reference orbits of arbitrary eccentricity [13–16]. Some of the undertaken efforts have included analyses in the mean orbit element space. As an example, Schaub and Alfriend [17] defined J_2-invariant orbits and Schaub et al. [18] analyzed formation control. In addition, Schaub [19] studied the problem in the classical orbit elements space under the assumption that all orbit parameter differences must be small, but without a time-explicit representation of relative motion. Finally, Fasano and D'Errico [20] developed a model in the classical mean orbit element space which offers a time-explicit representation of relative trajectories.

When one has to select or develop a dynamical model, a tradeoff is always needed between completeness and complexity, which is even more critical when the purpose is trajectory design. Complex models allow for a more realistic representation of the motion, with the advantage of defining a nominal trajectory which inherently accounts for a large number of perturbations. Therefore, the efforts needed to control the real motion in the neighborhood of the nominal trajectory are minimized. On the other hand, a strong modeling complexity can

spoil model usefulness because it makes difficult to explore all design possibilities by analytic or semi-analytic procedures (i.e. through an extensive parametric analysis). As an example, for orbit design of classical, monolithic Earth observation missions the current design practice builds on relatively simple analytical models which include the most relevant perturbations (e.g. J_2 secular to define sun-synchronous and/or repetitive orbits, $J_2 + J_3$ secular for frozen orbits) to define a preliminary orbit design to be successively verified by more complex numerical tools. Such models typically use classical mean orbit elements because, despite their weakness (singularities under special conditions), they offer a direct physical and geometrical interpretation of the design result.

With reference to relative trajectory design, it is a widespread opinion that Clohessy and Wiltshire models are largely inadequate for modern formation design because they do not capture relative J_2 perturbations and consequent relative drifts. If a relative trajectory is designed with this model, the relative orbit control will be in practice fighting (and expelling propellant) against J_2 differential perturbations. Such disproportionate procedure would be equivalent to a design of Earth orbits under Keplerian dynamics, with the orbit control designed to remove, among the other perturbations, ascending node and perigee precessions. Nonetheless, no standard formation design practice has been assessed and agreed up to now, probably because most of the huge literature contribution in the last decade, very partially summarized previously, has been mainly devoted to formation control rather than formation design. In this ambit, the first important contribution was proposed by Schaub and Alfriend [17, 21]. They introduced the concept of J_2-invariant relative orbits, i.e. obtained by means of deputy and chief orbital parameter differences (semi-major axis, inclination, and eccentricity) able to nullify mean relative drifts of ascending node right ascension and argument of latitude. Their analysis showed that such concept may have some critical points when implemented for near-polar orbits (when inclination is close to 90°) and for near circular-orbits (when eccentricity approaches 0). In such cases in fact, a desired inclination difference usually leads to a large requirement on eccentricity difference, which in turn produces large relative trajectories.

In order to describe relative motion, satellites are typically defined as chief and deputy. With this wording, it is implicitly assumed that the relative motion under analysis is the one of the deputy with respect to the chief. Earth orbits of both satellites are defined in the earth-centered inertial reference frame (Fig. 3.1), whereas to model the relative motion a reference frame attached to the chief is needed. The most typical choice is the Hill reference frame, which is rigorously defined for chief circular orbits only as: origin in the chief center of mass, y-axis directed along the velocity vector, x-axis along the radial direction, z-axis perpendicular to the angular plane (aligned with angular momentum to complete the right-handed reference frame). Such definition can be easily generalized for chief elliptic orbit by keeping the definition for the origin, x- and z-axis and by defining y-axis as the one completing the right-handed reference frame, which leads to an y-axis in the orbital plane and with positive projection along the chief velocity vector. The Hill reference frame is not to be confused with the widespread Local Horizon Local

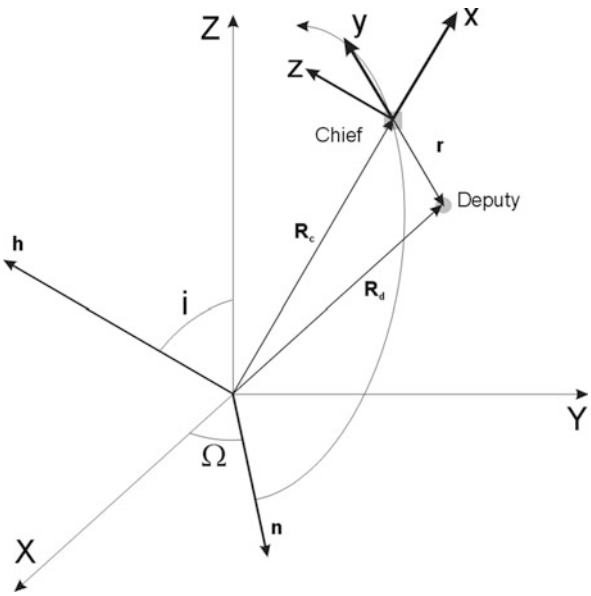


Fig. 3.1 Geometry and reference frames to describe relative motion (**r**) of the deputy satellite (position $\mathbf{R_d}$ in the geocentric inertial reference frame, XYZ) with respect to the chief satellites (position, $\mathbf{R_c}$; inclination, i; ascending node right ascension, Ω; angular momentum, **h**; line-of-node vector, **n**) in the Hill reference frame (xyz)

Vertical (LHLV) reference frame, since $x_{LHLV} \| y$, $y_{LHLV} \| -z$, and $z_{LHLV} \| -x$. To limit the complication of a reference frame moving with varying angular velocity and with y-axis forming a variable angle with respect to velocity vector, a fictitious satellite on a circular orbit with the chief's semi-major axis is sometimes introduced as a reference point to describe both chief and deputy local motions.

With this background, the description of relative dynamics consists in deriving the relative position of the deputy with respect to the chief as a function of time in the generalized Hill reference frame once the "absolute" motions are known (i.e. the functions $\mathbf{R_c}(t)$ and $\mathbf{R_d}(t)$). Instead, when formation is designed the issue is deriving $\mathbf{R_d}(t)$ on the basis of $\mathbf{R_c}(t)$ and a desired (application-dependent) $\mathbf{r}(t)$. Of course the "desired" $\mathbf{r}(t)$ is to be traded with Earth gravity field since, as it is well known in standard orbit design, fighting against mother nature is not cost effective. Relative motion models can be roughly grouped into two families depending if they are developed either using Cartesian coordinates in the Hill reference frame or orbital parameters differences (but a few hybrid approaches have also been proposed). If on the one hand it is extremely difficult to give general statements, on the other hand a logical rule of thumb consists in utilizing the latter models for formation design, whereas Cartesian coordinate models are better suited for control problems.

In the following, the formation design problem for small eccentricity chief orbits is discussed in Sect. 3.2, while an overview of the works developed for large eccentricity orbits is presented in Sect. 3.3. It is worth noting that Sect. 3.2 covers both the class of near-circular and small eccentricity orbits as defined by Schaub [19].

3.2 Relative Trajectory Modeling: Small Eccentricity Orbits

Relative motion in small eccentricity orbits has special interest for distributed space systems. From the application point of view, their majority require circular or near circular orbits: in typical remote sensing missions, the eccentricity is at most of the order of 10^{-3} and it is usually set to achieve sun-synchronous and/or frozen orbits [22]. In addition, from the dynamics point of view, relative trajectories in small eccentricity formations exhibit some simple features which are gradually lost for an increasing eccentricity. Finally, from the mathematics point of view, motion description is greatly simplified by the fact that the developments in power series of eccentricity can be arrested at first order. This aspect becomes even more important when dealing with trajectory design.

A key point for relative trajectory modeling is related to the distance (baseline) between satellites. As a rule of thumb, if it is of the order of 1/1,000 of the absolute orbit semi-major axis, one can likely consider a linearization of the problem, neglecting higher order effects. This is especially true for design purposes, where modeling complexity has to be carefully traded off against model applicability. In practical terms, in the case of typical Earth observation missions flying in LEO, a "close formation" is characterized by a baseline of the order of a few kilometers, while "large formations" foresee baselines of hundreds of kilometers. Even in low eccentricity cases, large formations require inclusion of second order terms for proper motion modeling. Close and large formations are treated in separate subsections in the following.

3.2.1 "Close" Formations

Within the huge amount of literature on the dynamics of close formations in small eccentricity and circular/quasi circular orbits, papers and research approaches are considered here, which can be most useful for the formation designer. As already stated in the introduction, the simplest model to describe deputy motion with respect to the chief in terms of relative position and velocity coordinates is represented by Hill's equations, also called Clohessy-Wiltshire equations (HCW) [1–4]. HCW equations are based on the assumptions of central gravitational field as the only external force, Keplerian circular orbit for the chief satellite, small distance between chief and deputy. They can be derived by writing the dynamic equation of the deputy with respect to the chief and then linearizing the gravitational acceleration around the chief position. In typical cases, HCW modeling error tends to increase with time because of the underlying assumptions. Hill's equations can be derived in differential form as follows:

$$\begin{cases} \ddot{x} - 2n\dot{y} - 3n^2 x = 0 \\ \ddot{y} + 2n\dot{x} = 0 \\ \ddot{z} + n^2 z = 0 \end{cases} \tag{3.1}$$

where n is the Keplerian mean motion of the chief satellite.

These linear constant coefficients differential equations admit an analytical solution that is, considering all the terms:

$$\begin{cases} x(t) = (\dot{x}_0/n)\sin(nt) - (3x_0 + 2\dot{y}_0/n)\cos(nt) + 4x_0 + 2\dot{y}_0/n \\ y(t) = (2\dot{x}_0/n)\cos(nt) + (6x_0 + 4\dot{y}_0/n)\sin(nt) - (6nx_0 + 3\dot{y}_0)t - 2\dot{x}_0/n + y_0 \\ z(t) = (\dot{z}_0/n)\sin(nt) + z_0\cos(nt) \end{cases} \tag{3.2}$$

The main features of relative motion can be straightforwardly understood from the solution. First of all, relative trajectory is a plane trajectory and its projection on the radial/along-track plane is a 2 × 1 ellipse, i.e., an ellipse with fixed eccentricity. All the oscillating terms are at orbital frequency, and in particular while the z coordinate exhibits a simple harmonic oscillation, an offset term appears in the radial coordinate, and both an offset and a linear secular drift appear in the y coordinate. It is worth noting that the along-track drift is directly proportional to the radial offset.

In HCW modeling framework, stable relative motion can be obtained if

$$\dot{y}_0 = -2x_0 n$$

The origin of this behavior can be better appreciated if relative trajectory is described in terms of orbital parameters and their differences.

Several approaches can be found in literature to derive the equations describing relative motion in the case of close formations: for example, Schaub [19] starts from a linear mapping between Cartesian Hill coordinates and orbit elements differences [23], Fasano and D'Errico [20, 24] and Vadali et al. [25] first write the exact relative motion equations in terms of differences in orbital parameters, then apply a series of simplifying assumptions.

Let us first consider the case of chief moving on a circular orbit.

In the following, the subscript D will refer to deputy, while no subscripts will be used for chief parameters and the symbol δ will represent the difference between deputy and chief parameters.

The different approaches are all based on the assumption that all the differences in orbit parameters are very small, i.e., $\delta a/a$, $\delta\Omega$, δi, δe (i.e. e_D) and $\omega_D + M_{D0} - u_0$ are $<<1$. u_0 represents the initial value of chief anomaly computed with respect to its ascending node (that is, chief's initial argument of latitude). In the following, in general u will represent the mean argument of latitude, i.e., $u = \omega + M$.

Relative motion equations in terms of orbital parameters, under Keplerian dynamics and with the chief true anomaly as independent variable can be written as proposed by Schaub [19]:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cong a \begin{bmatrix} \frac{\delta a}{a} - \delta e \cos(v) \\ 2\delta e \sin(v) + \delta\omega + \delta M_0 + \delta\Omega \cos i - \frac{3}{2}(v - v_0)\delta a \\ \sqrt{\delta i^2 + \sin^2 i \delta\Omega^2} \cos(\omega + v - \phi) \end{bmatrix} \tag{3.3}$$

Where

$$\phi = \tan^{-1}\left(\frac{\delta i}{-\sin i \delta\Omega}\right) \tag{3.4}$$

Under the same assumptions, the equations can be written in time explicit form and in a form that easily allows inclusion of Earth oblateness terms as done by Fasano and D'Errico [24]:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cong a \begin{bmatrix} \dfrac{\delta a}{a} - \delta e \cos\left(M_{D0} + \dot{M}_D t\right) \\ 2\delta e \sin\left(M_{D0} + \dot{M}_D t\right) + (\omega_{D0} + M_{D0} - u_0) + \delta\Omega_0 \cos i + \left(\delta\dot{u} + \delta\dot{\Omega}\cos i\right)t \\ -\left(\delta\Omega_0 + \delta\dot{\Omega}t\right)\sin i \cos(\omega_{D0} + M_{D0} + \dot{u}_D t) + \delta i \sin(\omega_{D0} + M_{D0} + \dot{u}_D t) \end{bmatrix} \tag{3.5}$$

In Keplerian dynamics, (3.3) coincide with (3.5) at first order, and (3.5) coincide with HCW equations (3.2). Comparison of these equation sets puts in evidence the origin of the HCW terms, In particular:

1. The radial offset is due to the difference in semi-major axes;
2. The harmonic in-plane oscillations are linearly related to deputy orbit eccentricity ($\delta e = e_D$);
3. The along-track offset depends on the initial difference in (mean) argument of latitude;
4. The cross-track oscillation can be activated by means of a difference in right ascension of the ascending node or inclination. The latter allows to achieve maximum separation at the poles and minimum on the equator, the other way round for the former;
5. In general (Earth oblateness effects included), the along track drift can be due to the differential effects $\delta\dot{u}$ and $\delta\dot{\Omega}$. These terms will be further discussed later.

The equations written in terms of orbit parameters clearly show that the in-plane motion is phased relative to the perigee (mean anomaly), while the out of plane motion is phased relative to the ascending node (argument of latitude). Thus, the in-plane and the out-of-plane oscillations will have the same frequency as long as perigee precession is neglected (Keplerian dynamics) or nullified.

Let us now consider the equations of relative motion which are applicable when the chief does not move on a circular orbit but on a low eccentricity orbit. From an application point if view, this means that the nominal orbit eccentricity is of the order of 10^{-3} (consider for example sun-synchronous frozen orbits [22]), which allows a linear approximation of terms related to eccentricity.

One interesting aspect to be considered is that, as noticed in Ref. [24] and then by Vadali et al. [25], in a low eccentricity close formation case the difference in mean anomaly and in argument of perigee is not necessarily small, although the sum of their differences shall be.

In time explicit form one can derive the equations [20]:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cong a \begin{bmatrix} \frac{\delta a}{a} - \delta e \cos(M_0 + \delta M_0 + \dot{M}_D t) + 2e \sin\left(\frac{\delta M_0}{2}\right) \sin\left(M_0 + \dot{M}_D t + \frac{\delta M_0}{2}\right) \\ 2\delta e \sin(M_0 + \delta M_0 + \dot{M}_D t) + 4e \sin\frac{\delta M_0}{2} \cos\left(M_0 + \frac{\delta M_0}{2} + \dot{M}_D t\right) \\ \quad + \delta(\omega_0 + M_0) + \delta\Omega_0 \cos i + t\left(\delta\dot{u} + \delta\dot{\Omega}\cos i\right) \\ \quad -(\delta\Omega_0 + \delta\dot{\Omega}t)\sin i \cos(\omega_{D0} + M_{D0} + \dot{u}_D t) + \delta i \sin(\omega_{D0} + M_{D0} + \dot{u}_D t) \end{bmatrix} \tag{3.6}$$

Vadali et al. [25] derive similar equations, with the in-plane terms arranged in a different way and with some second order terms in the cross-track coordinate:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cong a \begin{bmatrix} \frac{\delta a}{a} + [(e_D \sin\delta M)\sin M + (e - e_D \cos\delta M)\cos M] \\ [\delta u + \delta\Omega \cos i - e(e_D \sin\delta M)] + 2[-(e - e_D\cos\delta M)\sin M + (e_D \sin\delta M)\cos M] \\ \delta i \sin u_D - \sin i \delta\Omega \cos u_D - \frac{3}{2}e(\delta i \sin\omega_D - \sin i \delta\Omega \cos\omega_D) \end{bmatrix} \tag{3.7}$$

If one compares the in-plane terms in (3.6) and (3.5), at first order a new harmonic term appears both in the radial and in the along-track coordinate which is due to the difference in mean anomaly between satellites and is linearly related to chief eccentricity. For example, considering the along-track coordinate, this oscillation has an amplitude of $4e \sin\frac{\delta M_0}{2}$ while the phase difference with respect to the other harmonic term related to δe is $\frac{\pi}{2} - \frac{\delta M_0}{2}$. Thus, it is important to underline that contrary to the δe-induced oscillation, δM impacts both the amplitude and the phase (with respect to the perigee) of this harmonic motion. When a small anomaly difference between the satellites exists, the δM-dependent term is of higher order and thus can be neglected. Thus, the in-plane motion is a combination of two harmonic oscillations both in the radial and in the along-track coordinate, which can be graphically explained if one considers a virtual reference point which moves on a circular orbit sharing all chief orbital parameters except eccentricity (and of course, in time, true anomaly). Figure 3.2 describes chief and deputy motion with respect to this virtual platform. Since at first order the Hill's reference frame of the reference point is aligned with the chief's Hill reference frame, one can easily find out the two harmonic contributions.

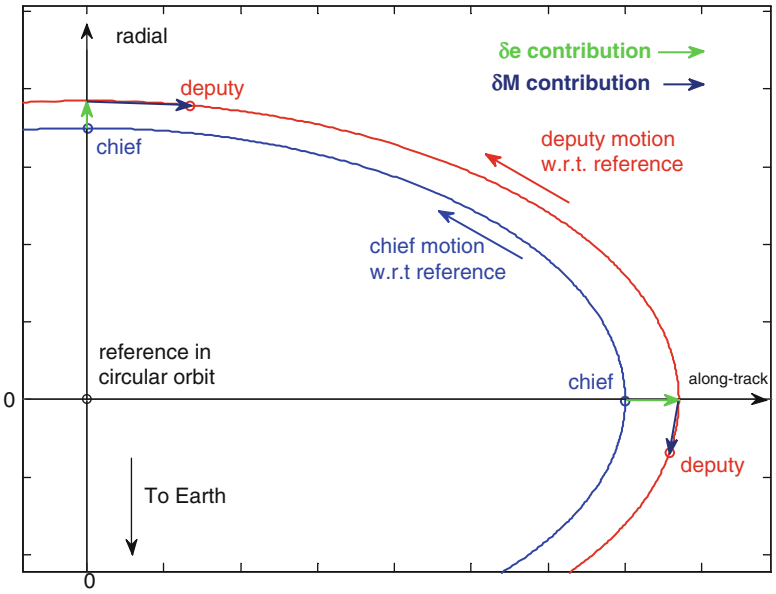


Fig. 3.2 Graphical explanation of the different contributions to in-plane relative motion when the chief satellite is moving on a low eccentricity orbit

Equation 3.6 also shows how the deputy perigee position influences the way the in-plane motion is phased relative to the cross track oscillation: for example, Fig. 3.3 shows what happens if the chief perigee is placed at 0°, at 45° or at 90°, and a difference in eccentricity is combined with a difference in right ascension of the ascending node. Modifying the relative phasing of in-plane and out-of-plane oscillations changes the xz and yz projections, which can be ellipses, circles or lines, while the relative trajectory always lies on the surface of an elliptic cylinder with axis along the cross-track direction [26].

Considering mean orbit parameters and secular Earth oblateness effects, both absolute and differential J_2 effects on relative trajectory are of particular interest to the formation designer.

From Ref. [27] we get:

$$\dot{\Omega} = -\frac{3}{2} J_2 \frac{R_{\oplus E}^2}{p^2} \dot{M} \cos i \tag{3.8}$$

$$\dot{\omega} = \frac{3}{2} J_2 \frac{R_{\oplus E}^2}{p^2} \dot{M} \left(2 - \frac{5}{2} \sin^2 i \right) \tag{3.9}$$

$$\dot{M}_p = \frac{3}{2} J_2 \frac{R_{\oplus E}^2}{p^2} n \sqrt{1 - e^2} \left(1 - \frac{3}{2} \sin^2 i \right) \tag{3.10}$$

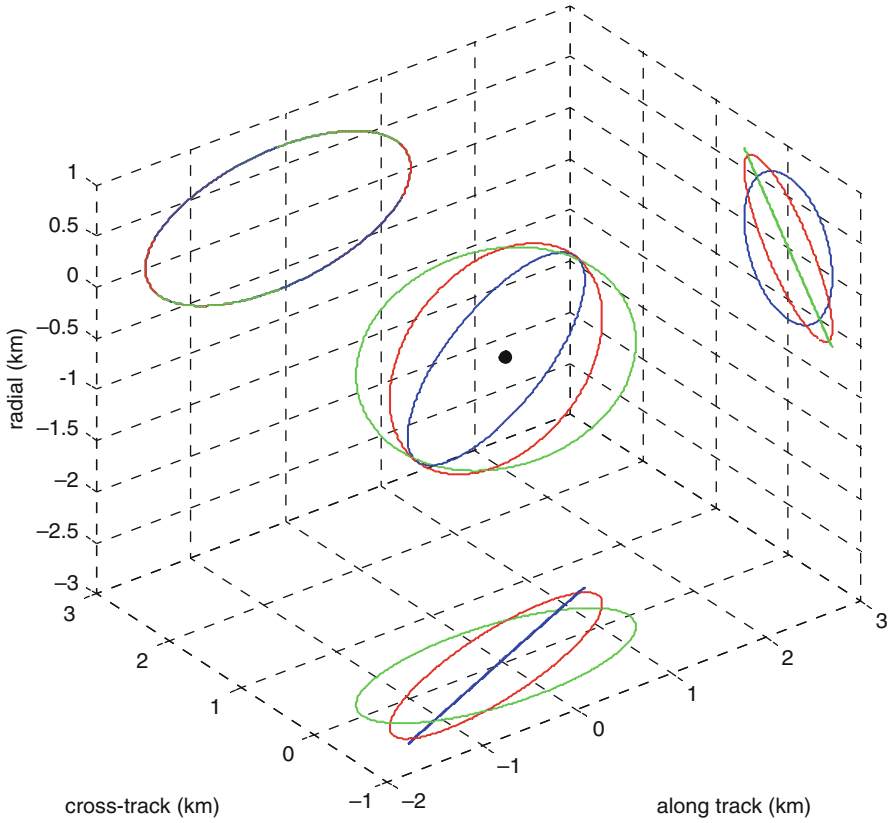


Fig. 3.3 3d relative motion and its projection on the coordinate planes for a formation with e = 0.001, a = 7,000 km, $\delta e = 1 \times 10^{-4}$, $\delta\Omega = 0.005°$, and different choices for the two perigees (0°, 45°, 90°)

First of all, if perigee precession rate calculated from (3.9) is non zero, then the phasing between in-plane and out-of-plane motion will change. As a result, even in absence of differential effects, the relative trajectory as seen from ground will modify.

This conclusion was first pointed out by Sabol et al. [28] heuristically, referring to the circular formation, and then by Schweighart and Sedwick [11], Fasano and D'Errico [24], Vadali et al. [25]. In other terms, since as stated above the radial/along-track motion is phased relative to the perigee, while the cross-track motion is phased relative to the nodal crossings, in general out-of-plane and in-plane coordinates will oscillate with different periods. D'Amico and Montenbruck [29] also noted such effect which is defined as the rotation of the relative eccentricity vector.

Among other things, the difference in the two frequencies also emphasizes the fact that the relative trajectory does not depend only on the difference of mean orbital parameters, but also on the absolute parameters of the chief satellite. As a result of this, controlling the difference of mean orbital parameters between deputy

and chief does not imply that relative trajectory remains unchanged, but only that single coordinates do not drift.

As for differential effects, for design purposes it is useful to express differential secular rates as linear functions of δa, δe, and δi as follows (adaptation from Refs. [19, 20]):

$$\begin{aligned} \delta\dot{\Omega} &= C_{\Omega,a}\frac{\delta a}{a} + C_{\Omega,e}\delta e + C_{\Omega,i}\delta i \\ \delta\dot{u} &= C_{u,a}\frac{\delta a}{a} + C_{u,e}\delta e + C_{u,i}\delta i \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} C_{\Omega,a} &= \varepsilon n\frac{7}{4}\cos i \\ C_{\Omega,e} &= -\varepsilon n\frac{2e}{1-e^2}\cos i \\ C_{\Omega,i} &= \frac{1}{2}\varepsilon n\sin i \\ C_{u,a} &= C_{u,a,kep} + C_{u,a,J2} = -\frac{3}{2}n - \frac{7}{8}\varepsilon n\left[\left(5+3\sqrt{1-e^2}\right)\cos^2 i - \sqrt{1-e^2} - 1\right] \\ C_{u,e} &= \varepsilon n\frac{e}{\sqrt{1-e^2}}\left[\cos^2 i\left(\frac{5}{\sqrt{1-e^2}}+\frac{9}{4}\right) - \frac{1}{\sqrt{1-e^2}} - \frac{3}{4}\right] \\ C_{u,i} &= \varepsilon n\sin(2i)\left(-\frac{5}{4} - \frac{3}{4}\sqrt{1-e^2}\right) \end{aligned} \tag{3.12}$$

and

$$\varepsilon = 3J_2\left[\frac{R_{eq}}{a(1-e^2)}\right]^2 \tag{3.13}$$

It is useful to analyze the orders of magnitude of the different terms for a typical near circular LEO orbit. Since $\varepsilon = O(10^{-3})$, then we have $C_{\Omega,a} = C_{\Omega,i} = O(10^{-3}\,n)$, $C_{\Omega,e} = O(10^{-6}\,n)$, $C_{u,a} = O(n)$, $C_{u,i} = O(10^{-3}\,n)$, $C_{u,e} = O(10^{-6}\,n)$.

Of course, $C_{u,a}$ is the largest coefficient because of the Keplerian term included. In terms of the along-track coordinate, a residual δa implies an along-track drift of the order of $-3\pi\delta a$ per orbit. This term is explicitly cited in (3.3). Eccentricity coefficients are the smallest ones and nullify if the chief is on circular orbit. In practical terms this means that really small differential effects are activated by differences in eccentricity.

Finally, node precession rate and mean angular velocity show a significant dependency on δi, and the chief orbit inclination determines the relative weight of $C_{\Omega,i}$, $C_{\Omega,a}$, $C_{u,i}$. In fact, in the case of near polar orbits, precession rate is very sensitive to the inclination while $C_{\Omega,a}$, $C_{u,i}$ tend to vanish.

Equations (3.6) and ((3.11), (3.12), (3.13)) can be used in a twofold way by the orbit designer. First, it is possible to quantitatively estimate formation stability in case of non J_2-invariant formations. On the other hand, the coefficients allow the trajectory designer to understand the most efficient ways of modifying relative trajectory by taking advantage of Earth oblateness effects.

Mean orbit parameters and secular J_2 effects have been considered in this discussion. Of course, actual formation dynamics is also driven by short-periodic and long-periodic effects [29]. These effects usually have to be taken into account when dealing with formation control, and analytical approaches exist which allow mapping osculating parameters to the corresponding mean elements [8, 26, 29]. On the other hand, it is also worth noting that relative motion as estimated on the basis of mean orbit parameters is different from averaged relative motion [30], due to the non linear relations involved. In practical terms, however, both these effects can be considered as higher order terms which can be likely neglected in the formation design phase.

It is useful to arrange the relative motion equations in order to put in evidence the design parameters for relative motion geometry. In detail, (3.6) can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cong \begin{bmatrix} x_{off} + A_x \sin\left(\dot{M}_D t + \varphi_x\right) \\ y_{off} + 2A_x \cos\left(\dot{M}_D t + \varphi_x\right) + t \cdot y_{dr} \\ A_z(t) \sin[\dot{u}_D t + \varphi_z(t)] \end{bmatrix} \tag{3.14}$$

Where

$$\begin{aligned}
x_{off} &= \delta a \\
A_x &= a\sqrt{\left[\delta e + 2e\sin^2\left(\frac{\delta M_0}{2}\right)\right]^2 + [e\sin(\delta M_0)]^2} \\
\varphi_x &= \tan^{-1}\left(-\frac{\delta e + 2e\sin^2\left(\frac{\delta M_0}{2}\right)}{e\sin(\delta M_0)}\right) \\
y_{off} &= a[\delta(\omega_0 + M_0) + \delta\Omega_0 \cos i] \\
y_{dr} &= a\left(\delta\dot{u} + \delta\dot{\Omega}\cos i\right) \\
&= a\left[C_{u,a}\frac{\delta a}{a} + C_{u,e}\delta e + C_{u,i}\delta i + \left(C_{\Omega,a}\frac{\delta a}{a} + C_{\Omega,e}\delta e + C_{\Omega,i}\delta i\right)\cos i\right] \\
A_z(t) &= a\sqrt{\delta i^2 + \left[\left(\delta\Omega_0 + \delta\dot{\Omega}\, t\right)\sin i\right]^2} \\
&= a\sqrt{\delta i^2 + \left\{\left[\delta\Omega_0 + \left(C_{\Omega,a}\frac{\delta a}{a} + C_{\Omega,e}\delta e + C_{\Omega,i}\delta i\right)t\right]\sin i\right\}^2} \\
\varphi_z(t) &= \tan^{-1}\left\{-\frac{\left[\delta\Omega_0 + \left(C_{\Omega,a}\frac{\delta a}{a} + C_{\Omega,e}\delta e + C_{\Omega,i}\delta i\right)t\right]\sin i}{\delta i}\right\}
\end{aligned} \tag{3.15}$$

Correlation with HCW equations also permits to derive (under HCW assumptions) the first order relations between initial conditions on relative position and velocity, and initial differences in orbit parameters. Specifically, under the Keplerian circular (chief) orbit assumption ($\dot{M}_D = \dot{u}_D = n_D = \sqrt{\frac{\mu_\oplus}{a_D{}^3}}$), from (3.2) and (3.6), the following relations hold:

$$
\begin{aligned}
x_{off} &= \delta a = 4x_0 + 2\frac{\dot{y}_0}{n} \\
A_x &= a\delta e = \sqrt{\left(\frac{x_0}{n}\right)^2 + \left(3x_0 + 2\frac{\dot{y}_0}{n}\right)^2} \\
\varphi_x &= -\frac{\pi}{2} \\
y_{off} &= a[\delta(\omega_0 + M_0) + \delta\Omega_0 \cos i] = -2\frac{\dot{x}_0}{n} + y_0 \\
y_{dr} &= a\delta\dot{u} = C_{u,a,kep}\delta a = -\frac{3}{2}n\delta a = -(6nx_0 + 3\dot{y}_0) \\
A_z &= a\sqrt{\delta i^2 + (\delta\Omega_0 \sin i)^2} = \sqrt{z_0{}^2 + \left(\frac{\dot{z}_0}{n}\right)^2} \\
\varphi_z &= \tan^{-1}\left\{-\frac{\delta\Omega_0 \sin i}{\delta i}\right\} = \tan^{-1}\left\{-\frac{z_0 n}{\dot{z}_0}\right\}
\end{aligned}
\tag{3.16}
$$

3.2.2 *"Large" Formations*

In the case of typical Earth observation missions flying in LEO, large formations foresee baselines of hundreds of kilometers thus needing second order modeling. The orbital parameters approach can still be used for relative motion analysis. Following [20], proper hypotheses have to be considered for the initial conditions.

As in the close formation case, the basic assumption is that δa, δΩ, δi, δe (i.e. e_D) and $\omega_D + M_{D0} - u_0$ are relatively small.

However, if second order terms have to be included in trajectory description, then the order of magnitude of considered approximations must be different for the different orbital parameters. In fact, Keplerian and J_2 gravity field harmonics (included in the model) should not disrupt the formation on a short timescale. Therefore, the semi-major axis is the most critical parameter since its variation modifies both Keplerian mean motion and J_2 effects (ascending node and perigee precession rates, and perturbed mean motion). Then, if nearly circular orbits are considered, it is worth noting that inclination is the parameter with major impact on J_2 effects, while eccentricity has a minor influence as demonstrated in (3.12). In addition, J_2 effects do not depend on argument of perigee, mean anomaly, and right ascension of the ascending node. The problem can be also analyzed from another point of view. Differential J_2 effects cause drifts on ascending node,

perigee, and mean anomaly, with absolute values of δΩ, δω, and δM increasing with time. If the model does not allow δΩ, δω, and δM to be large enough, it could be useless after a relatively short time. As a consequence, the non-dimensional difference in semi-major axis (δa/a) must be designed as the smallest parameter, while all the other parameter differences are larger, with δi in-between.

On the basis of these considerations the following assumptions can be made:

- Retain only linear and quadratic terms for $\delta\Omega$, $\omega_{d0} + M_{d0} - u_0$, $\delta\dot{u}t = \delta\left(\dot{M} + \dot{\omega}\right)t$
- Retain only linear terms for δi
- All mixed terms given by the product of two orbital parameter differences taken into account, apart from the ones containing δa/a.

Considering these assumptions, Fasano and D'Errico derived the following set of equations [20]:

$$\begin{aligned}\frac{x}{a} \cong\ & \frac{\delta a}{a} - \delta e \cos\left(M_{D0} + \dot{M}_D t\right) + 2e \sin\left(\frac{\delta M_0}{2}\right) \sin\left(M_0 + \frac{\delta M_0}{2} + \dot{M}_D t\right) \\ & + e\left(\delta\dot{M}_D t\right) \sin\left(M_0 + \dot{M}_D t\right) - \frac{(\delta u_0)^2}{2} - \frac{(\delta\dot{u}\, t)^2}{2} - \delta u_0 \cdot \delta\dot{u}\, t \\ & - e_D(\delta u_0 + \delta\dot{u}\, t) 2 \sin\left(M_{D0} + \dot{M}_D t\right) - \delta\Omega \cos i \cdot \\ & \left[\delta u_0 + \delta\dot{u}t + 2e_D \sin\left(M_{D0} + \dot{M}_D t\right)\right] + \left(-\frac{\delta\Omega^2}{2} + \frac{\delta\Omega^2}{4} \sin^2 i\right) \\ & - \frac{\delta\Omega^2}{4} \sin^2 i \cos\xi + \frac{1}{2}\delta\Omega\delta i \sin i \sin\xi + \left[2e \sin\left(M_0 + \dot{M}t\right)\right] \\ & (\delta u_0 + \delta\dot{u}\, t + \delta\Omega \cos i)\end{aligned} \tag{3.17}$$

$$\begin{aligned}\frac{y}{a} \cong\ & \delta u_0 + \delta\dot{u}\, t + \delta\Omega \cos i + 2\delta e \sin\left(M_{D0} + \dot{M}_D t\right) + 4e \sin\left(\frac{\delta M_0}{2}\right) \\ & \cos\left(M_0 + \frac{\delta M_0}{2} + \dot{M}_D t\right) + 2e\left(\delta\dot{M}\, t\right) \cos\left(M_0 + \dot{M}_D t\right) \\ & - e_d(\delta u_0 + \delta\dot{u}\, t) \cos\left(M_{d0} + \dot{M}_D t\right) - \delta\Omega \cos i\, e_d \cos\left(M_{d0} + \dot{M}_D t\right) \\ & - \frac{1}{2}\delta\Omega\delta\, i \sin i + \frac{\delta\Omega^2}{4} \sin^2 i \cdot \sin\xi + \frac{1}{2}\delta\Omega\delta i \sin i \cos\xi\end{aligned} \tag{3.18}$$

$$\begin{aligned}\frac{z}{a} \cong\ & -\delta\Omega \sin i \cos\Psi + \delta i \sin\Psi - \delta\Omega e_D \sin i \\ & \left[\frac{1}{2}\cos\left(2M_{D0} + \omega_{D0} + \dot{M}_D t + \dot{u}_D t\right) - \frac{3}{2}\cos(\omega_{D0} + \dot{\omega}_D t)\right] \\ & + \delta\, i\, e_D \left[\frac{1}{2}\sin\left(2M_{D0} + \omega_{D0} + \dot{M}_D t + \dot{u}_D t\right) - \frac{3}{2}\sin(\omega_{D0} + \dot{\omega}_D t)\right] \\ & + \frac{\delta\Omega^2}{2} \sin i \cos i \sin\Psi\end{aligned} \tag{3.19}$$

where δΩ is to be considered as the actual value, i.e. $\delta\Omega = \delta\Omega_0 + \delta\dot{\Omega}\,t$, and

$$\xi = \omega_{D0} + M_{D0} + u_0 + 2\dot{u}_D t \tag{3.20}$$

$$\Psi = \omega_{D0} + M_{D0} + \dot{u}_D t \tag{3.21}$$

On the basis of these equations, the following considerations can be formulated:

- The second order terms basically have a coupling effect on the coordinates. In the first order model (3.6), δa, δe, δω and δM appear explicitly in x and y (in-plane coordinates), whereas δΩ and δi only appear in the cross-track coordinate, apart from the fact that δΩ contributes to the along-track offset and linear drift. This is consistent with Hill's differential equations, where in plane motion is independent from the cross-track coordinate. When the satellites' separation grows, as shown in the second order model all coordinates are coupled. On the other hand, δa only appears as an offset in the radial coordinate because of previously reported assumptions;
- Secular terms also appear in the radial coordinate at second order;
- All the coordinates are the sum of different functions of time: constant offsets, linear and quadratic terms, long period oscillations due to perigee precession and other periodic terms with constant or time-varying amplitude. As for the latter ones, it is interesting to note that some terms appear in x and y whose frequency is twice the orbital one (3.17 and 3.18);
- In typical cases, not all the terms in (3.17), (3.18), and (3.19) are of the same order of magnitude, and some can be neglected. For example, if the relative trajectory is "slowly" modified by differential J_2 effects (deviation in position over tens of orbits non significant in percentage), all the second order terms containing secular factors can be likely neglected.

3.3 Relative Trajectory Modeling: Elliptic Orbits with Arbitrary Eccentricity

Alfriend et al. [26] give an overview and comparisons of some proposed models of relative motion with respect to elliptic reference orbits, which are basically classified on the grounds of the selected independent variable. In particular, Melton's model [31] is developed in both a Cartesian and a cylindrical coordinate system attached to the chief. Relative position is expressed as a function of time thanks to an expansion in powers of eccentricity. Another family of relative motion models utilizes true anomaly as independent variable and it is based on the former independent developments by Lawden [32], de Vries [33], Tshauner and Hempel [7]. These latter models have been then additionally developed by Carter [13], who found new formulations of Lawden's solutions and succeeded in removing

singularities in true anomaly. In addition, Yamanaka and Ankersen [34] found an additional, independent solution of Lawden's equations.

Inalhan et al. [35] reconstruct relative dynamics between two eccentric orbits and identify a solution which, for homogeneous equation (i.e. in Keplerian dynamics), coincides with the solutions identified by Lawden/Carter. They also analyze a set of initial conditions able to guarantee periodicity of relative trajectories. The first proposed condition coincides with the well-known equal-energy (i.e. semi-major axis) requirement, which is also a clear outcome of HCW equations. In addition, Inalhan et al. also propose different conditions to avoid relative drift which depends on a set of four constants. Finally, a simulation is presented, which shows that neglecting chief eccentricity (of the order of 10^{-3}) has an effect on in-plane drift larger than neglecting J_2 differential perturbation. The simulations are performed at an inclination of 52°, which could partly explain such result.

Ketema [36] also proposes an analytical model for Keplerian relative motion between two elliptic orbits. Method peculiarity consists in the definition of a set of relative orbital parameters of the deputy in the chief perifocal reference frame, whereas the time dependence of true anomaly and time is solved numerically. Modeling J_2 effects on the new set of orbital parameters is needed if the method needs to be extended.

Lane and Axelrad [37] present a geometrical method to describe relative motion with time as independent variable, building on previous derivations by Broucke [38]. The proposed method is based on the description of the relative motions as a function of chief's true anomaly and o the orbital parameter differences, which are all assumed small. Applications of the method are presented to design typical formations (in-track, follower, in-track/cross-track), with an approach similar to the one presented in Sect. 3.2. Relative motion is modeled for Keplerian dynamics, which makes the model accurate over short time intervals.

Relative position of the deputy with respect to the chief is expressed by Schaub [19] as a function of orbital parameter differences, with eccentricity and perigee generalized in order to avoid singularities, and chief's true anomaly. Schaub [19] provides relative motion models for either arbitrary or small eccentricity (discussed in Sect. 3.2) chief's orbits. Therefore, the sixth relative parameter is selected as the difference between chief and deputy mean anomalies, which is constant for orbits with the same energy. J_2 effects can be incorporated adequately modeling time dependence or orbital parameter differences.

In their analysis, Sengupta and Vadali [39] working with curvilinear coordinates in the LHLV reference frame of the chief and expanding a previous work by Sengupta et al. [40], show relations between Lawden's solutions and the ones obtained with the orbital parameter difference approach. They extensively analyze the effect of chief eccentricity putting in evidence that it introduces a number of variations with respect to the classical solution of the HCW equations: (a) oscillations at frequency higher than the orbital one; (b) enlargement of relative trajectory in cross-track and shrinkage in along-track; (c) phase shift; (d) displacement of relative trajectory center with respect to the LHLV reference frame origin.

In conclusion, a chief's orbit with an arbitrary eccentricity adds complications to the description of relative motion even in Keplerian dynamics. Therefore, simplification of the relative motion model to allow a straightforward design process is still under debate, with Schaub [19] probably offering the best option. Nonetheless, since the vast majority of current ideas for distributed missions for Earth observations utilize low eccentricity orbits, this limitation has limited practical impact in the present.

3.4 Relative Trajectory Design

3.4.1 Synthetic Aperture Radar Interferometry

As described in detail in Chaps. 1 and 2, single pass cross-track interferometry [41, 42] builds on two antennas, one transmitting and both receiving the backscattered echoes (transmission can be switched from one antenna to the other at each impulse), and allows one to build a three-dimensional model of observed areas.

The basic formation parameter which determines observation geometry and achievable performance in Digital Elevation Model (DEM) generation is the so-called "effective baseline", which is the projection of the baseline vector onto the normal of the radar viewing direction (Fig. 3.4).

Given a radar look angle θ the effective baseline is a linear combination of the vertical (along the x-axis, radial) and the horizontal (along the z-axis, cross-track) baseline ($|x \sin \vartheta - z \cos \vartheta|$).

The effective baseline is usually chosen on the basis of a trade-off between height reconstruction accuracy and height of ambiguity [42], thus depending on radar wavelength. The height of ambiguity is defined as the altitude difference that generates an interferometric phase change of 2π after interferogram flattening. In fact, although larger baselines allow better height measurement sensitivity, spatial decorrelation determines larger phase measurement noise, and moreover height of ambiguity decreases, which is a drawback especially in areas with steep slopes.

For given system parameters and effective baseline, both height accuracy and height of ambiguity are roughly proportional to radar wavelength. This is clearly shown in Fig. 3.5 where X-band, C-band and L-band SAR are considered (the chief satellite is considered as transmitting/receiving, while the deputy is receiving only), height accuracy models in literature are used [43, 44], and realistic values are assumed for the other parameters (altitude 600 km, signal-to-noise ratio 15 dB, 4 looks, radar range resolution 5 m).

It is worth underlining that only phase noise has been considered in evaluating height accuracy. Uncertainties in slant range, baseline components, and attitude also affect height accuracy but are not considered here because they depend on the

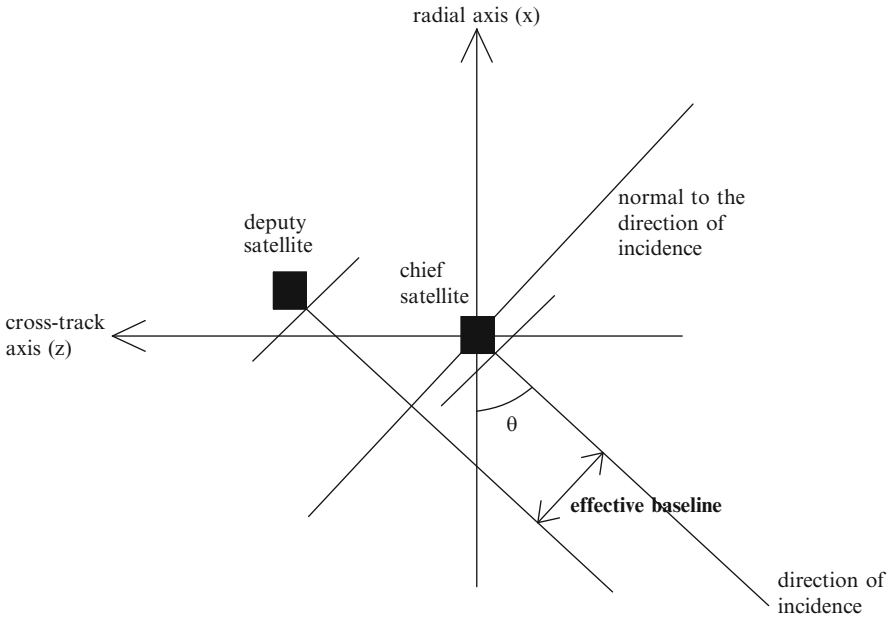


Fig. 3.4 Geometry to evaluate effective baseline (Reproduced with kind permission from Springer Science+Business Media B.V. © 2009 Springer Published in Ref. [20])

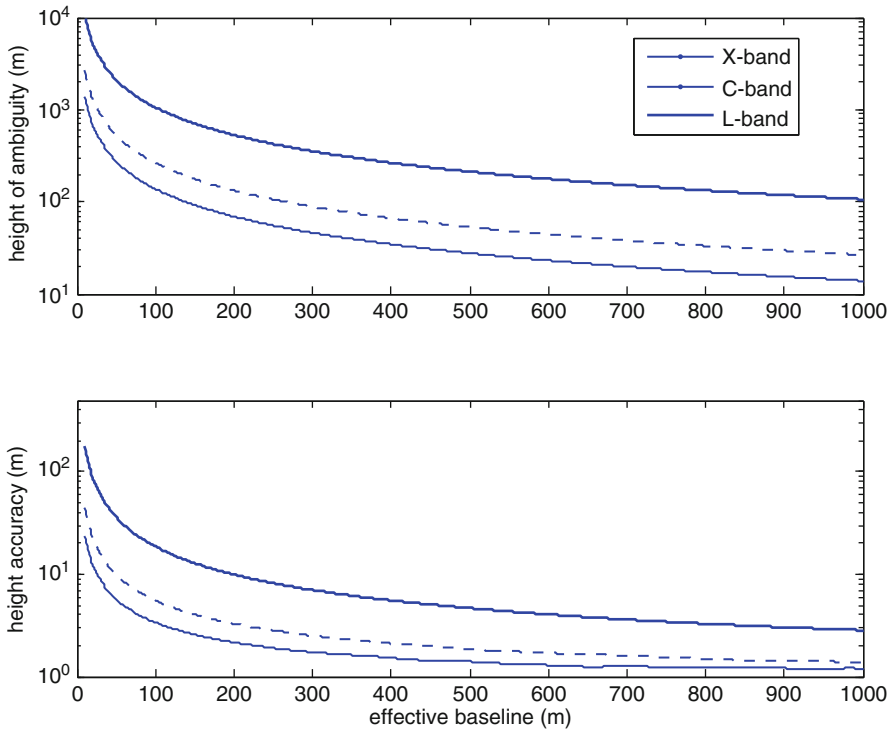


Fig. 3.5 Basic performance parameters in SAR interferometry as a function of effective baseline

navigation system. Nevertheless it must be noted that phase noise effect is the main source of error, provided that baseline is very accurately estimated (millimetric level).

On the basis of this type of budgets, an "optimal" value (or, more practically, an acceptable range) of the effective baseline can be usually identified, and orbit design is aimed at achieving this baseline range at the required latitudes. In fact, orbital motion produces a continuous change of effective baseline and a different range of latitudes where the given performance is achieved, depending on formation geometry. Thus, different candidate formation geometries must designed and traded against effective baseline performance and orbit stability.

Let us refer to the dynamics of close formations on small eccentricity orbits (3.6). First of all, it is interesting to note that two sinusoidal oscillations (in general at slightly different frequency) with zero offset have to be combined. In fact, it is not realistic to impose an offset in the radial coordinate which would result in formation instability over short timescales. Thus, for any choice of the orbital parameters, the effective baseline has a sinusoidal trend as a function of time. The degrees of freedom for the designer are the amplitude and the phase of the oscillation with respect to latitude. If cross-track component is designed by means of δi, an undesired cross-track instability (3.12) is a drawback. Thus, horizontal motion has to be obtained by means of $\delta\Omega$, which leads to an effective baseline variation from the equator (maximum) to the poles (zero), depending on the radar off-nadir angle. On the other hand, the same trend is obtained in both the ascending and descending phases of the orbit. Because of the oscillation of the deputy around an axis parallel to the cross-track axis, this formation is usually known as "pendulum". Usually, an offset in the along-track coordinate is also foreseen for safety reasons, which can be obtained by a δM (or a $\delta\omega$). Both orbits can have some (common) eccentricity without effects on relative motion.

If $[B_{min}\ B_{max}]$ is the required range of effective baseline, then from (3.6) the $\delta\Omega$ to maximize the useful orbit fraction can be calculated as

$$a\delta\Omega \sin i \cos \vartheta = B_{\max} \tag{3.22}$$

In terms of latitude coverage, this type of formation does not allow achieving the desired baselines at high latitudes. An example of this effect is shown in Fig. 3.6.

An alternative option is the exploitation of the vertical oscillation only, which can be achieved by utilizing either a δe or opposite differences in δM and $\delta\omega$ with non zero chief eccentricity. Again, the orbit fraction where the required range of effective baselines is achieved can be maximized. For example, if a difference in eccentricity is selected, we can impose

$$a\delta e \sin \vartheta = B_{\max} \tag{3.23}$$

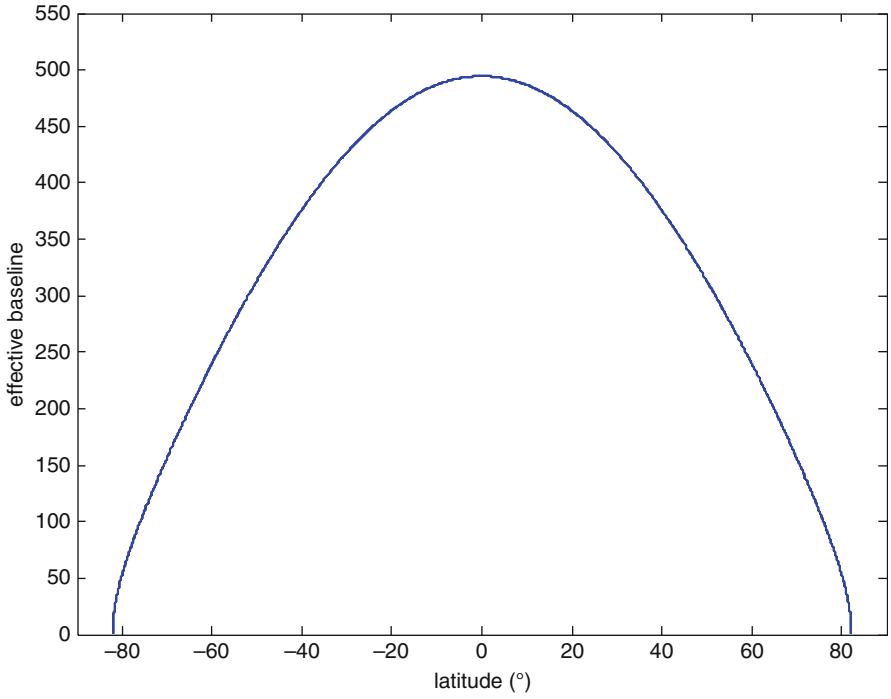


Fig. 3.6 Effective baseline as a function of latitude for a pendulum formation (COSMO/SkyMed considered as chief orbit, $\delta\Omega = 4.9 \times 10^{-3}$)

While, if we impose differences in mean anomaly and argument of perigee (for small eccentricity orbits), we have

$$2ae \sin\left(\frac{\delta M_0}{2}\right) \sin\vartheta = B_{\max} \tag{3.24}$$

In both cases, the motion is phased relative to the perigee. Thus the latitude trend depends on the argument of perigee, with perigee precession modifying the latitude coverage. This type of formation is usually known as "cartwheel" [45, 46], and the relative trajectory is on the 2×1 in plane ellipse of Hill's equations (Fig. 3.7) while the effective baseline trend as a function of latitude is depicted in Fig. 3.8 for the two different values for the argument of perigee.

If a δe-based cartwheel is considered, it is interesting to observe that when the argument of perigee is 90°, the same effective baseline trend is achieved in the ascending and in the descending phase of the orbit, which is complementary with respect to the pendulum effect: the effective baseline is now maximum at the poles and nullifies at the equator. Instead, when the perigee lies at intermediate latitudes (e.g., 45°), then different trends are achieved in the different phases of the orbit, and

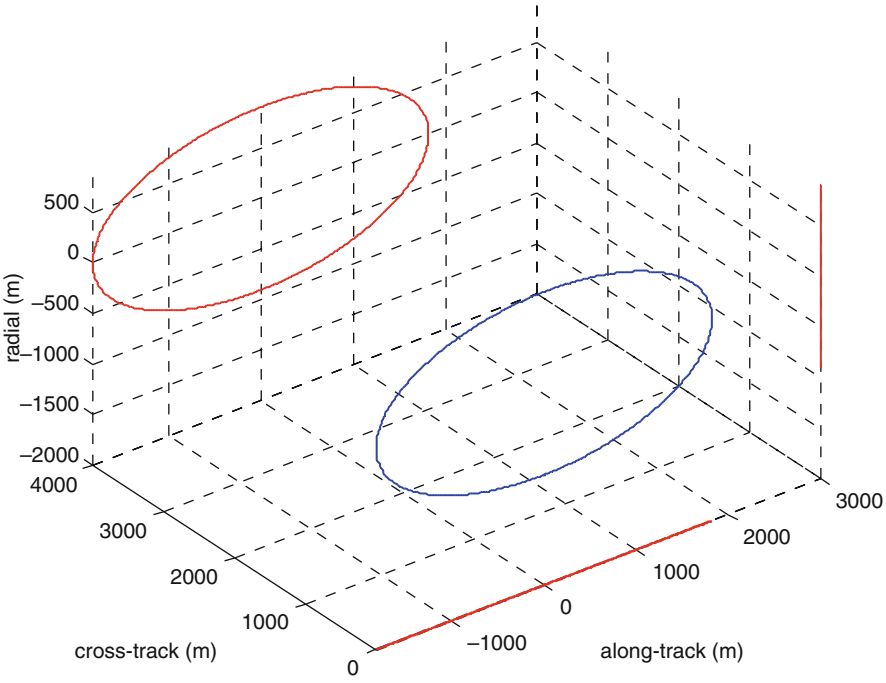


Fig. 3.7 3d relative trajectory and projections on the coordinate planes for a cartwheel formation

all the achievable latitudes can be covered with a relatively small effective baseline percentage variation.

Finally, vertical and horizontal oscillations can be combined. Since horizontal motion has to be necessarily obtained by means of a difference in right ascension of the ascending node, the argument of perigee and the choice of eccentricity, δe, δM and $\delta\omega$, determines the vertical motion and the relative phasing of the two oscillations, which in turn influence the latitude trend of the effective baseline and the trajectory shape in the radial/cross-track plane. A common choice in remote sensing orbits is to place the perigee at 90° to achieve sun-synchronous frozen orbits [22]. If this is the case, then imposing a slight difference in eccentricity and no differences in perigee and mean anomaly allows obtaining a vertical oscillation with a phase difference of 90° with respect to the horizontal motion. This implies that the resulting trajectory in the radial-cross-track plane is an ellipse whose principal directions coincide with the coordinate axes and whose semi-minor and semi-major axes depend linearly on $\delta\Omega$ and δe, respectively. This formation is usually called "helix" and has found application in the TanDEM-X mission (Chap. 13). Helix formations are characterized by the fact that the separation in the radial/cross-track plane never nullifies. This means that the baseline between the satellites is never purely an along-track separation, which is an advantage for collision avoidance and relaxes formation control requirements.

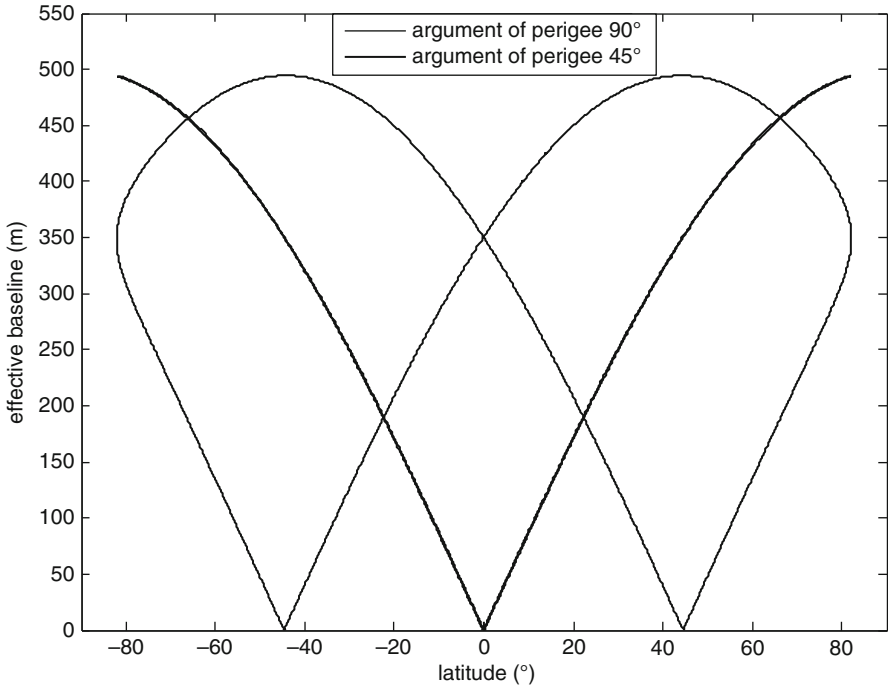


Fig. 3.8 Effective baseline as a function of latitude for two cartwheel formations (based on δe)

Then, as in the cartwheel case for intermediate values of the argument of perigee, all the latitudes can be observed with similar effective baseline, provided that the difference in right ascension of the ascending node and eccentricity are properly selected.

Effective baseline can be expressed as a function of time as:

$$\begin{aligned} B_{eff} &= a|x\sin\vartheta - z\cos\vartheta| \\ &\cong a\left|-\delta e\cos\left(M_0+\dot{M}_D t\right)\sin\vartheta + (\delta\Omega_0+\delta\dot{\Omega}t)\sin i\cos(90^\circ+M_0+\dot{u}_D t)\cos\vartheta\right| \end{aligned} \tag{3.25}$$

Considering that the difference between $\dot{M}_D$ and $\dot{u}_D$ is $\dot{\omega}_D$ (three orders of magnitude less than $\dot{M}_D$), B_{eff} is the sum of two cosines with different amplitude, 90° phase shift, and very near frequency. Thus, B_{eff} is nearly the absolute value of a sinusoid whose amplitude is given by $a\sqrt{\delta e^2\sin^2\vartheta + (\delta\Omega\sin i\cos\vartheta)^2}$ while the phase is related to the latitude where maximum effective baseline is achieved. If the amplitude is imposed as large as the maximum baseline limit, the orbit fraction where the required effective baseline range is achieved is maximized. As a second condition, if the ratio of semi-major to semi-minor axis equals the

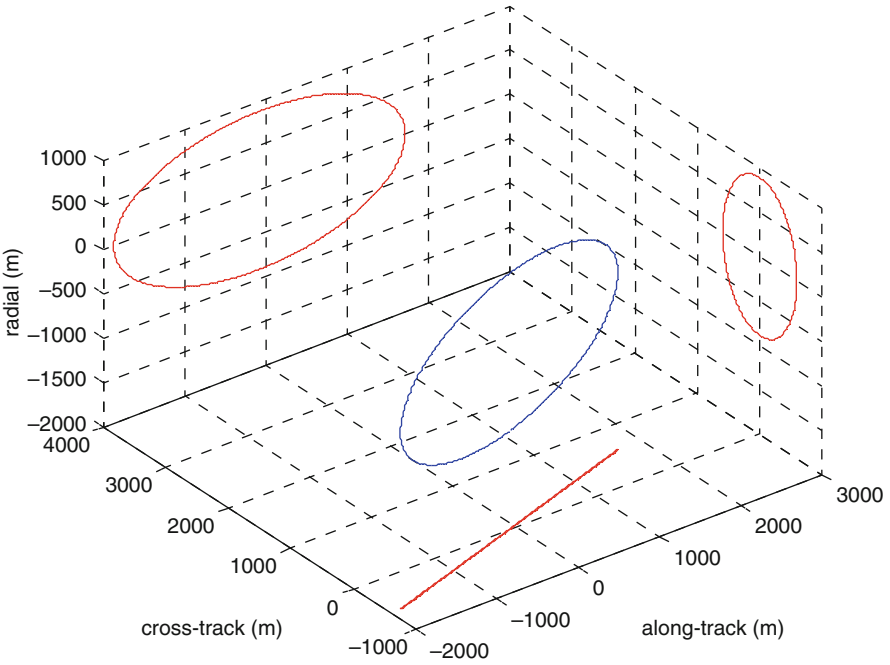


Fig. 3.9 3d relative trajectory and projections on the coordinate planes for a helix formation

tangent of the off-nadir angle, then maximum effective baseline is achieved at intermediate latitudes and all the latitudes can be observed with a maximum effective baseline variation of about 30% (Figs. 3.9 and 3.10).

An important difference among the described formations is the effect of perigee precession. In the pendulum case, perigee precession has no effect either on latitude coverage or on formation geometry, whereas for cartwheel formations, perigee precession changes the latitude belts where given performance is achieved, without changing the relative trajectory. Finally, in the helix case, perigee precession changes both the effective baseline values and the latitude dependence. In fact, if perigee moves away from 90°, the minimum separation in the radial-cross track plane decreases and then nullifies for $\omega = 0°$ or $\omega = 180°$. That is, the helix advantage of non intersecting orbits is then lost. Figure 3.11 describes the effect of perigee precession on a helix formation, pointing out the modifications in relative trajectory and effective baseline trend.

Table 3.1 summarizes the different characteristics of the described formations. Previous considerations about interferometry requirements and formation geometries can be used together in an example of trajectory design. Let us assume as a requirement that height accuracy has to be of the order of 1 m and height of ambiguity has to be larger than 20 m. Assuming for the sake of concreteness Cosmo/SkyMed as chief orbit, from budgets similar to the ones shown in

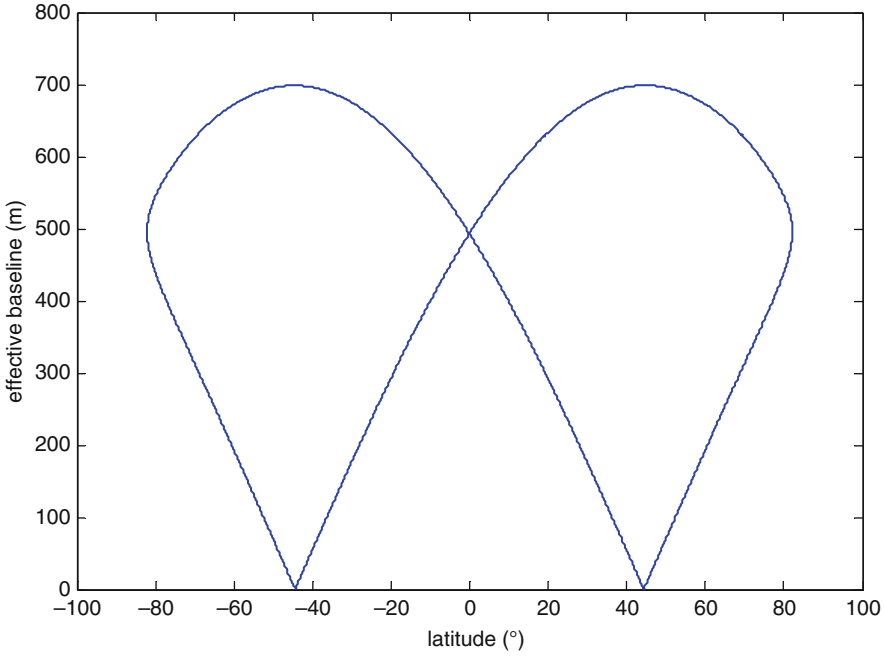


Fig. 3.10 Effective baseline as a function of latitude for a helix formation

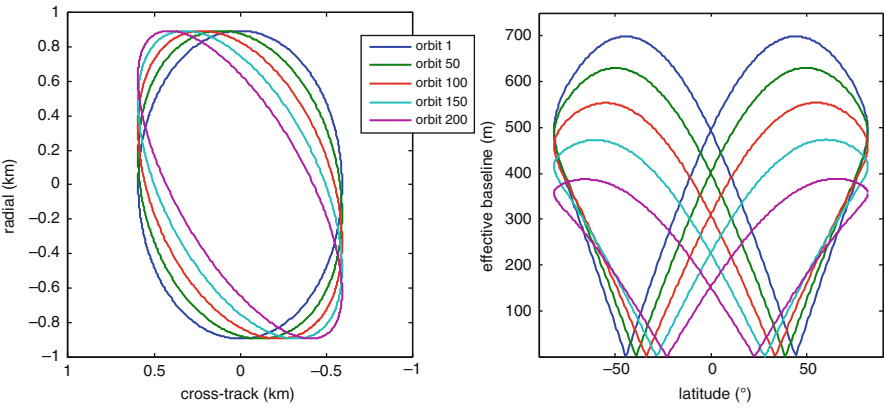


Fig. 3.11 Effect of perigee precession on a helix formation (Reproduced with kind permission from Springer Science+Business Media B.V. © 2009 Springer Published in Ref. [20])

Fig. 3.5, we get for the different bandwidths the effective baseline ranges shown in Table 3.2. Then, from (3.23), (3.24), and (3.25) we can calculate the needed differences in orbital parameters for the three bands. The results are shown in Table 3.2.

Table 3.1 Characteristics of different formation geometries

Formation geometry	Differences in orbital parameters	J_2 invariant	Latitude coverage	Collision risk	Effect of perigee precession
Pendulum	$\delta\Omega$ (δM for collision avoidance)	Yes	Effective baseline nullifies at high latitudes	Based on along-track constant baseline	None
Cartwheel	δe (and/or $\delta M = -\delta\omega$ and $e \neq 0$)	Yes if $\delta e = 0$	Depends on argument of perigee	Based solely on along-track baseline when vertical distance nullifies	Change of latitude coverage
Helix	$\delta\Omega$ and δe	No	If $\omega = 90°$, all the latitudes can be covered with maximum effective baseline percentage variation of 30%	The orbits do not intersect, separation is never solely in along-track	Change of relative trajectory, of effective baseline maximum value and its dependency on latitude

Table 3.2 Example of optimization of formation geometries for SAR interferometry

Band	Effective baseline range (m)	Height accuracy range (m)	Height of ambiguity range (m)	Pendulum	Cartwheel	Helix
X	400–700	0.97–1.24	21.5–36	$\delta\Omega = 6.9 \times 10^{-3}$ ($\delta M = 8 \times 10^{-3}$)	$\delta e = 1.8 \times 10^{-4}$	$\delta\Omega = 4.9 \times 10^{-3}$ $\delta e = 1.3 \times 10^{-4}$
C	700–1,300	0.99–1.3	21.5–39.5	$\delta\Omega = 1.3 \times 10^{-2}$ ($\delta M = 8 \times 10^{-3}$)	$\delta e = 3.4 \times 10^{-4}$	$\delta\Omega = 9.1 \times 10^{-3}$ $\delta e = 2.4 \times 10^{-4}$
L	2,800–5,000	1–1.3	22–39.5	$\delta\Omega = 5.0 \times 10^{-2}$ ($\delta M = 8 \times 10^{-3}$)	$\delta e = 1.3 \times 10^{-3}$	$\delta\Omega = 3.5 \times 10^{-2}$ $\delta e = 9.1 \times 10^{-4}$

3.4.2 Synthetic Aperture Radar Tomography

Synthetic aperture radar tomography is a technique to achieve focused 3-D SAR images [47] that relies on an aperture synthesis along the direction normal to the slant range within the slant plane, i.e. the elevation direction. That aperture is synthesized by processing properly several SAR images of the same area gathered with (slightly) different off-nadir/incidence angles (Chap. 2). Since aperture synthesis procedures strongly rely on uniform data sampling [47, 48] appropriate resampling techniques [49] have to be applied to estimate a set of samples on a uniform spacing grid and their performance degrades if heavily non-uniform samples are utilized. Bistatic/multistatic SAR can be properly exploited to acquire

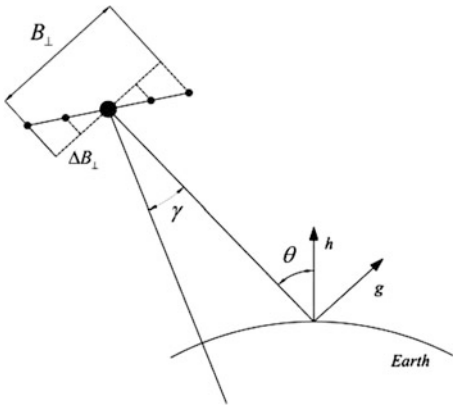


Fig. 3.12 Tomographic geometry for a satellite formation of bistatic SAR sensors (*small dots*) teamed with a monostatic SAR (*large dot*)

suitable tomographic data with accurate baselines and without being affected by temporal decorrelation, which severely degrades performance in repeat-pass scenarios. Spaceborne SAR tomography is currently an open research theme, and a quantitative analysis of the relevant trade-offs is under investigation by the scientific community.

As in cross-track interferometry, the basic geometry parameter for SAR tomography is the effective baseline. In fact, the receivers have to properly sample the effective baseline along an interval that defines the tomographic aperture (Fig. 3.12).

Using mathematical models developed in Refs. [48, 50] and on the basis of the scenario depicted in Fig. 3.12, defining λ as the wavelength, R_0 as the slant range, and θ as the local incidence angle (Eqs. 3.26–3.29 only), one can derive that the maximum effective baseline is determined by the desired height resolution, Δh:

$$B_\perp = \frac{\lambda R_0 \sin\theta}{\Delta h} \tag{3.26}$$

while the distance among the satellites is limited by the necessity to avoid ambiguities in the tomographic reconstruction. Hence, if H is the desired unambiguous height, it must be:

$$\Delta B_\perp = \frac{\lambda R_0 \sin\theta}{H} \tag{3.27}$$

H and Δh also determine the number of required satellites in a single-pass scenario:

$$N = 1 + \frac{B_\perp}{\Delta B_\perp} = 1 + \frac{H}{\Delta h} \tag{3.28}$$

Finally it is important to remark that the coherent sum of the received signals is exploited by SAR tomography, and therefore the maximum effective baseline is also limited by the necessity to keep phase coherence among the signals [46, 51, 52]

$$B_{\perp} \leq \frac{\lambda R_0}{2 \cos\theta \Delta r} \tag{3.29}$$

where Δr is the ground range resolution.

When translating these requirements into the formation design framework, it comes out that the basic requirement is the uniform sampling of the effective baseline in a given interval, provided that this interval satisfies (3.29).

Following the same line of reasoning of the previous section, different formation types can be exploited, such as multi-cartwheel, multi-pendulum, and multi-helix formations (and their combinations). Since at first order the effective baseline is a linear function of $\delta\Omega$ (pendulum), δe (cartwheel) or $\delta\Omega$ and δe if $\delta\Omega/\delta e$ is constant (helix), in all cases it can be uniformly sampled at all latitudes along the orbit by properly scaling the differences in orbit parameters. Of course, as in the case of cross-track interferometry, the difference among the different formation types lies in the latitude coverage and the collision avoidance aspects.

As an example, let us assume X-band as operating frequency, radar range resolution of 5 m and off nadir angle of 30°, operating altitude of 600 km, chief orbit inclination of 97.79° (sun-synchronous) and chief orbit eccentricity of 10^{-3}. Then, from the tomography point of view, let us assume a vertical resolution of 5 m and a vertical dimension of 30 m. From (3.26) to (3.27), it comes out that an effective baseline of 2,400 m needs to be sampled in six intervals of 400 m, leading to the need for seven receiving apertures in a single pass scenario.

In the multi-pendulum case, assuming that the optimal observation geometry is required at the equator, from (3.6) it results

$$\delta\Omega = K \frac{\Delta B_{\perp}}{a \cos\theta \sin i} \qquad \text{with K} = [-3, -2, -1, \ 1, \ 2, \ 3] \tag{3.30}$$

and in the case under analysis the second member equals $K \times 0.0038°$. Since the baseline nullifies at the poles, tomographic reconstruction accuracy is lost at high latitudes. In addition, proper along-track offsets have to be foreseen to minimize collision risk. The effective baseline trend and the behavior of the out-of-plane coordinate is shown in Fig. 3.13, assuming one of the receivers as the chief satellite (in the center of the formation) for 3 deputies only (due to the symmetry).

Let us now consider a multi-cartwheel formation based on slight differences in eccentricity. Again, on the basis of (3.6), the required differences in eccentricity can be calculated as

$$\delta e = K \frac{\Delta B_{\perp}}{a \sin\theta} \qquad \text{with K} = [-3, -2, -1, \ 1, \ 2, \ 3] \tag{3.31}$$

and it results $\delta e = \text{K} \times 1.15 \times 10^{-4}$.

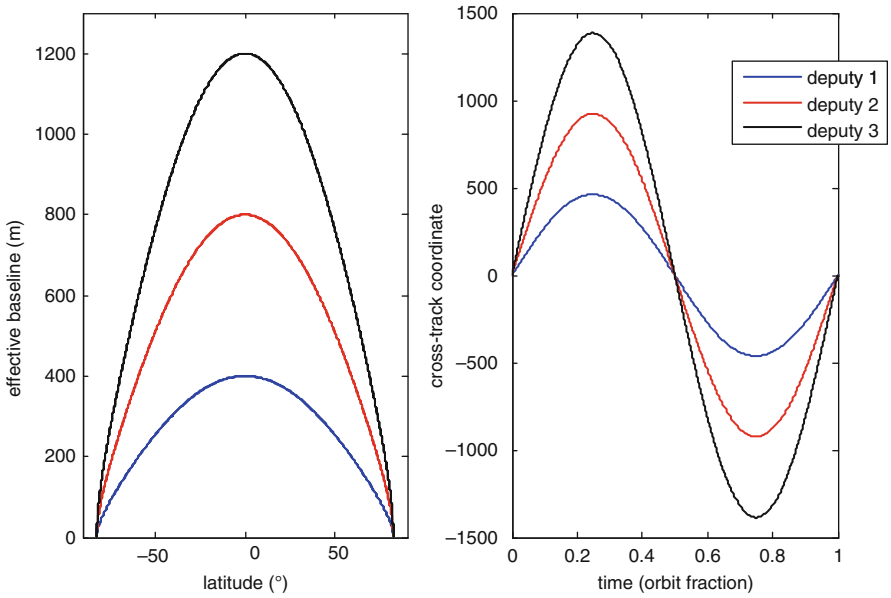


Fig. 3.13 Effective baseline as a function of latitude, and cross-track coordinate along the orbit, for a multi-pendulum formation (3 deputies)

The effective baseline trend as a function of latitude depends on the argument of perigee. If the perigees are at 90°, the trend shown in Fig. 3.14 is obtained.

In the case of a multi-helix formation, it is possible to uniformly sample the effective baseline achieving a tomographic accuracy of the same order at all the observable latitudes along the orbit. In fact, assuming as a first condition that the ratio between the vertical and horizontal oscillations equals the tangent of the radar off-nadir angle θ, $\delta\Omega$ and δe can be calculated from the further condition

$$\sqrt{\delta e^2 \sin^2\theta + \delta\Omega^2 \sin^2 i \cos^2\theta} = K\frac{\Delta B_\perp}{a} \quad \text{with K} = [-3, -2, -1, 1, 2, 3] \tag{3.32}$$

and in the considered case it results $K \times 5.7 \times 10^{-5}$.

The obtained trend of effective baseline resembles the one obtained with a multi-cartwheel with the perigees at 45°, while collision risk is minimized by the 90° phase difference between radial and cross-track oscillations (Figs. 3.15 and 3.16).

It is very interesting to consider the effect of perigee precession on geometry of tomographic observations. Again, the multi-pendulum geometry and latitude coverage are not affected by perigee precession. In the multi-cartwheel case, the formation geometry is not affected, rather the formation changes its latitude coverage although keeping the capability to uniformly sample the same effective baseline interval, provided that differential effects are counteracted by formation control. Instead, in the multi-helix case, not only does the latitude coverage change, but the maximum

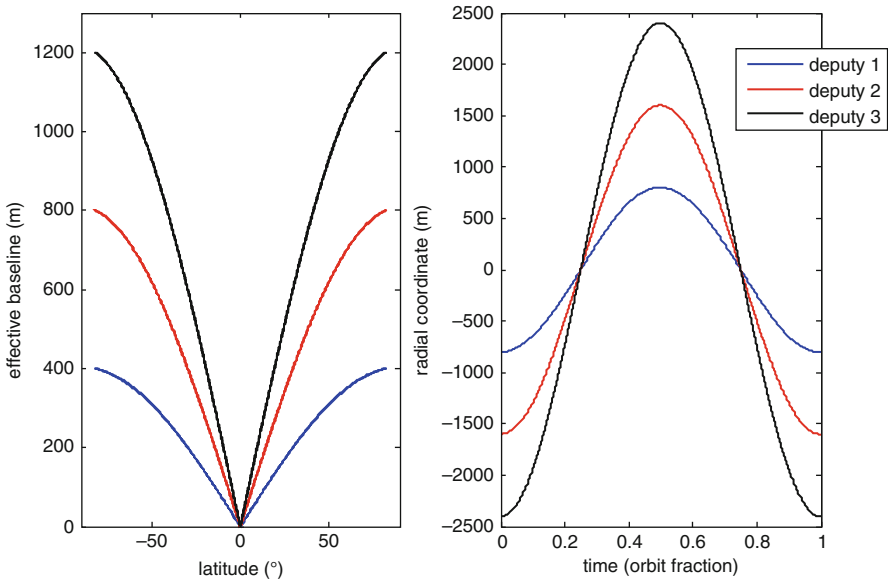


Fig. 3.14 Effective baseline and radial coordinate for a multi-cartwheel formation (3 deputies)

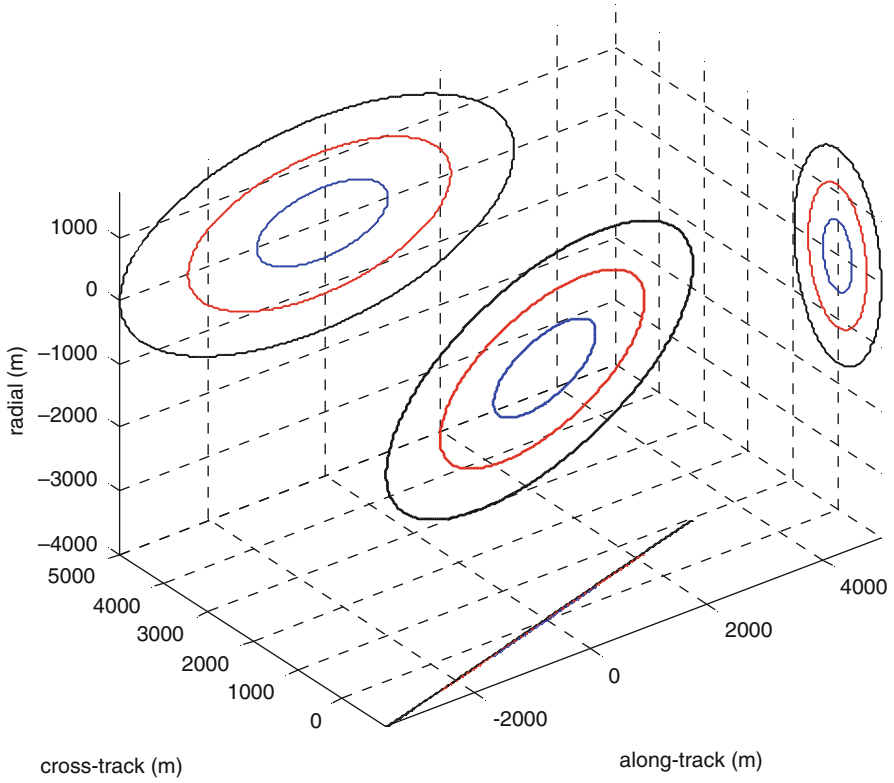


Fig. 3.15 3d geometry for a multi-helix formation (3 deputies)

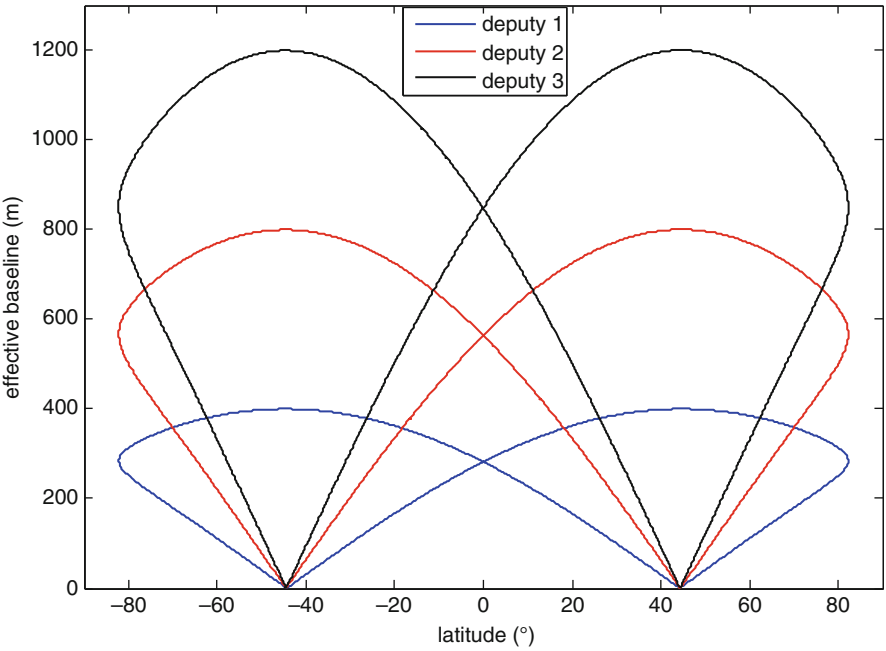


Fig. 3.16 Effective baseline as a function of latitude for a multi-helix formation (3 deputies)

achievable effective baseline is also modified, though at each latitude sampling uniformity is kept. These considerations are important if a hybrid repeat/single pass scenario is considered in order to reduce the number of satellite platforms, accepting the drawback of temporal decorrelation. In this case, ground track repeat cycle can be synchronized with perigee precession in order to optimize the different passages of the formation, but eventual changes in formation geometry and baseline sampling have to be properly addressed. Furthermore, performance of absolute orbit control becomes a key factor for ensuring sampling uniformity in the different formation passages.

3.4.3 Large Baseline Synthetic Aperture Radar

Large baseline bistatic SAR data are gathered under relatively large angle difference to achieve uncorrelated information on radar cross section. This means that relatively large differences in orbital parameters are required. However, if on the one hand two satellites must be widely separated, on the other hand the passive receiver must be capable of collecting the scattered radar signal. Thus, it must be located somewhere near the elevation plane of the transmitting radar. In addition, in most cases a large satellite distance implies a large relative orbital motion with

respect to small relative displacement typical of interferometric configurations. Therefore, the need of adequate pointing to maintain the overlap between the two radar swaths along the orbit comes up as an additional peculiarity.

Another point to be considered is that large differences in orbital parameters can induce different effects of orbit perturbations, thus requiring an unacceptable orbit control effort. Following this line of reasoning, a large separation in along-track direction, though easy to achieve imposing a difference in mean anomaly, is not advantageous from the points of view of both SNR and attitude/pointing angles.

In addition, the required bistatic angle can be hardly obtained by a large vertical separation because significant differential drag effect and along-track relative motion are generated.

Thus, cross-track (horizontal) separation remains as the only possible solution, and the basic parameter to be exploited to achieve the required bistatic angle is a difference in right ascension of the ascending node. The main drawback of this choice is that the baseline will change along the orbit and nullify at orbit poles, thus loosing the bistatic geometry at high latitudes. On the other hand, since Ω does not affect J_2 secular effects, large baselines can be obtained with no impact on formation stability.

In practical terms, large baseline bistatic SAR can be achieved by a pendulum formation where $\delta\Omega$ depends on the maximum required bistatic angle, while as shown by D'Errico and Moccia [53] the δM can be set imposing that at the equator the receiver satellite lies in the elevation plane of the transmitter, which minimizes attitude/pointing requirements. In other terms, δM is used to compensate for the along-track offset $\delta\Omega\cos i$ (see 3.18) and the eventual yaw steering maneuver performed by the transmitter.

This formation is J_2 invariant and, if eccentricities and inclinations are properly chosen and the perigees are placed at 90°, the sun-synchronous frozen orbit condition can be satisfied both for the transmitter and the receiver. However, the two orbits intersect at the poles and collision avoidance is based on the control of the along-track offset.

As in helix formations, an additional (small) difference in eccentricity can be used to separate the orbits and introduce a residual vertical baseline at orbit poles. J_2 invariancy is then lost, but the differential effects (see 3.12) are very small and can be counteracted without a significant propellant expense.

For example, considering a transmitter satellite on a sun-synchronous circular orbit at an altitude of 600 km, and imposing for the transmitter a constant off-nadir angle of 42° and for the receiver an off-nadir angle variable in the interval [5°, 23°], using models in Ref. [53], one obtains for $\delta\Omega$ and δM the trends shown in Fig. 3.17. Moreover, a difference in eccentricity of the order of 1×10^{-4} (with the perigee at 90°) can be chosen to ensure a safety separation of about 1 km at the poles.

Second order equations (3.17), (3.18), and (3.19) can be used to understand the relative trajectory. In particular, considering that, as shown in Fig. 3.17, the differences in right ascension of the ascending node and mean anomaly are of the order of 1°, and evaluating the order of magnitude of the different terms, at a first

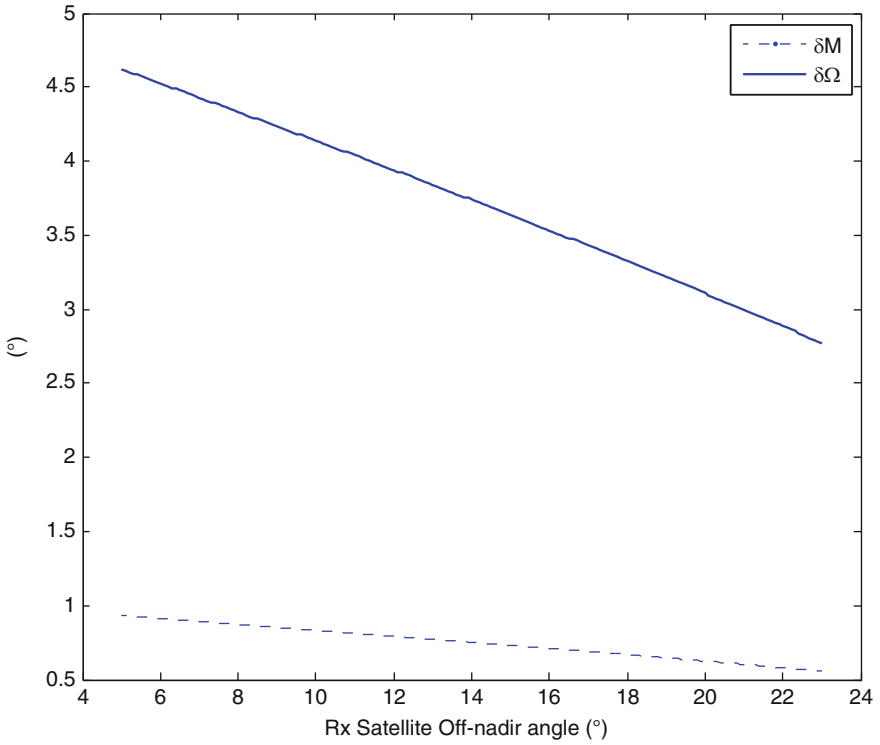


Fig. 3.17 Differences in mean anomaly and right ascension of the ascending node in a large baseline bistatic SAR formation, as a function of the off-nadir angle of the receiver

level of approximation it is possible to neglect most of the contributions thus obtaining:

$$\frac{x}{a} \cong -\delta e \cos\left(M_{D0} + \dot{M}_D t\right) - \frac{\delta\Omega^2}{2} + \frac{\delta\Omega^2}{4}\sin^2 i - \frac{\delta\Omega^2}{4}\sin^2 i \cos(\omega_{D0} + M_{D0} + u_0 + 2\dot{u}_D t) \tag{3.33}$$

$$\frac{y}{a} \cong 2\delta e \sin\left(M_{D0} + \dot{M}_D t\right) + \delta u_0 + \delta\Omega\cos i + \frac{\delta\Omega^2}{4}\sin^2 i \cdot \sin(\omega_{D0} + M_{D0} + u_0 + 2\dot{u}_D t) \tag{3.34}$$

$$\frac{z}{a} \cong -\delta\Omega \sin i \cos(\omega_{D0} + M_{D0} + \dot{u}_D t) \tag{3.35}$$

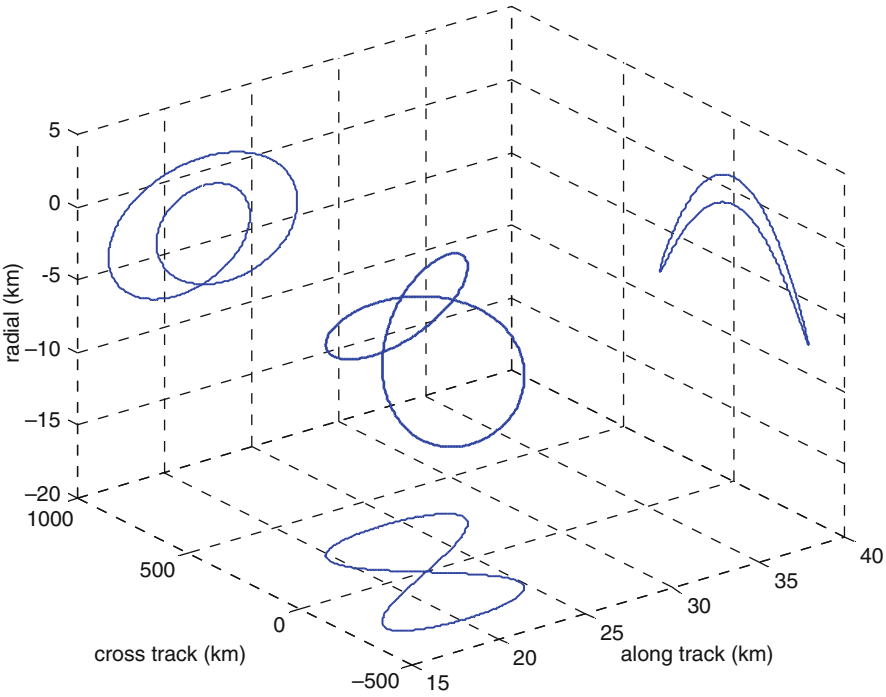


Fig. 3.18 3d relative trajectory and projections for a large baseline bistatic SAR formation

Equation 3.33, in particular, clearly shows that the effect of the difference in eccentricity is to avoid nullification of radial separation at the orbit poles.

The cross-track coordinate oscillates at orbital frequency as in the interferometric case, while the main terms contributing to the in-plane coordinates are at a frequency that is twice the orbital one. Offsets are generated both for the radial and the along-track coordinate, the former depending on $\delta\Omega$, the latter on the yaw steering attitude maneuver. The resulting 3d relative trajectory, and the projections on the coordinate planes, are shown in Fig. 3.18 (different scales along the axes) which is obtained in the example described above (Fig. 3.17) by selecting an off-nadir angle of 23° for the receiver ($\delta\Omega = 2.77°$, $\delta M = 0.56°$, $\delta e = 1.4 \times 10^{-4}$). The combination of oscillations at different frequencies produces motion patterns which differ from the case of close formations such as the ones used for SAR interferometry. This formation geometry produces a stable large baseline bistatic SAR formation, which allows generating constant trends of bistatic angle as a function of latitude.

It is also interesting to investigate "dynamic" large baseline formations, where relative trajectory is modified in a controlled way to obtain SAR acquisitions with varying bistatic angle. As shown in Refs. [54, 55], the basic concept to ensure mission feasibility it to exploit differential J_2 effects to the mission advantage. On the grounds of (3.12), a differential nodal drift can be activated by imposing a difference in inclination. This will also cause an in-plane drift which in general

differs from the one required for minimization of attitude/pointing requirements. To counteract this drawback, two solutions can be envisaged. The first one consists in applying periodical in-plane corrections to keep the optimal along-track separation. Whereas, the second one builds on an ad-hoc δa. In fact, considering the order of magnitude of the coefficients in (3.12), it comes out that due to the high sensitivity of the along-track drift on the difference of semi-major axis, along-track motion can be finely tuned by a very small δa (of the order of a few meters) which has no significant impact on the geometry of relative trajectory. Therefore, apart from the manoeuvre necessary to establish the required δa, no additional propellant expense are foreseen for along-track relative displacement control.

Finally, if the arguments of perigees equal 90° in nominal conditions, a difference in eccentricity can be again used to generate a safety separation at the orbit poles. What is obtained is a drifting bistatic formation that exploits differential J_2 effects to modify the observation geometry in a controlled way.

Required differences in orbital parameters can be calculated by straightforward analytical relations if differential J_2 effects are linearized as in (3.11), (3.12), and (3.13).

In fact, assuming a required nodal drift $\delta\dot{\Omega}_{req}$, the optimal in-plane drift can be calculated as [54]

$$\delta\dot{u}_{req} \cong K_{req}\delta\dot{\Omega}_{req} \tag{3.36}$$

where K_{req} is a coefficient of order 1.

Neglecting J_2 dependancy on eccentricity, differences in inclination and semi-major axes can be calculated by solving the linear system

$$\begin{cases} \delta\dot{\Omega}_{req} = C_{\Omega,a}\frac{\delta a}{a} + C_{\Omega,i}\delta i \\ \delta\dot{u}_{req} = C_{u,a}\frac{\delta a}{a} + C_{u,i}\delta i \end{cases} \tag{3.37}$$

Indeed, given the order of magnitude of the coefficients (see Sect. 3.2.1), it comes out that $\frac{\delta a}{a} \cong O\left(10^{-3}\delta i\right)$, and equations' solution can be approximated as follows:

$$\begin{cases} \delta i \cong \dfrac{\delta\dot{\Omega}_{req}}{C_{\Omega,i}} \\ \dfrac{\delta a}{a} = \dfrac{\delta\dot{u}_{req} - C_{u,i}\delta i}{C_{u,a}} \end{cases} \tag{3.38}$$

As an example, Fig. 3.19 shows the time evolution of relative coordinates for a drifting bistatic formation over a time span of 2,000 orbits (under the sole J_2 effect), starting from a helix close formation dedicated to SAR interferometry (δa ≈ 10 m, δi ≈ 0.1°).

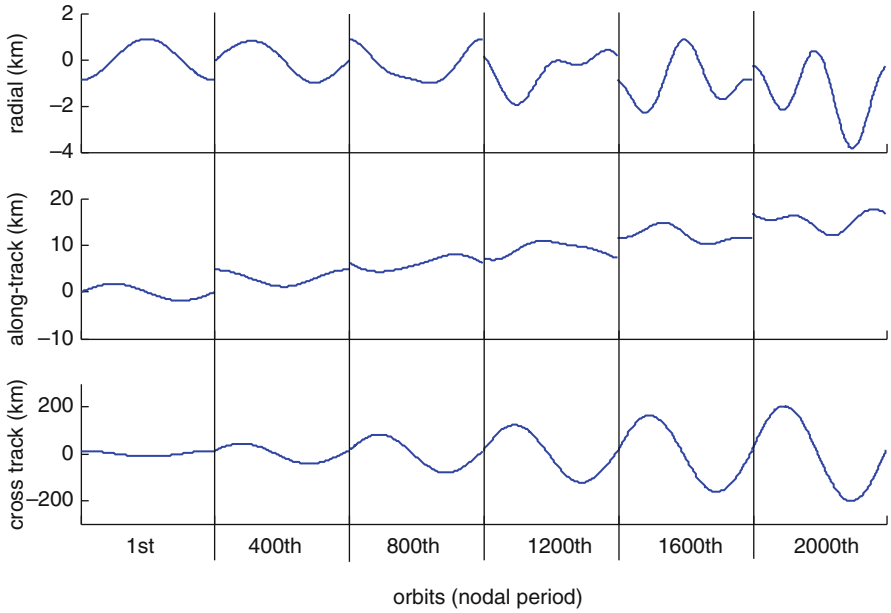


Fig. 3.19 Example of relative motion components for a slowly drifting bistatic formation (Reproduced with kind permission from Springer Science+Business Media B.V. © 2010 Springer Published in Ref. [55])

3.5 Conclusion

Simple but effective relative motion model is a strong need of formation design. In the last decade, relative motion models have been strongly improved with respect to the initial Hill's schematization in two main fields: inclusion of orbit eccentricity and J_2 effects. This worldwide research effort has in fact produced a number of relative models which, relying on linearized orbital parameter differences, bear many similarities in their final forms. They now allow producing more effective design particularly for close (separation of a few kilometers) formations and small eccentricities. Large formations (separations till a few hundreds of kilometers) require additional, quadratic terms to be included, which would lead to a complex fully general model with application difficulties. Indeed, models have been developed for some special cases. Relative motion models have also been developed for large eccentricities of the chief's orbit, although their effective application to formation design is questionable due to their complexity and the inherent difficulties to derive design parameters from requirements.

It has also been shown how relative motion models can capture the different dynamic effects and it is possible to directly establish a cause/effect relation between relative trajectory components and orbital parameters for a number of

effects: offset, oscillation at orbital frequency, oscillation at multiple of the orbital frequency, etc. This possibility has a positive outcome on formation design, as it was pointed out in previous pages. In particular, it was shown that one can derive desired trajectory features from requirements in the case of SAR interferometry and then utilize relative motion model to determine the needed differences in the orbital parameters, selecting the most appropriate relative trajectory shape. Such approach was also satisfactory for SAR tomography. In addition, inclusion of some quadratic terms proved to be effective to design large formations for bistatic SAR applications.

References

1. Hill GW (1878) Researches in the lunar theory. Am J Math 1(1):5–26
2. Hill GW (1878) Researches in the lunar theory. Am J Math 1(2):129–147
3. Hill GW (1878) Researches in the lunar theory. Am J Math 1(3):245–260
4. Clohessy WH, Wiltshire RS (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 27(9):653–658
5. London HS (1963) Second approximation to the solution of the rendezvous equation. AIAA J 1 (7):1691–1693
6. Anthony ML, Sasaki FT (1965) Rendezvous problem for nearly circular orbits. AIAA J 3 (9):1666–1673
7. Tshauner J, Hempel P (1965) Rendezvous zu einem in Elliptischer Bahn Umlaufenden Ziel. Acta Astronaut 11(2):104–109
8. Schaub H, Junkins JL (2003) Spacecraft formation flying. In: Analytical mechanics of space systems. AIAA, Reston, pp 593–674
9. Vadali SR, Alfriend KT, Vaddi S (2000) Hill's equations, mean orbit elements, and formation flying of satellites. In: The Richard H. Battin astrodynamics conference, College Station, TX
10. Wiesel WE (2002) Relative satellite motion about an oblate planet. AIAA J Guid Contr Dyn 25 (4):776–785
11. Schweighart S, Sedwick R (2002) High fidelity linearized J2 model for satellite formation flight. AIAA J Guid Contr Dyn 25(6):1073–1080
12. Halsall M, Palmer PL (2007) Modelling natural formations of LEO satellites. Celest Mech Dyn Astron 99(2):105–127
13. Carter TE (1998) State transition matrices for terminal tendezvous studies: brief survey and new examples. AIAA J Guid Contr Dyn 21(1):148–155
14. Gim D-W, Alfriend KT (2003) State transition matrix of relative motion for the perturbed noncircular reference. AIAA J Guid Contr Dyn 26(6):956–971
15. Gim D-W, Alfriend KT (2005) Satellite relative motion using differential equinoctial elements. Celest Mech Dyn Astron 92(4):295–336
16. Sengupta P, Vadali SR, Alfriend KT (2007) Second order state transition for relative motion near perturbed, elliptic orbits. Celest Mech Dyn Astron 97(2):101–129
17. Schaub H, Alfriend KT (2001) J_2 invariant orbits for spacecraft formations. Celest Mech Dyn Astron 79:77–95
18. Schaub H, Vadali SR, Alfriend KT (2000) Spacecraft formation flying control using mean orbit elements. J Astronaut Sci 48(1):69–87
19. Schaub H (2004) Relative orbit geometry through classical orbit element differences. AIAA J Guid Contr Dyn 27(5):839–848
20. Fasano G, D'Errico M (2009) Modeling orbital relative motion to enable formation design from application requirements. Celest Mech Dyn Astron 105(1–3):113–139

21. Alfriend KT, Schaub H (2000) Dynamic and control of spacecraft formations: challenges and some solutions. J Astronaut Sci 48(2–3):249–267
22. Vallado DA (2003) Fundamentals of astrodynamics and applications, 2nd edn. Microcosm, Inc., El Segundo, 2001
23. Schaub H, Alfriend KT (2002) Hybrid Cartesian and orbit element feedback law for formation flying spacecraft. J Guid Contr Dyn 25(2):387–393
24. Fasano G, D'Errico M (2006) Relative motion model including J_2: derivation and application to INSAR. In: 2006 I.E. aerospace conference, Big Sky, MT, USA, March 2006. DOI: 10.1109/AERO.2006.1655771
25. Vadali SR, Sengupta P, Yan H, Alfriend KT (2008) Fundamental frequencies of satellite relative motion and control of formations. J Guid Contr Dyn 31(5):1239–1248
26. Alfriend KT, Vadali SR, Gurfil P, How JP, Breger LS (2010) Formation flying. Butterworth-Heinemann, Oxford, pp 103–112
27. Chobotov VA (ed) (2002) Orbital mechanics, 3rd edn. AIAA, Reston
28. Sabol C, Burns R, McLaughlin CA (2001) Satellite formation flying design and evolution. J Spacecraft Rockets 38(2):270–278
29. D'Amico S, Montenbruck O (2006) Proximity operations of formation flying spacecraft using an eccentricity/inclination vector separation. J Guid Contr Dyn 29(3):554–563
30. Sengupta P, Vadali SR, Alfriend KT (2008) Averaged relative motion and applications to formation flight near perturbed orbits. J Guid Contr Dyn 31(2):258–272
31. Melton RG (2000) Time-explicit representation of relative motion between elliptical orbits. J Guid Contr Dyn 23(4):604–610
32. Lawden DF (1963) Optimal trajectories for space navigation. Butterworths, London
33. de Vries JP (1963) Elliptic elements in terms of small increments of position and velocity components. AIAA J 1(11):2626–2629
34. Yamanaka K, Ankersen F (2002) New state transition matrix for relative motion on an arbitrary elliptical orbit. J Guid Contr Dyn 25(1):60–66
35. Inalhan G, Tillerson MJ, How J (2002) Relative dynamics and control of spacecraft formations in eccentric orbits. J Guid Contr Dyn 25(1):48–59
36. Ketema Y (2005) An analytic solution for relative motion with an elliptic reference orbit. J Astronaut Sci 53(4):373–389
37. Lane C, Axelrad P (2006) Formation design in eccentric orbits using linearized equations of relative motion. J Guid Contr Dyn 29(1):146–160
38. Broucke RA (2003) Solution of the elliptic rendezvous problem with the time as independent variable. J Guid Contr Dyn 26(4):615–621
39. Sengupta P, Vadali SR (2007) Relative motion and the geometry of formations in Keplerian elliptic orbits. J Guid Contr Dyn 30(4):953–964
40. Sengupta P, Vadali SR, Alfriend KT (2004) Modeling and control of satellite formations in high eccentricity orbits. J Astronaut Sci 52(1):149–168
41. Gens R, van Genderen JL (1996) SAR interferometry-issues, techniques applications. Int J Rem Sens 17:1803–1835
42. Bamler R, Hartl P (1998) Synthetic aperture radar interferometry. Inverse Probl 14:R1–R54
43. Li FK, Goldstein RM (1990) Studies of multibaseline spaceborne interferometric synthetic aperture radars. IEEE Trans Geosci Rem Sens 28(1):88–97
44. Moccia A, Vetrella S (1992) A tethered interferometric synthetic aperture radar (SAR) for a topographic mission. IEEE Trans Geosci Rem Sens 30(1):103–109
45. Massonnet D (2001) Capabilities and limitations of the interferometric cartwheel. IEEE Trans Geosci Rem Sens 39(3):506–520
46. Moccia A, Fasano G (2005) Analysis of spaceborne tandem configurations for complementing COSMO with SAR interferometry. EURASIP J Appl Signal Proces 2005(20):3304–3315
47. Reigber A, Moreira A (2000) First demonstration of SAR tomography using multibaseline L-band data. IEEE Trans Geosci Rem Sens 38(5):2142–2152

48. Richards JA (2009) Interferometric and tomographic SAR. In: Remote sensing with imaging radar. Springer, Berlin/Heidelberg, pp 209–215
49. Fornaro G, Serafino F (2006) Imaging of single and double scatterers in urban areas via SAR tomography. IEEE Trans Geosci Rem Sens 44(12):3497–3505
50. Renga A, Fasano G (2011) Analysis of formation geometries for multistatic SAR tomography. Internal note, Department of Aerospace Engineering, University of Naples "Federico II"
51. Rosen PA, Hensley S, Joughin IR, Li FK, Madsen SN, Rodriguez E, Goldstein RM (2000) Synthetic aperture radar interferometry. Proc IEEE 88(3):333–382
52. Krieger G, Moreira A, Fiedler H, Hajnsek I, Werner M, Younis M, Zink M (2007) TanDEM-X: a satellite formation for high-resolution SAR interferometry. IEEE Trans Geosci Rem Sens 45(11):3317–3341
53. D'Errico M, Moccia A (2003) Attitude and antenna pointing design of bistatic radar formations. IEEE Trans Aerosp Electron Sys 39(3):949–960
54. D'Errico M, Fasano G (2008) Design of interferometric and bistatic mission phases of COSMO/SkyMed constellation. Acta Astronaut 62(2–3):97–111
55. D'Errico M, Fasano G (2010) Relative trajectory design for bistatic SAR missions. In: Sandau R, Roeser H-P, Valenzuela A (eds) Small satellite missions for earth observation. Springer, Berlin/Heidelberg, pp 145–154

Chapter 4
Formation Establishment, Maintenance, and Control

Srinivas R. Vadali and Kyle T. Alfriend

Abstract This chapter presents continuous and impulsive control methods for formation initialization, maintenance, and reconfiguration. For two-body, circular reference orbits, elementary impulsive control schemes are developed based on the available state transition matrix for relative motion. Formation propagation and control models are presented in the space of differential orbital elements and Cartesian/curvilinear coordinate systems. The J_2-perturbation effects are conveniently modeled with the mean elements and their secular drift rates. Methods for accommodating the disturbance due to the J_2 by modification of the relative orbit initial conditions are discussed. Examples provided include multi-impulse optimal formation initialization maneuvers and a novel inter-satellite fuel balancing concept.

4.1 Introduction

Neighboring spacecraft in unperturbed two-body orbits of equal periods remain in closed relative trajectories, constituting a formation. Formations with small separation distances can be described by the periodic solutions of the Clohessey-Wiltshire (CW) [4] and Tschauner-Hempel (TH) [10, 19] equations, for circular and elliptic reference orbits, respectively. These classic solutions form the basis for the design of formations and methods for their control, which have to be modified to account for the unmodeled effects of geometric nonlinearities and perturbations.

Persistent disturbances must be accommodated and cannot be countered with control action. Hence, the formation geometry must confirm to natural solutions

S.R. Vadali (✉) • K.T. Alfriend
Department of Aerospace Engineering, Texas A&M University, College Station, USA
e-mail: svadali@tamu.edu; alfriend@tamu.edu

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_4,

dictated by physical reality as closely as possible. This chapter presents a description of the geometry of satellite formations and the initial conditions required to establish them. In particular, the mean orbital element differences required to accommodate the J_2-perturbation are provided, along with a brief discussion of methods for continuous and impulsive control for formation establishment, maintenance, and reconfiguration.

4.2 Formation Geometry for Circular Orbits

Relative motion of a satellite (deputy) in the small, with respect to a circular reference orbit (of a chief, either real or virtual), with mean motion n, is described by the CW equations:

$$\ddot{x} - 2n\dot{y} - 3n^2 x = 0 \tag{4.1}$$

$$\ddot{y} + 2n\dot{x} = 0 \tag{4.2}$$

$$\ddot{z} + n^2 z = 0 \tag{4.3}$$

where x, y, and z represent, respectively, the radial, along-track, and cross-track separations. The matrix representation of the solution to the CW equations can be expressed as

$$\mathrm{x}(\theta) = \Phi(\theta)\ \mathrm{x}(0) \tag{4.4}$$

where the state vector is $\mathrm{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$ and the state transition matrix (STM) is

$$\Phi = \begin{bmatrix} 4 - 3\cos\theta & 0 & 0 & \dfrac{\sin\theta}{n} & \dfrac{2(1-\cos\theta)}{n} & 0 \\ -6(\theta - \sin\theta) & 1 & 0 & \dfrac{-2(1-\cos\theta)}{n} & \dfrac{4\sin\theta - 3\theta}{n} & 0 \\ 0 & 0 & \cos\theta & 0 & 0 & \dfrac{\sin\theta}{n} \\ 3n\sin\theta & 0 & 0 & \cos\theta & 2\sin\theta & 0 \\ -6n(1-\cos\theta) & 0 & 0 & -2\sin\theta & -3 + 4\cos\theta & 0 \\ 0 & 0 & -n\sin\theta & 0 & 0 & \cos\theta \end{bmatrix} \tag{4.5}$$

The variable $\theta = nt$ is an orbit reference angle or equivalently, the nondimensional time. Secular terms present in the STM can be eliminated by selecting initial conditions satisfying

$$\dot{y}(0) = -2nx(0) \tag{4.6}$$

With the application of Eq. (4.6), the periodic solutions of the CW equations can be represented by

$$x = \rho_x \sin(\theta + \alpha_x) \tag{4.7}$$

$$y = \rho_y + 2\rho_x \cos(\theta + \alpha_x) \tag{4.8}$$

$$z = \rho_z \sin(\theta + \alpha_z) \tag{4.9}$$

where ρ_x and ρ_z are amplitudes of motion along the respective directions and α_x and α_z are the corresponding phase angles. The parameter ρ_y determines the along-track bias or the mean standoff distance. Relative trajectories can be shaped and oriented by choosing appropriate amplitudes and phase angles. For the special case of $\alpha_x = \alpha_z$, the radial and cross-track separations are in phase, i.e., they cross zero at the same time. On the other hand, if $\alpha_x = \alpha_z \pm \frac{\pi}{2}$, radial and cross-track separations are maximally out of phase; one is at its extremum when the other is zero. The out-of-phase option is preferable for collision avoidance, in the presence of along-track drift or uncertainty in the parameter ρ_y.

The *Projected Circular Orbit* (PCO) in the *y*-*z* plane is obtained by selecting

$$\alpha_x = \alpha_z,\ \rho_y = 0,\ \rho_z = 2\rho_x \tag{4.10}$$

Similarly, the following choice of parameters

$$\alpha_x = \frac{\pi}{2} + \alpha_z,\ \rho_y = 0,\ \rho_z = \rho_x \tag{4.11}$$

results in the *x*-*z* PCO. The *x*-*y*-*ellipse* formation is the result of setting $\rho_y = \rho_z = 0$.

Although Eqs. (4.7)–(4.9) are valid only under the ideal assumptions behind the CW model, they provide a preliminary design for the formation, which can be modified to accommodate perturbations. The CW STM also proves useful for formation control under impulsive thrust applications.

4.3 Formation Establishment by Impulsive Control Based on CW Equations

Multiple satellites launched from a common launch vehicle are often inserted into the same orbit, with some degree of separation along the reference orbit, resulting in an in-line or leader-follower formation. The separation distances or phase angles between satellites can be altered by phasing maneuvers to initialize or establish a

formation. This section treats elementary formation establishment problems based on the CW STM and impulsive thrusting, with the number of impulses and their application times selected a priori.

4.3.1 Initialization of In-plane Elliptic Formations

Satellites in an in-line formation can be initialized into the *x-y ellipse* by a variety of maneuvers, involving radial and tangential impulses. It is assumed that the in-line state vector of the CW equations is $\mathbf{x}(0) = [0,\ y(0),\ 0,\ 0,\ 0,\ 0]^T$. The cross-track components of the state vector are ignored for the planar examples considered.

4.3.1.1 Single Radial Impulse

Application of the selected initial conditions to Eqs. (4.4)–(4.6) shows that the requirement for achieving the desired formation with a single radial impulse is

$$\begin{bmatrix} 0 \\ y(0) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \dfrac{\sin\theta}{n} \\ \dfrac{-2(1-\cos\theta)}{n} \\ \cos\theta \\ -2\sin\theta \end{bmatrix} \Delta v = \rho_x \begin{bmatrix} \sin(\alpha_x+\theta) \\ 2\cos(\alpha_x+\theta) \\ n\cos(\alpha_x+\theta) \\ -2n\sin(\alpha_x+\theta) \end{bmatrix} \tag{4.12}$$

Equation (4.12) is satisfied for $\alpha_x = 0$, $\Delta v = n\rho_x$, and $y(0) = 2\rho_x$. Hence, the satellite achieves the *x-y ellipse* formation initial conditions immediately after the application of the impulse, but with the restriction that the semimajor axis of the ellipse be equal to $y(0)$. Depending on the thrust limit, the required impulse may not be realizable with a single burn if $y(0)$ is large; multiple burns of smaller durations may have to be considered.

4.3.1.2 Two Tangential Impulses

It is impossible to achieve the desired initialization with a single tangential impulse. The constraint resulting from the application of two tangential impulses, the first at the initial point and the second, after a reference orbit angle of θ, is

$$\begin{bmatrix} 0 \\ y(0) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \dfrac{2(1-\cos\theta)}{n} \\ \dfrac{4\sin\theta - 3\theta}{n} \\ 2\sin\theta \\ -3+4\cos\theta \end{bmatrix} \Delta v_1 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta v_2 \end{bmatrix} = \rho_x \begin{bmatrix} \sin(\alpha_x+\theta) \\ 2\cos(\alpha_x+\theta) \\ n\cos(\alpha_x+\theta) \\ -2n\sin(\alpha_x+\theta) \end{bmatrix} \tag{4.13}$$

The set of parameters satisfying Eq. (4.13) is

$$\alpha_x = \frac{-\theta}{2} \tag{4.14}$$

$$\Delta v_1 = -\Delta v_2 = \frac{ny(0)}{3\theta} \tag{4.15}$$

$$\rho_x = \frac{2\sqrt{2(1-\cos\theta)}}{3\theta} y(0) \tag{4.16}$$

Hence, α_x and ρ_x are not independent, but functions of θ and $y(0)$. Furthermore, $3\rho_x < 2y(0)$ and for very small θ, the impulse magnitude is excessive. For the special case of $2\rho_x = y(0)$, $\theta \approx 2.55$ and the total impulse $|\Delta v| \approx 0.523 n\rho_x$.

4.3.1.3 Three Tangential Impulses

Assuming a fixed time interval between the three impulses, e.g., corresponding to $\theta = \pi$, the final state at the end of the third maneuver can be expressed as

$$\mathrm{x}(2\pi) = \Phi^2(\pi) \begin{bmatrix} 0 \\ y(0) \\ 0 \\ \Delta v_1 \end{bmatrix} + \Phi(\pi) \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Delta v_3 \end{bmatrix} \tag{4.17}$$

As previously noted, only the in-plane portion of the STM is being considered for this example. Evaluation of the STM for $\theta = \pi$ results in the constraint

$$\begin{bmatrix} \dfrac{4\Delta v_2}{n} \\ y(0) - \dfrac{3\pi(2\Delta v_1 + \Delta v_2)}{n} \\ 0 \\ \Delta v_1 - 7\Delta v_2 + \Delta v_3 \end{bmatrix} = \rho_x \begin{bmatrix} \sin\alpha_x \\ 2\cos\alpha_x \\ n\cos\alpha_x \\ -2n\sin\alpha_x \end{bmatrix} \tag{4.18}$$

The above equation is satisfied for $\alpha_x = \pi/2 \, or \, 3\pi/2$ with

$$\begin{bmatrix} \Delta v_1 \\ \Delta v_2 \\ \Delta v_3 \end{bmatrix} = \begin{bmatrix} \dfrac{ny(0)}{6\pi} - \dfrac{1}{8}\rho_x n \sin\alpha_x \\ \dfrac{\rho_x n}{4}\sin\alpha_x \\ -\dfrac{ny(0)}{6\pi} - \dfrac{1}{8}\rho_x n \sin\alpha_x \end{bmatrix} \tag{4.19}$$

The second impulse is independent of $y(0)$ and for the case of $2\rho_x = y(0)$, the total impulse required is $|\Delta v| = 0.5n\rho_x$.

A number of examples of reconfiguration schemes based on the CW equations can be found in Ref. [11]. A brief discussion of optimal maneuvers is presented in Sect. 4.9.

4.4 Description of Formations with Differential Orbital Elements

The classical orbital elements are

- a, semimajor axis
- e, eccentricity
- i, inclination
- Ω, right ascension of the ascending node (RAAN)
- ω, argument of perigee
- M, mean anomaly

In the following the argument of latitude is defined as $\theta = \omega + f$. For near-circular orbits, it is convenient to replace e, ω, and M with the nonsingular elements $q_1 = e\cos\omega$, $q_2 = e\sin\omega$, and the mean argument of latitude, $\lambda = \omega + M$.

It has been shown [5, 7, 20] that in a curvilinear coordinate representation, the relative motion variables can be expressed in terms of the differential orbital elements as

$$x = \delta r \tag{4.20}$$

$$y = r_0(\delta\theta + \delta\Omega\cos i_0) \tag{4.21}$$

$$z = r_0(-\sin i_0 \delta\Omega\cos\theta_0 + \delta i\sin\theta_0) \tag{4.22}$$

where δ indicates a differential and $(\cdot)_0$ refers to variables pertaining to the reference orbit, e.g., r_0 is the radius of the chief. In the curvilinear coordinate system, radial separation is the difference between the radii of the two satellites; y is the arc length from the chief to the radius vector of the deputy along an imaginary circular orbit of radius r_0 in the plane of the chief's orbit, and similarly, z is the cross-track arc length to the radius vector of the deputy. Linearization in the curvilinear coordinate system captures some of the nonlinear effects neglected by the CW model.

Relationships between the differential elements and the CW parameters are obtained by comparing Eqs. (4.21) and (4.22) with Eqs. (4.8) and (4.9):

$$\delta i = \frac{\rho_z}{a_0}\cos\alpha_z \tag{4.23}$$

$$\delta\Omega = -\frac{\rho_z}{a_0}\frac{\sin\alpha_z}{\sin i_0} \tag{4.24}$$

$$\delta q_1 = -\frac{\rho_x}{a_0}\sin\alpha_x \tag{4.25}$$

$$\delta q_2 = -\frac{\rho_x}{a_0}\cos\alpha_x \tag{4.26}$$

$$\delta\lambda = \frac{\rho_y}{a_0} - \delta\Omega\cos i_0 \tag{4.27}$$

Equations (4.25)–(4.27) can be manipulated to obtain the following classical differential orbital elements:

$$\delta e = \delta q_1\cos\omega_0 + \delta q_2\sin\omega_0 \tag{4.28}$$

$$\delta\omega = \frac{1}{e_0}(-\delta q_1\sin\omega_0 + \delta q_2\cos\omega_0) \tag{4.29}$$

$$\delta M = \delta\lambda - \delta\omega \tag{4.30}$$

These equations provide initial conditions for establishing formations; in general, they are not solutions to the relative motion equations.

4.5 Formation Initialization Accounting for the Effects of J_2

The formation initialization constraints accounting for J_2 can be stated in terms of the mean differential orbital elements. The mean elements can be obtained from the osculating elements by means of a canonical transformation [3]. Under the effect of J_2, the mean a, e, and i remain constant and the mean Ω, ω, and M vary linearly. The secular rates of the angles are

$$\dot{M} = n\left[1 - J_2\left(\frac{3{R_e}^2}{4a^2\eta^3}\right)\left(1 - 3\cos^2 i\right)\right] \tag{4.31}$$

$$\dot{\omega} = -nJ_2\left(\frac{3{R_e}^2}{4a^2\eta^4}\right)\left(1 - 5\cos^2 i\right) \tag{4.32}$$

$$\dot{\Omega} = -nJ_2\left(\frac{3{R_e}^2}{2a^2\eta^4}\right)\cos i \tag{4.33}$$

$$\dot{\lambda} = n - nJ_2\left(\frac{3{R_e}^2}{4a^2\eta^4}\right)\left[\eta\left(1 - 3\cos^2 i\right) + \left(1 - 5\cos^2 i\right)\right] \tag{4.34}$$

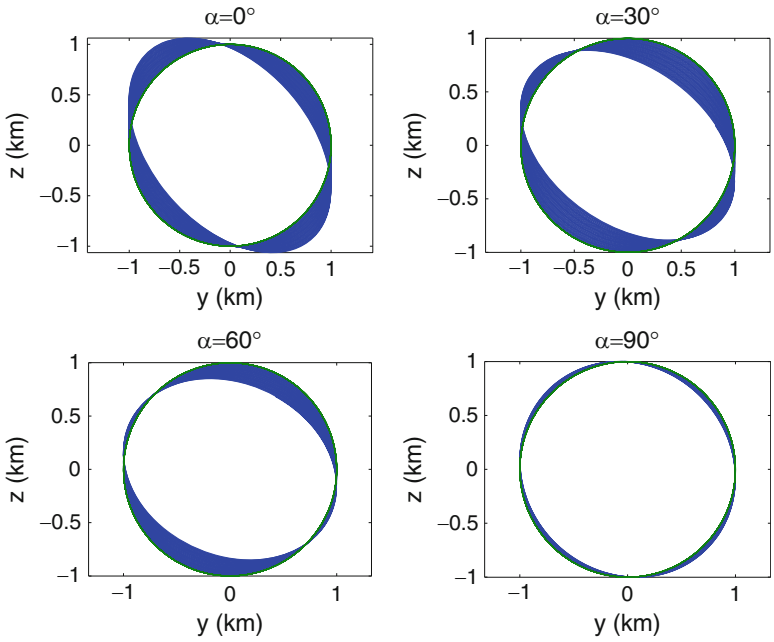


Fig. 4.1 Uncontrolled PCOs for various $\alpha(0)$ values (50 orbits shown)

where R_e is the equatorial radius of the Earth and $\eta = \sqrt{1-e^2}$. Note that the angular rates depend on the three parameters a, e, and i.

Satellites in a formation will drift apart if their respective angular rates do not match. In general, it is impossible to simultaneously null the three primary differential angular rates: $\delta\dot{M}$, $\delta\dot{\omega}$, and $\delta\dot{\Omega}$, by selecting the three parameters δa, δe, and δi [1, 14]. The condition for eliminating cross-track drift is $\delta\dot{\Omega} = 0$ and that for along-track drift is $\delta\dot{\lambda} + \delta\dot{\Omega}\cos i = 0$ [22]. The condition for eliminating along-track drift can be reduced to a condition on δa [1]:

$$\delta a = -0.5 J_2 a_0 \left(\frac{R_e}{a_0}\right)^2 \left(\frac{3\eta_0 + 4}{\eta_0^4}\right) \times \left[\left(1 - 3\cos^2 i_0\right) \frac{q_{10}\delta q_1 + q_{20}\delta q_2}{\eta_0^2} + \sin 2i_0 \delta i\right] \tag{4.35}$$

Equations (4.23)–(4.27) and (4.35) are essential for initializing relative orbits. As an example, consider the initialization of y-z PCOs with the following mean elements of the chief:

$$\begin{aligned} a_0 = 7{,}092\,km, \quad e_0 = 0, \quad i_0 = 70^\circ \\ \Omega_0 = 45^\circ, \quad \omega_0 = 0^\circ, \; M_0 = 0^\circ \end{aligned} \tag{4.36}$$

Relative orbits with $\rho_z = 1$ km are shown in Fig. 4.1, for several α ($= \alpha_x = \alpha_z$) values and for 50 orbits of the chief. Although along-track motion is bounded for

each case, cross-track drift is not. Moreover, the relative orbits precess due to non-commensurate in-plane and out-of-plane frequencies, with the precession rate depending on α.

As shown in Ref. [24], the in-plane frequency of relative motion is $\dot{M}_0$ and the cross-track frequency is given by

$$n_z \approx \dot{M}_0 + \dot{\omega}_0 - \frac{\sin i_0 \delta\dot{\Omega}\delta i}{\delta i^2 + (\sin i_0 \delta\Omega(0))^2} \tag{4.37}$$

Besides the perigee rotation rate, the cross-track frequency depends also on the differential nodal precession rate. It is interesting to note that α_z also gets affected by J_2 and is given by

$$\dot{\alpha}_z = \frac{-\sin i_0 \delta\dot{\Omega}\delta i}{\delta i^2 + (\sin i_0 \delta\Omega)^2} \tag{4.38}$$

A matching condition relating the orbit inclination and the initial phase angle for formations in near-circular orbits is [13, 24]

$$i_0 = \sin^{-1}\left(\sqrt{\frac{4}{5 + 2\cos^2\alpha_z(0)}}\right) \tag{4.39}$$

Equation (4.39) is valid only over the short-term, since certain assumptions to this effect have been made in its derivation. For $\alpha_z(0) = 90^\circ$ and 270°, the results are the two critical inclinations, $i_0 = 63.\,43^\circ$ and $116.\,57^\circ$, respectively. For $\alpha_z(0) = 0$ and 180°, the inclination values are, respectively, $i_0 = 49.\,11^\circ$ and $130.\,89^\circ$. Four frequency-matched y-z PCOs satisfying Eq. (4.39) are shown in Fig. 4.2 for different inclinations. Since a frequency-matched relative orbit adheres to the classical CW solution reasonably well, it requires relatively small control effort to counter the precession due to J_2.

4.6 Formation Maintenance by Continuous Control

The problem of formation maintenance with continuous control is formulated in this section. The CW equations have been modified in [23, 24] to account for the J_2-induced disturbance as

$$\ddot{x} - 2\dot{M}_0\dot{y} - 3\dot{M}_0^2 x = -3n_0^2\delta a + u_{x_c} \tag{4.40}$$

$$\ddot{y} + 2\dot{M}_0\dot{x} = u_{y_c} \tag{4.41}$$

$$\ddot{z} + \dot{M}_0^2 z = -2n_0\dot{\omega}_0 z - 2\rho_z(0)kn_0 \sin i_0^2 \sin\lambda_0 \cos\alpha_z(0) + u_{z_c} \tag{4.42}$$

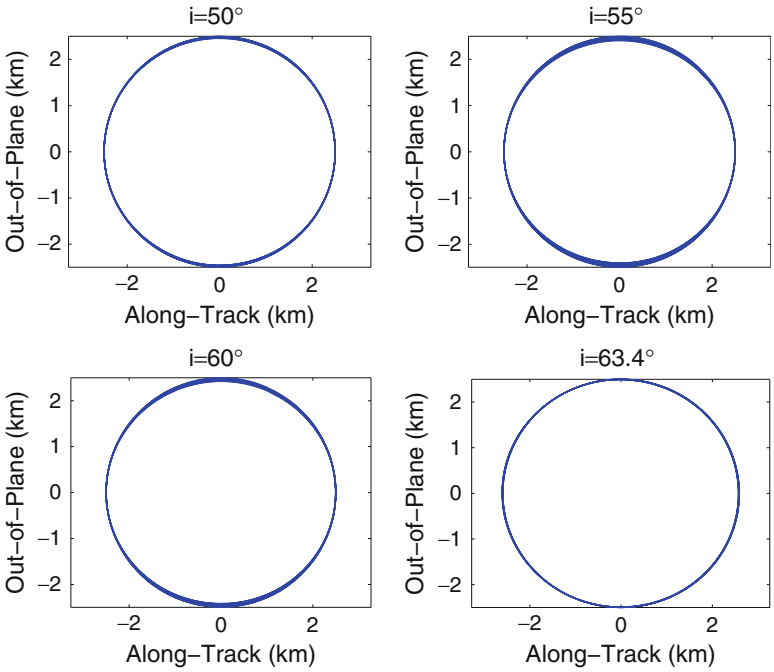


Fig. 4.2 Frequency matched relative orbits for various inclinations

where

$$k = -1.5J_2\left(\frac{R_e}{a_0}\right)^2 n_0 \tag{4.43}$$

and u_{x_c}, u_{y_c}, and u_{z_c} are the control accelerations along their respective axes. These equations resemble the CW equations somewhat, if n is replaced by $\dot{M}_0$. The forcing terms present in the radial and cross-track directions are functions of δa, $\rho_z(0)$, and $\alpha_z(0)$. This model, albeit approximate, accounts for the differential nodal precession caused by a differential inclination. Equation (4.38) shows that α_z is time-varying in the presence of J_2 and indeed, so is ρ_z.

Figure 4.3a shows a snapshot of an ideal eight-satellite y-z PCO formation when the chief is at the ascending node of its own orbit. Satellites 1 and 5 have the maximum δi and $\delta\Omega = 0$, whereas satellites 3 and 7 have the maximum $\delta\Omega$ and $\delta i = 0$. For near-circular orbits, the fuel required to negate differential J_2 effects is dominated by δi. Hence, satellites 1 and 5 will require more fuel than the pair 3 and 7. A formation maintenance approach, motivated by the flight patterns of migratory birds, is to command a slow time variation in the phase angle of each satellite, given by $\alpha = \alpha(0) + \dot{\alpha}t$, where $\frac{\dot{\alpha}}{n_0}$ is $O(J_2)$. As a result, the formation appears to regress at a subsequent ascending nodal crossing of the chief, as shown by the small white circles in Fig. 4.3b.

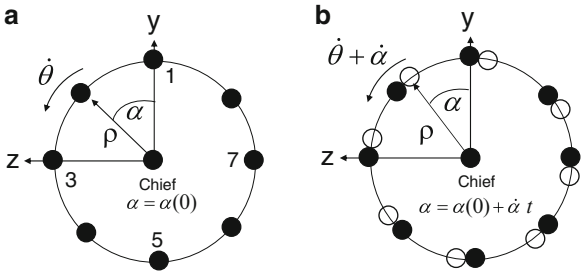


Fig. 4.3 (**a**) PCO formation with $\alpha = \alpha(0)$ when the chief is at the ascending node. (**b**) PCO formation with $\alpha = \alpha(0) + \dot{\alpha}t$ when the chief is at the ascending node

4.6.1 Fuel Minimization and Balancing

Fuel balancing among identical satellites in a formation results in a common ballistic coefficient, virtually eliminating perturbations due to differential aerodynamic drag. In the discussion following, it is assumed that radial thrust, which has been proven to be inefficient for this application [24], is inhibited. In the absence of radial thrust, the mean square control acceleration, averaged over an orbit period is

$$\mathscr{J} = \frac{n_0}{2\pi} \int_0^{2\pi/n_0} (u_{y_c}^2 + u_{z_c}^2) dt \tag{4.44}$$

The averaged cost per satellite in the formation, considering an infinite number of satellites with a uniform distribution of $\alpha(0)$ is

$$\mathscr{J}_{Formation} = \frac{n_0}{4\pi^2} \int_0^{2\pi/n_0} \int_0^{2\pi} (u_{y_c}^2 + u_{z_c}^2) d\alpha(0) dt \tag{4.45}$$

The formation maintenance cost is evaluated for the following reference trajectory, specifically designed by accounting for the lack of radial control:

$$x_r(t) = 0.5\left[1 + \frac{0.5(\dot{\omega}_0 + \dot{\alpha})}{n_0}\right] \rho(0) \sin(\lambda_0(0) + \alpha(0) + (\dot{\lambda}_0 + \dot{\alpha})t) \tag{4.46}$$

$$y_r(t) = \rho(0) \cos(\lambda_0(0) + \alpha(0) + (\dot{\lambda}_0 + \dot{\alpha})t) \tag{4.47}$$

$$z_r(t) = \rho(0) \sin(\lambda_0(0) + \alpha(0) + (\dot{\lambda}_0 + \dot{\alpha})t) \tag{4.48}$$

The control inputs required to achieve the reference trajectory, while satisfying Eqs. (4.40)–(4.42) are

$$u_{y_c} \approx -0.5\rho(0) n_0 (\dot{\omega}_0 + \dot{\alpha}) \cos(\lambda_0(0) + \alpha(0) + (\dot{\lambda}_0 + \dot{\alpha})t) \tag{4.49}$$

$$\begin{aligned} u_{z_c} \approx & -2n_0\dot{\alpha}\rho(0)\sin(\lambda_0(0)+\alpha(0)+(\dot{\lambda}_0+\dot{\alpha})t) \\ & +2\rho(0)kn_0\sin^2 i_0\cos(\alpha(0))\sin\lambda_0 \end{aligned} \tag{4.50}$$

Evaluation of the integral Eq. (4.45) with the obtained reference control accelerations results in

$$\mathscr{J}_{Formation} = (\rho n_0)^2\left[\frac{1}{8}(\dot{\omega}_0+\dot{\alpha})^2+2\dot{\alpha}^2+k^2\sin^4 i_0-2\dot{\alpha}k\sin^2 i_0\right] \tag{4.51}$$

Minimization of the above expression with respect to $\dot{\alpha}$ yields

$$\dot{\alpha} = \frac{1}{17}(-\dot{\omega}_0+8k\sin^2 i_0) < 0 \tag{4.52}$$

where k is given by Eq. (4.43).

4.6.2 Implementation of the Control Law

A formation maintenance control law is designed based on the steady-state Linear-Quadratic-Regulator (LQR) approach and the unperturbed CW model. The feedback controls for the y and z axes are augmented with the feedforward controls given by Eqs. (4.49) and (4.50):

$$\begin{bmatrix} u_y \\ u_z \end{bmatrix} = \begin{bmatrix} u_{y_c} \\ u_{z_c} \end{bmatrix} - \mathbf{K}\begin{bmatrix} \mathrm{x}-\mathrm{x}_r \\ \dot{\mathrm{x}}-\dot{\mathrm{x}}_r \end{bmatrix} \tag{4.53}$$

where, $\mathbf{K}$ is the LQR gain matrix and $\mathrm{x} = \begin{bmatrix} x & y & z \end{bmatrix}^T$. The reference position vector $\mathbf{x}_r$ is obtained from Eqs. (4.46)–(4.48).

Beginning with a set of mean elements for the chief, the initial mean elements for the deputies are setup as per Eqs. (4.23)–(4.27) and (4.35). The mean elements of the satellites are converted into the respective osculating elements via the procedure given in [7]. The osculating orbital elements can be converted into inertial position and velocity vectors to obtain the initial conditions required for a numerical integration process. The relative position and velocity states are obtained from the inertial states of the satellites by using the transformations presented in Ref. [21]. Next, the relative position and velocity states are filtered [17] to remove the short-periodic oscillations and then feedback via the control law given by Eq. (4.53).

A formation of seven satellites in a 1 km, PCO configuration with initial phase angles ranging from 0° to 90° is considered in Ref. [24]. The chief is assumed to be in a circular orbit with $a_0 = 7{,}100$km and $i_0 = 49.11^\circ$. For this example, $\dot{\alpha}_{opt} = -2.07^\circ/day$. The control weight matrix is diagonal, with entries $[1,\ 1]/{n_0}^4$;

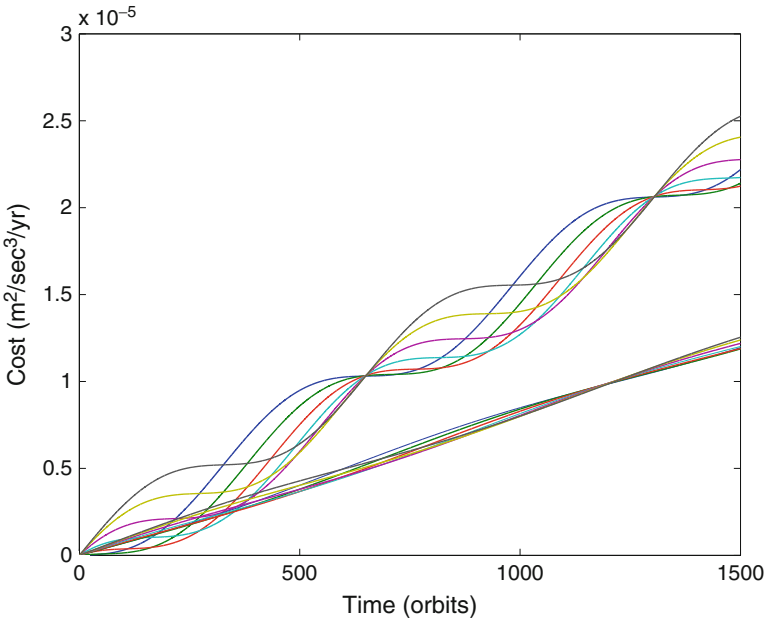


Fig. 4.4 Cost vs. time with and without $\dot{\alpha}$, $i_0 = 49.\ 11^\circ$, 1,500 orbits

the state weight matrix is also diagonal with unit weights for the position coordinates and ${n_0}^{-2}$ for the rate errors.

Figure 4.4 shows the cost of formation maintenance using a mean-element based feedback controller, with and without fuel-balancing, obtained by evaluating Eq. (4.44) for each satellite. The balanced cost curves are close to each other and have a smaller average slope compared to the unbalanced cost curves. The total formation maintenance cost with fuel balancing is 8.49×10^{-5} m^2/s^3/year, compared to 1.59×10^{-4} m^2/s^3/year obtained by holding $\alpha(0)$ constant for each satellite.

An alternative *virtual center* approach [18] for formation maintenance with fuel balancing is to exert control on each satellite in proportion to its fuel content with respect to the formation average. The satellite with the least fuel is designated the uncontrolled chief, with respect to which the remaining satellites are commanded to maintain formation. A method for controlling formations by establishing the required differential mean elements via a control Lyapunov function approach is presented in Ref. [16].

4.7 Gauss' Variational Equations Including Mean J_2 Effects

Gauss' variational equations (GVE) provide a convenient model for computing solutions to multi-impulse orbit correction problems. These equations, originally derived for propagating the osculating orbital elements under the action of

perturbations, can be augmented with the mean secular rates given by Eqs. (4.31)–(4.33) [2, 15], thus providing an approximate method for mean element propagation. The modified GVE are

$$\dot{\mathbf{oe}} = \mathbf{A} + \mathbf{Bu} \tag{4.54}$$

where the vector of orbital elements $\mathbf{oe} = [a\ e\ i\ \Omega\ \omega\ M]^T$, the external control vector $\mathbf{u} = [u_x u_y u_z\ ]^T$,

$$\mathbf{A} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -nJ_2\left(\dfrac{3R_e{}^2}{2a^2\eta^4}\right)\cos i \\ -nJ_2\left(\dfrac{3R_e{}^2}{4a^2\eta^4}\right)\left(1-5\cos^2 i\right) \\ n\left[1-J_2\left(\dfrac{3R_e{}^2}{4a^2\eta^3}\right)\left(1-3\cos^2 i\right)\right] \end{bmatrix} \tag{4.55}$$

and

$$\mathbf{B} = \begin{bmatrix} \dfrac{2a^2 e\sin f}{h} & \dfrac{2a^2 p}{hr} & 0 \\ \dfrac{p\sin f}{h} & \dfrac{(p+r)\cos f + re}{h} & 0 \\ 0 & 0 & \dfrac{r\cos\theta}{h} \\ 0 & 0 & \dfrac{r\sin\theta}{h\sin i} \\ \dfrac{-p\cos f}{he} & \dfrac{(p+r)\sin f}{he} & \dfrac{-r\sin\theta\cos i}{h\sin i} \\ \dfrac{\eta(p\cos f - 2re)}{he} & \dfrac{-\eta(p+r)\sin f}{he} & 0 \end{bmatrix} \tag{4.56}$$

The term $\mathbf{A}$ in Eq. (4.54) does not appreciably influence the maneuver calculations over a short duration. Propagation of the elements during a coast phase is quite simple, the three angle variables being linear functions of time. The orbital elements undergo jumps at the time of application of an impulse, given by

$$\Delta\mathbf{oe} = \mathbf{B}\begin{bmatrix} \Delta v_x \\ \Delta v_y \\ \Delta v_z \end{bmatrix} \tag{4.57}$$

where $[\Delta v_x, \Delta v_y, \Delta v_z]^T$ is the impulse vector. Solution to Kepler's equation is a requirement of the above procedure for elliptic orbits if time is the independent variable.

4.7.1 Gauss' Equations for Circular Orbit Formation Control

For mean circular reference orbits, the effects of differential eccentricity on the J_2-induced drift rates are negligible. The first-order approximations to drifts in the mean differential nonsingular elements are [1]:

$$\delta\dot{a} = 0 \tag{4.58a}$$

$$\delta\dot{i} = 0 \tag{4.58b}$$

$$\delta\dot{\Omega} = \left(\frac{\partial\dot{\Omega}}{\partial i}\delta i\right)t \tag{4.58c}$$

$$\begin{Bmatrix}\delta\dot{q}_1\\ \delta\dot{q}_2\end{Bmatrix} = \begin{bmatrix}\cos(\dot{\omega}_0 t) & -\sin(\dot{\omega}_0 t)\\ \sin(\dot{\omega}_0 t) & \cos(\dot{\omega}_0 t)\end{bmatrix}\begin{Bmatrix}\delta q_1\\ \delta q_2\end{Bmatrix} \tag{4.58d}$$

$$\delta\dot{\lambda} = \left(\frac{\partial\dot{\lambda}}{\partial i}\delta i + \frac{\partial\dot{\lambda}}{\partial a}\delta a\right)t \tag{4.58e}$$

where $t = 0$ is assumed to be the initial time and

$$\frac{\partial\dot{\lambda}}{\partial i} = -6J_2\left(\frac{R_e}{a_0}\right)^2 n_0 \sin 2i_0 \tag{4.59a}$$

$$\frac{\partial\dot{\lambda}}{\partial a} = -\frac{3n_0}{2a_0} \tag{4.59b}$$

$$\frac{\partial\dot{\Omega}}{\partial i} = \frac{3}{2}J_2\left(\frac{R_e}{a_0}\right)^2 n_0 \sin i_0 \tag{4.59c}$$

The jump conditions at an impulse application time are

$$\Delta(\delta a) \approx \frac{2}{n_0}\Delta v_y \tag{4.60a}$$

$$\Delta(\delta i) \approx \gamma \cos\theta_0 \Delta v_z \tag{4.60b}$$

$$\Delta(\delta\Omega) \approx \frac{\gamma \sin\theta_0}{\sin i_0}\Delta v_z \tag{4.60c}$$

$$\Delta(\delta q_1) \approx \gamma \sin\theta_0 \Delta v_x + 2\gamma \cos\theta_0 \Delta v_y \tag{4.60d}$$

$$\Delta(\delta q_2) \approx -\gamma \cos\theta_0 \Delta v_x + 2\gamma \sin\theta_0 \Delta v_y \tag{4.60e}$$

$$\Delta(\delta\lambda) \approx -2\gamma \Delta v_x - \gamma \sin\theta_0 \cot i_0 \Delta v_z \tag{4.60f}$$

where $\gamma = \sqrt{a_0/\mu}$. Control problems involving mean circular orbits can be formulated with Eqs. (4.58) and (4.60).

4.8 Gim-Alfriend State Transition Matrix

The Gim-Alfriend State transition matrix (GA STM) propagates the relative states (position and velocity vectors) under the influence of J_2. Included in this method is a conversion between the mean and osculating differential orbital elements. The curvilinear state vector, $\mathbf{x} = [x, \dot{x}, y, \dot{y}, z, \dot{z}]^T$ is propagated as

$$\mathrm{x}(t) = \Sigma(t)\phi(t, t_0)\Sigma^{-1}(t_0)\ \mathrm{x}(t_0) \tag{4.61}$$

where $\Sigma(t)$ is the transformation matrix relating the instantaneous osculating, nonsingular differential orbital element vector $\delta\mathbf{\text{œ}}(t)$ and $\phi(t, t_0)$ is the STM for the relative mean elements. The nonsingular orbital element vector is defined as $\mathbf{\text{œ}} = [a, \theta, i, q_1, q_2, \Omega]^T$. The elements of $\Sigma(t)$ and $\phi(t, t_0)$ have been derived in [1, 7].

The GA STM is an analytical propagation method and hence, is faster than numerical integration procedures. Generally, it is more accurate than the mean-J_2-augmented-GVE approach. Of course, the most accurate method for relative motion prediction is to difference the results of numerical integration of the equations of motion of the individual satellites.

4.9 Fuel-Optimal Control

In this section an impulsive thrust approximation is used for fuel-optimal formation control. If each impulse vector is produced by a single thruster which can be pointed in the desired direction, by attitude control of the satellite, then the appropriate performance index is

$$\mathscr{J}_1 = \sum_{j=1}^{N} \|\Delta \mathrm{v}_j\| = \sum_{j=1}^{N} \sqrt{\Delta v_{x_j}^2 + \Delta v_{y_j}^2 + \Delta v_{z_j}^2} \tag{4.62}$$

where N is the number of impulses. Alternatively, if the scalar components of an impulse vector are independently produced, the appropriate cost function is

$$\mathcal{J}_2 = \sum_{j=1}^{N} \left|\Delta v_{x_j}\right| + \left|\Delta v_{y_j}\right| + \left|\Delta v_{z_j}\right| \tag{4.63}$$

The problem of minimization of Eq. (4.63) can be simplified if each impulse is represented as follows [9]:

$$\Delta v = \Delta v^{+} - \Delta v^{-} \tag{4.64}$$

with Δv^{+} and Δv^{-} limited by

$$0 \leq \Delta v^{+} \leq \Delta v_{max} \tag{4.65a}$$

$$0 \leq \Delta v^{-} \leq \Delta v_{max} \tag{4.65b}$$

where Δv_{max} is the maximum impulse magnitude. With the above representation, the performance index can be written as

$$\mathcal{J}_2 = \sum_{j=1}^{N} \left(\Delta v_{x_j}^{+} + \Delta v_{x_j}^{-} + \Delta v_{y_j}^{+} + \Delta v_{y_j}^{-} + \Delta v_{z_j}^{+} + \Delta v_{z_j}^{-}\right) \tag{4.66}$$

with the understanding that the solution for each impulse will be characterized by either Δv^{+} or Δv^{-} being zero.

The impulse application times and the Δv magnitude/directions become the free parameters for optimization. For example, the computation of a two-impulse maneuver requires the determination of six impulse components (three per impulse), the two impulse application times, and a final coast interval, leading to nine parameters. Although the number of impulses required for an optimal transfer is unknown a priori, two to four impulses are adequate for most orbit establishment maneuvers. Collision avoidance between satellites and compensation for thrust implementation errors are important factors, not considered in this chapter.

4.9.1 Formation Establishment and Reconfiguration

In this section, an example of a formation establishment/reconfiguration problem is presented. The fuel-optimal control problem is formulated with $\mathcal{J}_2$ as the performance measure. The simulation model is based on numerical integration of the J_2-perturbed equations of motion of the satellites in the Earth Centered Inertial (ECI) frame. Relative motion variables are transformed into the chief-centered

Table 4.1 Formation establishment results

Imp	Radial (m/s)	Along-track (m/s)	Cross-track (m/s)	$\mathscr{J}_2$ (m/s)	t_j (orbits)
2	$\begin{bmatrix} 0.812 \\ -0.243 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -0.002 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -2.007 \end{bmatrix}$	3.065	$\begin{bmatrix} 0 \\ .5 \end{bmatrix}$
4	0	$\begin{bmatrix} -0.072 \\ 0 \\ 0.263 \\ -0.193 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -2.007 \\ 0 \\ 0 \end{bmatrix}$	2.535	$\begin{bmatrix} 0.238 \\ 0.5 \\ 0.762 \\ 1.239 \end{bmatrix}$

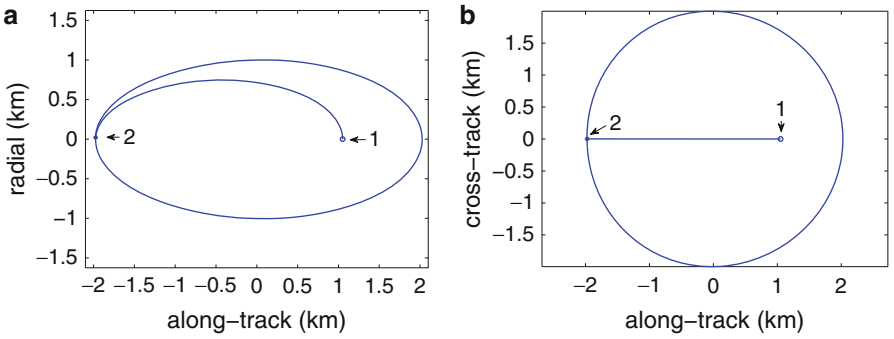


Fig. 4.5 Establishment of a y-z PCO with $\alpha = \alpha(0)$ starting from an in-line formation by a two-impulse maneuver

rotating frame for visualization, the process for which is given in Ref. [1]. The chief is assumed to be uncontrolled, with initial classical mean elements

$$\mathbf{oe}_0 = [7{,}100\,\text{km}\ .05\ 60^\circ\ 00\ 0]^T \tag{4.67}$$

The deputy is assumed to be in the chief's orbit, but with a separation distance of 1 km. Hence, all the differential mean elements are zero, except for $\delta M = \rho_y/a_0$, with $\rho_y = 1$ km. The deputy is to be initialized into a y-z PCO with $\rho = 2$ km and $\alpha = 0$, for which the constraints are given by Eqs. (4.27) and (4.35). The optimal solutions for the impulse magnitudes and the application times (t_j) obtained from the software package SNOPT [6] are presented in Table 4.1. The four-impulse problem has been solved without radial thrust, resulting in a reduction of 50% in the in-plane thrust.

The two-impulse maneuver is primarily executed by radial and cross-track impulses, along-track impulses being negligible. The impulses are separated by a half orbital period. The solution to this problem is non-unique. A similar example, with the use of Eq. (4.54), is provided in Ref. [1]. There are slight differences in the two results due to the model used for orbit propagation. Figure 4.5 provides two views of the relative trajectory of the deputy, identifying the locations of

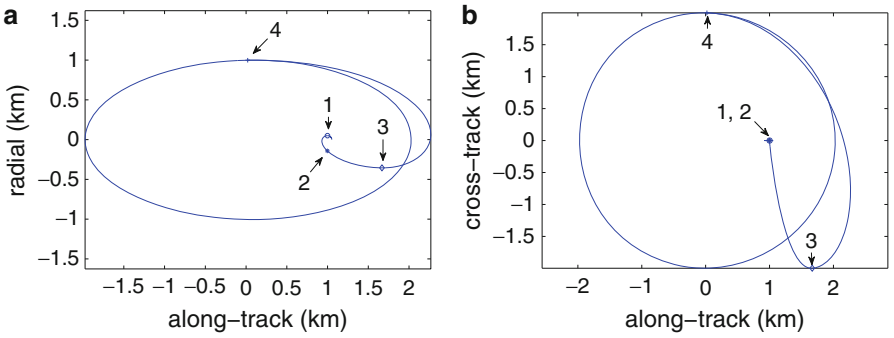


Fig. 4.6 Establishment of a *y*-*z* PCO with $\alpha = \alpha(0)$ starting from an in-line formation by a four-impulse maneuver

the impulses. Figure 4.5a shows that the minimum separation distance during the course of the two-impulse maneuver is ≈0.7km in the radial direction. Figure 4.6 shows that the four-impulse maneuver maintains a higher clearance between the two satellites.

Formation reconfiguration is a special case of initialization. An example of a two-impulse PCO reconfiguration problem, based on the analytical solution to Eqs. (4.60), is presented in Ref. [26]. In general, J_2 does not significantly alter the results for maneuvers performed over the course of one or two orbits. The impulse magnitudes and maneuver times can be well estimated by two-body models. However, even the implementation of the impulse data obtained from one J_2-model into another equivalent simulation model can produce unacceptable results. Hence, it cannot be overemphasized that in practice, maneuvers in close proximity to other satellites require careful planning and accurate execution. Chaps. 11 and 12 of Ref. [1] and [8, 12] deal with the practical realities of dealing with off-nominal thruster performance and navigation errors.

4.9.2 Formation Maintenance

Long-term formation maintenance and inter-satellite fuel balancing is a challenging problem, especially with impulsive control. For mean circular orbits, a two-impulse-per-orbit scheme based on Eqs. (4.60) has been developed in Ref. [25]. Impulse magnitudes and their application times are obtained analytically by decoupling in-plane and cross-track dynamics. Results are presented for the following choices of the mean nonsingular elements of the chief:

$$\begin{aligned} a_0 = 7{,}092\,\text{km},\ \theta_0 = 0\,\text{rad},\ i_0 = 70^\circ \\ q_{10} = 0, \quad q_{20} = 0, \quad \Omega_0 = 45^\circ \end{aligned} \tag{4.68}$$

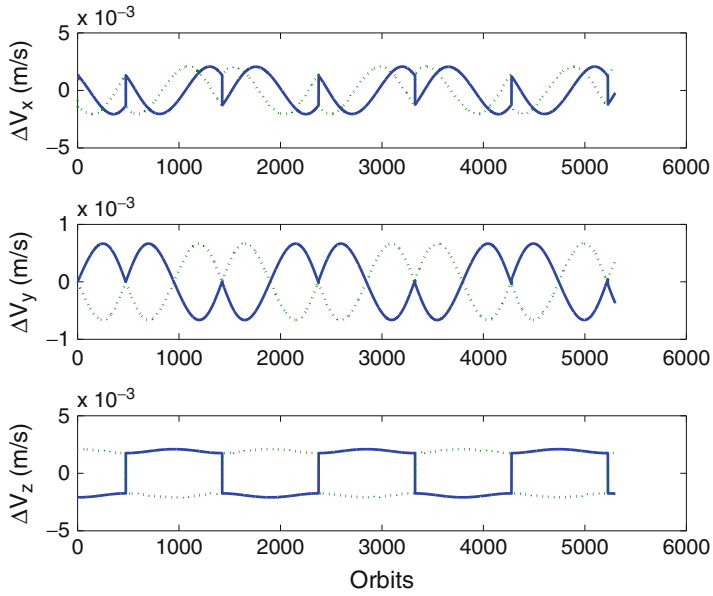


Fig. 4.7 Impulse components for formation maintenance, $\alpha(0) = 0$ and $\dot{\alpha} = -2.723^\circ/\text{day}$

A reference *y-z* PCO given by Eqs. (4.7)–(4.9) is selected with $\rho = 1$ km. Without fuel balancing, the maximum cross-track and in-plane impulse requirements are 41 m/s/year and 6 m/s/year, respectively. Since the cross-track control component dominates in this example, it is used to determine $\dot{\alpha}$ for fuel balancing. For a *y-z* PCO, with a uniform distribution of $\alpha(0)$ values and corresponding to the selected mean elements of the chief, $\dot{\alpha} \approx -2.723^\circ/\text{day}$. In-plane and out-of-plane impulse components are combined into a single impulse, each of which is applied at an interval of half the orbital period. The performance measure $\mathscr{J}_1$ is applicable to this problem.

Figure 4.7 shows the impulses for the satellite with $\alpha(0) = 0$. The average cost per satellite with fuel balancing is 25.8 m/s/year, close to the average of the maximum cross-track and in-plane control costs. The error in the radius of PCO formation is shown in Fig. 4.8. The maximum error in the radius of the PCO due to the effect of J_2 is approximately 3 m. As shown in Ref. [25], the error bound increases to 40 m if the impulses are applied during every tenth orbit of the chief.

4.10 Conclusions

This chapter presents a geometric description of formations in circular and elliptic orbits using Cartesian/curvilinear coordinates and differential orbital elements. The initial conditions for unperturbed formations are modified to accommodate

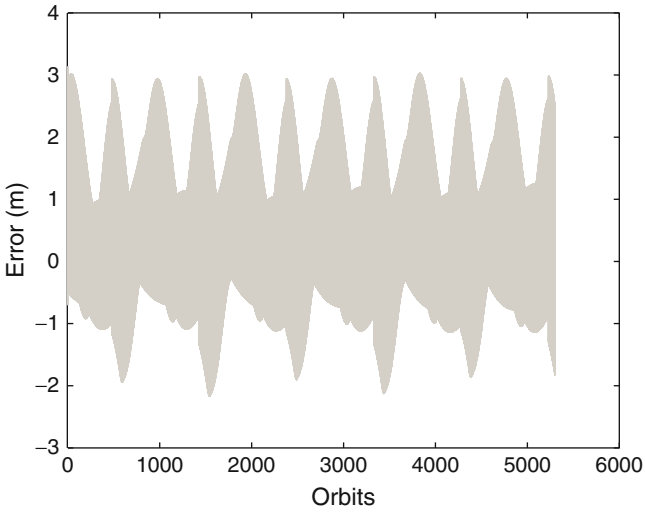


Fig. 4.8 Error in the radius of the PCO for $\alpha(0) = 0$ and $\dot{\alpha} = -2.723^{\circ}/$ day

the effect of J_2. Methods for continuous and impulsive control for formation establishment, reconfiguration, and maintenance are discussed, including multi-impulse, fuel-optimal maneuvers. Also treated in conjunction with formation maintenance is the concept of inter-satellite fuel balancing. The emphasis in this chapter has been on the use of mean orbital elements and the accommodation of the J_2 disturbance.

References

1. Alfriend KT, Vadali SR, Gurfil P, How JP, Breger LS (2010) Spacecraft formation flying: dynamics, control, and navigation. Elsevier astrodynamics series. Elsevier, Amsterdam/Boston/London
2. Breger L, How J (2005) J2-modified GVE-based MPC for formation flying spacecraft. In: Proceedings of the AIAA GNC conference, AIAA, San Francisco, CA. AIAA–2005–5833
3. Brouwer D (1959) Solution of the problem of artificial satellite theory without drag. Astron J 64:378–397
4. Clohessy WH, Wiltshire RS (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 27:653–658, 674
5. Garrison JL, Gardner TG, Axelrad P (1995) Relative motion in highly elliptical orbits. Adv Astron Sci 89(2):1359–1376. Also Paper AAS 95-194 of the AAS/AIAA Space Flight Mechanics Meeting
6. Gill PE, Murray W, Saunders MA (2008) User's guide to SNOPT version 7: software for large-scale nonlinear programming. Department of Mathematics/University of California, San Diego/La Jolla, CA, pp 92093–920112
7. Gim DW, Alfriend KT (2003) State transition matrix of relative motion for the perturbed noncircular reference orbit. J Guidance Control Dyn 26(6):956–971

8. How JP, Tillerson M (2001) Analysis of the impact of sensing noise on formation control. In: American control conference (ACC), Arlington, VA
9. Kumar R, Seywald H (1995) Fuel-optimal station keeping via differential inclusions. J Guidance Control Dyn 18(5):1156–1162
10. Lawden DF (1963) Optimal trajectories for space navigation. Butterworths, London
11. McLaughlin CA, Alfriend KT, Lovell TA (2002) Analysis of reconfiguration algorithms for formation flying experiments. In: International symposium on formation flying missions and technologies, Toulouse
12. Montenbruck O, D'Amico S, Ardaens JS, Wermuth M (2011) Carrier phase differential gps for leo formation flying – the prisma and tandem-x flight experience. In: Paper AAS 11–489,AAS astrodynamics specialist conference, Girdwood, AK
13. Sabatini M, Izzo D, Bevilacqua R (2008) Special inclinations allowing minimal drift orbits for formation flying satellites. J Guidance Control Dyn 31(1):94–100
14. Schaub H, Alfriend KT (2001) J_2 invariant relative orbits for formation flying. Celest Mech Dyn Astron 79(2):77–95
15. Schaub H, Junkins JL (2003) Analytical mechanics of space systems. AIAA education series. American Institute of Aeronautics and Astronautics, Reston
16. Schaub H, Vadali SR, Junkins JL, Alfriend KT (2000) Spacecraft formation flying using mean orbital elements. J Astron Sci 48(1):69–87
17. Sengupta P, Vadali SR, Alfriend KT (2008) Averaged relative motion and applications to formation flight near perturbed orbits. J Guidance Control Dyn 31(2):258–272
18. Tillerson M, Breger L, How JP (2003) Distributed coordination and control of formation flying spacecraft. In: Proceedings of the IEEE American control conference, Denver, CO, pp 1740–1745
19. Tschauner JFA, Hempel PR (1965) Rendezvous zu einemin elliptischer bahn umlaufenden ziel. Astronautica Acta 11(2):104–109
20. Vadali SR (2002) An analytical solution for relative motion of satellites. In: Proceeding of the 5th dynamics and control of systems and structures in space conference, Cranfield University, Cranfield
21. Vadali SR, Schaub H, Alfriend KT (1999) Initial conditions and fuel-optimal control for formation flying of satellites. In: Proceedings of the AIAA GNC conference, AIAA, Portland, OR. AIAA 99–4265
22. Vadali SR, Alfriend KT, Vaddi SS (2000) Hill's equations, mean orbital elements, and formation flying of satellites. Adv Astron Sci 106:187–204. Also Paper AAS 00–258 of the Richard H. Battin Astrodynamics Symposium
23. Vadali SR, Vaddi SS, Alfriend KT (2002) An intelligent control concept for formation flying satellite constellations. Int J Nonlinear Robust Control 12:97–115
24. Vadali SR, Sengupta P, Yan H, Alfriend KT (2008) Fundamental frequencies of satellite relative motion and control of formations. J Guidance Control Dyn 31(5):1239–1248
25. Vadali SR, Yan H, Alfriend KT (2008) Formation maintenance and reconfiguration using impulsive control. In: AIAA/AAS astrodynamics specialists meeting, Honolulu, HI. Paper No. AIAA-08-7359
26. Vaddi SS, Alfriend KT, Vadali SR, Sengupta P (2005) Formation establishment and reconfiguration using impulsive control. J Guidance Control Dyn 28(2):262–268

Chapter 5
GPS Based Relative Navigation

Oliver Montenbruck and Simone D'Amico

Abstract The use of Global Positioning System (GPS) measurements provides the primary technique for determining the relative position of cooperative, formation-flying satellites in low Earth orbit. Similar to terrestrial applications, the relative navigation benefits from a high level of common error cancellation. Furthermore, the integer nature of double-difference carrier phase ambiguities can be exploited in carrier phase differential GPS (CDGPS). Both aspects enable a substantially higher relative accuracy than can be achieved in single-spacecraft navigation. Following an overview of spaceborne GPS receivers, the dynamical and measurement models for relative navigation using single- or dual-frequency measurements are presented along with a discussion of estimation schemes for real-time and offline applications. Actual flight results from the TanDEM-X and PRISMA missions are presented to demonstrate the feasibility of mm-level post-facto baseline determination and cm-level real-time navigation using CDGPS.

5.1 Introduction

Originally conceived as a purely military system, the Global Positioning System (GPS) has long evolved into a ubiquitous and indispensable tool for civil navigation and timing [1]. From car navigation to precision farming and from network time synchronization to aircraft landing, a vast range of applications has evolved, which are enabled or at least supported by GPS. However, use of the Global Positioning System is not restricted to the immediate vicinity of the Earth alone. Similar signal coverage can be achieved up to altitudes of at least 1,500 km, which covers most of the low Earth orbit (LEO) and opens the way for GPS navigation in space.

O. Montenbruck (✉) • S. D'Amico
German Aerospace Center (DLR), German Space Operations Center, Münchner Straße 20, Wessling 82234, Germany
e-mail: oliver.montenbruck@dlr.de; simone.damico@dlr.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_5,

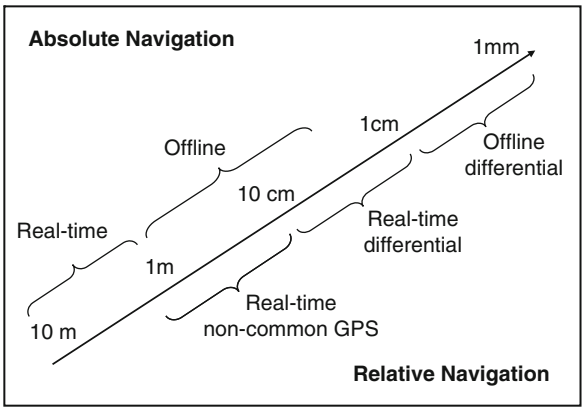


Fig. 5.1 Range of achievable GPS navigation accuracies using absolute and relative positioning techniques

Starting from first experiments in the 1980s, the use of GPS receivers has long proven to be of great benefit for satellite orbit determination and control, in particular when onboard autonomy is desired. A wide range of space-capable GPS receivers is offered today and in fact most LEO satellites launched these days are equipped with GPS receivers for either spacecraft operations or as science instruments. While GPS tracking at high altitudes remains a persistent challenge, the increased demand for GPS use in space has already resulted in the definition of a dedicated Space Service Volume [2] for the next generation of GPS satellites.

Formation flying naturally depends on an accurate knowledge of the relative motion of multiple spacecraft to acquire and maintain a given formation geometry. For single-satellite navigation, absolute positioning accuracies in the 10–5 cm range can presently be achieved with GPS depending on the timeliness requirements (onboard real-time versus ground based offline processing), the hardware capabilities (single versus dual-frequency receiver) and the sophistication of the processing techniques. On the other hand, the full potential of GPS can best be exploited in relative navigation, where mm level positioning accuracies can be achieved through the use of differential carrier phase measurements (Fig. 5.1). This is well known in surveying, deformation monitoring and geodesy, but likewise applicable to formation flying missions in space [3].

As a rule of thumb, maintaining a formation of a given size requires a 10 times better control accuracy [4], which in turn requires a navigation accuracy that is better by yet another factor of ten. Many "loose" formations can thus be controlled based on absolute orbit determination results for the individual spacecraft, but a relative navigation technique is clearly advisable for close formation flight at separations of less than 100 m.

Evidently, the use of GPS based navigation is of primary relevance for formation flying missions in low Earth orbits, where one can fully benefit from a simultaneous tracking of numerous GPS satellites. Even though the subsequent presentation is focussed on LEO applications, it should be mentioned that GPS-based relative navigation is likewise considered for formation flying in highly elliptical orbits [5]

or collocation of geostationary satellites. However, the achievable performance in these orbits is certainly far below that in LEO formation flying missions due to a heavily constrained GPS visibility.

The present chapter starts with a discussion of spaceborne GPS technology and an overview of receivers that can be applied for spacecraft formation flying. Following a presentation of general principles and models for GPS-based orbit determination, the specific aspects of real-time relative navigation of formation flying spacecraft and the precise post-facto baseline reconstruction are discussed in more detail. Flight data from the PRISMA and TanDEM-X formation flying missions are used in the respective sections to illustrate the current status of carrier phase differential GPS navigation in onboard and ground applications.

5.2 Spaceborne GPS Receivers

5.2.1 SGPS Technology Aspects

While the basic functionality of a spaceborne GPS (SGPS) receiver is the same as that of a terrestrial or aeronautical receiver, its design has to properly account for the high signal dynamics and the more hostile environment [6] in which these receivers are operated. Whereas the signal tracking aspects can largely be handled by suitable adaptations of the core receiver software, the environmental robustness of space equipment is a continued source of concern and dedicated engineering and qualification standards have been established by relevant space agencies [7] and satellite manufacturers. Key issues to be considered in this context include the resistance to thermal-vacuum conditions, vibration and shock loads as well as ionizing radiation and single event effects. Besides a cost driving test and qualification effort that is implied by these standards, suitably qualified electronic components are often less powerful and require higher resources (mass, power) than state-of-the-art consumer electronics. Aside from technical issues, it has to be emphasized that the threat of a possible abuse of spaceborne GPS receivers for military purposes imposes stringent export restrictions under both national and international export control laws [8, 9]. A free trade is only permitted for GPS receivers which inhibit operation above altitudes of 18 km (60,000 ft) and velocities of greater than 515 m/s (1,000 nautical miles per hour), which clearly excludes their use in orbit.

For a continued use in the space, the receiver components (baseband processor, CPU, memory, power conditioning) need to be sufficiently tolerant against the influence of space radiation. Radiation may cause both long-term effects, i.e. a gradual component aging, as well as single event effects (SEE) such as upsets (SEU) or latch-ups (SEL). Long-term effects are typically caused by ionization in the electronics components which is described by the total ionization dose (TID). For use in low Earth orbits (i.e. altitudes of several hundreds of kilometers) a TID of

about 10 krad should be tolerated, but substantially higher limits may apply for long duration missions with stringent reliability requirements. A large TID will, among others, result in increased power consumption and an eventual failure of the device. The sensitivity of individual components or an entire electronics board can be tested through irradiation with a radioactive Co-60 source. Board-level tests for various types of commercial-off-the-shelf (COTS) GPS receivers have shown a representative TID tolerance of 5–15 krad [10], which often justifies their use in low-cost micro- and nano-satellite missions.

Single-event effects are caused by individual high energy particles causing a temporary build-up of free charges in the semi-conductor material. These may cause bit flips and software errors disturbing the proper function of a GPS receiver. More severely, they may also trigger short circuits (through parasitic transistors) resulting in the release of excessive heat and often a permanent destruction of the affected chip area. While total dose effects can be reduced by external shielding, other means of protection are required for single event effects. In a fully space grade design, specific manufacturing processes, such as Silicon-on Insulator (SOI) technology, may be applied to obtain radiation tolerant components. Memory contents may further be protected by an error detection and correction (EDAC) code supported through hard- or software. For COTS GPS receivers that do not incorporate such features, a latch-up protection must be considered as a minimum to avoid the risk of fatal damages. Fundamentally, a latch-up protection is a rapid electronic fuse, which shuts down the receiver power, if a non-nominal over-current is sensed. The design of a proper latch-up protection and its overall feasibility depend critically on the specific design of the employed receiver. Besides the nominal power consumption, the TID-induced current increase, the in-rush current and the energy stored in capacitors have to be accounted for. SEL (and SEU) testing for space electronics can only be performed in large accelerator labs and may comprise both heavy-ion and proton irradiation (see e.g. [11]). So far, board level SEL tests have only been conducted for a very limited number of COTS GPS receivers (such as the DLR Phoenix receiver [12] and the Javad Triumph receiver [13]) to demonstrate the survivability of the respective devices in combination with specific power supply and latch-up protection circuitry.

In addition to radiation, other, but generally less critical, environmental conditions must be taken care of, when selecting a GPS receiver for use in space. Among others, the outgassing of components, vibration stress during launch and the combined effects of temperature and vacuum need to be considered. Temperature ranges of -30 to $+60°$C that might apply in a satellite are generally well compatible with consumer electronic components, but care must be taken in the hardware design to enable adequate heat dissipation in the absence of convection.

Besides hardware related effects of the space environment, the signal tracking and navigation software has to properly account for the specific conditions of an orbiting platform. Even though the signal levels encountered in LEO are almost identical to those of a terrestrial or aeronautical receiver, the constellation of visible satellites changes much more rapidly and the signal is subject to much more pronounced line-of-sight velocities and accelerations. Also, mission specific,

sub-optimal GPS antenna orientations may affect the capability of a spaceborne receiver to properly track or acquire GPS signals. Concerning the signal dynamics, peak Doppler shifts of up to 45 kHz may be encountered by a LEO satellite, which results in representative cold start times of about 10 min for a traditional sequential signal search [14]. Modern receivers make use of dedicated fast acquisition units with massively parallel correlators or Fourier-transform processing [15] to achieve a substantially improved acquisition performance even under high dynamics. As an advanced example, a cold-start time-to-first-fix of 60 s has been obtained for LEO orbits with the Javad Triumph receiver selected for the ACES experiment onboard the ISS [13].

For use in formation flying missions with demanding navigation and control requirements, particular care must be taken to ensure a high overall accuracy of the code and carrier phase measurements provided by the employed GPS receiver. As a rule of thumb, pseudorange measurements in a single-frequency receiver should exhibit an average noise level of 0.5 m or better, while carrier phases should be accurate to about 1 mm. In case of dual-frequency measurements, the ionosphere-combination amplifies errors by a factor of three, which results in somewhat tighter budgets for the individual measurements. To avoid systematic measurement errors, use of a 3rd-order phase-locked-loop (PLL) is recommended for tracking the carrier phase of the C/A-code signal, while aided loops of lower order can typically be used for code tracking and tracking of other signals. Further more, great care is required in the formation of measurements of a spaceborne GPS receiver envisaged for formation flying missions. Given the orbital velocity of a LEO satellite, even a timing inconsistency of only 1 μs causes a 7 mm position error, which may not be tolerable for precise navigation. Such timing offsets have been identified in a variety of geodetic space receivers [16] and a thorough receiver calibration in a signal simulator testbed is therefore strongly advisable.

When forming differential measurements from two spaceborne receivers (Sect. 5.3.1) it is likewise mandatory to synchronize the measurement epochs to roughly 0.1 μs, if mm-level relative navigation accuracies are desired. This is automatically ensured by receivers that steer their measurements to integer seconds of GPS time. Otherwise, the synchronization can in part be achieved by inter- or extrapolation, but even then the physical measurement epochs should not differ by more than 1 ms. Receivers with free-floating measurements epochs require highly stable oscillators and extensive post-processing to achieve synchronized measurements.

Finally, the GPS antenna system and its accommodation on the spacecraft must be designed in accord with mission specific accuracy requirements. In order to fully exploit the high precision of differential carrier phase measurements for relative navigation of satellites, use of a choke ring antenna is mandatory. Otherwise, multipath and nearfield distortions of the antenna phase pattern may cause systematic carrier phase measurement errors of up to several centimeters. Even when using a choke ring antenna, phase center variations of up to 10 mm have been observed in geodetic space missions that call for a dedicated ground and inflight calibration [17, 18].

As concerns the frequency bands and signals required by a GPS receiver for formation flying applications in low Earth orbit, both single- and dual-frequency receivers may be considered and have successfully been applied in actual missions. The choice is ultimately driven by the dimension of the formation, the targeted accuracy and the available (technical and financial) budgets. Evidently, dual-frequency receivers enable a rigorous elimination of ionospheric path delays and allow for a highly accurate (5 cm level) absolute navigation. On the other hand, ionospheric path delays are effectively cancelled when differencing observations over short baselines [19]. Depending on solar activity and accuracy requirements, single-frequency receivers can thus be used for sub-centimeter level relative orbit determination of two spacecraft across distances of up to a few kilometers. With respect to carrier phase ambiguity resolution, dual-frequency receivers offer additional advantages through the availability of two independent carrier phase measurements. This enables, for example, the formulation of a wide-lane combination, which exhibits a large effective wavelength of 86 cm and thus allows for a very simple ambiguity resolution. However, the trade-off between single-and dual-frequency ambiguity resolution in formation flying missions with short and moderate baselines remains a topic of further research and final conclusions may only be possible after thorough analysis of flight data from the recently launched PRISMA and TanDEM-X missions (see Chaps. 21 and 13).

Even though a variety of new GPS signals and new Global Navigation Satellite Systems (GNSS) are presently emerging, their impact on future formation flying missions remains to be assessed. Evidently, the availability of new civil signals (such as GPS L2C, GPS L5, or Galileo E1 and E5) will ultimately replace the need for semi-codeless tracking of the encrypted GPS P(Y)-code signals on the L1 and L2 frequencies and enable measurements with a presumably lower noise level. Also, the use of tri-carrier measurements (GPS L1/L2/L5 or Galileo E1/E6/E5) is considered to enable an improved ambiguity resolution in differential navigation applications by various authors (see, e.g., [20]). It is unclear, though, when and to what extent the potential benefits of such new signals can be materialized and extensive research is pending to ensure a consistent measurement and processing quality across GNSS constellations and signals [21].

5.2.2 Receivers

As illustrated by an annual review of the "GPS World" magazine, roughly 500 different GPS receiver models are presently offered by more than 70 manufacturers world wide [22]. Compared by number of receiver units and total market volume, spaceborne GPS receivers certainly constitute a niche market even though the number of providers appears still large in view of the limited range of potential customers. Following Ref. [23] a total of 480 LEO satellites is expected to be launched between 2007 and 2016, which ultimately places a natural limit of roughly 50 units per year to the demand for space grade GPS receivers.

The small market segment and high specialization of SGPS receivers as well as the associated test and qualification effort inevitably results in high unit cost ranging from roughly 100 k€ to 1 M€. Various companies and research institutes have therefore made efforts to come up with low cost solutions based on the use of commercial-off-the-shelf (COTS) components.

Within Europe, ESA has continuously supported and promoted the independent development of dual-frequency receivers for space applications through the AGGA correlator chip. The Advanced GPS/GLONASS ASIC (AGGA) was initially developed (version 0) by the Institute of Navigation, University of Leeds (UK) from 1993–1995. Funded by ESA initiatives, the chip has evolved to version 2, which is currently used in the GRAS and Lagrange receivers for MetOp and GOCE and will likewise be employed for the upcoming SWARM and Sentinel missions. As a successor to the AGGA-2, the next generation AGGA-4 chip will support new signals (GPS L2C &L5, Galileo O/S), provide a larger number of channels (30–36), and exhibit a built-in 32-bit LEON microprocessor [24].

An overview of relevant GPS receivers currently in use or considered for space missions is provided in Tables 5.1 and 5.2 for single- and dual-frequency GPS receivers, respectively. Both spacegrade receiver designs and COTS receivers are considered. So far, none of the listed receivers can support Galileo tracking, with the exception of the Javad Triumph receiver selected for the ACES payload onboard the ISS.

Out of the listed receivers, the MosaicGNSS, IGOR and Phoenix receivers are presently in use for actual formation flying missions (cf. Fig. 5.2). Other receivers such as the Viceroy and Tensor receivers have been used for relative navigation in short term rendezvous missions to the MIR space station and the ISS.

5.3 GPS Orbit Determination Principles

While each GPS receiver computes its own navigation solution, it is generally desirable in space applications to process GPS raw measurements in a separate navigation filter or orbit determination process to achieve a higher accuracy and robustness. The present section introduces the basic concepts of this processing and describes relevant models and algorithms.

5.3.1 Observation Types and Measurement Models

GPS receivers provide two basic measurement types that can be used for navigation purposes, namely pseudorange and carrier phase measurements [1, 38]. Both of them provide essentially information on the satellite-to-receiver distance but differ in their accuracy and specific processing needs.

Table 5.1 Single-frequency GPS receivers for space applications

Manufacturer (Country)	Receiver	Channels	Ant	Power weight	TID (krad)	Missions, References
EADS Astrium (D)	MosaicGNSS	6–8 C/A	1	10 W, 1 kg	>30	SARLupe, TerraSAR-X [25], TanDEM-X, (Aeolus)
General Dynamics (US)	Viceroy	12 C/A	1–2	4.7 W, 1.2 kg	15	MSTI-3, Seastar, MIR [26], Orbview, Kompsat-1
SSTL (UK)	SGR-07	12 C/A	1	1.6 W, 0.5 kg	>10	Deimos-1, UK-DMC-2
	SGR-10/20	2/4 × 6 C/A	2/4	6.3 W, 1 kg	>10	PROBA-1, UOSat-12, BILSAT, AlSAT-1 [27]
Thales Alenia Space (I/F)	Tensor	9 C/A	1–4	15 W, 4 kg	100	Globalstar, SAC-C, ATV [28]
	TopStar 3000	12–16 C/A	1–4	1.5 W, 1.5 kg	>30	Demeter, Kompsat-2 [29]
DLR (D)	Phoenix-S	12 C/A	1	0.9 W*, 20 g*	15	Proba-2 [14], PRISMA [30], X-Sat (TET, FLP)
NovAtel (CA)	OEMV-1	14 C/A	1	1.0 W*, 25 g*		RAX [31] (CanX-4/5, NEOSSat)

Physical parameters for standalone receiver boards are indicated by an asterisk. Planned missions are given in brackets

Table 5.2 Dual frequency GPS receivers for space applications

Manufacturer (Country)	Receiver	Channels	Ant	Power weight	TID (krad)	Missions, References
JPL (US)/BRE (US)	BlackJack/IGOR	16×3 C/A, P1/2	4	10 W, 3.2/ 4.6 kg	20	GRACE, Jason-1/2, COSMIC, TerraSAR-X [16, 17], TanDEM-X
RUAG (A)	POD Receiver	8×3 C/A, P1/2	1	10 W, 2.8 kg		(SWARM, Sentinel-3) [32]
SAAB (S)	GRAS/GPSOS	12 C/A, P1/2	3	30 W, 30 kg		METOP [33]
Thales Alenia Space (IT,FR)	Lagrange	16×3 C/A, P1/2	1	30 W, 5.2 kg	20	ENEIDE [34], Radarsat-2, Oceansat-2, GOCE
	TopStar 3000 G2	6×2 C/A, L2C	1			PROBA-2 [35]
BRE (US)	Pyxis POD	16–64 C/A, P1/2, L2C, L5	1–2	20 W, 2 kg		Under development
EADS Astrium (D)	LION	32; GPS, GAL	1			Under development
Javad (US/RUS)	Triumph DG3TH	216; GPS, GAL, GLO	1	2.5 W*, 100 g*	10	(ISS/ACES) [13]
NovAtel (CA)	OEM4-G2L	12×2 C/A, P2	1	1.5 W*, 50 g*	6	CanX-2, (CASSIOPE) [36]
Septentrio (B)	PolaR×2	16×3 C/A, P1/2	1(3)	5 W*, 120 g*	9	(TET) [16, 37]

Physical parameters for standalone receiver boards are indicated by an asterisk. Planned missions are given in brackets

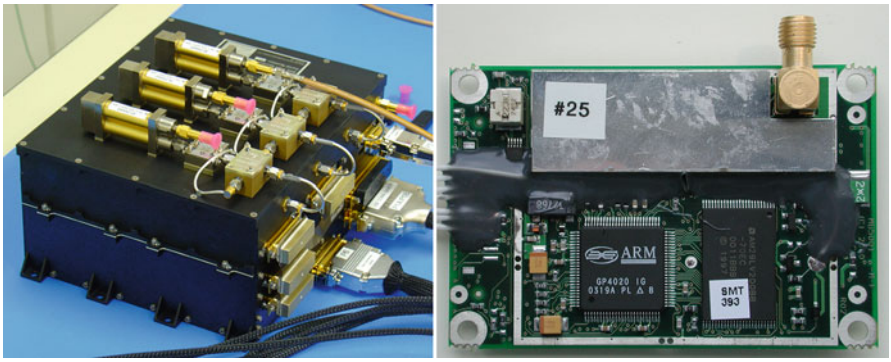


Fig. 5.2 Examples of spaceborne GPS receivers for formation flying missions: IGOR dual-frequency receiver on TerraSAR-X/TanDEM-X (*left*) and Phoenix GPS receiver of PRISMA (*right*)

A pseudorange (or code) measurement is obtained by comparing the reception time of a GPS signal (referred to the local receiver clock) with the transmit time (referred to the GPS satellite clock) and multiplying the result with the speed of light. If all clocks were perfectly aligned with the GPS system time, the pseudorange would be identical to the actual distance of the signals traveled between transmission and reception. In practice the pseudorange differs from the geometric range by additional clock offset contributions. In addition, the signal propagation time is affected by atmospheric delays, whenever the signal traverses the ionosphere or troposphere. For orbiting spacecraft, the latter contribution is of no relevance and the pseudorange measurement ρ can thus be modeled as

$$\rho = |\mathbf{r} - \mathbf{r}_{\mathrm{GPS}}| + c\delta t - c\delta t_{\mathrm{GPS}} + I \tag{5.1}$$

Here $\boldsymbol{r}$ and $\boldsymbol{r}_{\mathrm{GPS}}$ denote the position of the user and GPS satellite antenna, δt and δt_{GPS} are the offsets of the receiver and satellite clocks from the GPS system time and I denotes the ionospheric path delay. For the ease of presentation, measurement errors such as multipath and receiver noise are ignored in this equation.

Carrier phase measurements are obtained by mixing the incoming signal with a carrier replica on the nominal signal frequency and measuring the phase of the resulting beat signal. Any variation in the distance between the GPS satellite and the receiver causes a Doppler shift of the received signal and likewise of the beat frequency. The carrier phase (or integrated Doppler) therefore constitutes a measure of the distance change over time (but not the absolute distance). For convenience, the sign of the carrier phase measurement is chosen such as to match the variation of the pseudorange. Depending on the application, the carrier phase measurement may either be expressed in range units (φ) or cycles (Φ). Both representations differ by a wavelength factor λ and can be modelled as

$$\varphi = \lambda\Phi = |\mathbf{r} - \mathbf{r}_{\mathrm{GPS}}| + c\delta t - c\delta t_{\mathrm{GPS}} - I + \lambda A + \lambda\Psi \tag{5.2}$$

where A is a (dimensionless) phase ambiguity that depends on the initial value assigned to the carrier phase measurement at start of tracking. The last term, finally describes the so-called phase wind-up that describes the change in the observed carrier phase due to a time varying rotation angle Ψ of the transmit and receiver antenna boresights with respect to the line of sight vector [39]. Similar to the pseudorange, the carrier phase model includes two clock offset terms since the measured phase varies with the integrated frequency error of the GPS satellite and receiver clocks. Comparing (5.1) and (5.2), it may further be noted that the propagation through the ionosphere results in a carrier phase change that is of equal size but different sign as compared to the ionospheric path delay affecting the pseudorange measurement. This difference is caused by the fact that the ionosphere is an electrically charges plasma, where the group and phase velocity of a modulated signal are affected in opposite manners [38]. Since the ionospheric delay varies with the inverse square of the signal frequency f, it can be rigorously eliminated by forming the ionosphere-free linear combination

$$\rho_{\mathrm{IF}} = \frac{f_{\mathrm{L1}}^2}{f_{\mathrm{L1}}^2 - f_{\mathrm{L2}}^2} \cdot \rho_{\mathrm{L1}} - \frac{f_{\mathrm{L2}}^2}{f_{\mathrm{L1}}^2 - f_{\mathrm{L2}}^2} \cdot \rho_{\mathrm{L2}} \tag{5.3}$$

of pseudorange measurements (or likewise carrier phases) on the L1 and L2 frequency. For single-frequency receivers, a ionosphere-free linear combination can be obtained from the arithmetic mean of the pseudorange and carrier phase

$$\rho_{\mathrm{GRAPHIC}} = \frac{1}{2}(\rho + \varphi), \tag{5.4}$$

which is known as GRAPHIC (GRoup and PHase Ionospheric Correction) measurement [40].

A course model of ionospheric path delays for orbital vehicles has been proposed by Lear [41]. Ignoring regional variations of the vertical total electron content (VTEC), the delay of an individual pseudorange measurement may be expressed as the product $I = I_0 \cdot m(E)$ of the vertical path delay

$$I_0 = 0.162\ \mathrm{m} \cdot \left(f_{\mathrm{L1}}^2 / f^2\right) \cdot \left(\mathrm{VTEC} / 10^{16} \mathrm{e}^- / \mathrm{m}^2\right) \tag{5.5}$$

and a mapping function $m(E)$

$$m(E) = \frac{2.037}{\sqrt{\sin^2 E + 0.076} + \sin E} \tag{5.6}$$

describing the path delay variation with elevation E. The model is frequently employed in simulations but can also be employed for coarse measurement corrections in GPS based orbit determination or relative navigation [19, 42].

While orbit determination of a single spacecraft may be affected by numerous error sources (GPS orbit and clock errors, unmodelled atmospheric delays, antenna phase center variations and multipath) relative navigation over sufficiently short baselines usually benefits from a high level of common error cancellation. This is most obvious when working with receiver-receiver differenced observations. These can be modeled as

$$\begin{aligned} \Delta\rho_{AB} &= \rho_B - \rho_A = \Delta|\mathbf{r} - \mathbf{r}_{\mathrm{GPS}}|_{AB} + c\Delta\delta t_{AB} + \Delta I_{AB} \\ \Delta\varphi_{AB} &= \varphi_B - \varphi_A = \Delta|\mathbf{r} - \mathbf{r}_{\mathrm{GPS}}|_{AB} + c\Delta\delta t_{AB} - \Delta I_{AB} + \lambda\Delta A_{AB} + \lambda\Delta\Psi_{AB} \end{aligned} , \tag{5.7}$$

where $\Delta(\bullet)_{AB} = (\bullet)_B - (\bullet)_A$ denotes the single difference (SD) of quantities for spacecraft B and A, respectively. While the satellite clock offset contribution cancels immediately, the impact of GPS satellite ephemeris errors is strongly suppressed for short separations. Similar considerations hold for the differential ionospheric path delays that can typically be neglected for separations of 1 km or less. The differential phase wind-up effect vanishes on short baselines, if both receiver antennas maintain the same attitude with respect to the transmitting GPS satellite, but will have to be considered for spacecraft with notably different antenna orientations [43]. However, this would likewise imply a restricted number of commonly tracked GPS satellites and is therefore avoided in missions with stringent relative navigation requirements.

Beyond the receiver-receiver difference, a satellite-satellite difference $\nabla(\bullet)^{ij} = \nabla(\bullet)^j - \nabla(\bullet)^i$ may be formed to further eliminate the satellite clock offset. Ignoring the differential ionospheric path delays and phase wind-up, the resulting double difference (DD) depends primarily on the relative geometry of the GPS satellites and the two receivers:

$$\begin{aligned} \nabla\Delta\rho^{ij}_{AB} &= \Delta|\mathbf{r} - \mathbf{r}_{\mathrm{GPS}}|^{ij}_{AB} \\ \nabla\Delta\varphi^{ij}_{AB} &= \Delta|\mathbf{r} - \mathbf{r}_{\mathrm{GPS}}|^{ij}_{AB} + \lambda\nabla\Delta A^{ij}_{AB} \end{aligned} \tag{5.8}$$

In case of carrier phase measurements, the observation model always involves the phase ambiguities, which are not known beforehand but have to be determined as part of the data processing. While the ambiguities are float values quantities, in general, their double difference can be shown to be of integer nature for a properly designed GPS receiver [43]. Once the true integer value $\nabla\Delta A^{ij}_{AB} = N^{ij}_{AB}$ of the DD ambiguity is known, the carrier phase measurements are effectively translated into pseudoranges with mm noise level and can thus be used to obtain high precision differential positions.

Conceptually, the integer ambiguities can indeed be obtained from the comparison of DD code and phase measurements (cf. (5.8)) and rounding to the nearest integer:

$$N^{ij}_{AB} = \left[\frac{1}{\lambda}\left(\nabla\Delta\varphi^{ij}_{AB} - \nabla\Delta\rho^{ij}_{AB}\right)\right]_{\mathrm{round}} \tag{5.9}$$

However, more advanced methods are required in practice for a robust ambiguity fixing, particularly if large baselines result in non-negligible ionospheric path delays. These include the use of special linear combinations as well as general search techniques [38, 44].

With dual-frequency measurements, a wide-lane combination

$$\varphi_{\mathrm{WL}} = \frac{f_{\mathrm{L1}}}{f_{\mathrm{L1}} - f_{\mathrm{L2}}} \cdot \varphi_{\mathrm{L1}} - \frac{f_{\mathrm{L2}}}{f_{\mathrm{L1}} - f_{\mathrm{L2}}} \cdot \varphi_{\mathrm{L2}} \tag{5.10}$$

can be formed from the L1 and L2 carrier phases. Due to the larger effective wavelength of $\lambda_{\mathrm{WL}} = c/(f_{\mathrm{L1}} - f_{\mathrm{L2}}) \approx 0.86$ m, the determination of the wide-lane ambiguity $N_{\mathrm{WL}} = N_{\mathrm{L1}} - N_{\mathrm{L2}}$ using the integer rounding (5.9) is less sensitive to pseudorange noise than that of the L1 and L2 measurements alone. Once the wide-lane ambiguity is known, the associated narrow-lane ambiguity $N_{\mathrm{NL}} = N_{\mathrm{L1}} + N_{\mathrm{L2}}$ and, finally, the individual L1 and L2 carrier phase ambiguities can be determined using only carrier phase measurements. With the upcoming availability of three signal frequencies in modernized GPS and the new Galileo system, the wide-laning technique may be extended by forming extra-widelane combinations (from L2 and L5 or E6 and E5 measurements) with an even longer effective wavelength.

In another class of ambiguity resolution method, the double-difference ambiguities are considered as unknowns in a adjustment problem for the relative antenna positions at a larger set of measurement epochs. After linearization, a standard least squares technique can be used to find the relative position and the float-valued DD ambiguities that minimize the residuals between the observations and the modeled measurements. However, what is actually desired is a solution, in which the permitted ambiguities are constrained to integer values. Different search techniques have been developed for this type of mixed real/integer problems, among which the Least-squares Ambiguity Decorrelation Adjustment (LAMBDA; [45]) is probably the most popular and widely applied. It is based on an integer-preserving transformation of the search space. This enables an efficient determination of the optimum ambiguities in the presence of pronounced correlations, which are naturally encountered in the dual-frequency processing. Furthermore, the LAMBDA method provides the algorithmic framework for validating the estimated ambiguities.

In the context of spacecraft formation flying the LAMBDA method and related variants have been adopted by various authors both for simulation studies [46, 47] and actual flight data processing [48, 49]. A potential limitation in its application results from the large number of ambiguity parameters (typically 1,000 per 24 h arc) that need to be handled in a batch orbit determination for a satellite in low Earth orbit and may result in excessive computing times. Traditional wide-lane/narrow-lane ambiguity resolution techniques have therefore been preferred by other researchers [50, 51] and are considered for on-board real-time applications [52]. Methods for bounding the number of past integer ambiguity states in a Kalman filter/smoother for spacecraft relative navigation have, furthermore, been studied by M.L. Psiaki [53] in an effort to limit the computational effort while retaining a near-optimal estimator.

5.3.2 Dynamical Models

For close formations in near-circular orbits around the Earth, the relative motion of two spacecraft can conveniently be described by the well known Clohessy-Wilthshire (CW) equations [54]. In the absence of perturbations, the relative motion decomposes into a (periodic plus linear) in-plane motion and a harmonic oscillation perpendicular to the orbital plane [55]. As an alternative to a purely Cartesian formulation, relative orbital elements can be employed to describe the solution of the CW equations, which facilitates the incorporation of Earth oblateness perturbations and extends the applicability of the relative motion model for larger along-track separations [56, 57]. To overcome the constraining simplifications of the unperturbed CW equations, considerable effort has been made to develop more sophisticated relative motion models for eccentric orbits [58, 59] as well as different types of orbital perturbations [60–63]. While the resulting analytical orbit models provide a high level of insight into the basic properties of relative motion and form an excellent basis for formation design and controller design, more rigorous, numerical models are typically desired in GPS-based relative navigation to comply with the high measurement accuracy.

It is therefore common to revert to a separate treatment of the individual satellite trajectories using accurate models of the relevant accelerations $a(t,\mathbf{r},\mathbf{v})$ as a function of time t, position $\mathbf{r}$ and velocity $\mathbf{v}$ rather than expanding the differential acceleration in terms of the relative separation. The resulting relative motion can then be obtained by differencing the trajectories obtained from the numerical integration of the equation of motion for each individual satellite in the formation:

$$\begin{pmatrix} \Delta\mathbf{r}_{AB} \\ \Delta\mathbf{v}_{AB} \end{pmatrix} = \begin{pmatrix} \Delta\mathbf{r}_{AB} \\ \Delta\mathbf{v}_{AB} \end{pmatrix}_0 + \int_{t_0}^{t} \begin{pmatrix} \mathbf{v}_B \\ a_B(t',\mathbf{r}_B,\mathbf{v}_B) \end{pmatrix} dt' - \int_{t_0}^{t} \begin{pmatrix} \mathbf{v}_A \\ a_A(t',\mathbf{r}_A,\mathbf{v}_A) \end{pmatrix} dt' \tag{5.11}$$

Alternatively, the difference of the individual accelerations can be formed and integrated along with the equation of motion of a reference satellite:

$$\begin{pmatrix} \mathbf{r}_A \\ \mathbf{v}_B \\ \Delta\mathbf{r}_{AB} \\ \Delta\mathbf{v}_{AB} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_A \\ \mathbf{v}_B \\ \Delta\mathbf{r}_{AB} \\ \Delta\mathbf{v}_{AB} \end{pmatrix}_0 + \int_{t_0}^{t} \begin{pmatrix} \mathbf{v}_A \\ a_A(t',\mathbf{r}_A,\mathbf{v}_A) \\ \Delta\mathbf{v}_{AB} \\ a_B(t',\mathbf{r}_B+\Delta\mathbf{r}_{AB},\mathbf{v}_B+\Delta\mathbf{v}_B) - a_A(t',\mathbf{r}_A,\mathbf{v}_A) \end{pmatrix} dt' \tag{5.12}$$

The motion of each individual satellite is governed by various types of gravitational and non-gravitational forces that need to be accounted in the numerical

trajectory modeling. Key contributions (in a rough order of importance) include the aspherical gravitational field of the Earth, air drag, luni-solar gravitational perturbations, solar radiation pressure, solid Earth and ocean tides, relativistic perturbations, and Earth albedo. Relevant models are well documented in the astrodynamics literature and the reader is referred to relevant textbooks and articles (see Refs. [55, 64] and references therein) for detailed information.

Gravitational forces can, in general, be described by high-fidelity physical models with known parameters. The choice of models in a specific application is therefore solely based on a trade-off between accuracy requirements and computational complexity. While the full range of perturbations is considered in GPS precise orbit determination tools [65–68], notable simplifications are commonly made in real-time, onboard navigation systems [69, 70]. Non-gravitational force models, in contrast, suffer to a large extent from a limited knowledge of physical conditions (such as air density [71] or reflected Earth radiation) and satellite surface/body properties (drag coefficients, reflectivity and absorption, etc.).

To achieve a high overall navigation accuracy despite force modeling limitations and to fully exploit the high geometric strength of GPS measurements, the concept of reduced dynamics orbit determination has been introduced in Ref. [72]. Here, potential deficiencies in the a priori dynamical model are compensated by empirical accelerations that are estimated along with other adjustment parameters. In view of a more direct physical interpretation, a frame aligned with the radial (R), along-track (T) and normal (N) direction is commonly used to describe these empirical accelerations. For GPS-based precise orbit determination of scientific satellites in low Earth orbit the adjustment of piece-wise constant RTN accelerations at intervals of 6–15 min has been adopted in many least-squares batch orbit determination systems [67, 73]. For sequential estimation, in contrast, the empirical accelerations are best treated as exponentially correlated random variables [67, 70, 72].

Orbital maneuvers represent a special case of non-gravitational "perturbation" that needs to be taken care of within the orbit determination process. Similar to the empirical accelerations, the RTN-frame represents a natural choice for the description of in-plane or out-of-plane maneuvers. For a generic model, a constant thrust level can be assumed, even though actual thruster systems may employ a pulse width modulation and varying net thrust across the total burn duration. While orbital maneuvers are rarely exercised in single satellite missions (once per week or month), the need for tight formation keeping may result in a much higher maneuver frequency (from several burns per orbit to quasi-continuous thrusting). Even though some a priori information on the actual thrust or velocity increment is generally available (e.g., from the maneuver planning or onboard accelerometers) the remaining uncertainties need to be covered as part of the navigation process. This can be accomplished by a suitable increase of process noise in a sequential navigation process or by a rigorous incorporation of maneuver states into the estimation parameter vector [74–77].

5.3.3 Numerical Integration

A wide range of numerical integration methods that can be used for satellite orbit modeling is known in the literature (see Refs. [64, 78] and references therein) and each method exhibits different advantages and disadvantages. As such, there is no unique method that is ideal for all types of problems and the choice is largely dependent on the requirements of a specific application.

As a rule of thumb, higher order methods provide the best efficiency (in terms of computational cost versus accuracy) when integrating the orbital motion of a satellite over large arcs without intermediate state updates or force model discontinuities (for example maneuvers). Multi-step methods provide the added advantage that interpolated values of the trajectory can easily be obtained from pre-computed data, which allows a decoupling of the integrator step size from the desired output step size. A general purpose Adams-Bashforth-Moulton method that supports variable order and stepsize and is well suitable for a wide range of orbital mechanics problems has been developed by Shampine and Gordon [64, 79]. The method is used for GPS-based precise orbit determination in the GHOST software [80], but a variety of alternative multi-step methods or collation methods has been adopted with great success in other software packages [65, 66, 68].

Low order methods, in contrast are preferable if the required time step is only a small fraction of the orbital period. This situation is typically encountered in real-time navigation systems where filter state updates are performed at time intervals in the 1–30 s range. Following a measurement update, the trajectory integration has to be restarted with a new state vector, which becomes highly inefficient for a high order integration method. In this case, the classical fourth-order Runge–Kutta (RK4) method represents a well justified choice, which combines the ease of use with reasonable short term accuracy. Given a differential equation $\mathbf{y} = \mathbf{f}(t, \mathbf{y}(t))$ for the state vector

$$\mathbf{y}(t) = \begin{pmatrix} \mathbf{r}(t) \\ \mathbf{v}(t) \end{pmatrix} \tag{5.13}$$

and the initial state at time t, the solution at time step $t + h$ is obtained from the approximation

$$\mathbf{y}(t+h) \approx \eta(t+h) = \mathbf{y}(t) + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{5.14}$$

with

$$\begin{aligned} \mathbf{k}_1 &= \mathbf{f}(t, \mathbf{y}(t)) \\ \mathbf{k}_2 &= \mathbf{f}(t + h/2, \mathbf{y}(t) + h\mathbf{k}_1/2) \\ \mathbf{k}_3 &= \mathbf{f}(t + h/2, \mathbf{y}(t) + h\mathbf{k}_2/2) \\ \mathbf{k}_4 &= \mathbf{f}(t + h, \mathbf{y}(t) + h\mathbf{k}_3). \end{aligned} \tag{5.15}$$

An alternative approach, which has originally been proposed in Ref. [81], combines the basic RK4 method with the concept of Richardson extrapolation to construct a fifth-order integration method with a lower number of derivative computations than would be required by any existing fifth-order RK method. Given the initial value $\boldsymbol{y}(t)$, approximate solutions $\boldsymbol{\eta}_1^h$ and $\boldsymbol{\eta}_2^h$ at $t + h$ and $t + 2h$ are obtained from two consecutive RK4 steps of step size h. In addition a single RK4 step of size $2h$ is performed that delivers an independent approximation $\boldsymbol{\eta}_2^{2h}$. Knowing that the error of the RK4 method grows with the fourth power of the step size, the difference of the two solutions at $t + 2h$ can be used to extrapolate the solution to a fictitious step size of zero. This yields a fifth-order approximation

$$\mathbf{y}(t+2h) \approx \boldsymbol{\eta}_2^h + \frac{\boldsymbol{\eta}_2^h - \boldsymbol{\eta}_2^{2h}}{2^4 - 1} \tag{5.16}$$

of the solution at $t + 2h$. Combining state vectors $\mathbf{y}_i = \mathbf{y}(t + ih)$ and derivatives $\boldsymbol{f}_i = \mathbf{f}(\mathbf{y}_i)$ from the individual integration steps, a quintic Hermite polynomial

$$\begin{aligned}\mathbf{y}(t+\theta h) = {} & d_0(\theta)\mathbf{y}_0 + d_1(\theta)h\mathbf{f}_0 + d_2(\theta)\mathbf{y}_1 \\ & + d_3(\theta)h\mathbf{f}_1 + d_4(\theta)\mathbf{y}_2 + d_5(\theta)h\mathbf{f}_2\end{aligned} \tag{5.17}$$

with coefficients

$$\begin{aligned} d_0 &= \frac{1}{4}(\theta-1)^2(\theta-2)^2(1+3\theta) & d_3 &= (\theta-1)\theta^2(\theta-2)^2 \\ d_1 &= \frac{1}{4}\theta(\theta-1)^2(\theta-2)^2 & d_4 &= \frac{1}{4}\theta^2(\theta-1)^2(7-3\theta) \\ d_2 &= \theta^2(\theta-2)^2 & d_5 &= \frac{1}{4}(\theta-2)\theta^2(\theta-1)^2 \end{aligned} \tag{5.18}$$

can be constructed that provides accurate interpolation over the entire time interval $[t,t + 2h]$. The method has successfully been applied in various real-time navigation systems for single and multiple satellites [76, 77, 82, 83], where it enables dense trajectory output without constraining the integrator and filter update interval.

5.3.4 Estimation

The combination of the numerical trajectory model and the GPS measurement model enables a prediction of the measurements for a given set of initial conditions and model parameters. By comparing the difference of the actual observation and the modeled measurements, corrections to the assumed parameters may be inferred to obtain an optimal estimate of the spacecraft position and related parameters.

A wide range of estimation methods has been developed that may be divided into batch and sequential estimators. By their very nature, batch least-squares estimation methods are primarily employed in post-processing, where data collected over a larger time span can be jointly processed to find the trajectory that best fits the entire set of measurements. Recursive estimation methods, in contrast, are favored in real-time applications, where they can provide new estimates of the instantaneous state vector at each measurement epoch.

The batch least-squares orbit determination using GPS observations is typically characterized by a large number of estimation parameters. The need to adjust epoch wise clock offsets results already in about 3,000 unknowns per spacecraft when processing a 24 h data arc at a 30 s sampling interval. Empirical accelerations add another 500 parameters per spacecraft considered in the orbit determination while the number of ambiguity parameters may range from a minimum of 500 to more than 1,000 (depending on the number of frequencies processed and the amount of phase discontinuities encountered in the data). Fortunately, the dimension of the resulting normal equations can be notably reduced by a pre-elimination of the clock parameters. To this end, the normal equations are partitioned into clock offset parameters and non-clock parameters (i.e., dynamical and ambiguity parameters). The sub-block of the normal matrix related to clock offset parameters is a pure diagonal matrix and can therefore be directly inverted and allows a formal elimination of these parameters from the normal equations (see, e.g., Ref. [84]). The reduced system then has a much smaller dimension and can thus be solved with reasonable computational effort. For precise relative navigation (or "baseline determination"), batch estimation strategies have so far been adopted by different researchers using the Bernese [50], EPOS-OC [85] and ZOOM [51] software packages. Here, trajectory parameters and float valued ambiguities for the ionosphere-free L1/L2 carrier phase combination of each individual spacecraft in the formation are first determined similar to a single satellite precise orbit determination. Thereafter double-difference ambiguities between GPS satellites and the formation flying spacecraft are resolved using a wide-lane/narrow-lane technique. Fixed ambiguities are then treated as additional constraints in a final orbit determination step that re-adjusts the (relative) trajectory and any unresolved ambiguity parameters.

Sequential estimation methods are a primary choice for onboard navigation systems requiring continuous information on the instantaneous formation geometry. Nevertheless, the concept has also been adopted for offline processing in Refs. [48, 49], since it reduces the number of ambiguity parameters that need to be simultaneously adjusted and fixed to the number of currently tracked channels. In the sequel, we briefly review the concept of the extended Kalman filter (EKF, [86]), which is the best known and most widely applied sequential estimation method, before discussing the different choices for the set of estimation parameters in relative navigation applications. For completeness, we mention that alternative filter concepts such as the sigma point or unscented Kalman filter (UKF) have received increasing attention in recent years [87–89]. They promise a higher robustness and improved convergence in case of non-linearities compared to the

EKF. However, practical experience with actual flight data shows that the EKF design is fully suitable and sufficient for GPS-based relative navigation of LEO satellites when working with reasonably accurate dynamical models and dense GPS measurements.

The extended Kalman filter is a recursive estimation method made up of a time-update step and a measurement update step. It starts with an initial value of the estimation state vector $\mathbf{y}_i^+$ and the associated covariance $\mathbf{P}_i^+$ at epoch t_i. The individual components of the state vector, which comprises the spacecraft position and velocity along with various other adjustable force and measurements model parameters, is then propagated to the epoch t_{i+1} of the subsequent measurement. The new position and velocity are obtained from a numerical integration of the equation of motion, while more simple models (assuming constancy or exponential damping) are used to propagate force model coefficients, clock parameters and ambiguities. Along with the propagated estimation state $\mathbf{y}_{i+1}^-$, the state transition matrix $\mathbf{\Phi}_{i+1} = \partial \mathbf{y}_{i+1}/\partial \mathbf{y}_i$ is computed, from which the covariance

$$\boldsymbol{P}_{i+1}^- = \mathbf{\Phi}_{i+1} \boldsymbol{P}_{i+1}^+ \mathbf{\Phi}_{i+1}^T \tag{5.19}$$

of the propagated state can then be obtained. Here, the superscript "−" is used to indicate quantities based on the propagation, while "+" denotes improved estimates obtained from the measurements processing. Statistical uncertainties in the state propagation can, furthermore, be taken into account by adding a process noise matrix $\mathbf{Q}_{i+1}$ on the right hand side of (5.19). This completes the time-update step.

Within the measurement-update of the Kalman filter, a refined estimate

$$\mathbf{y}_{i+1}^+ = \mathbf{y}_{i+1}^- + \boldsymbol{K}_{i+1} \cdot (\mathbf{z}_{i+1} - \boldsymbol{h}(\mathbf{y}_{i+1}^-)) \tag{5.20}$$

of the estimation state at the new epoch is computed from the residuals, i.e., the difference of actual observations $\mathbf{z}_{i+1}$ and modeled measurements $\mathbf{h}(\mathbf{y}_{i+1}^-)$. The Kalman gain

$$\boldsymbol{K}_{i+1} = \boldsymbol{P}_{i+1}^- \boldsymbol{H}_{i+1}^T \cdot \left(\mathrm{Cov}(\mathbf{z}_{i+1}) - \boldsymbol{H}_{i+1} \boldsymbol{P}_{i+1}^- \boldsymbol{H}_{i+1}^T\right)^{-1} \tag{5.21}$$

with $\mathbf{H}_{i+1} = \partial \mathbf{h}(\mathbf{y})/\partial \mathbf{y}|_{i+1}$ maps the residuals from the measurement space into the state vector space, taking into account the covariance of the measurements and the propagated state. The blending of information from the predicted state and the measurements, furthermore results in a decreased covariance

$$\boldsymbol{P}_{i+1}^+ = (1 - \boldsymbol{K}_{i+1} \boldsymbol{H}_{i+1}) \cdot \boldsymbol{P}_{i+1}^- \tag{5.22}$$

of the state vector following the measurement-update. Subsequently, the Kalman filter proceeds recursively with the next time- and measurement-updates.

The equations given above describe only the most elementary form of the EKF and some aspects deserve further mentioning. First of all, the covariance update equation

(5.22) is prone to numerical errors and numerically stable variants of the filter equations (such as UD factorization [90]) may be required in case of limited arithmetic precision. Secondly, the "tuning" of the filter, i.e. the optimal choice of measurement weights, a priori covariances and process noise, may pose considerable challenges when working with actual flight data. While software simulations or hardware-in-the-loop simulations with known truth states can provide a good starting point, a need to re-adjust filter settings is typically encountered in real-world applications. To increase the robustness of the filter in the presence of environmental uncertainties the use of an adaptive filter has been proposed and successfully demonstrated for relative navigation of formation flying satellites by F.D. Busse [91]. A final remark concerns the use of the extended Kalman filter in offline processing for precise relative navigation (see Refs. [48, 49]). Other than in a batch least-squares processing, the state estimate in the EKF is only based on past observations when processing in the order of increasing time. In the offline navigation, the filter may instead be used to process the data in both a forward and backwards direction and a smoothed solution can then be obtained from the weighted average at each epoch [84]. The filter/smoother solution is typically more accurate and robust than the individual solutions and does not suffer from degradation in the start-up phase.

Besides the use of a specific filter concept, the set of estimation parameters considered in the filter remains to be decided in the implementation of a navigation system. A wide range of different options has been considered by different authors and the following discussion aims to highlight a few aspects of general interest. Fundamentally, the filter may comprise the following types of parameters:

- the position and velocity ($\mathbf{y} = (\mathbf{r},\mathbf{v})$) of the spacecraft under consideration;
- parameters related to the dynamical motion model, such as drag (C_D) and radiation pressure coefficients (C_R), empirical accelerations ($\mathbf{a}_{emp}$), or maneuvers ($\Delta\mathbf{v}$);
- the receiver clock offset ($c\delta t$);
- ambiguity parameters (A) for single or dual frequency carrier phase measurements;
- ionospheric parameters such as channel-wise path delays (I) (for dual-frequency processing) or a common vertical path delay (I_0) (for single-frequency processing; cf. (5.5)–(5.6)).

Most of these parameters can either be handled as absolute quantities (i.e., referring to a single spacecraft) or relative quantities (i.e., describing the difference of the respective parameter for one s/c relative to a reference spacecraft in the formation) and both representations can be rigorously transformed into each other. As mentioned in Sect. 5.3.2, the lack of suitably accurate relative motion models requires the use of independent trajectory propagation for each spacecraft. Irrespective of the dynamical modeling, it can, nevertheless, be beneficial to consider only the relative position and velocity in the filter to keep the number of state parameters at a minimum. Such "relative-only" filters have, for example been proposed in Refs. [91–93] and [48], while a fully symmetric handling using absolute state vectors for all spacecraft in the formation is preferred in

Refs. [76, 77, 94] and [49]. Both approaches are perfectly legitimate and yield consistent results, if a priori information on the relative dynamics is incorporated in a proper and comparable manner. As an example, the relative empirical accelerations can be constrained more tightly in a formation involving nearly identical spacecraft than the empirical acceleration for the absolute motion of each individual spacecraft. Similar considerations apply for constraints on ionospheric path delays in a close formation, which favor a parameterization in terms of relative quantities. Clock offset parameters, on the other hand, are entirely uncorrelated for different receivers and a relative formulation provides no advantages over the estimation of individual biases.

Altogether, a minimum of about 18 parameters [92], to a maximum of about 80 parameters [49] has to be considered in the various filter concepts assuming a representative number of 12 tracking channels per receiver. Since each filter step involves extensive vector–matrix operations that scale with the dimension of the filter state, the computational effort for the measurement update may well exceed that of the trajectory propagation and time update phase. Specific care in the choice of an optimum parameter set is therefore required for onboard navigation systems with tight real-time requirements. Evidently there is no unique choice that can be recommended for all types of applications and a trade-off between accuracy, robustness and computational load will have to be made in each individual case.

5.4 Mission Results

5.4.1 Rendezvous Missions

Relative navigation using differential GPS has first been applied in the context of various rendezvous missions and precursor experiments.

In Sep. 1995, a first relative navigation experiment was conducted on the STS-69 flight of the Space Shuttle. It made use of a Rockwell-Collins 3M GPS receiver on the Shuttle and a Turbo-Rogue receiver of JPL that was installed on the Wake Shield Facility (WSF) free-flyer. While the envisaged real-time navigation demonstration could not be performed as planned, the data were re-processed offline and used to demonstrate the feasibility for relative navigation at the 10–100 m level using as little as four pseudorange measurements [95].

As part of the ATV Rendezvous Precursor (ARP) project for the European Automated Transfer Vehicle (ATV), a total of three flight experiments were conducted between 1996 and 1998. Flight Demonstration 1 (FD1) was performed in Nov./Dec. 1996 during the STS-80 flight of the Space Shuttle. It made use of a 9-channel Laben Tensor GPS receiver on the free-flying Shuttle PAllet Satellite (SPAS) and a 6-channel Tans Quadrex receiver on the Shuttle. GPS data were recorded for ground analysis but suffered from irregular sampling that inhibited a direct double-differencing [96]. Following Ref. [97], an 8 m position accuracy and

a 3 cm/s velocity accuracy were achieved in a relative navigation filter processing pseudoranges along with a Clohessy-Wiltshire orbit model. Two further experiments (FD2, FD3) were conducted as part of the STS-84 (May 1997) and STS-86 (Sep.–Oct. 1997) flights of the Shuttle to the Russian MIR space station. Data from the Laben receiver on the Shuttle and the Motorola Viceroy receiver [26] of the MOMSNAV payload on MIR were recorded onboard and later processed on ground [98].

Based on the experience gained in the ARP program, a GPS based real-time relative navigation system was implemented for the autonomous approach of the ATV to the International Space Station (ISS). The system is intended for approach navigation in the 30–0.3 km distance range [28] and received its flight qualification on the Jules Verne maiden flight of the ATV in April 2008. It processes data from the 9-channel Laben Tensor GPS receiver on the ATV and the Russian 12-channel ASN-M receiver onboard the ISS in a relative navigation filter (RGPS) based on the Clohessy-Wiltshire dynamical model. Besides pseudorange or code phase data the filter makes use of Doppler or carrier phase measurements to improve the velocity knowledge. Compared to the Flight Control Monitor (FCM) which employs a filtering of the GPS receiver positioning fixes, the RGPS module of ATV achieved accuracies in the 10 m and 2 cm/s range [99].

Already 10 years before the ATV, a very first in-flight demonstration of GPS guided autonomous approach was successfully conducted in Aug. 1998 as part of the Japanese ETS-VII mission [100]. The 2.5 t chaser spacecraft and the 0.4 t target satellite were each equipped with a 6-channel GPS receiver that offered a pseudorange accuracy of about 7 m and delta range accuracy of 1.5 cm (rms). Based on a filtering of differential pseudoranges and carrier phases a relative position and velocity accuracy of 10 m and 3 cm/s, respectively, was obtained in comparison to a post-processed trajectory. This is well within the specified limits of 21 m and 5 cm/s established for the far range approach down to a 150 m distance. Further improvements down to 5 m and 1 cm/s were later achieved offline by incorporation of carrier phase measurements [101].

While all of the above results stay well below the accuracy potential of carrier phase based differential GPS navigation, they represent important milestones on the way to more advanced systems and important achievements based on the technology available at that time.

5.4.2 Precise Baseline Determination

A precise, a posteriori reconstruction of the relative motion of two spacecraft is required in many formation flying missions to support the scientific mission goals or to assess the real-time navigation and control performance. The most stringent accuracy requirements in this context have so far been formulated for SAR interferometry missions such as TanDEM-X (see Chap. 13). A primary goal of TanDEM-X consists in the generation of a global digital elevation model

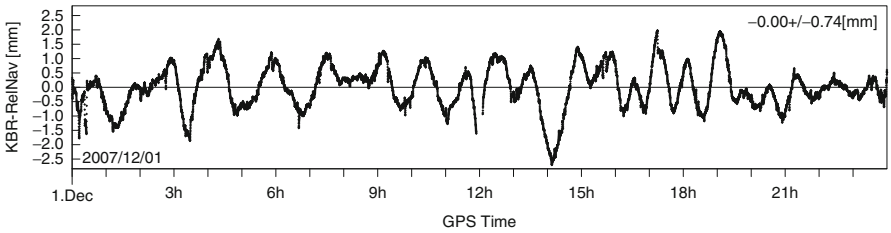


Fig. 5.3 Comparison of GPS-derived distance of the GRACE satellites with K-band ranging measurements for a sample day in December 2007 based on processing with the FRNS baseline processing module of the GHOST software (see Ref. [84])

(DEM) using cross-track interferometry [102]. Here, complex-valued SAR images are simultaneously taken by two spacecraft and combined into an interferogram from which the terrain height can be derived. In this process, the line-of-sight component of the relative position (or "baseline") of the two SAR antennas must be known with utmost accuracy. At an X-band wavelength of only 3 cm and a representative SAR incidence angle of about 30°, a line-of-sight baseline error of 1 mm will result in vertical and horizontal DEM shifts of 1 and 2 m, respectively [103].

The feasibility of precise baseline reconstruction based on GPS measurements has first been demonstrated within the GRACE mission (see Chap. 19). Orbiting the Earth at a 450 km altitude and a mutual separation of about 200 km, the two GRACE satellites form a large gradiometer in space [104]. By measuring distance variations between the satellites with a precision in the 10 μm range, the mission enables a detailed study of the Earth's gravity field and its variation in time. Even though the primary instrument on GRACE is a K-band ranging system, the satellites make use of high-grade dual-frequency GPS receivers for precise orbit determination and timing [105]. As part of the mission operations, orbits of each individual satellite with an estimated accuracy at the few cm level are determined on a routine basis by the Jet Propulsion Laboratory. While these exhibit a certain level of common errors, the relative position obtained from a differencing of the absolute trajectories still exhibits representative errors of 10–20 mm (1 σ) in comparison to the high precision K-band measurements of the inter-satellite distance [84]. Early efforts for a fully differential orbit determination exploiting the integer nature of double difference carrier phase ambiguities resulted in baseline estimates with precision of several mm in along-track direction [106] and an even better performance of <1 mm (cf. Fig. 5.3) was later achieved by researchers at TU Delft, JPL, and the University of Bern [48–50]. Due to the large separation of the GRACE satellites, ionospheric path delays do not cancel and a rigorous resolution of the L1 and L2 ambiguities is therefore required to achieve relative navigation solutions of this quality. Following Ref. [48], success rates of about 85% could be achieved in the dual-frequency ambiguity resolution.

An interesting opportunity to assess relative navigation concepts for shorter spacecraft separations occurred in late 2005 when the leading and trailing

GRACE satellite were exchanged to balance the surface erosion of the K-band horn by atomic oxygen [57]. For a period of about 12 h on Dec. 10 the spacecraft stayed within a distance of less than 10 km, thus enabling a study of differential ionospheric path delays at short baselines [19] and the first flight demonstration of precise relative navigation using single-frequency carrier phase differential GPS. Accuracies of better than 2 mm (3D rms) were obtained for the single-frequency solution (in comparison to the dual-frequency reference) during the proximity phase when using the Lear mapping function (5.6) along with an adjustable VTEC parameter to model the differential ionospheric path delays. Aside from a lower noise level of the individual carrier phase measurements, the single-frequency processing benefits from a higher ambiguity fixing rate of up to 98% near the encounter.

Unfortunately, the K-band radar does not measure the absolute distance between the two GRACE satellites but only distance changes. Therefore, it can only be used to assess the precision but not the accuracy of the GPS-based baseline determination and provides no hint at all on the quality of the radial and cross-track component. Inter-comparisons of GRACE baseline products derived by different institutions revealed the risk of systematic biases of up to several mm in the individual baseline solutions [18]. While the nature of these biases is not finally understood, their presence illustrates the natural limitations of a microwave-based measurement system operating at a wavelength that is substantially larger than the desired accuracy.

Even though the GRACE mission itself never required a precise relative navigation it paved the way for SAR interferometry missions that rely on GPS-based baseline products of utmost accuracy. Key lessons learned from the GRACE data processing include the benefit of low noise carrier phase measurements from the C/A code tracking branch (rather than from the semi-codeless P(Y) tracking) and the improvements achievable through empirical phase pattern corrections of the GPS antennas [17, 18, 50].

The TanDEM-X project builds directly on experience gained from the GRACE GPS data processing. The two spacecraft of the TerraSAR-X/TanDEM-X formation are equipped with high-grade dual-frequency GPS receivers contributed by the GeoForschungszentrum (GFZ). The Integrated GPS Occultation Receiver (IGOR; see Fig. 5.2) selected for the mission represents a commercial rebuild of the BlackJack receiver flown earlier on numerous other science missions. On the TerraSAR-X and TanDEM-X satellites it provides pseudorange and carrier phase measurements with a mean noise level of 15 cm and 0.7 mm, or, equivalently, 21 cm and 1 mm for the receiver-receiver single-difference (Fig. 5.4). Aside from precise orbit and baseline determination, the GPS receivers are also used as radio science instruments for atmospheric sounding. Such radio-occultation measurements of GPS satellites close to the Earth's limb enable the reconstruction of temperature and density profiles of the troposphere and are a key input for global weather models [107].

As part of the TanDEM-X mission requirements, a 1 mm (1D rms) baseline accuracy has been specified to avoid tilts and shifts in individual DEMs and to

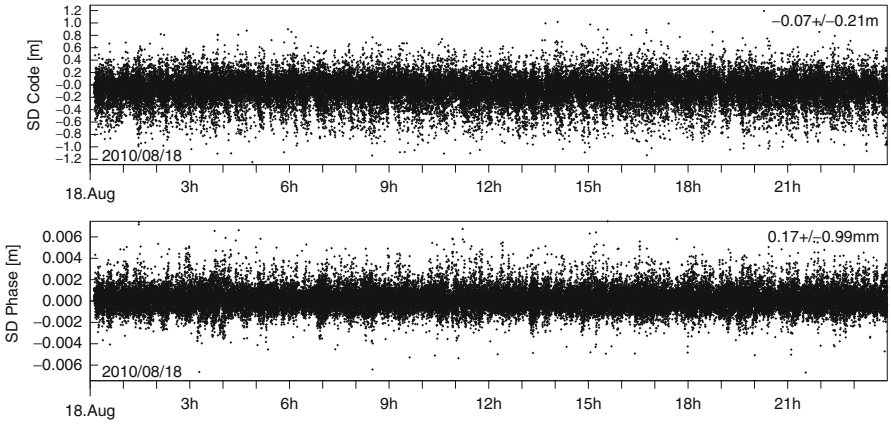


Fig. 5.4 Residuals of L1 single-difference pseudorange and carrier phase measurements from the IGOR receivers on TerraSAR-X/TanDEM-X

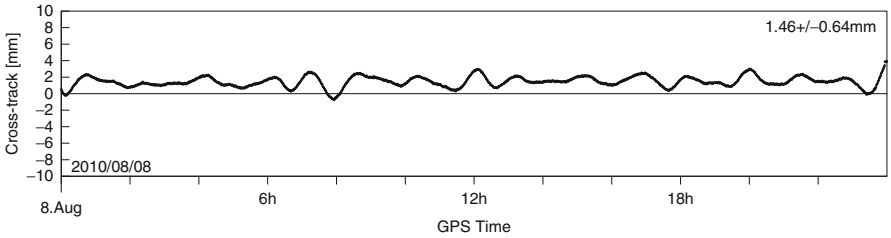


Fig. 5.5 Difference of single- and dual-frequency solutions for the cross-track baseline of the TanDEM-X formation

enable a flawless mosaicking. In view of the criticality of accurate baseline products for the overall mission performance, the baseline generation is routinely performed by two processing centers at GFZ and the German Space Operations Center (DLR/GSOC). The two solutions are generated with different algorithms and tool chains to ensure maximum independence and to facilitate a basic consistency check. Furthermore, a merged baseline product is generated from a weighted average of the individual solutions, which is then employed in the DEM generation [108].

Due to the lack of an independent measurement system providing a truth reference, the inter-comparison of independently generated solutions provides the primary means for assessing the quality of the TanDEM-X baseline products. As an example, Fig. 5.5 shows the difference of two baseline products generated in the single- and dual-frequency processing, respectively, during the early mission phase, when the spacecraft were still separated by 20 km. The two solutions show a high level of consistency except for systematic biases at the 1–2 mm level. These can in part be attributed to phase pattern distortions that are corrected in the dual-frequency solution but not the single-frequency processing.

A comparison of TanDEM-X baseline products generated by various agencies has been presented in Ref. [109]. Here, differences with a standard deviation of 0.5–1.0 mm per axis have been obtained between individual solutions, but systematic biases of similar size could likewise be identified. These biases reflect the impact of different processing concepts and are of similar order as the biases observed earlier in the GRACE mission.

Beyond biases in the GPS-derived relative motion of the two spacecraft, the interferometric SAR processing is likewise affected by uncertainties in the SAR antenna phase centers and instrumental delays. To assess the cumulative impact of all system specific biases on the measured instrument phase, dedicated calibration data takes are performed on a routine basis in the TanDEM-X mission over flat target sites with well known altitude profiles. By comparison with uncalibrated raw DEMs from the SAR interferometry systematic biases may be identified and corrected. Initial tests documented in Ref. [103] suggest residual biases at the level of a few mm, which is well within the pre-mission expectations. However, a combined analysis of calibration data takes in different observations modes will be required to arrive at a consolidated system calibration.

Other than the TerraSAR-X/TanDEM-X satellites, which are equipped with high-grade, dual-frequency GPS receivers, the PRISMA mission (Chap. 21) relies exclusively on single-frequency measurements for GPS-based relative navigation. Complementary to the real-time use (Sect. 5.4.3) the baseline between the two PRISMA satellites (named Mango and Tango) is reconstructed on ground from the Phoenix GPS (Fig. 5.2) raw measurements to provide a reference for assessing other sensors and the formation control performance [110]. Absolute orbits of the individual satellites are first determined in a reduced dynamic orbit determination using the ionosphere-free GRAPHIC combination (cf. (5.4); [67]) of the pseudorange and carrier phase measurements. The subsequent precise baseline reconstruction makes use of an extended Kalman filter/smoother which processes single-difference code and carrier phase measurements and resolves double difference ambiguities. The filter builds on earlier work described in Ref. [84] but is specifically adapted for single-frequency measurements. Instead of channel-wise ionospheric path delays and L1/L2 ambiguities, the vertical path delay and L1-only ambiguities are adjusted in this mode of the relative orbit determination software.

The differential carrier phase measurements exhibit post-fit residuals at the level of 6 mm, which clearly exceeds the value of 2 mm observed in pre-mission tests using Phoenix GPS measurements from a signal simulator test bed [111]. This higher error level is mainly attributed to lacking knowledge of the GPS antenna phase pattern distortions and a reduced attitude knowledge for the TANGO satellite. Carrier phase ambiguities can be fixed to integer values with success-rates of up to 95% in smooth mission phases with (near)-zenith-pointing GPS antennas and limited thruster activity. During periods with heavy orbit and attitude maneuvers, in contrast, processing reverted to a float-ambiguity-only solution.

In the absence of a truth reference only limited possibilities exist to assess the quality of the ground based relative orbit determination of PRISMA. One such

opportunity was given during the early mission, when the two spacecraft were still operated in a clamped configuration. Despite a partial obstruction of the TANGO antenna field of view the estimated baseline was found to match the theoretical value with a better than 1 cm 3D rms accuracy when ambiguity fixing was employed. In case of a float ambiguity solution, a degraded performance of 5 cm 3D rms was obtained [110].

During specific experiment campaigns, the separation of the two satellites could independently be measured with the Formation Flying Radio Frequency sensor (FFRF; Chaps. 21 and 6) on PRISMA. Aside from a systematic bias reflecting an incomplete system calibration, the FFRF measurements agree with the GPS-derived distance at a noise level of less than 1 cm [112], which demonstrates a good overall performance of both measurement systems.

5.4.3 Real-Time Navigation

The first demonstration of precise onboard relative navigation based on carrier-phase differential GPS has been carried out by the PRISMA formation flying mission (Chap. 21). The GPS relative navigation system employs DLR's Phoenix receivers (Sect. 5.2) and represents the primary formation flying sensor on-board the PRISMA satellites. Its key objective is the continuous provision of accurate and reliable relative position and velocity information at all times during the mission. The relative navigation state is used on-board the Mango spacecraft to autonomously control its motion with respect to Tango throughout a multitude of experiment scenarios including, among others, formation keeping and reconfiguration of passive relative orbits and quasi-continuous control of forced motion trajectories. In the absence of a relative navigation system characterized by higher technology readiness level, relative GPS represents also the safe mode sensor of the PRISMA formation. In particular the GPS relative navigation is used on-board Mango to support fault detection isolation and recovery (FDIR) functionalities like relative motion monitoring and collision avoidance. The twofold scope of the GPS navigation onboard PRISMA poses challenging requirements to the system design since ultimate accuracy is desired for precision control purposes whereas robustness and reliability are needed for safe mode activities.

In order to fulfill these objectives, a tailored GPS hardware/software system has been designed for PRISMA [76, 77, 113]. The cold-redundant hardware architecture is identical on Mango and Tango. It is based on two single-frequency (L1) Phoenix GPS receivers and two passive GPS antennas which are located at opposite sides of the spacecraft structure to provide a nearly omni-directional GPS coverage. Only one antenna is active at any time, but it is possible to select the desired GPS antenna in use either autonomously on-board or via telecommand from ground. The measurements collected on Tango are provided to Mango through an Ultra-High-Frequency (UHF) Inter-Satellite-Link and processed together with the Mango measurements by the navigation software which is integrated into the FPGA-based LEON-3 Mango

on-board computer clocked at 24 MHz. The estimation approach employs a common EKF for the absolute states of the two satellites, which accounts for the interdependency of absolute and relative navigation without the need of an explicit relative state. Thank to the sequential processing of GRAPHIC data types and single-difference carrier-phase measurements (Sect. 5.3.1), common GPS satellite visibility does not represent a prerequisite to estimate the relative state of the formation. In the case of rotating spacecraft with non-zenith pointing antennas, a relative navigation solution can be obtained through the processing of un-differenced ionosphere-free measurements alone.

The PRISMA satellites have been launched aboard a Dnepr launcher from Yasny, Russia, on June 15, 2010. The two satellites were released, clamped together in launch configuration, into a nominal dusk-dawn orbit at a mean altitude of 757 km, 0.004 eccentricity and 98.28° inclination. After the successful accomplishment of the launch and early operations phase, on June 17, 2010, PRISMA entered a 57-days long commissioning phase. This phase was characterized by the careful verification and checkout of on-board equipment, essential on-board functions and by the calibration of navigation algorithms such as attitude, rate estimators, and GPS-based navigation. Most of the commissioning phase operates PRISMA as a combined spacecraft where Tango is still mated to Mango. However, the last part of this phase included the Tango separation from Mango (on August 11, 2010) and the subsequent GPS relative navigation calibration campaign (5 days from August 16, 2010). The flight results from the initial and final commissioning phase of the GPS navigation system have been presented in Refs. [30, 110].

The adopted calibration process is based on a comparison of the real-time on-board navigation estimates available in the telemetry stream with the post-facto precise orbit determination products (Chap. 21) and consists of basically three steps. First the navigation flight software is replayed on-ground and fed with the same data inputs received on-board. To this end native GPS messages produced by the Phoenix receivers, spacecraft attitude estimates of Mango and Tango, as well as maneuver delta-v estimates from accelerometers are recorded on-board and downloaded as telemetry. Once the data output of the flight software's playback match the on-board results, a tuning process can start where dedicated EKF parameters like e.g. measurements standard deviations, a-priori standard deviations of state parameters, data editing thresholds are changed in order to improve the overall behavior of the flight software in comparison with the precise orbit determination. The third and final step of the calibration process is the upload of the newly derived EKF parameters to the spacecraft. This approach proved to be extremely useful to improve the robustness of the navigation software in unfavorable scenarios characterized for example by frequent data gaps, or spacecraft tumbling during safe modes. Furthermore the playback process enabled the rigorous analysis and debugging of the navigation software behavior during two anomalies which resulted in the release of a new flight software version and its upload to the spacecraft.

The overall outcome of the extensive numerical and hardware-in-the-loop testing performed before launch [113] and the in-flight calibration process resulted in

three possible configurations for the PRISMA GPS relative navigation, each given by dedicated settings for normal performance, frequent thrusting, and robust behavior. Due to the satisfactory results obtained in orbit, no changes have been made to the normal performance and frequent thrust configurations selected before launch during testing and validation. As compared to the normal performance, the frequent thrust settings apply a ca. 10 times higher standard deviation to pseudorange and carrier-phase measurements. This reduces the weight of measurements and increases the trust in the force model thus providing more dynamic features to the EKF. This approach can be better understood if we consider that a set of GPS measurements is processed every 30 s (sample time of the EKF) and frequent orbit control phases activate 1-N thrusts at typical rates between two minutes and 10 s. These are considered as extreme conditions for the navigation filter which augments its state to estimate delta-v maneuvers in real-time using a marginal number of observations.

On the contrary the robust behavior configuration is a specific outcome of the flight commissioning phase. As recognized in orbit, the purely absolute formulation of the EKF does not allow the constraining of the relative empirical accelerations between co-orbiting spacecraft to the desired level without disrupting the estimates of the absolute empirical accelerations (Sect. 5.3.4). This inherent weakness of the filter design became transparent during occasional phases characterized by frequent and consecutive data gaps of typically 20 min duration where the Phoenix receivers were switched off to reduce the risk of radiation induced single event effects (Sect. 5.2). The application of the normal performance settings results in excessively large estimates of the empirical accelerations which act as a force during the propagation phase and consequently degrades the navigation accuracy. The replay and tuning on-ground of the navigation filter during the calibration phase in August 2010 showed that ca. 100 times smaller standard deviation and noise of the empirical accelerations provide the desired robustness to the filter during contingency scenarios at the cost of a reduced absolute navigation accuracy. As a consequence the so-called robust EKF settings have replaced the normal performance settings as default telecommands on-board and are usually applied in scenarios characterized by data gaps, spacecraft tumbling and in general during safe mode activities.

The successful completion of the commissioning phase and the consolidation of the real-time navigation filter parameters paved the way to the start of the nominal PRISMA mission. At the time of writing, nearly one year of primary experiments have been successfully conducted without interruptions. The on-board navigation solution is routinely used as trustworthy information for numerous autonomous tasks like relative motion monitoring, collision avoidance, formation keeping and reconfiguration as well as forced motion control. The GPS-based relative navigation has been demonstrated to properly perform under all possible formation flying scenarios within the range of the Inter-Satellite-Link, i.e., between virtually 0 m (in clamped configuration) and 30 km (maximum separation achieved so far). Key flight results from the PRISMA GPS-based navigation system obtained during the conduction of various experiments have been presented in Refs. [114–116].

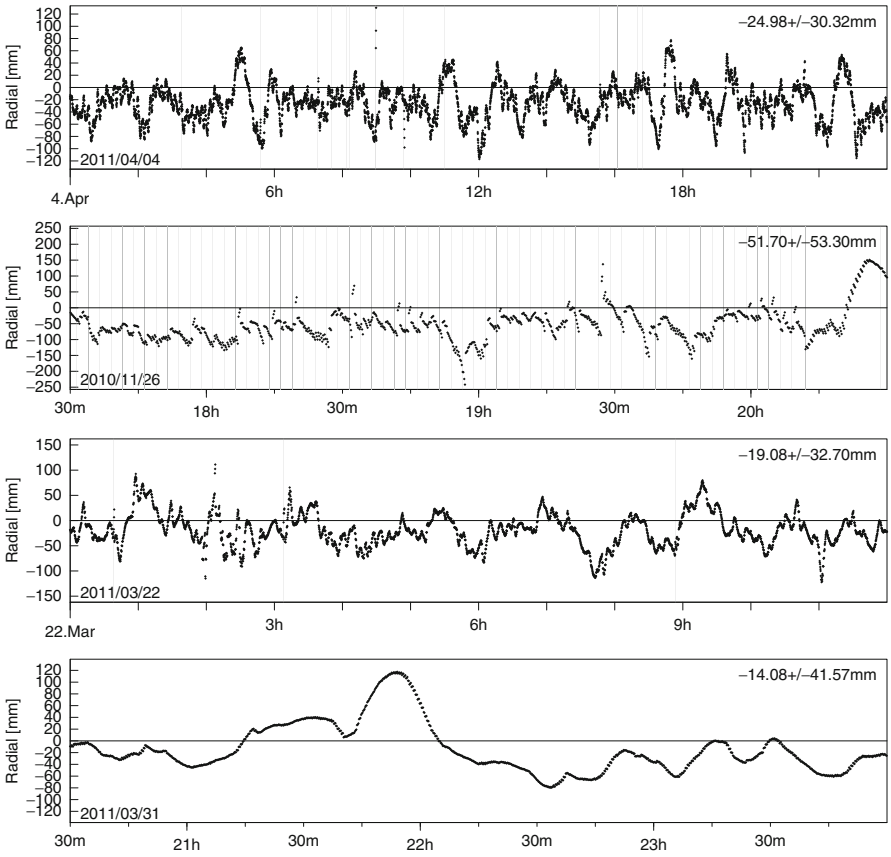


Fig. 5.6 Representative onboard GPS relative navigation errors in radial direction during sparse and dense orbit control activity with near-Zenith pointing antennas (*top* and *middle-top*), and during small and large pitch rotations of the Mango spacecraft (*middle-bottom* and *bottom*). Vertical lines represent orbit control maneuvers

Figure 5.6 illustrates representative on-board navigation errors obtained from the comparison with the post-facto precise orbit determination. In order to show how the navigation system behaves over the comprehensive range of scenarios offered by PRISMA, four different time arcs have been selected.

First, sparse orbit control activities on April 4, 2011 are shown over 24 h. Here normal performance settings are applied, the GPS antennas are near-zenith pointing, and the spacecraft separation is loosely maintained at a 100 m hold point in along-track direction. Secondly, dense orbit control activities on November 26, 2010 are considered over 3 h. Here frequent thrust settings are applied, the GPS antennas are near-zenith pointing, and the spacecraft separations are between 10 and 20 m in forced motion. Third, a scenario characterized by periodic pitch oscillations of $\pm 30^\circ$ from Zenith of the Mango GPS antennas is considered on March 22, 2011 over 12 h. Here normal performance settings are applied and

Table 5.3 Real-time onboard GPS-based relative navigation accuracy during representative operational scenarios including sparse/dense orbit maneuvering (first two rows) and small/large attitude rotations (last two rows)

Operational scenario (date) and *applied navigation settings*	Rel. position (cm), 3D			Rel. velocity (mm/s), 3D		
	Mean	Std	Rms	Std	Mean	Rms
Sparse orbit control (Apr. 4, 2011) *Normal performance settings*	5.00	2.00	5.39	0.13	0.15	0.20
Dense orbit control (Nov. 26, 2011) *Frequent thrust settings*	8.78	3.89	9.61	0.69	0.58	0.90
Small attitude rotations (Mar. 22, 2011) *Normal performance settings*	5.49	2.39	5.99	0.14	0.14	0.20
Large attitude rotations (Mar. 31, 2011) *Robust settings*	6.83	2.94	7.44	0.08	0.04	0.09

Mango is permanently pointing at the Tango spacecraft in a relative elliptic motion centered at 500 m mean along-track separation with 200 m amplitude oscillations in radial and cross-track direction. Finally, a slow tumbling 360° pitch motion of Mango is considered on March 31, 2011 over 3.5 h where the robust settings are applied and the usage of only one GPS antenna is forced during the once-per-orbit rotations of the spacecraft (ca. 0.06 deg/s).

For simplicity, the plots in Fig. 5.6 provide only the real-time navigation accuracy in the radial direction, which is characterized by the largest errors. However, full statistics of the 3D relative position and velocity errors are listed in Table 5.3 for completeness.

In most of the encountered scenarios, the demonstrated overall performance is well below 10 cm and 1 mm/s (3D, rms) for relative position and velocity respectively. The navigation error budget is dominated by maneuver executions, attitude estimation errors, multi-path, and phase pattern distortions. As discovered during the commissioning phase, the relative navigation accuracy does not degrade substantially during large attitude motions of the Mango spacecraft [30]. Due to a lacking ground plane, the GPS antennas at the end of the solar panels are notably sensitive to signals received from the antenna backside and offer a near omni-directional visibility. Despite the lower signal-to-noise ratio of signals tracked at negative elevations, the availability of commonly tracked satellites by the Phoenix receivers on Mango and Tango strengthens the relative navigation whose accuracy is not dominated by the meter level noise of the GRAPHIC data types as expected from the system performance testing conducted before launch.

5.5 Conclusion

GPS-based relative positioning is an enabling technology for distributed sensors for Earth system monitoring. Especially in the context of formation flying missions, which are commonly considered to offer new or enhanced science opportunities,

differential GPS provides high-accuracy relative (and absolute) navigation at adequate cost.

The intensive research conducted in the past decade has demonstrated the suitability of carrier-phase differential techniques for high-precision relative navigation of spacecraft in low Earth orbit. Carrier phase measurements offer a noise two to three order of magnitude lower than pseudorange measurements and, especially in formation flying applications, benefit from a high level of common error cancellation. If properly combined with a rigorous orbit dynamics model, differential carrier-phase enables mm to cm positioning accuracies in post-processing and real-time applications.

Since the availability of an adequate relative navigation system is a prerequisite for a successful formation flying mission, this chapter has presented the state-of-the-art of three key aspects which need to be carefully considered throughout the design and development: GPS hardware, estimation approach, and flight experience.

A wide selection of spaceborne GPS receivers is offered by numerous suppliers, and some of them are presently in use for actual formation flying missions or have been used for relative navigation in short-term rendezvous. This includes fully space qualified dual-frequency hardware for high-end projects as well as low-cost miniaturized receivers built around commercial-off-the-shelf components for technology demonstrators.

In advanced space applications it is generally desirable to process GPS raw measurements in a dedicated on-ground or on-board orbit determination process to achieve a higher accuracy and robustness as compared to the receiver navigation solution. To this end an introduction of the basic processing concepts has been provided, including observation types, relevant dynamical models, and available estimation schemes.

Finally, a summary of key flight results obtained from precursor experiments as well as rendezvous and formation flying missions has been presented. Although the very first in-flight demonstrations of GPS relative navigation with meter-level accuracy were already performed in the 1990s as part of various national rendezvous programs, it took more than a decade to exploit the full potential of differential carrier-phase techniques. Certainly much better performances could later be achieved in hardware-in-the-loop simulations using GPS signal simulators, but it is only since the GRACE mission that the feasibility of reconstructing the relative position between co-orbiting spacecraft with mm-level accuracy has been proven.

Even though the understanding and removal of all system specific biases still represents a tremendous challenge, the 3D baseline between the TanDEM-X satellites, the first synthetic aperture radar interferometer in orbit, is nowadays routinely determined with a consistency of 1–2 mm on-ground. Moreover the high technology readiness level reached by real-time relative GPS is clearly demonstrated by the PRISMA satellites whose relative motion is autonomously controlled on a regular basis using a cm-accurate GPS-based relative navigation system as primary sensor.

Given the wide range of orbiting and proposed formation flying missions, a single navigation system will never be able to fulfill all functional and performance

requirements. In fact, the use of differential GPS is typically restricted to cooperative missions in low Earth orbit, whereas independent optical or radiometric sensors are required for the approach to non-cooperative targets, for deep-space formation flying and for interferometric space telescopes. Nevertheless, the use of GPS has already been considered for high-altitude missions and even lunar navigation despite severe performance restrictions that have to be expected at large distances from the Earth.

References

1. Kaplan ED, Hegarty CJ (eds) (2006) Understanding GPS – principles and applications, 2nd edn. Artech House, Boston/London
2. Bauer FH, Moreau MC, Dahle-Melsaether ME, Petrofski WP, Stanton BJ, Thomason S, Harris GA, Sena RP, Parker Temple L III (2006) The GPS space service volume. ION-GNSS-2006, Fort Worth, 26–29 Sept 2006
3. Corazzini T, Robertson A, Adams JC, Hassibi, A, How JP (1997) GPS sensing for spacecraft formation flying. ION-GPS-1997, Kansas, Sept 1997
4. Fehse W (2003) Automated rendezvous and docking of spacecraft. Cambridge University Press, New York
5. Kelbel D, Lee T, Long A, Carpenter R, Gramling C (2001) Evaluation of relative navigation algorithms for formation-flying satellites using GPS. In: Proceedings of the 2001 flight mechanics symposium. NASA CP-2001-209986, Greenbelt, 19–21 June 2001
6. Space Engineering – Space Environment (2008) ECSS-E-ST-10-04C. European cooperation for space standardization, Noordwijk, 15 Nov 2008
7. Collaboration website of the European cooperation for space standardization. http://www.ecss.nl. Last accessed 8 July 2012
8. Council Regulation (EC) No. 428/2009 of 5 May 2009 setting up a Community regime for the control of exports, transfer, brokering and transit of dual-use items. European Commission, Brussels (2009). http://trade.ec.europa.eu/doclib/docs/2009/june/tradoc_143390.pdf. Last Accessed 8 July 2012
9. US Department of State, Directorate of Defense Trade Controls (2011) International Traffic in Arms Regulations. http://www.pmddtc.state.gov/regulations_laws/itar_official.html. Last accessed 8 July 2012
10. Renaudie C, Markgraf M, Montenbruck O, Garcia-Fernandez M (2007) Radiation testing of commercial-off-the-shelf GPS technology for use on low earth orbit satellites. RADECS 2007. In: Proceedings of 9th European conference radiation and its effects on components and systems, Deauville, 10–14 Sept 2007
11. Virtanen A (2006) The use of particle accelerators for space projects. EPS Euroconference XIX nuclear physics divisional conference. J Phys: Conf Ser 41:101–114. DOI:10.1088/1742-6596/41/1/008
12. Zadeh A, Santandrea S, Landstroem S, Markgraf M (2010) DLR Phoenix GPS receiver radiation characterisation campaign proton irradiation testing at PSI – June 2010 test report. TEC-SYV/81/2011/REP/SS, ESA/ESTEC, Noordwijk, 8 Feb 2011
13. Helm A, Hess M-P, Minori M, Gribkov A, Yudanov S, Montenbruck O, Beyerle G, Cacciapuoti L, Nasca R (2009) The ACES GNSS subsystem and its potential for radio-occultation and reflectometry from the International Space Station. In: 2nd international colloquium on scientific and fundamental aspects of the Galileo program, Padua, 14–16 Oct 2009

14. Markgraf M, Montenbruck O, Santandrea S, Naudet J (2010) Phoenix-XNS navigation system onboard the PROBA-2 spacecraft – first flight results. Small satellites systems and services – the 4S symposium, Madeira, Portugal, 31 May–4 June 2010
15. Winternitz LMB, Bamford WA, Heckler GW (2009) A GPS receiver for high-altitude satellite navigation. IEEE J Sel Top Signal Process 3(4):541–556. doi:10.1109/JSTSP.2009.2023352
16. Montenbruck O, Garcia-Fernandez M, Williams J (2006) Performance comparison of semi-codeless GPS receivers for LEO satellites. GPS Solutions 10:249–261. doi:10.1007/s10291-006-0025-9
17. Montenbruck O, Garcia-Fernandez M, Yoon Y, Schön S, Jäggi A (2009) Antenna phase center calibration for precise positioning of LEO satellites. GPS Solutions 13(1):23–34. doi:10.1007/s10291-008-0094-z
18. Jäggi A, Dach R, Montenbruck O, Hugentobler U, Bock H, Beutler G (2009) Phase center modeling for LEO GPS receiver antennas and its impact on precise orbit determination. J Geodesy 83(12):1145–1162. doi:10.1007/s00190-009-0333-2
19. van Barneveld P, Montenbruck O, Visser P (2008) Differential ionospheric effects in GPS based navigation of formation flying spacecraft. In: Proceedings of 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, Noordwijk, 23–25 Apr 2008
20. Julien O, Cannon ME, Alves P, Lachapelle G (2004) Triple frequency ambiguity resolution using GPS/Galileo. Eur J Navig 2(2):51–57
21. Van der Marel H (2010) Combining GNSS signals – bias and calibration issues. IGS analysis workshop, Newcastle, June 28–July 1, 2010
22. GPS World Receiver Survey. GPS World. Jan 2012, 1–23
23. García-Rodríguez A (2008) Onboard radio navigation receivers. Technical Dossier – European Space Technology Harmonization, TEC-ETN/2007.65, 16 Apr 2008
24. Roselló Guasch J, Silvestrin P, Aguirre M, Massotti L (2010) Navigation needs for ESA's earth observation missions. In: Sandau R, Röser H-P, Valenzuela A (eds) Small satellite missions for earth observation – new developments and trends. Springer, Berlin, Heidelberg, pp 457–466. doi:10.1007/978-3-642-03501-2_41
25. Montenbruck O, Yoon Y, Ardaens J-S, Ulrich D (2008) In-flight performance assessment of the single frequency MosaicGNSS receiver for satellite navigation. In: Proceedings of 7th international ESA conference on guidance, navigation and control systems, ESA WPP-288, Tralee, 2–5 June 2008
26. Föckersperger S, Hollmann R, Dick G, Reigber C (1997) On board MIR: orienting remote images with MOMSNAV. GPS World 8:32–39
27. Ebinuma T, Rooney E, Gleason S, Unwin M (2005) GPS receiver operations on the disaster monitoring constellation satellites. J Navig 58:227–240
28. Pinard D, Reynaud S, Delpy P, Strandmoe SE (2007) Accurate and autonomous navigation for the ATV. Aerosp Sci Technol 11:490–498
29. Hwang Y, Lee B-S, Kim J, Jung O-C, Chung D-W, Kim H (2010) KOMPSAT-2 orbit determination status report. AIAA 2010–8260; AIAA guidance, navigation, and control conference, Toronto, 2–5 Aug 2010
30. D'Amico S, Ardaens J-S, De Florio S, Montenbruck O, Persson S, Noteborn R (2010) GPS-based spaceborne autonomous formation flying experiment (SAFE) on PRISMA: initial commissioning. AIAA/AAS astrodynamics specialist conference, Toronto, 2–5 Aug 2010
31. Spangelo S, Kleshy A, Cutler J (2010) Position and time system for the RAX small satellite mission. AIAA-2010-7980, AIAA/AAS astrodynamics specialist conference, Toronto, 2–5 Aug 2010
32. Sust M, Carlström A, Garcia-Rodriguez A (2009) European spaceborne dual frequency GPS receiver for science and earth observation. ION-GNSS-2009, Savannah, GA, 22–25 Sept 2009

33. Montenbruck O, Andres Y, Bock H, van Helleputte T, van den IJssel J, Loiselet M, Marquardt C, Silvestrin P, Visser P, Yoon Y (2008) Tracking and orbit determination performance of the GRAS instrument on MetOp-A. GPS Solutions 12(4):289–299. doi:10.1007/s10291-008-0091-2
34. Zin A, Landenna S, Conti A, Marradi L, Di Raimondo MS (2006) L1/L2 Integrated GPS/WAAS/EGNOS Receiver. ENC-GNSS- 2006 - The European Navigation Conference, 8-10 May 2006, Manchester, UK.
35. Serre S, Mercier F, Garcia A, Grondin M, Boyer C, Favaro H, Gerner J-L, Issler J-L (2010) First In-orbit Results of the L2C-L1C/A GPS receiver on board the PROBA-2 microsatellite. ION-GNSS-2010, Portland, Oregon, 22–24 Sept 2010
36. Langley RB, Montenbruck O, Markgraf M, Kang CS, Kim D (2004) Qualification of a commercial dual-frequency GPS receiver for the e-POP platform onboard the Canadian CASSIOPE spacecraft. NAVITEC'2004, Noordwijk, The Netherlands, 8–10 Dec 2004
37. Markgraf M, Renaudie C, Montenbruck O (2008) The NOX payload – flight validation of a low-cost dual-frequency GPS receiver for micro- and nano-satellites. Small satellites systems and services – the 4S symposium, Rhodes, Greece, 26–30 May 2008
38. Misra P, Enge P (2006) Global positioning system (GPS): signals, measurements and performance, 2nd edn. Ganga-Jamuna, Lincoln
39. Wu JT, Wu SC, Hajj GA, Bertiger WI, Lichten SM (1993) Effects of antenna orientation on GPS carrier phase. Manuscr Geodaet 18:91–98
40. Yunck TP (1993) Coping with the atmosphere and ionosphere in precise satellite and ground positioning. In: Valance-Jones A (ed) Environmental effects on spacecraft trajectories and positioning. AGU Monograph, Washington, DC
41. Lear WM (1987) GPS navigation for low-Earth orbiting vehicles. NASA 87-FM-2, Rev. 1, JSC-32031, Lyndon B. Johnson Space Center, Houston, TX
42. Tancredi U, Renga A, Grassi M (2010) GPS-based relative navigation of LEO formations with varying baselines. AIAA/AAS astrodynamics specialist conference, Toronto, 2–5 Aug 2010
43. Psiaki ML, Mohiuddin S (2007) Modeling, analysis, and simulation of GPS carrier phase for spacecraft relative navigation. J Guid Contr Dyn 30(6):1628–1639. doi:10.2514/1.29534
44. Hofmann-Wellenhoff B, Lichtenegger H, Collins J (2001) GPS: theory and practice, 5th edn. Springer, New York
45. Teunissen PJG (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. J Geodesy 70(1–2):65–82. doi:10.1007/BF00863419
46. Cox DB, Brading JDW (2000) Integration of LAMBDA ambiguity resolution with Kalman filter for relative navigation of spacecraft. Navig: J ION 47(3):205–210
47. Mohiuddin S, Psiaki ML (2005) Satellite relative navigation using carrier-phase differential GPS with integer ambiguities. AIAA guidance, navigation, and control conference, San Francisco, 15–18 Aug 2005
48. Kroes R, Montenbruck O, Bertiger W, Visser P (2005) Precise GRACE baseline determination using GPS. GPS Solutions 9:21–31. doi:10.1007/s10291-004-0123-5
49. Wu S-Ch, Bar-Sever Y.E (2006) Real-time sub-cm differential orbit determination of two low-Earth orbiters with GPS bias fixing. ION GNSS 2006, Fort Worth, TX, 26–29 Sep 2006
50. Jäggi A, Hugentobler U, Bock H, Beutler G (2007) Precise orbit determination for GRACE using undifferenced or doubly differenced GPS data. Adv Space Res 39:1612–1619. doi:10.1016/j.asr.2007.03.012
51. Laurichesse D, Mercier F, Berthias JP, Broca P, Cerri L (2009) Integer ambiguity resolution on undifferenced GPS phase measurements and its application to PPP and satellite precise orbit determination. Navig: J ION 56(2):135–149
52. de Ruiter A, Lee J, Ng A, Kim Y (2008) Orbit determination and relative positioning techniques for JC2Sat-FF. In: Proceedings of the 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, ESA SP-654, Noordwijk, The Netherlands, 23–25 Apr 2008

53. Psiaki ML (2010) Kalman filtering and smoothing to estimate real-valued states and integer constants. J Guid Contr Dyn 33(5):1404–1417
54. Clohessy WH, Wiltshire RS (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 270:653
55. Vallado DA (1997) Fundamentals of astrodynamics and applications. McGraw-Hill, New York
56. Schaub H-P (2002) Spacecraft relative orbit geometry description through orbit element differences. In: 14th US national congress of theoretical and applied mechanics, Blacksburg, 23–28 June, 2002
57. Montenbruck O, Kirschner M, D'Amico S, Bettadpur S (2006) E/I-vector separation for safe switching of the GRACE formation. Aerosp Sci Technol 10(7):628–635
58. Tschauner J, Hempel P (1965) Rendezvous zu einem in Elliptischer Bahn Umlaufenden Ziel. Astronaut Acta 11:104–109
59. Yamanaka K, Ankersen F (2002) New state transition matrix for relative motion on an arbitrary elliptical orbit. J Guid Contr Dyn 25(1):60–66. doi:10.2514/2.4875
60. Gim D-W, Alfriend KT (2001) The state transition matrix of relative motion for the perturbed non-circular reference orbit. Paper no. 01–222, AAS/AIAA space flight mechanics meeting, Santa Barbara, Feb 2001
61. Carter T, Humi M (2002) Clohessy–Wiltshire equations modified to include quadratic drag. J Guid Contr Dyn 25(6):1058–1063
62. Schweighart SA, Sedwick RJ (2002) High-fidelity linearized J2 model for satellite formation flight. J Guid Contr Dyn 25(6):1073–1080
63. Hamel JF, De LaFontaine J (2007) Linearized dynamics of formation flying spacecraft on a J2-perturbed elliptical orbit. J Guid Contr Dyn 30(6):1649–1658. doi:10.2514/1.29438
64. Montenbruck O, Gill E (2000) Satellite orbits – models, methods and applications. Springer Verlag, Heidelberg
65. Dach R, Hugentobler U, Fridez P, Meindl M (2007) Bernese GPS software version 5.0 – user manual. University of Bern, Jan 2007
66. Rowlands D, Marshall JA, McCarthy J et al (1995) GEODYN system description II, vol 1–5. Hughes STX, Greenbelt
67. Montenbruck O, van Helleputte T, Kroes R, Gill E (2005) Reduced dynamic orbit determination using GPS code and carrier measurements. Aerosp Sci Technol 9(3):261–271. doi:10.1016/j.ast.2005.01.003
68. Springer T (2009), NAPEOS Mathematical Models and Algorithms. DOPS-SYS-TN-0100-OPS-GN, Issue 1.0, Nov. 2009, ESA/ESOC, Darmstadt
69. Goldstein DB, Born GH, Axelrad P (2001) Real-time, autonomous precise orbit determination using GPS. Navig: J ION 48(3):155–168
70. Montenbruck O, Ramos-Bosch P (2008) Precision real-time navigation of LEO satellites using global positioning system measurements. GPS Solutions 12(3):187–198. doi:10.1007/s10291-007-0080-x
71. Vallado DA, Finkleman D (2008) A critical assessment of satellite drag and atmospheric density modeling. AIAA-2008-6442, AIAA/AAS astrodynamics specialist conference, Honolulu, Hawaii, 18–21 Aug 2008
72. Wu SC, Yunck TP, Thornton CL (1991) Reduced-dynamic technique for precise orbit determination of low Earth satellites. J Guid Contr Dyn 14(1):24–30
73. Jäggi A, Hugentobler U, Beutler G (2006) Pseudo-stochastic orbit modeling techniques for low-Earth orbiters. J Geodesy 80(1):47–60. doi:10.1007/s00190-006-0029-9
74. Montenbruck O, Hauschild A, Zangerl F, Zsalcsik W, Ramos-Bosch P, Klein U (2009) Onboard real-time navigation for the sentinel-3 mission. ION-GNSS-2009, Savannah, USA, 22–25 Sept 2009
75. Montenbruck O, Wermuth M, Kahle R (2010) GPS based relative navigation for the TanDEM-X mission – first flight results. ION-GNSS-2010 conference, Portland, Oregon, 21–24 Sept 2010

76. D'Amico S, Gill E, Garcia-Fernandez M, Montenbruck O (2006) GPS-based real-time navigation for the PRISMA formation flying mission. 3rd ESA workshop on satellite navigation user equipment technologies, NAVITEC'2006, Noordwijk, 11–13 Dec 2006
77. D'Amico S, Ardaens J-S, Montenbruck O (2009) Navigation of formation flying spacecraft using GPS: the PRISMA technology demonstration. ION-GNSS-2009, Savannah, USA, 22–25 Sept 2009
78. Montenbruck O, Gill E (2001) State interpolation for on-board navigation systems. Aerosp Sci Technol 5:209–220. doi:10.1016/S1270-9638(01)01096-3
79. Shampine LF, Gordon MK (1975) Computer solution of ordinary differential equations. Freeman and Comp., San Francisco
80. Wermuth M, Montenbruck O, van Helleputte T (2010) GPS High precision Orbit determination software tools (GHOST). 4th international conference on astrodynamics tools and techniques, Madrid, 3–6 May 2010
81. Hairer E, Nørsett SP, Wanner G (1987) Solving ordinary differential equations I. Springer, Berlin-Heidelberg/New York
82. Gill E, Montenbruck O, Kayal H (2001) The BIRD Satellite Mission as a milestone towards GPS-based autonomous navigation. Navig: J ION 48(2):69–75
83. Montenbruck O, Markgraf M, Santandrea S, Naudet J, Gantois K, Vuilleumier P (2008) Autonomous and precise navigation of the PROBA-2 spacecraft. AIAA 2008–7086, AIAA astrodynamics specialist conference, Honolulu, Hawaii, 18–21 Aug 2008
84. Kroes R (2006) Precise relative positioning of formation flying spacecraft using GPS. Ph.D. thesis, TU Delft
85. Zhu S, Reigber Ch, König R (2004) Integrated adjustment of CHAMP, GRACE, and GPS data. J Geodesy 78:103–108. doi:10.1007/s00190-004-0379-0
86. Brown RG, Hwang PYC (1992) Introduction to random signals and applied Kalman filtering, 2nd edn. Wiley, New York
87. Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. IEEE Trans Autom Contr 92(3):401–422. doi:10.1109/JPROC.2003.823141
88. Ilyas M, Lim J, Lee JG, Park CG (2008) Federated unscented Kalman filter design for multiple satellites formation flying in LEO. International conference on control, automation and systems, Seoul, Korea, 14–17 Oct 2008
89. Pardal PCPM, Kuga HK, Vilhena de Moraes R (2011) Robustness assessment between sigma point and extended Kalman filter for orbit determination. In: 22nd international symposium on spaceflight dynamics, Sao Jose dos Campos, Brazil, 28 Feb–4 Mar 2011
90. Bierman GJ (1977) Factorization methods for discrete sequential estimation. Academic, New York
91. Busse FD, How JP, Simpson J (2002) Demonstration of adaptive extended Kalman filter for low earth orbit formation estimation using CDGPS. In: Proceedings of ION-GPS-2002, Portland, OR, 24–27 Sept 2002
92. Leung S, Montenbruck O (2005) Real-time navigation of formation-flying spacecraft using global positioning system measurements. J Guid Contr Dyn 28(2):226–235
93. Roth N, Urbanek J, Johnston-Lemke B, Bradbury L, Armitage S, Leonard M, Ligori M, Grant C, Damaren C, Zee R (2011) System-level overview of CanX-4 and CanX-5 formation flying satellites. 4th international conference on spacecraft formation flying missions and technologies, St-Hubert, Quebec, 18–20 May 2011
94. Ebinuma T, Bishop RH, Lightsey G (2001) Spacecraft rendezvous using GPS relative navigation. In: AAS 01–152, 11th annual AAS/AIAA space flight mechanics meeting, Santa Barbara, CA
95. Park YW, Brazzel JP Jr, Carpenter JR, Hinkel HD, Newman JH (1996) Flight test results from real-time relative global positioning system flight experiment on STS-69, vol 104824, NASA technical memorandum. National Aeronautics and Space Administration, Washington, DC

96. Moreau G, Marcille H(1997) RGPS postflight analysis of ARP-K flight demonstration. 12th international symposium on spaceflight dynamics, Darmstadt, Germany, ESA SP-403, 2–6 June 1997, pp 97–102
97. Highsmith D, Axelrad P (1999) Relative state estimation using GPS flight data from co-orbiting spacecraft. ION-GPS-1999, Nashville TN, 14–17 Sept 1999
98. Moreau G, Marcille H (1998) On-board precise relative orbit determination. 2nd European symposium on global navigation satellite systems, Toulouse, France, 20–23 Oct 1998
99. Cavrois B, Personne G, Stramdmoe S, Reynuad S, Narmada Zink M (2008) Two different implemented relative position/velocity estimations using GPS sensors on-board-ATV. 7th ESA conference on guidance, navigation and control systems, Tralee, Ireland, 2–5 June 2008
100. Kawano I, Mokuno M, Kasai T, Suzuki T (1999) First autonomous rendezvous using relative GPS navigation by ETS-VII. ION-GPS-1999, Nashville TN, 14–17 Sept 1999
101. Kawano I, Mokuno M, Miyano T, Suzuki T (2000) Analysis and evaluation of GPS relative navigation using carrier phase for RVD experiment satellite of ETS-VII. ION-GPS-2000, Salt Lake City, Utah, 19–22 Sept 2000
102. Krieger G, Moreira A, Fiedler H, Hajnsek I, Werner M, Younis M, Zink M (2007) TanDEM-X: a satellite formation for high resolution SAR interferometry. IEEE Trans Geosci Rem Sens 45(11):3317–3341. doi:10.1109/TGRS.2007.900693
103. Wermuth M, Montenbruck O, Wendleder A (2011) Relative navigation for the TanDEM-X mission and evaluation with DEM calibration results. 22nd international symposium on spaceflight dynamics, Sao Jose dos Campos, Brazil, 28 Feb–4 Mar 2011
104. Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: mission overview and early results. Geophys Res Lett 31(9):L09607
105. Bertiger W, Bar-Sever Y, Desai S, Dunn C, Haines B, Kruizinga G, Kuang D, Nandi S, Romans L, Watkins M, Wu S, Bettadpur S (2002) GRACE: millimeters and microns in orbit. In: Proceedings of ION-GPS-2002, Portland, Oregon, pp 2022–2029, 24–27 Sept 2002
106. Svehla D, Rothacher M (2004) Formation flying of LEO satellites using GPS. EOS Trans. AGU, 85(47), Fall Meeting Supplement Abstract SF53A-0735. American Geosciences Union 2004, San Francisco, 13–17 Dec 2004
107. Wickert J, Arras Ch, Beyerle G, Heise S, Jakowski N, Rothacher M, Schmidt Th, Stosius R (2009) Scientifc use of GPS signals in space. 7th IAA symposium on small satellite for earth observation, Berlin, 4–8 May 2009
108. Montenbruck O, Wermuth M, Kahle R (2010) GPS based relative navigation for the TanDEM-X mission – first flight results. ION-GNSS-2010 conference, Portland, Oregon, 21–24 Sep 2010
109. Jäggi A, Montenbruck O, König R, Wermuth M, Moon Y, Bock H, Bodenmann D(2012) Inter-agency comparison of TerraSAR-X and TanDEM-X baseline solutions; Advances in Space Research 50(2):260–271. doi:10.1016/j.asr.2012.03.027
110. Ardaens JS, D'Amico S, Montenbruck O (2011) Final commissioning of the PRISMA GPS navigation system. 22nd international symposium on spaceflight dynamics, Sao Jose dos Campos, Brazil, 28 Feb–4 Mar 2011
111. Montenbruck O, Delpech M, Ardaens J-S, Delong N, D'Amico S, Harr J (2008) Cross-validation of GPS and FFRF-based relative navigation for the PRISMA mission. 4th ESA workshop on satellite navigation user equipment technologies. NAVITEC'2008, ESA/ESTEC, Noordwijk, ESA WPP-297, 10–12 Dec 2008
112. Grelier T, Guidotti P-Y, Delpech M, Harr J, Thevenet J-B, Leyre X (2010) Formation flying radio frequency instrument: first flight results from the PRISMA mission. NAVITEC'2010, Noordwijk, The Netherlands, 8–10 Dec 2010
113. D'Amico S, De Florio S, Ardaens J-S, Yamamoto T (2008) Offline and hardware-in-the-loop validation of the GPS-based real-time navigation system for the PRISMA formation flying mission. 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, Noordwijk, 23–25 Apr 2008

114. D'Amico S, Ardaens J-S, Larsson R (2011) In-flight demonstration of formation control based on relative orbital elements. 4th international conference on spacecraft formation flying missions and technologies, St-Hubert, Quebec, 18–20 May 2011
115. Larsson R, Noteborn R, Bodin P, D'Amico S, Karlsson T, Carlsson A (2011) Autonomous formation flying in LEO – seven months of routine formation flying with frequent reconfigurations. 4th international conference on spacecraft formation flying missions and technologies, St-Hubert, Quebec, 18–20 May 2011
116. Larsson R, D'Amico S, Noteborn R, Bodin P (2011) GPS navigation based proximity operations by the PRISMA satellites – flight results. 4th international conference on spacecraft formation flying missions and technologies, St-Hubert, Quebec, 18–20 May 2011

Chapter 6
Radio Frequency-Based Relative Navigation

D. Maessen, E. Gill, T. Grelier, and M. Delpech

Abstract For distributed systems in space, knowledge of the relative position and velocity is required to maintain the relative geometry of the satellites within certain boundaries. This knowledge can be obtained using an autonomous relative navigation system based on radio frequency (RF) signals. The design aspects of such a system are detailed in this chapter. As all RF-based relative navigation systems are based on GNSS technology, the discussion is limited to this technology only. Navigation is performed by measuring the range (rate) between the satellites, which allows, in combination with a relative dynamics model and the exchange of data between the satellites, an onboard estimation of the relative state of the satellites. Obtaining accurate and unambiguous measurements requires a balanced signal design that minimizes measurement errors, but which also takes into account multiple access and formation safety considerations. Hardware-induced measurement biases should be minimized and hardware (self-) calibration is mandatory to achieve satisfactory performance in space. Details on the design, testing, and performance of one particular system, the FFRF, are provided.

A/D	Analog to digital
AC	Alternating current
ADCS	Attitude determination and control system
AFF	Autonomous formation flying
AFRL	Air force research laboratory

D. Maessen (✉) • E. Gill
Space Systems Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: D.C.Maessen@tudelft.nl; e.k.a.gill@tudelft.nl

T. Grelier • M. Delpech
CNES, Toulouse Space Center, 18 Av. E. Belin, 31401 Toulouse, France
e-mail: thomas.grelier@cnes.fr; michel.delpech@cnes.fr

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_6,

AMP	Amplifier
APL	Applied physics laboratory
ASIC	Application specific integrated circuit
BCS	Binary coded symbols
BER	Bit error rate
BOC	Binary offset carrier
BPF	Band pass filter
BPSK	Binary phase shift keying
CBPSK	Composite binary phase shift keying
CCNT	Constellation communications and navigation transceiver
CDMA	Code division multiple access
CLT	Crosslink transceiver
COTS	Commercial off the shelf
CW	Clohessy-Wiltshire
DLL	Delay lock loop
DSP	Digital signal processor
DSSS	Direct sequence spread spectrum
EKF	Extended Kalman Filter
FDIR	Fault detection isolation and recovery
FDMA	Frequency division multiple access
FFRF	Formation flying radio frequency
FPGA	Field programmable gate array
GNC	Guidance navigation, and control
GNS	GPS navigation system
GNSS	Global navigation satellite system
GPS	Global positioning system
GRACE	Gravity recovery and climate experiment
GRAIL	Gravity recovery and interior laboratory
GSFC	Goddard space flight center
IAR	Integer ambiguity resolution
IF	Intermediate frequency
IRAS	Intersatellite ranging and alarm system
ITU	International telecommunication union
JHU	Johns Hopkins University
JPL	Jet propulsion laboratory
KBR	K-band ranging
LGRS	Lunar gravity ranging system
LNA	Low noise amplifier
LOS	Line-of-sight
LPT	Low power transceiver
LVLH	Local vertical local horizontal
M2inT	Miniature multifunctional integrated terminal
MCS	Multilevel coded spreading symbols
MCXO	Microcomputer-compensated crystal oscillator

MLS	Maximum length sequence
MMS	Magnetospheric multiscale
NCLT	NanoSat crosslink transceiver
NCO	Numerically controlled oscillator
OBC	On-board computer
OCXO	Oven controlled crystal oscillator
PLL	Phase lock loop
PPS	Pulse-per-second
PRISMA	Prototype research instruments and space mission technology advancement
PRN	Pseudo-random noise
PSWF	Prolate spheroidal wave function
PVT	Position velocity, time
QPSK	Quadrature phase shift keying
RelNav	Relative navigation
RF	Radio frequency
RFE	RF front end
RTC	Real time clock
RTN	Radial-tangential-normal
RTU	Remote terminal unit
Rx	Receiver
SOC	Sinusoidal offset carrier
SPTC	Stanford pseudolite transceiver crosslink
SW	Saw filter
TC	Telecommand
TCXO	Temperature compensated crystal oscillator
TDMA	Time division multiple access
TEC	Total electron content
TH	Tschauner-Hempel
TPF	Terrestrial planet finder
TM	Telemetry
TT&C	Telemetry tracking and command
Tx	Transmitter
UHF	Ultra-high frequency
XO	Crystal oscillator
YA	Yamanaka-Ankersen

6.1 Introduction

For formations requiring extreme relative navigation and control accuracy, radio frequency (RF) based relative navigation is commonly considered to be the first element in a metrology chain, providing relatively coarse navigation information for the next element, which is typically an interferometric laser metrology system.

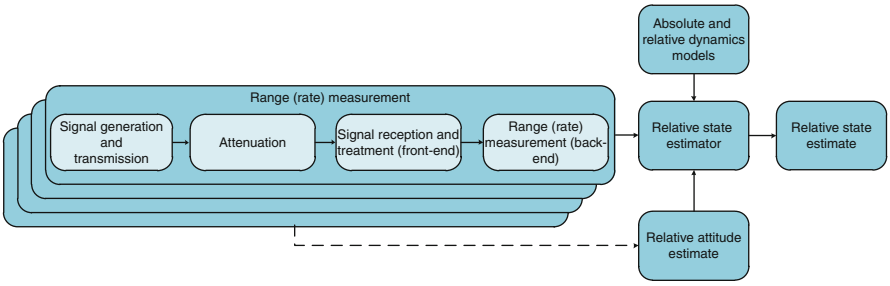


Fig. 6.1 RF-based relative navigation principle

This places specific requirements on the RF-based relative navigation system such as full sky coverage and robustness. Relative navigation can be performed in two ways: (1) using external systems like GNSS or tracking stations to determine absolute states that can be differenced or (2) using systems which are implemented on the individual satellites in the formation and which provide relative information only. This chapter deals with the latter option. This introductory section introduces the need for such a relative navigation system, its basic principles, and the outline for the remainder of the chapter.

6.1.1 Need

Chapter 5 has shown that highly accurate relative navigation for Earth-orbiting formation flying satellites can be achieved using differential GNSS measurements. However, there are missions for which this type of relative navigation is not sufficient to obtain the required performance. These can be missions that demand extremely accurate relative navigation or missions where the orbit altitudes are, entirely or for a certain time period, higher than those of the GNSS constellations. In addition, it might be desirable to have a supplementary system next to the relative GNSS system as a back-up or for cross-calibration purposes.

6.1.2 Basic Principles

Relative navigation deals with the estimation of a set of parameters which sufficiently defines the relative state of two platforms at a certain epoch. As shown in Fig. 6.1, physical measurements and both absolute and relative dynamics models are used to produce this estimate.

Here, the physical measurements constitute a number of RF range (rate) measurements between a transmitter and a receiver. Figure 6.1 also shows the

relative attitude estimate as an input for the relative state estimator since this information is needed to convert the measurements from their local reference frame to a co-rotating or inertial reference frame in case measurements are not obtained from external systems like GNSS. Relative attitude information can be obtained by differencing the inertial attitudes of the platforms, as determined by their Attitude Determination and Control System (ADCS), or by using multiple transmitters and receivers provided that these are present in sufficient quantity (hence the dashed line in Fig. 6.1). In addition, accurate time-tagging of the measurements and relative attitude estimates is essential to produce an accurate relative state estimate.

6.1.3 Chapter Outline

This chapter is structured as follows. First, the overall concept of relative navigation using locally generated ranging signals is discussed. This is followed by a description of the system design on signal and hardware level. Next, examples are given of existing system designs and their performances. The chapter ends with conclusions and an outlook for future system improvements.

6.2 Relative Navigation

The main objective of relative navigation is to determine the relative state (e.g., relative position and velocity) of two platforms at a certain epoch. The focus here is on methods commonly applied in relative navigation for formation flying satellites. The section first discusses the nature and the modeling of the observations which serve as inputs to the relative state estimation, which is detailed in the second part of the section.

6.2.1 Observations

In RF-based relative navigation, ranging signals are used to measure range or range rate between two objects. In the present discussion, we will mainly focus on range measurements as these provide information that is typically better suited for the relative navigation. Range rate measurements are assumed to be available as additional measurements. We first introduce the basic range (rate) measurement techniques and list typical error sources that need to be dealt with in order to achieve high ranging accuracy. Next, we model the measurements made using code or carrier observations. From these, we can construct range and clock

bias estimates. In case of multiple receiving antennas, angular measurements can be constructed based on the difference in their measured range.

6.2.1.1 Ranging Principles

When using RF signals to directly estimate the range between two points, two basic measurement types are possible, namely time-of-flight (requires knowledge of the time t_{Tx} when the signal is radiated) and carrier phase shift (requires knowledge of the signal phase when radiated). A direct estimate of the change in range can be obtained by measuring Doppler shift (requires knowledge of the transmitter frequency f_{Tx}). Equations 6.1, 6.2 and 6.3 provide the basic relationships for time-of-flight measurement, phase measurement and Doppler measurement respectively [1]

$$r_\tau = c\tau = c\Delta t = c(t_{\mathrm{Rx}} - t_{\mathrm{Tx}}) \tag{6.1}$$

$$r_\varphi = \frac{\lambda}{2\pi}\Delta\varphi = \frac{\lambda}{2\pi}(\varphi_{\mathrm{Rx}} - \varphi_{\mathrm{Tx}}) = \frac{c}{2\pi f}(\varphi_{\mathrm{Rx}} - \varphi_{\mathrm{Tx}}) \tag{6.2}$$

$$\dot{r} = -c\frac{\Delta f}{f_{\mathrm{Tx}}} = c\frac{f_{\mathrm{Tx}} - f_{\mathrm{Rx}}}{f_{\mathrm{Tx}}}. \tag{6.3}$$

In the above equations, r denotes range, $\dot{r}$ denotes range rate, Δ denotes arithmetic difference, c denotes the speed of light in vacuum, τ denotes time-of-flight, λ denotes carrier wavelength, φ denotes carrier phase, and f denotes frequency. The subscripts Rx and Tx denote the receiver and the transmitter, respectively. These observations can be performed using one-way or two-way ranging signals. In the former method, the transmitter is located on one platform and sends the ranging signal to a different platform where it is received and processed. In the latter method, the transmitter sends a signal to the other platform where it is either reflected (radar) or received, amplified, re-transmitted at a different frequency (transponder) and finally received and processed at the original transmitting platform. In case of a two-way ranging method, the result obtained from (6.1) to (6.3) needs to be divided by two to obtain the range (rate) between the platforms. Note that in case of using of a transponder the re-transmission delay at the transponder has to be known and needs to be stable over time. In addition, to obtain accurate results, the relative motion between the vehicles during the signal travel time needs to be taken into account. This is not done in (6.1), (6.2) and (6.3). Phase measurements can be used to construct highly accurate range and angular measurements, which are treated in Sect. 6.2.1.3.

For all ranging methods, the measurements are affected by errors. The most important error sources for space-based relative navigation are:

- *Ionosphere (if present)*: The free electrons in this dispersive medium cause a frequency dependent carrier phase advancement (i.e., shorter apparent range)

and group (code) retardation (i.e., longer apparent range) of a modulated RF signal. As the effect is inversely proportional to the square of the frequency, the induced error can be eliminated by performing measurements using multiple frequencies and adding these in a linear fashion to form a 'synthesized' ionosphere-free measurement. This comes however at the price of increased contributions from other error sources if these are uncorrelated with signal frequency. Alternatively, knowledge on the variation of total electron content (TEC) can be used to mitigate this effect in case of a single frequency ranging system. If measurements from different receiver antennas are differenced on the same satellite, this error is cancelled.

- *Multipath*: This phenomenon is caused by the reception of reflected replicas of the desired signal. Since the path traveled by a reflection is longer than the direct path, multipath signals are delayed with respect to the direct signal. In addition, the signal power of the multipath signals is less than that of the direct signal. However, since the reflection is caused by the receiving satellite structure, the delay is very small (order of ns) and is thus very difficult to separate from the direct signal. Currently, mitigation of this effect can only be achieved by means of calibration, which is further elaborated in Sect. 6.4.3.
- *Thermal noise in receiver hardware*: This results from random variations in current or voltage caused by the random movement of electrons due to thermal energy. Thermal noise can be reduced by reducing the temperature of the circuit.
- *Receiver resolution*: Due to the use of digital equipment, the measurement accuracy is limited by the granularity of the receiver equipment (e.g., correlator spacing, analog-to-digital quantization error).
- *Receiver bias*: Signals are delayed as they travel through the antenna, analog hardware (RF and intermediate frequency (IF) filters, low noise amplifiers (LNAs), mixers) and digital processing until the point where the measurements are physically made. This causes an artificial ranging bias which must be calibrated out. However, as this bias is not constant over time due to temperature differences and component aging, the system must be able to perform self-calibration once in orbit. Impedance mismatch causing hardware-induced multipath must also be prevented.
- *Phase windup*: A circularly polarized antenna's phase depends directly on the antenna's orientation with respect to the signal source. As a result, the observed carrier phase depends on the relative orientation of the transmitting and receiving antennas as well as the direction of the line of sight between them. Changing the receiver antenna orientation changes the reference direction and thus the measured phase. Similarly, changing the orientation of the transmitting antenna changes the direction of the electric field at the transmitting antenna and subsequently that at the receiving antenna. The result is also a change in the measured phase. As one or both of the antennas rotate, the phase change accumulates and is referred to as phase wind-up. In addition, the rotation of the receiving antenna causes an apparent change in the carrier frequency. It is distinguished from the normal Doppler shift in that phase wind-up is carrier-

frequency independent and does not affect ranging modulation group delay. The polarization-induced frequency shift is sometimes called rotational Doppler.

- *RF signal interference*: RF-based relative navigation in space will predominantly be affected by self-interference. Depending on the multiple access scheme chosen for the inter-satellite links, this interference can be in-band or out-of-band. When the received signals have similar power levels and are orthogonal to each other, the receiver can acquire the correct signal with little difficulty. Only when the power level of the interfering signal is much higher than the power level of the desired signal there is a chance that the receiver cannot lock onto the desired signal since it is 'drowned' in the noise caused by the interferer. Telemetry (TM) and telecommand (TC) signals can also interfere harmfully with the inter-satellite links, as is the case in the PRISMA mission. There, this is solved by applying appropriate filtering [2].
- *Antenna phase center location*: Since the true phase center of an antenna can be offset from its geometric center by several mm, this is a significant error source for very accurate range measurements. The antenna phase center location is also a function of pointing angle (azimuth and elevation) and will thus change if there is any significant relative motion. Uncertainties in the location of the antenna with respect to the satellite center of mass further contribute to this error. Extensive calibration is needed to minimize this error.
- *Relative clock drift*: As for navigation with GNSS, the clocks commonly applied on formation flying satellites exhibit non-negligible relative drift in time and are not actively synchronized. This causes a non-negligible bias in the range measurement in case of one-way ranging. This bias can be estimated using an additional range measurement. This is further elaborated on in Sect. 6.2.1.2.

6.2.1.2 Code and Carrier Observations

Code and carrier observations are both phase shift measurements and are therefore ambiguous in nature since it is unknown, without extra information, how many complete cycles are between the transmitter and the receiver. The only situation for which this ambiguity is not present is for code and carrier periods longer than the distance to be measured. For all other cases, extra information is needed to resolve the ambiguity. For any signal modulated with a ranging code, carrier phase observations, due to their shorter period, will result in more accurate range estimates than code observations. Therefore, it is highly desirable to resolve the ambiguity on the carrier phase. A typical phase measurement accuracy is 0.1 rad (1σ) [3] for code and carrier phase measurements, leading to measurement accuracies of ~1.6 m for code wavelengths of 100 m and ~2.2 mm for carrier frequencies of 2.2 GHz (S-band). Unfortunately, resolving the ambiguity on the carrier phase measurement is difficult. It is treated briefly in Sect. 6.2.2.3. For now, we assume this ambiguity to be known.

For space systems employing one-way ranging, due to non-synchronized clocks on the transmitter and the receiver, a relative clock bias will exist that will greatly

increase the error in the range measurement. Thus, this bias needs to be resolved. This is done using two range measurements, one from platform A to platform B and another one from platform B to platform A. This method is known as dual one-way ranging and is common practice in RF-based relative navigation. Throughout the remainder of this chapter, it will be assumed that this method is used to perform the inter-satellite range measurements. Following Ref. [4], dual one-way ranging results in the following fractional phase measurements

$$\begin{aligned}\varphi_{\mathrm{A}}^{\mathrm{B}}(t_1) &= [\varphi_{\mathrm{A}}(t_1) + \delta\varphi_{\mathrm{A}}(t_1)] - [\varphi_{\mathrm{B}}(t_1 - \tau_1) + \delta\varphi_{\mathrm{B}}(t_1 - \tau_1)] + E_{\mathrm{A}} \\ \varphi_{\mathrm{B}}^{\mathrm{A}}(t_2) &= [\varphi_{\mathrm{B}}(t_2) + \delta\varphi_{\mathrm{B}}(t_2)] - [\varphi_{\mathrm{A}}(t_2 - \tau_2) + \delta\varphi_{\mathrm{A}}(t_2 - \tau_2)] + E_{\mathrm{B}}\end{aligned} \tag{6.4}$$

where $\varphi_{\mathrm{A}}^{\mathrm{B}}(t_1)$ and $\varphi_{\mathrm{B}}^{\mathrm{A}}(t_2)$ represent the phase measurements at times t_1 and t_2 at receivers A and B respectively due to signals transmitted at times t_1–τ_1 and t_2–τ_2. These consist out of the difference between the received phase $\varphi(t)$ and the reference phase $\varphi(t–\tau)$, phase noise due to oscillator instability $\delta\varphi$ and a lumped error term E including ionospheric delay, signal multipath, hardware induced noise, etc. The signal travel time, or time-of-flight, is denoted as τ. Addition and subtraction of (6.4) provides estimations of the true phase difference $\Delta\varphi$ and clock bias $\Delta\delta\varphi$ as follows

$$\widehat{\Delta\varphi} = \frac{\varphi_{\mathrm{A}}^{\mathrm{B}}(t_1) + \varphi_{\mathrm{B}}^{\mathrm{A}}(t_2)}{2} = \frac{\Delta\varphi_1 + \Delta\varphi_2}{2} + \frac{\Delta\delta\varphi_1 - \Delta\delta\varphi_2}{2} + \frac{E}{2} \tag{6.5}$$

$$\widehat{\Delta\delta\varphi} = \frac{\varphi_{\mathrm{A}}^{\mathrm{B}}(t_1) - \varphi_{\mathrm{B}}^{\mathrm{A}}(t_2)}{2} = \frac{\Delta\delta\varphi_1 + \Delta\delta\varphi_2}{2} + \frac{\Delta E}{2} \tag{6.6}$$

where

$$\begin{aligned}\Delta\varphi_1 = \varphi_{\mathrm{A}}(t_1) - \varphi_{\mathrm{B}}(t_1 - \tau_1) \quad &, \quad \Delta\varphi_2 = \varphi_{\mathrm{B}}(t_2) - \varphi_{\mathrm{A}}(t_2 - \tau_2) \\ \Delta\delta\varphi_1 = \delta\varphi_{\mathrm{A}}(t_1) - \delta\varphi_{\mathrm{B}}(t_1 - \tau_1) \quad &, \quad \Delta\delta\varphi_2 = \delta\varphi_{\mathrm{A}}(t_2 - \tau_2) - \delta\varphi_{\mathrm{B}}(t_2) \\ E = E_{\mathrm{A}} + E_{\mathrm{B}} \quad &, \quad \Delta E = E_{\mathrm{A}} - E_{\mathrm{B}}.\end{aligned}$$

If $t_1 \approx t_2 = t$, $\tau_1 \approx \tau_2 = \tau$, and if oscillator instabilities during the signal travel time are neglected (i.e., $\delta\varphi(t–\tau) = \delta\varphi(t)$) the above reduces to

$$\widehat{\Delta\varphi} = \frac{\varphi_{\mathrm{A}}^{\mathrm{B}}(t) + \varphi_{\mathrm{B}}^{\mathrm{A}}(t)}{2} = \Delta\varphi + \frac{E}{2} \tag{6.7}$$

$$\widehat{\Delta\delta\varphi} = \frac{\varphi_{\mathrm{A}}^{\mathrm{B}}(t) - \varphi_{\mathrm{B}}^{\mathrm{A}}(t)}{2} = \Delta\delta\varphi + \frac{\Delta E}{2} \tag{6.8}$$

A method to mitigate ranging errors due to clock bias and drift if $t_1 \neq t_2$ is detailed in Ref. [5].

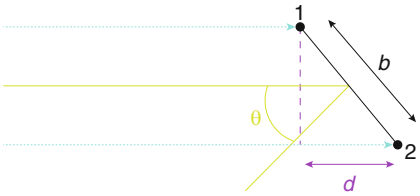


Fig. 6.2 Definition of LOS

The actual pseudorange ρ between the antennas consists out of the fractional phase as determined using (6.7) plus 2π times an unknown integer number of signal cycles N. For ambiguous code-based range measurements, N can be determined by providing data on the time of transmission of the ranging signal. When this is known, the phase measurement can be converted to meters by multiplying with the wavelength λ of the signal and dividing by 2π to obtain an estimate for the pseudorange ρ:

$$\hat{\rho} = \frac{\lambda_{\text{code}}}{2\pi}\widehat{\Delta\varphi} + \lambda_{\text{code}}N. \tag{6.9}$$

Similarly, the clock bias Δt is estimated as

$$\widehat{\Delta t} = \frac{\lambda_{\text{carrier}}}{2\pi c}\widehat{\Delta\delta\varphi}. \tag{6.10}$$

Note that in the above, we have made no distinction between code and carrier observations, which is made possible by collecting all non-oscillator related errors in a common error term E. The same approach can be applied to Doppler measurements to obtain estimates for the range rate and relative clock drift. The contribution of lumped errors practically vanishes for the clock bias estimation but is still significant for the pseudorange estimate. The effect of these errors can be reduced using techniques already treated in Sect. 6.2.1.1.

6.2.1.3 Line-of-Sight Observations

A very rough initial estimate of the Line-Of-Sight (LOS) between two satellites can be obtained by using code-based pseudorange observations and received signal power levels. This requires knowledge on the original power level of the transmitted signal and the variation of the receiving antenna's gain with the angle from the antenna boresight. The resulting LOS estimate typically has an accuracy of tens of degrees. When phase measurements are made at two antennas, it is possible to construct a much more accurate derived observation of the LOS of the antenna baseline with respect to the transmitter. This is depicted in Fig. 6.2 where the ranging signal of a faraway transmitter (dotted cyan arrows) arrives at antennas

1 and 2. When the path difference d (dashed purple line) and the antenna baseline length b (black arrow) are known, the LOS angle θ is computed as

$$\theta = \sin^{-1}\left(\frac{d}{b}\right). \tag{6.11}$$

With a third antenna, a second LOS angle can be obtained. These angles allow the creation of a unitary LOS vector and hence the direction of the transmitter is known in the local frame. Addition of a range measurement then allows estimation of the position of the transmitter in the local frame of the receiving satellite.

A difficulty of this method is that, for highly accurate LOS angle determination, the path difference needs to be determined using carrier phase measurements. Since only the fractional phase is measured, the integer number of carrier wave cycles in d is generally unknown. Relative attitude information from the ADCS cannot be used to solve this ambiguity. Therefore, there exists an integer ambiguity problem on both LOS and range. A long wavelength for the carrier signal can potentially remove the ambiguity problem, but then the phase measurement is relatively inaccurate, leading to a poor LOS estimate. Methods to solve the integer ambiguity problem are treated in Sect. 6.2.2.3.

6.2.2 Relative State Estimation

This section deals with all aspects involved in the estimation of the relative state of two formation flying satellites. We will start by briefly mentioning the states that are typically to be estimated. Next, some commonly used dynamic models that describe the relative motion in the local orbital frame are introduced. We then have a look at some methods that are utilized to resolve the integer ambiguity in the carrier phase measurements, which is vital for accurate navigation solutions. Lastly, we briefly touch upon the estimation itself.

6.2.2.1 State Vector Components

The estimation of the relative state of the satellites relies on the treatment of various measurements. These measurements can be obtained from the RF-based relative navigation system, but can also be provided by other sensors (GNSS sensors are not considered here). We will first discuss the states that can only be estimated by making use of the measurements from the RF-based relative navigation system. Thereafter we will treat the states that can be estimated using different sensors.

Assuming that the parameterization is performed using rectilinear coordinates, the relative state vector will contain at least three states x, y, z to describe the relative position $\mathbf{r} = (x, y, z)^T$, three states $\dot{x}, \dot{y}, \dot{z}$ to describe the relative velocity

$\mathbf{v} = (\dot{x},\ \dot{y},\ \dot{z})^T$, and a state Δt to describe the relative clock bias. Next to these seven 'basic' states, three states to describe non-modeled (empirical) relative accelerations a_x, a_y, a_z can be used to improve the estimation results. The integer ambiguities N in the phase measurements for all carrier signals used also need to be estimated if high accuracy is desired. The estimation of LOS biases Δx_{LOS} and Δy_{LOS} and a range bias Δr is also advisable as these can persist even after calibration. Furthermore, if the applied dynamic model allows for it, the differential atmospheric drag coefficient ΔC_D can also be estimated. The estimation of relative clock drift $\Delta \dot{t}$ will also improve the relative navigation solution. Depending on the orbital altitude, it can also be necessary to estimate the ionospheric path delay I. Thus, the number of states that can/must be estimated using an RF-based relative navigation system can be up to 18 (assuming a dual frequency ranging system) for a two-satellite formation.

The estimation of the relative position and velocity and the LOS bias requires an estimate of the relative attitude. This is needed to transform the measurements from the body frame to the local orbital frame. For the LOS biases this is needed since these are dependent on the relative attitude. In case the two satellites are equipped with multiple transmitters and receivers or if there are more than two satellites in the formation and if signals can be transmitted and received between all satellites, then it is in principle possible to estimate the relative attitude of the satellites using measurements from the RF-based relative navigation system [6]. If this is not the case, relative attitude must come from differencing the absolute attitude estimates of the satellites. To enhance the estimator's responsiveness to maneuvers, three states are needed to fully estimate all delta-V components. The estimation of these states requires information from onboard accelerometers.

6.2.2.2 Relative Dynamics Modeling

The relative motion of two satellites is commonly expressed in a co-orbiting local orbital reference frame. Many denominations exist for this frame but of these the Local Vertical Local Horizontal (LVLH), Radial-Tangential-Normal (RTN), and Hill frame are the most common. The orientation of this frame changes in time and can be inferred from the Earth-centered inertial position vector $\mathbf{R}$ and inertial velocity vector $\mathbf{V}$ of the origin of the reference frame. The reference frame is depicted in Fig. 6.3. The unit vectors in radial ($\mathbf{e}_R$), tangential/along-track ($\mathbf{e}_T$), and normal/cross-track ($\mathbf{e}_N$) directions are computed as

$$\mathbf{e}_R = \frac{\mathbf{R}}{\|\mathbf{R}\|} \quad , \quad \mathbf{e}_T = \mathbf{e}_N \times \mathbf{e}_R \quad , \quad \mathbf{e}_N = \frac{\mathbf{R} \times \mathbf{V}}{\|\mathbf{R} \times \mathbf{V}\|}. \tag{6.12}$$

In this subsection, two linearized relative dynamics models for the relative motion of two objects orbiting the same primary body are introduced. Both models assume unperturbed, i.e., Keplerian, relative motion, a small distance between the

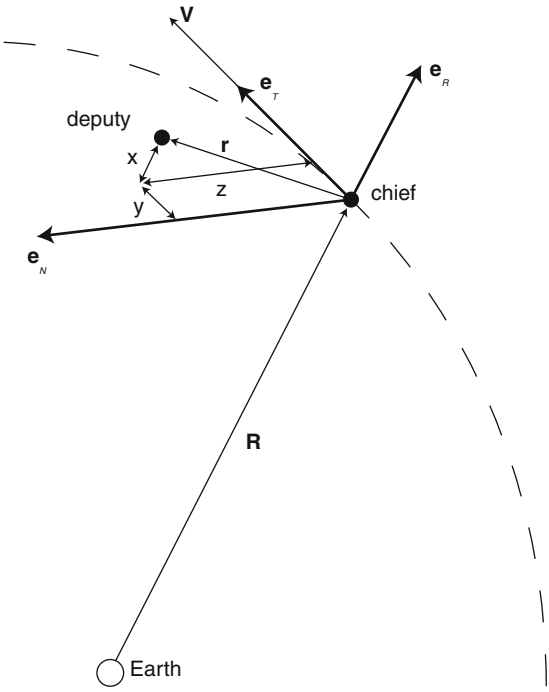


Fig. 6.3 The local orbital reference frame

two objects, and express the relative motion in the local orbital frame. For both models also one form of the homogeneous solution is presented.

The first relative motion model assumes a circular reference orbit, which leads to the well-known Clohessy-Wiltshire (CW) equations, presented here in state-space form as [7]

$$\dot{\mathbf{x}}(t) = \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \\ \ddot{x}(t) \\ \ddot{y}(t) \\ \ddot{z}(t) \end{pmatrix} = \mathbf{A}\mathbf{x}(t) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 3n^2 & 0 & 0 & 0 & 2n & 0 \\ 0 & 0 & 0 & -2n & 0 & 0 \\ 0 & 0 & -n^2 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \\ z(t) \\ \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} \tag{6.13}$$

where the single and double dot indicate, respectively, the first and second derivatives with respect to time and where x denotes radial direction, y denotes along-track direction, and z denotes cross-track direction, respectively. The variable n represents the mean motion of the chief, which is equal to $(\mu/a^3)^{1/2}$ where μ is the Earth's gravitational coefficient and a is the semi-major axis of the chief's orbit. As the system matrix $\mathbf{A}$ is time-invariant, the homogeneous solution of the CW equations may be formulated in terms of the state transition matrix $\mathbf{\Phi}_{\mathrm{CW}}(t, t_0)$ as

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) = \Phi_{\text{cw}}(t,t_0)\mathbf{x}(t_0) = \{t_0 = 0\}$$
$$= \begin{pmatrix} 4-3c_{nt} & 0 & 0 & s_{nt}/n & 2(1-c_{nt})/n & 0 \\ 6(s_{nt}-nt) & 1 & 0 & 2(c_{nt}-1)/n & 4s_{nt}/n-3t & 0 \\ 0 & 0 & c_{nt} & 0 & 0 & s_{nt}/n \\ 3ns_{nt} & 0 & 0 & c_{nt} & 2s_{nt} & 0 \\ 6n(c_{nt}-1) & 0 & 0 & -2s_{nt} & 4c_{nt}-3 & 0 \\ 0 & 0 & -ns_{nt} & 0 & 0 & c_{nt} \end{pmatrix} \begin{pmatrix} x(0) \\ y(0) \\ z(0) \\ \dot{x}(0) \\ \dot{y}(0) \\ \dot{z}(0) \end{pmatrix} \tag{6.14}$$

where $c_{nt} = \cos(nt)$ and $s_{nt} = \sin(nt)$. Equation 6.14 shows that, to first order, the in-plane motion is decoupled from the out-of-plane motion and that the along-track separation will drift linearly with time unless the semi-major axes of the two satellites are equal. Alternatively, it can be shown that the integration constants of the CW equations are equal to a set of relative states known as relative orbital elements. This allows expressing the homogeneous solution to the CW equations as a function of the relative orbital elements at $t = 0$, which is more intuitive than the rectilinear parameterization shown here. Treatment of this solution is beyond the scope of this work and the reader is referred to Refs. [8, 9] for an in-depth discussion.

The second relative motion model assumes an elliptic reference orbit, which leads to the Tschauner-Hempel (TH) equations [7]

$$\begin{aligned} \bar{x}'' &= \frac{3}{k}\bar{x} + 2\bar{y}' \\ \bar{y}'' &= -2\bar{x}' \\ \bar{z}'' &= -\bar{z}' \end{aligned} \tag{6.15}$$

where $(\cdot)'$ and $(\cdot)''$ indicate, respectively, the first and second derivatives with respect to the true anomaly f. The relative coordinates have been normalized with the factor $k = (1 + e\cos f)$, where e denotes orbit eccentricity. Several homogeneous solutions to the TH equations have been developed over time, but we will constrict ourselves to the solution provided by Yamanaka and Ankersen (YA), which is relatively compact and simple to program and valid for orbits with $0 \leq e < 1$. The YA state transition matrix $\mathbf{\Phi}_{\text{YA}}(t, t_0)$ can be represented as $\Phi_{\text{YA}}(t, t_0) = \Phi(f)\Phi^{-1}(f_0)$ where $\mathbf{\Phi}(f)$ and $\mathbf{\Phi}^{-1}(f_0)$ are given as

$$\Phi(f) = \begin{pmatrix} s & 2-3esI & 0 & c & 0 & 0 \\ c\left(1+\frac{1}{k}\right) & -3k^2I & 0 & -s\left(1+\frac{1}{k}\right) & 0 & 0 \\ 0 & 0 & \cos f & 0 & 0 & \sin f \\ s' & -3e\left(s'I+\frac{s}{k^2}\right) & 0 & c' & 0 & 0 \\ -2s & -3(1-2esI) & 0 & e-2c & 0 & 0 \\ 0 & 0 & -\sin f & 0 & 0 & \cos f \end{pmatrix} \tag{6.16}$$

$$\Phi^{-1}(f_0) = \frac{1}{\eta^2} \times \begin{pmatrix} -3s\frac{k+e^2}{k^2} & 0 & 0 & c-2e & -s\frac{k+1}{k} & 0 \\ 3k-\eta^2 & 0 & 0 & es & k^2 & 0 \\ 0 & 0 & \eta^2 c_f & 0 & 0 & -\eta^2 s_f \\ -3\left(e+\frac{c}{k}\right) & 0 & 0 & -s & -\left(c\frac{k+1}{k}+e\right) & 0 \\ -3es\frac{k+1}{k^2} & \eta^2 & 0 & -2+ec & -es\frac{k+1}{k} & 0 \\ 0 & 0 & \eta^2 s_f & 0 & 0 & \eta^2 c_f \end{pmatrix} \tag{6.17}$$

if the state vector is defined as $\bar{\mathbf{x}} = (\bar{x}, \bar{y}, \bar{z}, \bar{x}', \bar{y}', \bar{z}')^T$. In (6.16) and (6.17), $c = k\cos f$, $s = k\sin f$, $c_f = \cos f$, $s_f = \sin f$, $I = \frac{\mu^2}{h^3}(t - t_0)$, $\eta = \sqrt{1-e^2}$, and h is the magnitude of the orbital angular momentum vector. Note that if $e = 0$, $\Phi_{\mathrm{YA}}(t, t_0) = \Phi_{\mathrm{CW}}(t, t_0)$ [10].

6.2.2.3 Integer Ambiguity Resolution

As already discussed, obtaining the most accurate estimation of the inter-satellite range, LOS or relative attitude requires carrier phase measurements as these are much more accurate than code measurements. For the angular parameters, the carrier phase measurements need to be differenced between receiving antennas on the same satellite. A major complication here is that the measured phase differences represent a combination of geometry (the desired LOS of the remote transmitter) and instrumental offsets, including an integer-cycle ambiguity. As long as the geometry remains fixed, the two contributions are indistinguishable and the LOS indeterminate. In order to separate the two components of differenced phase, two integer ambiguity resolution (IAR) methods can be used and will be detailed in the following. The first method effectively adds extra information in the ranging signal which allows resolution of the integer ambiguities while the second method relies on a relative geometry change to separate the two contributions.

The first IAR method uses multiple signals with different frequencies, which allows the construction of artificial phase measurements by linearly combining the phase measurements of the individual signals. The goal here is to form multiple (if possible) so-called wide-lane signals, whose wavelength, λ_{WL}, is much longer than that of the original signals. The integer ambiguity on these long wavelength signals is much easier to resolve than for the shorter wavelength signals. Thus, if we use ranging signals S1 and S2 with frequencies $f_{\mathrm{S1}} = 2.25$ GHz ($\lambda_{\mathrm{S1}} = 13.3$ cm) and $f_{\mathrm{S2}} = 2.1$ GHz ($\lambda_{\mathrm{S2}} = 14.3$ cm) and with phase measurements φ_{S1} and φ_{S2} respectively, the resulting wide-lane wavelength λ_{WL} and the wide-lane phase measurement φ_{WL} are

$$\begin{aligned}\varphi_{\mathrm{WL}} &= m_1\varphi_{\mathrm{S1}} - m_2\varphi_{\mathrm{S2}} \\ \lambda_{\mathrm{WL}} &= \frac{c}{m_1 f_{\mathrm{S1}} - m_2 f_{\mathrm{S2}}}\end{aligned} \tag{6.18}$$

with $m_1, m_2 \in \mathbb{N}$. Choosing $m_1 = m_2 = 1$ results in $\lambda_{\mathrm{WL}} = 2\ \mathrm{m} \gg \lambda_{\mathrm{S1}}, \lambda_{\mathrm{S2}}$. Having two frequencies S1 and S2 allows the formation of two linearly independent wide-lane combinations WL1 and WL2 of S1 and S2. Using also a third frequency S3 increases the number of possible linearly independent wide-lane combinations to three. In addition, since the accuracy of the wide-lane measurement σ_{WL} scales as $\sigma_{\mathrm{WL}} = ((m_1\sigma_{\mathrm{S1}})^2 + (m_2\sigma_{\mathrm{S2}})^2 + \ldots + (m_{\mathrm{N}}\sigma_{\mathrm{SN}})^2)^{½}$, the magnitude of m should generally be kept small [11].

By now choosing the frequencies of the ranging signals in such a way that the wavelength of the different wide-lane combinations gradually reduces ($\lambda_{\mathrm{WL1}} > \lambda_{\mathrm{WL2}} > \ldots > \lambda_{\mathrm{S}}$), the integer ambiguity on the carrier phase of ranging signal S can be resolved in a cascaded approach. For example, assuming we want to remove the carrier phase integer ambiguity on a distance measurement, the process starts, if necessary, with smoothing the unambiguous but relatively inaccurate code-based pseudorange measurement with the much more accurate carrier phase measurements using a sufficiently long averaging time. When the accuracy of the smoothed pseudorange, $\sigma_{\rho,\mathrm{smoothed}}$, is better than half the wavelength of the wide-lane signal with the longest wavelength, i.e. $\sigma_{\rho,\mathrm{smoothed}} < ½\lambda_{\mathrm{WL1}}$, the ambiguity on this wide-lane signal can be resolved with high confidence. This unambiguous wide-lane signal is subsequently treated as a pseudorange measurement and smoothed using carrier phase measurements until its accuracy is better than half the wavelength of the next wide-lane signal ($\sigma_{\mathrm{WL1,smoothed}} < ½\lambda_{\mathrm{WL2}}$). This cascaded process continues all the way up to the carrier phase of the transmitted signal.

When choosing the signal frequencies to be used by the relative navigation system, a proper balance needs to be found between the magnitude of λ_{WL} and the difference between λ_{WL} and λ_{S}: A relatively large magnitude for λ_{WL} results in higher confidence for the IAR performed on the wide-lane, but results in a large difference between λ_{WL} and λ_{S}, which makes it hard to resolve the ambiguity on signal S with high confidence. The system design is further constrained by the frequencies that are allowed to be used for this application, to be discussed in Sect. 6.3.1.1, and by the increase in system complexity when the number of used signal frequencies is increased.

An example of an RF-based relative navigation sensor that applies this maneuver-free principle is the Autonomous Formation Flying (AFF) sensor, which utilizes a special "ultra-BOC" (Binary Offset Carrier) signal, cf. Fig. 6.4. The signal consists out of a central carrier that is modulated with a ranging code, two inner tones that are slowly modulated with data, and two unmodulated outer tones. First the pseudorange is determined using code measurements with accuracy of ~0.5 m. Then the closely spaced tones on one side of the carrier frequency are used to form a wide-lane (WL1) with 7.5 m wavelength. The code measurements can now be used to solve the ambiguity on this wide-lane. Now a second wide-lane (WL2) of 1 m wavelength

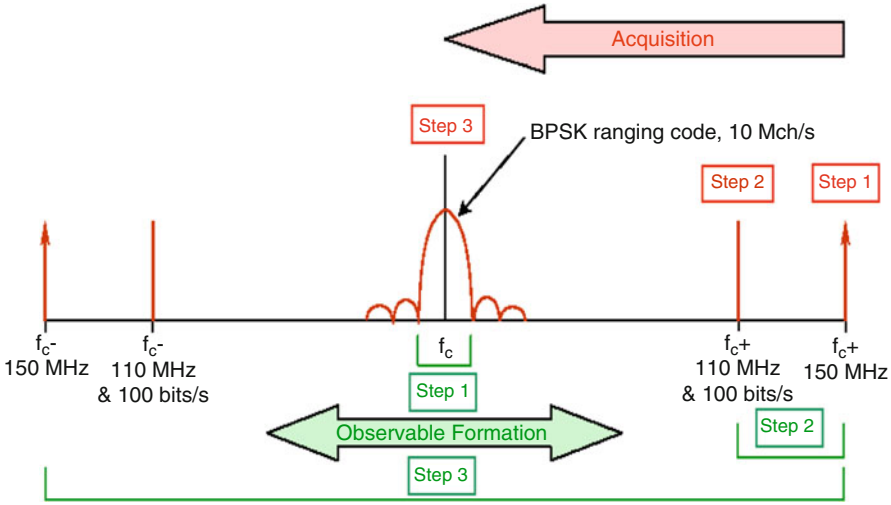


Fig. 6.4 Ultra-BOC signal structure and processing steps [12]

is formed using the two outermost tones. The unambiguous WL1 is now used to solve the ambiguity on WL2. In the last step, the unambiguous WL2 is used to resolve the phase ambiguity at the carrier phase itself [12].

The second method makes use of the fact that a precisely known change in the relative geometry should result in a predictable change in the measurements. Since multipath-induced bias and antenna phase center location are a function of the LOS, the geometry change is best not effected through a change in the LOS but through a real or apparent change in the inter-satellite distance along the LOS vector. An apparent change in the inter-satellite distance can be achieved by rotating the satellite around the boresight of the transmitting antenna in case of a circularly polarized signal. The change in the measured carrier phase will be different for different receiving antennas and is a function of the LOS. This allows removal of some candidates from the integer search space and selection of the correct integer for the LOS.

Naturally, IAR is preferably performed using a maneuver-free approach since maneuvers require precious time (usually minutes), energy, and need to be repeated whenever a receiver loses lock on a signal. This is still manageable for a small formation consisting out of just a few satellites, but not for formations consisting out of tens of satellites. The Formation Flying Radio Frequency (FFRF) instrument, which can support formations consisting of up to four satellites, uses a mixed approach: It uses two separate signals which can be used to form a wide-lane and thus allows for a cascaded approach. However, due to multipath biases, this strategy will not always result in high enough confidence levels to assure a correct IAR. Therefore, it initiates the IAR procedure using a satellite rotation to first solve the ambiguity on the LOS. The exact procedure for the FFRF is outlined in Sect. 6.4.2.

As a final note, IAR is widely recognized to be one of the most challenging aspects of highly accurate RF-based relative navigation, which is mainly due to the presence of various measurement biases, of which signal multipath is perhaps the most problematic. For example, differencing carrier phase measurements between antennas removes correlated noise sources (e.g., clock errors, ionospheric refraction effects) but amplifies uncorrelated noise sources (e.g., multipath). In addition, without careful design and calibration, instrumental biases can be different between antennas, resulting in a residual bias even after differencing measurements between antennas.

6.2.2.4 Estimation

Real time estimation of the relative states is typically performed using a reduced dynamics approach, which uses the highly precise measurements together with a dynamical model which is augmented with to-be-estimated empirical accelerations to cope with the imperfections of the dynamic model. A sequential filter such as an Extended Kalman Filter (EKF) performs relative and/or absolute orbit determination and prediction using the (smoothed) code and carrier measurements from both satellites. As there is a wealth of information on the general implementation of an EKF in the literature, the algorithms involved will not be discussed here. Note that in case of dual one-way ranging, the dual one-way measurements will be only available to the filter some time after the actual one-way measurement took place since the measurement data needs to be transmitted between the satellites. Thus, for real time onboard implementation, this needs to be accounted for in the filter innovation: The delayed measurements must be propagated to the time reference used by the EKF, taking into account relative dynamics if high accuracy is needed.

In case only inter-satellite range measurements are used as inputs for the EKF, the relative orbit estimation is hampered by the following issue: If a relative dynamics model is used in the estimator (e.g., CW, YA), the relative states are expressed in a local orbital frame, for which commonly the Hill frame is adopted. To enable conversion of the measured ranges and angles from the body frame to the Hill frame, the orientation of the satellite with respect to the Hill frame has to be known. Therefore, not only the satellite attitude, but also the orientation of the Hill frame has to be known on board. The latter can only be known when the absolute orbit of at least one of the satellites is known since the orientation of the Hill frame is a function of the absolute orbital parameters. This information can in principle be determined by the satellites themselves, but requires, in absence of GNSS measurements, a high-fidelity dynamic model and a specific attitude sensor suite. Therefore, it can be more practical to upload this information from ground.

The estimation of the relative states can be done centrally or distributed, depending on the number of satellites in the formation and the desired accuracy: If all measurements are collected and processed centrally in one large filter, the result will be more accurate than when many local solutions are computed which are based on limited information. Drawback of the centralized approach is that the

satellite which runs the filter requires a large amount of processing power (since matrix operations such as multiplication and inversion scale with the third power of the number of states to be estimated) and needs to be able to communicate with all other satellites in a timely fashion for command and control dissemination. In addition, in the centralized approach, if only one satellite is capable of performing the relative state estimation for the entire formation, this satellite represents a single point of failure for the formation and thus for the mission. This can be overcome by designing multiple satellites such that they can assume the role of 'leader', but comes at the cost of increased complexity of these satellites.

6.3 System Design Considerations

Relative navigation using RF signals imposes various design challenges and considerations on the complete formation system design. These are addressed on signal level and hardware level.

6.3.1 Signal Level Design Considerations

As the inter-satellite links between the satellites in the formation must enable both ranging and communication to allow relative navigation to be performed, it makes sense to combine these two functions in a single signal as this provides spectral efficiency and reduced satellite complexity. This subsection deals with the available frequencies for the signal, multiple access considerations, ranging signal design, and data modulation. It is noted that networking protocols are considered to be out of the scope of the current discussion.

6.3.1.1 Available Frequencies

RF frequency bands are regulated by the International Telecommunication Union (ITU). Table 6.1 lists the frequency bands between 1 and 100 GHz that are recommended for use in RF-based inter-satellite crosslinks. Frequency selection is also affected by the frequency used for the Telemetry, Tracking and Command (TT&C) subsystem. Additional filtering can be required when these frequencies do not have sufficient spectral separation. It is noted that the Ultra-High Frequency (UHF) band, due to its low propagation loss, is an interesting option for inter-satellite ranging if there are no high accuracy requirements, but requires regulatory changes to be put into effect to obtain the necessary frequency allocations. Conversely, higher frequency bands allow for higher ranging accuracy and lower multipath error, but require more transmitter power to compensate for increased free space loss, reduced effective C/N_0, and ultimately make the overall communication system more complex.

Table 6.1 Recommended frequencies for RF-based inter-satellite crosslinks [12, 13]

Band	S	Ku	Ka	W
Frequency range (GHz)	2.025–2.110	13.75–14.3	22.55–23.55 25.25–27.5	59–64
	2.200–2.290	14.5–15.35	32.3–33.4	65–71

6.3.1.2 Multiple Access

Accurate inter-satellite ranging requires the use of multiple access technologies to prevent mutual interference. The basic forms of multiple access are Frequency Division Multiple Access (FDMA), Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA). Hybrid forms of these can also be employed. Usage of FDMA and CDMA allows transmission and reception of ranging signals by multiple platforms at the same time, thereby preventing the need for scheduling as required when using TDMA. Downside is the potential for near-far interference which occurs when a high power signal transmitted by a platform close by completely drowns the signal transmitted by a platform far away. This problem is especially severe when the high power signal is produced by the receiving platform itself, which is known as self-interference, and is difficult to remove completely using hardware solutions (filtering, internal loop). FDMA is furthermore hampered by the need for a very wide frequency bandwidth for the entire formation in case the formation consists out of many satellites, complicating hardware design and ITU frequency filing.

The usage of TDMA will not lead to near-far interference problems, but it can be inefficient in case of highly varying separation distances between the platforms. In that case, the guard bands between different time slots must be large enough to prevent signals transmitted by far away platforms arriving after a nearby platform has started transmitting in its time slot. Then, there can be a large time interval between the dual one-way range measurements between two platforms, resulting in a relatively large clock drift and thus inaccurate results. In addition, to prevent constantly needing to reacquire a signal, the receiver needs to flywheel a certain channel in case the transmitter assigned to that channel is silent. However, when flywheeling is applied, if the TDMA duty cycle is too long, the clock will drift out of the delay lock loop (DLL) capture range, and signal reacquisition will need to be performed unless data is provided that allows dynamic compensation of the tracking loop [14]. The greatest challenge of TDMA is time synchronization with a maximum synchronization error that is equal to the propagation time of the ranging signal (e.g. smaller than 10 μs when the separation distance is less than 3 km). The usage of TDMA for multiple access is generally preferred for small formations (2–4 satellites) since it requires the least complicated hardware. Reference [12] presented variations on the traditional TDMA scheme which are deliberately non-synchronized, but still guarantee that all satellites receive each of

the others equally after a long period of time. These schemes remove the need for synchronization, but suffer from near-far interference (although self-interference is ruled out).

Hybrid forms of multiple access start to become attractive once the number of satellites in the formation gets larger than about five. A popular approach then applied is a combination of FDMA and CDMA [15]. The satellites within small sub-clusters in the formation then use the same frequency for their ranging signals and achieve spectral separation by means of CDMA. Spectral separation between the sub-clusters is achieved by means of FDMA. The choice of multiple access technology is also driven by the mission need. Sometimes half-duplex (TDMA) is sufficient, but full-duplex can be required for challenging missions.

6.3.1.3 Ranging Signal Design

RF-based inter-satellite ranging typically relies on a signal modulation technique called Direct Sequence Spread Spectrum (DSSS), which is also used for GNSS signals. Therefore, ranging signal design for formation flying satellites benefits tremendously from the work done in this field. DSSS uses a periodic high rate Pseudo-Random Noise (PRN) waveform to spread a low rate data signal modulated on a carrier wave over a wider bandwidth, reducing the effect of interference from other signals and allowing signal acquisition even when the signal power is well below the noise floor. The PRN waveform can be used for coarse range measurements by correlating a locally generated replica with the received PRN waveform. The finite sequence of bits used to generate the PRN waveform over one period is referred to as a PRN *sequence* or *code*. Assigning each transmitter with its own PRN code and using PRN codes with low cross-correlation properties allows unambiguous identification of the transmitter at the receiver and CDMA-based operation.

Ranging by means of a DSSS ranging signal occurs in two steps. First, a coarse two-dimensional search over frequency and time is performed over the entire PRN code length, with steps of one code chip, and a representative frequency range, also with fixed steps, to find the ranging signal by means of correlation with the local replica. Once the signal has been acquired, the tracking phase starts in which a DLL is used to maintain lock onto the signal and to obtain an accurate measurement of the code phase, also by means of correlation, and a phase lock loop (PLL) is used to measure the carrier phase and to aid the DLL. The PLL does this by forming a phase vector of inphase (I) and quadrature (Q) components of the signal. Due to the presence of data bits in the signal, of which the sequence is not known a priori, a Costas PLL loop has to be used to extract the data contained in the signal.

For code ranging, the phase measurement discussed in Sect. 6.2.1.2 can be made unambiguous by providing information on the time of transmission of the code. This is done by means of the data signal modulated on the same carrier. Therefore, code phase measurements can be treated as time delay measurements. New GNSS signals have one signal component containing a PRN code and data (data component) and one signal component with only a PRN code (pilot component). The latter allows for

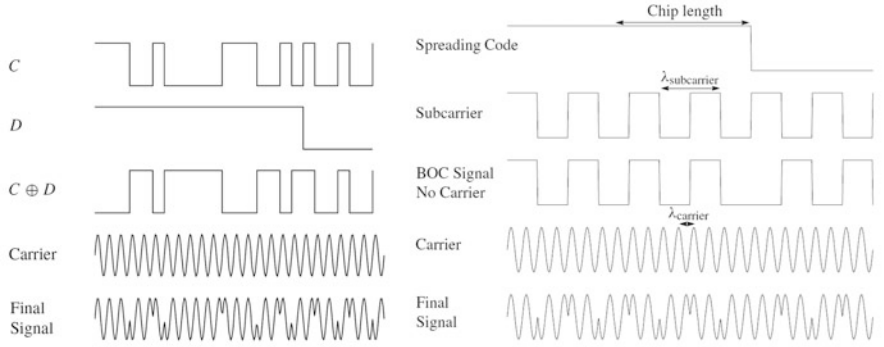


Fig. 6.5 DSSS modulation using BPSK-R (**a**) and BOC (**b**) (in (**a**), "C" is code and "D" is data and in (**b**), navigation data is not shown) [17] (With kind permission from Springer Science + Business Media B.V.)

longer integration times and therefore higher ranging accuracy. Typically, both components are assigned equal power levels in the combined signal.

The minimum time interval between transitions in the PRN waveform is known as the *chip period* T_c, the portion of the PRN waveform is referred to as a *chip* or *spreading symbol* and the reciprocal of the chip period is known as the *chipping rate* f_c. The traditional modulation used for the chips is Binary Phase Shift Keying (BPSK), often referred to as BPSK-R due to the rectangular shape of the spreading symbols. In recent years, BOC signals and their derivatives (alternate BOC (altBOC), Composite BOC (CBOC), Multiplexed BOC (MBOC)), have made their way to implementation on GNSS systems like Galileo and modernized GPS, but not yet to inter-satellite ranging. A BOC(m, n) signal is created by modulating a sine wave carrier with the product of a PRN spreading code and a square wave subcarrier having each binary ± 1 values. The parameter m stands for the ratio between the subcarrier frequency and the reference frequency, usually 1.023 MHz, and n stands for the ratio between the code rate and the reference frequency. Thus, BOC(10, 5) means a 10.23 MHz subcarrier frequency and a 5.115 MHz code rate.

For ranging signals, chip pulse design is limited by practical hardware limitations. Since the ranging accuracy depends on the ability of the hardware to faithfully reproduce the spreading waveform, the use of signals that can be generated using simple digital means is highly preferred. Spreading symbol shapes under consideration for future implementation are Binary Coded Symbols (BCS), Multilevel Coded Spreading Symbols (MCS), Composite BPSK (CBPSK), Sinusoidal Offset Carrier (SOC) signals, and Prolate Spheroidal Wave Functions (PSWF) [16]. Figure 6.5 depicts the principles of BPSK-R and BOC modulation.

The chipping rate of a ranging signal influences the ranging accuracy in that a higher chipping rate results in more accurate code ranging for the same pulse shape due to a sharper autocorrelation peak, cf. Fig. 6.6. This comes at the price of larger signal bandwidth. The pulse shape also affects the autocorrelation function. Figure 6.6 shows that BOC(1,1) has a much sharper correlation peak than the

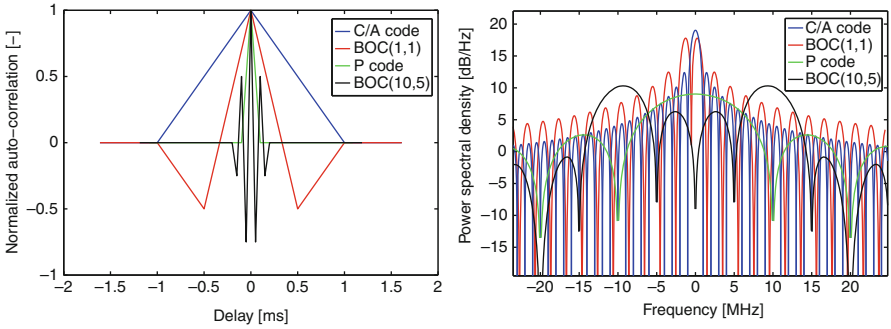


Fig. 6.6 Normalized auto-correlation and amplitude spectrum properties of the GPS C/A, GPS P (Y), BOC(1,1) and BOC(2,1) code

GPS C/A code, which uses a rectangular pulse, even though their chipping rate is the same. However, the correlation function of BOC signals has side extremes that are not present for rectangular pulses, negatively influencing their acquisition and tracking robustness since the receiver can potentially lock onto the wrong peak in case of for instance significant multipath.

Using the Cramér-Rao lower bound theory, it can be proven [18, 19] that for the best code ranging accuracy the signal energy should be concentrated as much as possible at the band edge of the signal This way, a gain (or actually a decrease) of several dB can be obtained in the required carrier to noise ratio (C/N_0) for a desired code ranging accuracy [18]. In that respect, out of the presently used chip pulse shapes, BOC modulation is inherently better than BPSK-R modulation. However, tracking and acquisition robustness must also be considered, which means that the amplitude of the side extremes of the autocorrelation function must be limited as much as possible. There, BPSK-R has a distinct advantage over BOC. Band limiting of the signal and spectral separation can also be important parameters for formation flying ranging signal design, but this depends greatly on the multiple access technology used in the formation. For instance, if FDMA is used, band limiting is of crucial importance to avoid large guard bands or complex filter designs, but when TDMA is used, this is of much less concern.

Multipath is an important performance limitation for current designs. Important parameters in this respect are the multipath-to-direct signal power ratio, the multipath carrier phase difference with the direct path, the multipath delay, and the signal frequency. The detrimental effects of multipath are best mitigated using small correlator spacing, large precorrelation bandwidths (to capture as much of the signal energy as possible) and a high code chipping rate (sharp correlation peak). For satellites, multipath delay is caused by the satellite structure and since the satellite dimensions are on the order of meters at most, the multipath delay is in the order of nanoseconds. Unfortunately, in that region, code and carrier multipath error is practically similar for different pulse shapes and chipping rates,

so a different signal design will hardly alleviate this problem. Current practice for multipath mitigation in space is calibration and will be detailed in Sect. 6.4.3.

The PRN code is, as the name suggests, a code designed mimic white noise but which is actually deterministic. The most famous example is the set of codes devised by Gold [20] which are used in the GPS C/A codes. There, the Gold codes have a length of 1,023 chips and have very good cross-correlation properties, making it possible to track multiple codes transmitted at the same time and with comparable C/N_0. Longer codes will approximate white noise better and will thus exhibit better cross-correlation properties. On the other hand, shorter codes lend themselves well for rapid acquisition since a smaller search space has to be covered. It is for that reason that GPS uses the short C/A code, which was initially not intended to be used for actual positioning, and the long P-code. This strategy is also applied in the Star Ranger instrument [21]. Downside of this strategy is that a relatively long search has to be done over time to acquire the longer code. To reduce this search time, so-called tiered codes have been developed. There, a primary code is modulated with a short secondary code, resulting in a long combined code, allowing the user to first quickly acquire the short primary code and then quickly switch to tracking of the long code. Some PRN code examples are Maximum Length Sequences (MLS, they form the basis for most of the codes used nowadays for GNSS), Kasami codes, Weil codes, and Random Memory codes [22].

6.3.1.4 Data Modulation and Bit Errors

Autonomous[1] formation flying requires the distribution of data between the satellites in the formation. Navigation data such as relative distance, velocity, attitude, and time information is essential for formation control purposes. Engineering data such as satellite health should also be exchanged to enable fault detection, isolation and recovery (FDIR). Even science data can potentially be distributed between satellites if for instance a star topology is used in which one master satellite transmits all science data to the ground. The volume and transmission frequency of the navigation data is tightly coupled to the nature of the mission. In case of tightly cooperating satellites in close formations (separation distances <1 km) with high positioning accuracy and tight control windows, the frequency of broadcast of navigation data can be on the order of seconds or even continuously [23]. Data rates can vary from kbit/s for missions with rather relaxed requirements to Mbit/s for challenging missions with a distributed control architecture.

Traditionally, this data is modulated on top of the PRN code using BPSK. Then, the achievable data rate is limited since the width of a data bit cannot be smaller

[1] In this chapter, "autonomous" implies that all the information which the system (the formation) requires for the guidance, navigation, and control (GNC) of the relative state of the satellites in the formation is obtained internally in the system itself. Thus, external information (e.g., GNSS measurements, telecommands) is not required for GNC of the formation.

than the code period because this would lead to failure of the acquisition process. In addition, acquisition is made easier when the data bits are at least several code periods wide since this allows for longer integration times and thus higher C/N_0. However, when separate data and pilot channels are used, this argument no longer holds. It is noted that the FFRF instrument performs data modulation differently than traditional GNSS signals as it uses quadrature PSK (QPSK) modulation for one of its two ranging signals. There, it modulates the PRN code on the inphase channel and the navigation and engineering data on the quadrature channel, thereby somewhat removing the constraint on the width of a data bit (which is still limited by the data demodulation capability of the digital hardware) [13].

Naturally, the transmitted data should be partitioned into frames and subframes of known length in order for the receiver to be able to recover the information. A preamble with a fixed sequence of zeroes and ones should be implemented to allow frame synchronization using a bit parity check and to solve the inherent 180° phase ambiguity of the Costas PLL. The subframe length, the data contained in each subframe, and the ordering of the subframes depend on many considerations and are considered to be out of scope. Data that can potentially be subdivided into separate subframes can be timing data, measurement data, absolute position and/or attitude data, spacecraft health, commands, and possibly payload data. To reduce the required transmission power but still achieve low bit error rate (BER) the data can be coded, for instance using convolutional coding.

Next to these considerations, it is beneficial when the signal transmitted from the satellite has a constant power envelope, i.e., the total transmitted power does not vary over time. Thus, the transmitted information is not contained in the signal amplitude and the transmitted signal amplitude becomes less critical. This is a very desirable property of the signal since it allows the use of efficient "class C"-like power amplifiers [17].

6.3.2 Hardware Level Design Considerations

Application of an RF-based relative navigation system on a satellite places specific demands on the hardware used on the satellite. Naturally, there are requirements on the hardware for the relative navigation system itself, but also the satellite design is affected. In addition, the accuracy of the signal generation and processing depends on the quality of the on-board clock. All these aspects will be explored in the next subsections.

6.3.2.1 Satellite Design

The RF-based relative navigation system will impact the satellite design mainly with respect to its layout. Protruding elements that are not part of the relative navigation system, e.g. solar panels and TTC antennas, should preferably be positioned such

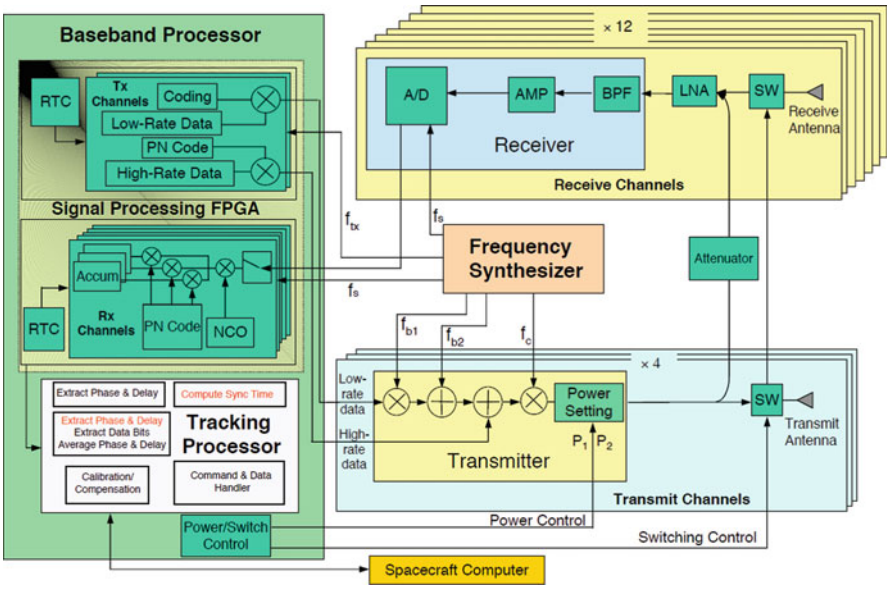
that multipath and RF interference contributions from these elements is reduced to a minimum. If a high level of onboard autonomy is required, the antennas of the relative navigation system itself should be mounted such that a 4π steradian field of view is obtained. This is mainly required to enable safe transition from a 'lost in space' configuration, where the relative positions of the satellites are unknown, to formation acquisition and finally formation maintenance. However, this can result in a very large amount of antennas needed: For the Terrestrial Planet Finder (TPF) mission, the antenna layout was complicated due to an unusual satellite shape caused by a large thermal shield, resulting in a total of 16 antennas on each satellite [12].

A trade-off has to be made between antenna beamwidth (i.e., antenna gain) and the total number of antennas: To reduce the influence of multipath for receiver antennas and to reduce the required transmission power for transmitting antennas, the antenna beam should be narrow (high gain). However, this increases the total number of antennas needed for full sky coverage. Purcell et al. [6] arrived at a circular beam with roughly constant gain out to 45° or 50° off axis and a sharp drop-off at larger angles. This result can be understood since in a formation, the satellites are commonly facing each other in a specific manner and the layout of the metrology system is tailored to that. Thus, the highest accuracy and therefore highest C/N_0 is usually required in a prescribed direction with little deviation from that direction. If one of the companion satellites is not in its desired position, the formation has not been acquired yet or has been lost. Then, the location and heading of the 'rogue' satellite does not need to be known with the highest accuracy possible, since the satellites are not in the desired configuration, and thus some loss in C/N_0 is tolerable.

Last but not least, the satellite platform should provide a mechanically and thermally stable environment to enable the most accurate measurements. This is put to the extreme in the GRACE (Gravity Recovery and Climate Experiment) mission where the entire satellite design is driven by the need for micrometer-level accuracy (~10^{-4} cycle) in the carrier phase measurements [24].

6.3.2.2 Transceiver Design

The 'heart' of the RF-based relative navigation system is the transceiver. There, the RF signals are generated to be transmitted and received RF signals are processed. A frequency standard, the local oscillator, and frequency synthesizers generate the high and low frequency signals needed to generate and process the desired RF signals. The receiver part consists out of an RF front-end where the received signals are filtered, down converted and digitized and a digital signal processor (DSP) where code and phase measurements are extracted from the digitized data. The transmitter part generates and modulates the PRN code and the data in baseband, modulates this onto a high frequency carrier in the front-end, and passes this through a power amplifier and finally an antenna. A complication in the transceiver design is that for each transmitting chain, the transmitted carrier and



Tx = Transmitter
RTC = Real Time Clock
PN = Pseudorandom Noise
Rx = Receiver
A/D = Analog-to-Digital
AMP = Amplifier
BPF = Band Pass Filter
SW = Saw Filter
LNA = Low Noise Amplifier
NCO = Numerically Controlled Oscillator

Fig. 6.7 Block diagram for the Acquisition Sensor of the TPF mission (After [12])

code phases are required to be coherent (to enable carrier-aided smoothing of the code measurements). Similarly, for each receiving chain, the local oscillator signal and the receiving channel digital clock are required to be coherent as well. Depending on the level of system integration, the transceiver can also contain a processing part that can handle various tasks such as relative state estimation and transmission power control. If the level of system integration is high, most if not all high-level processing needed for the relative navigation will be performed by the satellite's Guidance, Navigation, and Control (GNC) system. An example of a transceiver design is depicted in Fig. 6.7. This particular design is used to generate and receive the "ultra-BOC" signal discussed earlier.

The DSP can be implemented on a Field Programmable Gate Array (FPGA), which is relatively cheap and allows for a high level of modification in orbit, or an Application Specific Integrated Circuit (ASIC), which has higher performance than an FPGA, but is more expensive. Traditionally, ASICs were the preferred option, but present day there are commercially available radiation-tolerant high-density FPGAs that can rival ASICs in terms of performance.

When there are many satellites in the formation and if the multiple access technique chosen for the formation allows for it, adaptive power control can be

applied at the transmitter to minimize power consumption. Since the formation will be designed such that the ranging accuracy and data BER between the two most separated platforms meets the requirements, the C/N_0 between all the other platforms will be higher than required and thus inefficient. As C/N_0 can be measured by the receiver and relayed to the transmitter, the transmitter can, by propagating the expected relative motion forward in time, predict which output power should be produced to achieve the desired C/N_0 with the least amount of power. However, depending on the amount of power control, it is necessary to dynamically adapt the code and phase tracking loops. This is a result of varying signal power that is received and variations in the data rate requirements. Even with power control, initialization of the power control method and initial signal acquisition requires the ability to adapt the signal acquisition loops to the received power level. In addition, due to inherent transmitter inefficiencies, significant power savings can be achieved by transmitting at the minimum required power level [15]. Often, there is simply a low and a high power setting available.

Internal calibration is needed to compensate for biases resulting from satellite thermal, electrical, and mechanical variations. This can be achieved by feeding some of the transmitted signal to the receiver to construct an artificial range measurement which consists out of the hardware delays only. This measurement is then subtracted from the true range measurement to remove the hardware-induced delays. This can be seen in Fig. 6.7, where some of the transmitted signal is attenuated and fed to the receiver. Note that the frequencies of the self-generated signal and of the received signal should be similar since then the instrumental effects on the two signals are the same and thus the self-calibration will lead to the best result.

If all satellites are transmitting and receiving at the same time, part of the signal transmitted by a satellite will inevitably leak into the receiver on the same satellite and potentially saturate the receiver front-end or overwhelm the other signals. Two remedies are available to deal with this without resorting to TDMA [6]:

- *Frequency separation*: All satellites transmit at widely spaced frequencies such that an appropriate filter can reject most of the self-signal.
- *Active rejection*: The received signal is correlated against a replica of the self-signal. The result is used to control the amplitude and phase of an "anti-self" component, added to the received signal, which effectively cancels the unwanted input. This complicates the receiver, but is very effective.

The design of the relative navigation system should allow for on-orbit re-calibration to account for different multipath environment, different antenna phase patterns, instabilities (thermal, electrical, mechanical) of the satellite, etc. Thus, if a multipath correction table, generated via on-ground calibration, is stored onboard and used to correct measurements, it should be possible to update it after in-orbit re-calibration. In addition, the design should be scalable to meet the demands of various missions and to provide cost-effectiveness.

6.3.2.3 Clock Stability

As already mentioned in Sect. 6.2.1.2, the stability of the onboard clock is an important parameter if very accurate navigation solutions are required. It also provides time to the rest of the satellite using pulse-per-second (PPS) signals. Due to clock drift, the frequency of the PPS will also drift over time. Note that if there are very stringent requirements on the absolute time reference used by the satellites, the maximum error in the PPS can be a more stringent requirement for theclock stability than what is asked for by the ranging system. The short-term stability of the oscillator is not only important to limit the relative clock drift during signal travel time, but also to guarantee good PLL tracking characteristics for small PLL noise bandwidths, which is a necessary condition for reducing the thermal noise error [25].

Quartz crystal oscillators (XOs) are nowadays the standard for onboard time keeping on non-GNSS satellites and this discussion will therefore be limited to this type of clock only. The stability of a frequency source can be described by starting with an oscillator whose output voltage $V(t)$ is given by [25]

$$V(t) = (V_0 + \varepsilon_V(t))(\sin(2\pi f_0 t + \delta\varphi(t))) \tag{6.19}$$

where V_0 and f_0 are the nominal amplitude and frequency, respectively, with corresponding errors $\varepsilon_V(t)$ and $\delta\varphi(t)$. The argument of the sine function in (6.19) is equal to the instantaneous phase $\varphi(t)$. The amplitude error for the output voltage is in practice not an issue since this can be mitigated using an amplitude limiter. The clock frequency can be modeled as [26]

$$f(t) = f_0 + \Delta f + (t - t_0)\dot{f} + \tilde{f}(t) \tag{6.20}$$

where f_0 is the nominal frequency, Δf is the frequency bias or offset, $\dot{f}$ is the frequency drift, $\tilde{f}$ is the random frequency error, and t_0 is the reference epoch. The time error in this clock at time t_1 (disregarding sign) is

$$\Delta t(t_1) = \Delta t(t_0) + \frac{\Delta f}{f_0}(t_1 - t_0) + \frac{\dot{f}}{2f_0}(t_1 - t_0)^2 + \int_{t_0}^{t_1} \frac{\tilde{f}(t)}{f_0} dt. \tag{6.21}$$

Differencing (6.21) between two free running clocks provides us with the relative time error at time t_1. Dual one-way ranging using range and Doppler measurements allows for the measurement of the constant and linear term in this equation. The first three terms in (6.21) represent systematic effects while the last term represents random frequency fluctuations, the size of which depends upon environmental effects and ageing. They are characterized by their variance and auto-correlation function. The Allan variance is a commonly used measure of the

Table 6.2 Oscillator comparison

	TCXO	MCXO	OCXO
Accuracy[a]/year [−]	2×10^{-6}	5×10^{-8}	1×10^{-8}
Aging/year [−]	5×10^{-7}	2×10^{-8}	5×10^{-9}
Temp. stability [−]	5×10^{-7}	3×10^{-8}	1×10^{-9}
(range, °C)	(−55 to + 85)	(−55 to + 85)	(−55 to + 85)
Allan deviation [−] ($\tau = 1$ s)	1×10^{-9}	3×10^{-10}	1×10^{-12}
Size (cm^3)	10	30	20–200
Warmup time (min)	0.03 (to 1×10^{-6})	0.03 (to 2×10^{-8})	4 (to 1×10^{-8})
Power (W) (at lowest temp.)	0.04	0.04	0.6
Approximate price (US$)	10–100	<1,000	200–2,000

The first four rows are expressed in terms of the fractional frequency error $\Delta f/f_0$ [27]
[a]Including environmental effects

timekeeping ability of a clock with respect to random processes. The Allan variance is defined as

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle (y_{i+1} - y_i)^2 \right\rangle \tag{6.22}$$

where y_i is a measurement of relative frequency deviation ($y = \Delta f/f_0$) of an oscillator averaged over the period τ, $(y_{i+1} - y_i)$ are the differences between pairs of successive measurements of y, and, ideally, $\langle \cdot \rangle$ denotes a time average of an infinite number of $(y_{i+1} - y_i)^2$. The nature of the random process associated with the measurements of relative frequency deviations changes with the size of the averaging interval τ and the Allan variance converges for all of them. The square root of the Allan variance is called the Allan deviation $\sigma_y(\tau)$. The r.m.s. error of a clock after an interval τ is approximated as $\tau\, \sigma_y(\tau)$.

Important frequency stability influences are temperature, time (aging and short-term stability), line voltage, and warm up. Typical performances of some oscillator compensation techniques are provided in Table 6.2. Other oscillator frequency influences are drive energy, gravity, shock, vibration, electromagnetic signals physically close to the oscillator, retrace, and hysteresis [28]. Line voltage specifications for an instrument refer to changes in the AC line supply power. Warm up is a special case of temperature variation which is brought about by the temperature rise from oscillator turn-on until stable operating point is reached. It may not be apparent that a temperature compensated crystal oscillator (TCXO), treated in the following, would have a warm up specification, and in fact, it is typically not specified. However, any instrument when placed into operation will generate a certain amount of heat. This heat elevates the temperature surrounding the crystal, and therefore, causes a frequency change. As is clear from Table 6.2, oscillator stability is very sensitive to temperature fluctuations. Since mass provides inertia against temperature changes and is therefore beneficial, miniaturization of oscillators is difficult.

The stability of a standard crystal oscillator without any means of compensation is too low to use for relative navigation in space. As temperature has a major effect on the frequency stability of an oscillator, most compensating designs focus on limiting the frequency change induced by a temperature change as much as possible. In a temperature compensated crystal oscillator (TCXO), the output signal from a temperature sensor (a thermistor) is used to generate a correction voltage that is applied to a voltage-variable reactance (a varactor) in the crystal network. The reactance variations produce frequency changes that are equal and opposite to the frequency changes resulting from temperature changes. In other words, the reactance variations compensate for the crystal's frequency versus temperature variations. In an oven controlled crystal oscillator (OCXO), the crystal unit and other temperature sensitive components of the oscillator circuit are maintained at a constant temperature in an oven. The crystal is manufactured to have a frequency versus temperature characteristic which has zero slope at the oven temperature. The best oscillator stability is achieved when the operating point is 15–20° above the highest temperature to which the unit will be exposed. A special case of a compensated oscillator is the microcomputer-compensated crystal oscillator (MCXO). The MCXO overcomes the two major factors that limit the stability achievable with a TCXO: thermometry and the stability of the crystal unit. Instead of a thermometer that is external to the crystal unit, such as a thermistor, the MCXO uses a much more accurate, "self-temperature sensing" method: Two modes of the crystal are excited simultaneously in a dual-mode oscillator. The two modes are combined such that the resulting beat frequency is a monotonic (and nearly linear) function of temperature. The crystal thereby senses its own temperature. To reduce the frequency versus temperature variations, the MCXO uses digital compensation techniques: pulse deletion in one implementation, and direct digital synthesis of a compensating frequency in another. Other than in an TCXO, the frequency of the crystal in an MCXO is not "pulled," which allows the use of high-stability crystal units [29].

For relative navigation, the short-term stability of an oscillator is the most important parameter. There, the OCXO clearly outperforms all other quartz-based crystal oscillators. However, its size and power requirements can be much larger than that of a TCXO or MCXO. Thus, if the oscillator-induced measurement noise is not driving the overall navigation accuracy, a TCXO or MCXO can be a better choice.

6.4 System Examples and Performance

During the last two decades, almost a dozen RF-based relative navigation systems have been designed for various distributed system missions. However, the vast majority of these missions were never launched and therefore extremely little on-orbit experience has been gained with this technology. This section will first briefly introduce the systems that never made it to space. Then, it will discuss the

only system with actual flight heritage: the FFRF. Note that systems that are not capable of providing standalone relative navigation solutions, such as the K-band ranging (KBR) system on the GRACE mission or the Lunar Gravity Ranging System (LGRS) on the Gravity Recovery and Interior Laboratory (GRAIL) mission, are not considered here. Typical testing and validation approaches are also treated. Finally, on-orbit performance characteristics, necessarily limited to the FFRF, are also provided.

6.4.1 Systems Without Flight Heritage

The following list provides a brief summary of RF-based relative navigation systems without flight heritage. All these systems have been developed in the USA and are based on GPS technology.

- *AFF (Autonomous Formation Flying)*: The AFF sensor has been developed by the Jet Propulsion Laboratory (JPL) in the USA and is based on the TurboRogue GPS receiver. The AFF was originally developed for the discontinued Deep Space-3 (DS-3) mission, the discontinued Space Technology-3 (ST-3) mission, the discontinued StarLight mission, and the postponed TPF mission (TPF was postponed in 2007 due to budget cuts). The AFF is also referred to as Acquisition Sensor.
- *CCNT (Constellation Communications and Navigation Transceiver)*: JPL also developed the CCNT, which was based on the AFF sensor. It was developed for the ST-5 mission, but although the ST-5 mission flew successfully from March 22, 1996 to June 29, 1996, the CCNT did not fly on ST-5.
- *CLT (Crosslink Transceiver)*: The CLT was developed by the Johns Hopkins University Applied Physics Laboratory (JHU APL) in the USA and based on their space-qualified spaceborne GPS Navigation System (GNS). A derivative of the CLT, the NanoSat CLT (NCLT) was intended to fly in 2004 on the ION-F mission, consisting of three university satellites. Three flight-ready NCLTs were delivered, but the mission never flew.
- *LPT (Low Power Transceiver)*: Development of the LPT started in fall 1998 as a joint effort by NASA Goddard Space Flight Center (GSFC) and the Advanced Engineering and Sciences department (formerly Stanford Telecom) of ITT Industries, Inc. in the USA. At a later stage, also the Air Force Research Laboratory (AFRL) got involved in the development since the LPT had been selected to fly on its XSS-11 satellite. Initially, the LPT was not intended to be used as a relative navigation sensor since it was not foreseen to generate its own ranging signal. It was a modular, reprogrammable, transceiver capable of autonomous navigation based on GPS data. At a later stage in its development, an inter-satellite ranging capability was added. From 1998 to 2006, the LPT has evolved over three generations and has flown on the Space Shuttle (2003), on the XSS-11 mission (2005), and on the TacSat-2 satellite (2006), but never as a

relative navigation sensor. A fourth generation, the Miniature Multifunctional Integrated Terminal or M2inT (also referred to as the Miniature Transceiver or MinT), was planned to be developed, but no substantial M2inT/MinT hardware has been produced so far.

- *Star Ranger*: The Star Ranger was developed by the company AeroAstro situated in Solana Beach, California, USA. The funding source for the development was the AFRL. It had been specifically developed for nano- and micro-satellites and was intended to fly on AFRL's cancelled TechSat 21 mission (cancelled in 2003).
- *IRAS (Intersatellite Ranging and Alarm System)*: IRAS is currently under development by NASA GSFC in the USA for the Magnetospheric Multiscale (MMS) mission which is slated for launch in 2014. It is composed out of the Navigator weak-signal GPS receiver and an S-band transceiver with ranging capability.
- *SPTC (Stanford Pseudolite Transceiver Crosslink)*: The SPTC was developed at Stanford University in the USA around the turn of the century. The SPTC was designed using commercial off the shelf (COTS) devices such as modems, L1 pseudolites, and a GPS receiver (Mitel GP2000). No specific mission was targeted for the SPTC.
- *RelNav-DDF (Relative Navigation) sensor*: The RelNav-DFF sensor is currently under development by Tethers Unlimited, Inc. (TUI) in the USA under funding from DARPA. No specific mission is targeted for the RelNav-DDF sensor and practically no data on it is publically available.

6.4.2 Systems With Flight Heritage

At this point in time, the only RF-based relative navigation system with actual flight heritage is the FFRF (Formation Flying Radio Frequency). The FFRF has been developed by TAS-France, TAS-Spain and GMV and is derived from the TOPSTAR 3000 GPS receiver. The FFRF sensor is being flight tested and validated on the Swedish PRISMA (Prototype Research Instruments and Space Mission technology Advancement) mission, launched in June 2010, and is foreseen by ESA and CNES to be used as the first metrology stage on future FF missions like PROBA-3 and Darwin. The FFRF sensor is designed to provide relative positioning (ranging and LOS measurements) for up to 4 satellites and for an operational range from 10 m to 10 km. It includes one RF terminal and up to four sets of antennas on each satellite. A set of antennas can be either a triplet, which is composed of 1 Receive/Transmit (Rx/Tx) master antenna and 2 Rx-only slave antennas (the triplet is required for LOS computation), or a single Rx/Tx antenna (allows ranging only). Using multi-antenna bases and TDMA sequencing, each terminal transmits and receives a GPS C/A navigation signal modulated on two S-band carrier frequencies (S1 = 2,275 MHz and S2 = 2,105 MHz) [30].

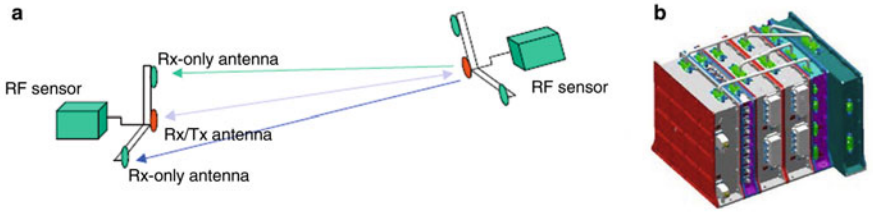


Fig. 6.8 FFRF ranging principle (**a**) and terminal architecture (**b**)

The inter-satellite ranging relies on two-way measurements to account for the relative clock drift of the platforms. First, range measurements are based on the unambiguous C/A code signal that provides meter-level accuracy. Next, centimeter-level accuracy can be achieved using carrier phase measurements as soon as the signal ambiguities ($\lambda_{S1} = 13.2$ cm) have been removed. Similarly, LOS measurements are obtained from the carrier phase differences between two antennas of the triplet antenna base and accuracy better than 1° can be achieved after removal of the signal ambiguity.

Carrier phase ambiguity resolution for distance relies on the dual frequency signal that allows to produce a larger wavelength synthetic signal (wide-lane, $\lambda_{WL} = 176$ cm). The procedure actually starts by removing the ambiguity on the LOS in order to get rid of the potential direction dependent biases due to multi-path effects. When the LOS is determined with the highest accuracy (<1°), biases can be compensated/mitigated and the ambiguity on the wide-lane can be solved by codefiltering. The last step consists in filtering the wide-line data over time to reduce the uncertainty level below the carrier phase wavelength and finally achieve the highest precision (1 cm). In the current design, the FFRF sensor cannot remove by itself the carrier phase ambiguity on the LOS. Ambiguity removal requires a rotation of the spacecraft along a direction roughly normal to the base of antennas. By processing the FFRF signals and using the satellite attitude information, the ambiguity can be removed after a rotation of about 50°.

The FFRF terminal is composed of the following hardware components, cf. Fig. 6.8b: RF front end (or RFE), Transmitter RF module, Receiver RF modules, Digital Processing Unit, DC/DC and OXCO unit. All modules are stacked together except for the RF filters (Rx and Tx) which can be accommodated separately. For the PRISMA mission, the FFRF terminal design differs for the two satellites (different RFE and Rx units) because they have different antenna configurations (the chaser, Mango, is connected to a unique antenna triplet whereas the target, Tango, carries three single Rx/Tx antennas). The masses of the terminals are 7 and 9 kg and their power consumption is between 23 and 30 W.

The FFRF sensor delivers at 1 Hz two types of metrology information: (1) RF signal raw data such as code-based pseudorange, carrier phase, and delta carrier phase from each terminal and clock bias, (2) Position, Velocity, Time (PVT) and LOS snapshot of the companion satellite. Each terminal implements software in

charge of computing the relative PVT from FFRF raw measurements (when IAR is completed, this data represents a measurement snapshot involving mostly geometric relations). The high level filtering functionality based on an EKF is implemented inside the GNC software running on the satellite On-Board Computer (OBC) and constitutes the main part of the relative navigation function. The PVT software removes the signal ambiguities with a specific algorithm combining signal smoothing techniques and use of external aiding data provided by the GNC system. The aiding data is twofold: (1) spacecraft attitude quaternions to solve the LOS ambiguity (which requires a rotation of the satellite) and (2) relative position data for signal acquisition and navigation: this allows speeding up the RF signal acquisition at sensor power-up or after tracking loss.

6.4.3 Testing and Validation

The testing and validation approach for the FFRF sensor is detailed in the following. Examples for other relative navigation sensors could also be provided, but would be very similar as the testing and validation approach taken is often very comparable to what has been done for the FFRF. Therefore, as the previous and next subsections are devoted to the FFRF, it has been chosen to limit the subject of this subsection to the FFRF as well.

End to end ground testing within the full operational range is a very challenging exercise for several reasons: (1) the need to provide long range mobility, (2) multipath effects induced by the ground testing configuration have to be avoided; (3) reference measurements for performance characterization must be provided. An ambitious approach would consist in accommodating the FFRF sensor on air vehicles like balloons or helicopters equipped with differential GPS to remove the ground-induced multipath effects. A more cost effective solution is to perform the validation in two successive steps described hereafter:

- *Build a high fidelity model through specific FFRF sensor characterization*: Sensor performances are measured inside an anechoic chamber in order to determine residual biases due to multipath effects but also uncertainties in the internal path delays and the position of the antenna phase centers – the collected data allows to build an accurate FFRF sensor numeric model (introduction of attitude dependent bias maps in the RF signal generator).
- *Extensive validation through simulations based on the high fidelity model*: Simulate the behavior of the RF navigation system by using a numeric simulator including a realistic model of the PRISMA satellites and the environment, the updated FFRF sensor model and the GNC algorithms – the navigation performance is characterized over the whole dynamic and geometric range.

The sensor characterization relies on engineering models that are mounted on representative satellite mock-ups from the radio-electric point of view (conducted and radiated tests). Both satellites are accommodated inside a 20 m anechoic chamber

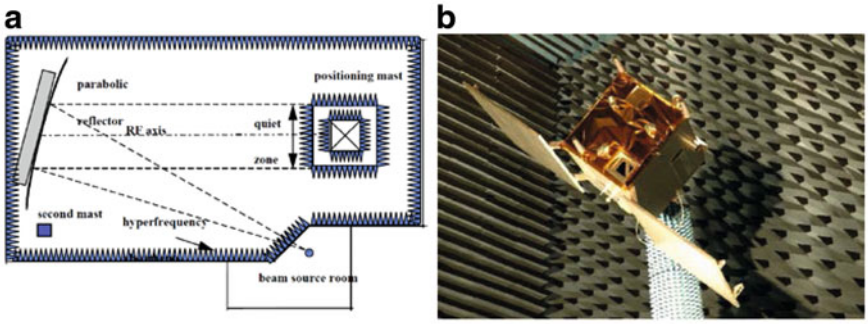


Fig. 6.9 The CNES anechoic chamber (**a**) and a radio-electric mockup of one of the PRISMA satellite (Mango) (**b**)

that is equipped with a parabolic reflector to generate parallel radio-electric rays, cf. Fig. 6.9. The radio reflector, along with signal attenuators, allows the simulation of long range relative configurations. Each satellite is successively mounted on a robotic arm to scan attitudes over a half hemisphere and thus measure its specific bias cartography. This phase allows also characterization of critical functionalities (signal acquisition conditions, IAR success rate). Afterwards, the FFRF sensor model, composed of a numeric RF signal/data processing simulator and the real sensor algorithms, is updated by feeding it with the bias maps.

The RF navigation validation has been performed in two types of simulator environments:

- A numeric simulator implemented in a Matlab/Simulink® environment that includes high fidelity models of the various PRISMA components, the space environment and the FFRF sensor. This flexible and computationally efficient simulator is dedicated to performance assessment over the full operational range through extensive test campaigns including Monte Carlo tests.
- The PRISMA avionics testbench that is a real time environment devoted to the satellites' flight software validation. This test-bench is composed of the satellites' avionics with the complete data handling architecture. Some real equipment is present and missing equipment is replaced by emulators that are interfaced via Remote Terminal Units (RTU). An emulator of the FFRF sensor built from the Simulink model is integrated in the testbench with interfaces to the data handling system and also the satellite simulator. This configuration allows validating the navigation software (built by auto-coding from Simulink models) in dynamic conditions and in relation with all critical architecture components. The purpose of the tests is to verify the full functional behaviour and identify peculiarities that are implemented afterwards in the numeric simulator to improve its fidelity level.

The final performance assessment is performed in flight. It relies on reference measurements provided by GPS receivers on both satellites which are later

Table 6.3 Conducted and radiated errors for the FFRF in a 30° elevation cone

FFRF measurement	Distance (mm)	LOS (mm)
Conducted	1.5	1.6
Radiated (1σ)	4	6
Total	5.5	7.6

post-processed on the ground for maximum relative positioning accuracy. The level of achievable accuracy is a few millimeters on three axes and is consistent with the validation purposes.

6.4.4 Achievable Performance

The achievable performance for the FFRF sensor is detailed in the following. The performance of the overall RF navigation system depends directly on the capability to solve the ambiguities during the FFRF sensor initialization phase. Characterization tests performed in the anechoic chamber have allowed verification of the success rate of the IAR process and determination of the level of potential residual errors:

- LOS ambiguity resolution with a rotation maneuver turns out to be very robust to multipath effects even for initial misalignments up to 80° (success rate = 100%).
- Distance ambiguity resolution is conversely very sensitive to residual electrical biases and multipath which produces a rather low success rate. The remaining bias error that is a half wavelength multiple can reach up to 80 cm.

In addition to the distance bias potentially provoked by a faulty IAR, the FFRF sensor is subject to conducted and radiated errors that reach in a 30° elevation cone the values reported in Table 6.3.

The achievable navigation performance is strongly dependent on the orbital environment. Low Earth Orbits (LEO) allow to reach the best performance since the orbital mechanics produce a strong coupling between the Cartesian components which speed up EKF convergence. This coupling offers also the possibility to estimate quasi-constant LOS biases due to position uncertainties (errors in the antenna mounting or phase center model) and multipath errors. Typical performance is shown in Table 6.4 for quasi stationary configurations in the PRISMA orbital environment (700 km altitude, circular orbit). They are given as statistical values (mean, standard deviation) computed over an orbital period after some convergence time (6,000 s).

The along track performance shows some degradation with distance due to ionospheric errors that are not compensated (a possible 80 cm bias due to distance IAR error must be added also to these figures). The estimation is quite efficient in the orbital plane with accuracy better than 1 cm below 1 km. Conversely, filter

Table 6.4 Relative navigation error statistics for the FFRF

	Distance (m)	300	1,000	3,000	6,000	9,000
Along track (m)	Mean	0.023	0.010	0.166	0.393	0.684
	Std. dev.	3e-4	2.4e-3	9e-4	0.004	0.013
Cross track (m)	Mean	0.03	0.046	0.131	0.220	0.030
	Std. dev.	0.050	0.335	1.792	6.824	16.524
Radial (m)	Mean	0.001	1.2e-4	0.002	0.004	0.006
	Std. dev.	6.9e-3	8.8e-3	0.011	0.022	0.036

convergence in cross-track is much slower due to the lack of dynamic coupling and the error reaches large values for a distance of several km. Fortunately, performance improves dramatically at short range with a 5 cm accuracy below 300 m.

6.5 Summary and Conclusions

The current section briefly summarizes the topics discussed in this chapter and then provides some conclusions. Some suggestions for future developments and focus points are also provided.

6.5.1 Summary

For distributed systems in space, knowledge of the relative position and velocity is typically required to maintain the relative geometry of the satellites within certain boundaries. This can be achieved using relative navigation based on GNSS, but it can be desired to (also) implement an autonomous relative navigation system based on RF signals. These signals are used to measure the range (rate) between the satellites, which allows, in combination with a relative dynamics model and the exchange of data between the satellites, an onboard estimation of the relative position and velocity of the satellites.

This process is however fraught with difficulties due to the challenging environment and due to hardware limitations. First of all, the signal design should allow for reliable and unambiguous range measurements but also allow for data exchange. When the system is composed out of many satellites, multiple access considerations become important since the signals should not interfere with each other but at the same time measurements should be conducted at a sufficiently high rate to guarantee high relative navigation accuracy and safety.

The hardware necessary to facilitate this has to be rugged enough to survive the rigors of space, but also flexible and adaptable whilst in orbit. Hardware-induced measurement biases should be minimized and, since a range measurement is basically a measurement of time-of-flight of a signal, the onboard clock has to be very

stable. On-ground calibration of these devices is therefore mandatory to achieve satisfactory performance in space. In the last two decades, systems have been developed that can autonomously perform RF-based relative navigation in space, but up to now actual operation of these systems in space has been very limited.

6.6 Conclusions

This chapter started with a discussion on the overall concept of relative navigation using locally generated radio frequency (RF) ranging signals known as Direct Sequence Spread Spectrum (DSSS) signals. With these signals, the phase shift of a code modulated on top of a carrier signal or the phase shift of the carrier signal itself between a transmitter and a receiver can be measured. This phase shift can be translated to a range if the integer number of code or carrier cycles between the transmitter and the receiver is known, which is trivial for the code-based measurements but hard for the carrier phase measurements. This technique is also applied in position determination using GNSS signals and is therefore well-established. The difference between autonomous relative navigation between satellites and position determination using GNSS signals is that the latter requires only one-way signals to be sent between the GNSS satellite and the user while autonomous relative navigation between satellites typically requires a two-way signal or dual one-way signals to be exchanged between the satellites. The dynamical models used for the relative state estimation usually are linearized relative motion models such as those from Clohessy-Wiltshire or Tschauner-Hempel, although absolute motion models are sometimes applied as well. A suitable estimator for the relative navigation is the Extended Kalman Filter (EKF). This can be applied centrally or distributed, depending on the system requirements.

The design of the ranging signal itself is firstly limited by the available frequencies. In addition, for systems with many satellites, multiple access solutions are required to prevent signal interference and to guarantee the timely exchange of data (e.g., measurements, housekeeping, command and control). For the coding of the signal, many options are present amongst which there are the spreading code type and length, the spreading symbol shape and width, the width of a data bit, the type of data modulation, and the type of error correction coding.

The hardware design has to be such that multipath and hardware-induced biases are minimized. Also, the relative navigation sensor should have a 4π steradian field of view if the system is to be fully autonomous. This results in a trade-off between antenna gain and total number of antennas. The transceiver used to generate and receive the ranging signals shall have a stable reference source and shall accurately reproduce the ranging signal even in the presence of component aging. Finally, it shall also have low mass and low power consumption. Depending on the level of system integration, the transceiver can also be required to process the measurements (smoothing) or even to perform some level of estimation.

In the last two decades, almost a dozen, mainly U.S., relative navigation sensors have been designed and developed. However, only one sensor, the FFRF (Formation Flying Radio Frequency) has actually been operated in space. Most of these systems are modified GPS receivers and provide roughly cm-level relative navigation performance and have a mass of several kilograms. Testing and validation is mandatory to chart measurement biases due to multipath and hardware delays. Emulation of the entire system, preferably using flight hardware, is also necessary to gain confidence in the performance of the system and to allow early detection of mistakes/issues.

6.6.1 Way Forward

There are many alternative paths that can be taken to advance the performance of RF-based relative navigation systems. In the following list, suggestions are given for areas in which further research and development can be beneficial in the future.

1. RF-based relative navigation system designs should take advantage of the experience gained with modernized GNSS signals. New spreading symbol shapes (e.g. BOC), signals with a pilot and a data component, and tiered codes are examples that can be explored to improve performance.
2. Carrier phase integer ambiguity resolution has to be facilitated by the ranging signal itself. This removes the need for a geometry change to solve the integer ambiguity problem, which will be infeasible for a system consisting of many satellites. For that, the "ultra-BOC" signal devised for the AFF sensor is an excellent starting point.
3. Short-delay multipath is currently an important source for measurement bias and can cause failure of the integer ambiguity resolution process. This should either be removed as much as possible, for instance by using multipath-limiting antennas, or be made detectable through smart signal design and/or via software.
4. Current relative navigation devices make maximum use of GPS heritage. However, alternatives such as designs based on ultra-wideband (UWB) signals or radar with active targets, which have been developed in the last couple of years, should also be explored as these might lead to system improvements.
5. Terrestrial networking solutions such as the internet protocol and delay-tolerant networks should be explored for their merit in autonomous space systems with a large number of satellites.

Furthermore, high data rates can be expected for future autonomous formations, this should be facilitated in future designs. Also, radiation-hard high-density FPGAs allow a shift from analog to reconfigurable digital designs. This allows a large amount of on-orbit re-programmability for mission extension/adaptation (e.g. different signal waveforms). Designs should also be modular to support various missions with a single instrument tailored to the specific mission needs.

References

1. Fehse W (2003) Automated docking and rendezvous of spacecraft. Cambridge University Press, Cambridge, UK
2. Lestarquit L, Harr J, Grelier T, Peragin E, Wilhelm N, Mehlen C, Peyrotte C (2006) Autonomous formation flying sensor development for the PRISMA mission. In: Proceedings of ION GNSS 19th international technical meeting of the Satellite Division of the Institute of Navigation, Fort Worth, TX, USA, 25–26 Sep 2006
3. Montenbruck O, Gill E (2005) Satellite orbits – models, methods, and applications, 1st edn. Springer, Berlin/Heidelberg
4. Kim J, Lee S (2009) Flight performance analysis of GRACE K-band ranging instrument with simulation data. Acta Astronaut 65:1571–1581. doi:10.1016/j.actaastro.2009.04.010
5. Winter J, Harlacher M, Draganov A, Schuchman D, Haas L (2003) Evolution of the low power transceiver for microsatellites. AIAA space 2003, Long Beach, CA, USA, 23–25 Sep 2003
6. Purcell G, Kuang D, Lichten S, Wu S-C, Young L (1998) Autonomous formation flying (AFF) sensor technology development. 21st annual AAS guidance and control conference, Breckenridge, CO, USA, 4–8 Feb 1998
7. Alfriend KT, Vadali SR, Gurfil P, How JP, Breger LS (2010) Spacecraft formation flying – dynamics, control and navigation, 1st edn. Butterworth-Heinemann, Kidlington
8. D'Amico S (2010) Autonomous formation flying in low earth orbit. Ph.D. dissertation, Delft University of Technology, Delft
9. D'Amico S (2005) Relative orbital elements as integration constants of Hill's equations. Technical note TN 05–08, Version 1.0, DLR GSOC, 15 Dec 2005
10. Yamanaka K, Ankersen F (2002) New state transition matrix for relative motion on an arbitrary elliptical orbit. J Guid Contr Dyn 25:60–66. doi:10.2514/2.4875
11. Jung J (2000) High integrity carrier phase navigation using multiple civil GPS signals. Ph.D. dissertation, Stanford University, Palo Alto
12. Tien JY, Purcell Jr GH, Srinivasan JM, Young LE (2004) Relative sensor with 4π coverage for formation flying missions. 2nd international symposium on formation flying missions & technologies, Washington, DC, 14–16 Sep 2004
13. Bourga C, Mehlen C, López-Almansa J-M, García-Rodríguez A (2002) A formation flying RF subsystem for DARWIN and SMART-2. 1st international symposium on formation flying missions & technologies, Toulouse, 29–31 Oct 2002
14. Stadter PA, Asher MS, Kusterer TL, Moore GT, Watson DP, Pekala ME, Harris AJ, Bristow JO (2004) Half-duplex relative navigation and flight autonomy for distributed spacecraft systems. In: Proceedings of the 2004 I.E. aerospace conference, Big Sky, 6–13 Mar 2004, pp 565–573
15. Stadter PA, Heins RJ, Chacos AA, Moore GT, Kusterer TL, Bristow JO (2001) Enabling distributed spacecraft systems with the crosslink transceiver. In: Proceedings of the AIAA Space 2001 conference and exhibition, Albuquerque, 28–30 Aug 2001
16. Avila-Rodriguez J-A, Wallner S, Hein GW, Eissfeller B, Irsigler M, Issler J-L (2007) A vision on new frequencies, signals and concepts for future GNSS systems. In: Proceedings of the ION GNSS 20th international technical meeting of the Satellite Division of The Institute of Navigation, Fort Worth, TX, 25–28 Sep 2007
17. Borre K, Akos DM, Bertelsen N, Rinder P, Holdt Jensen S (2007) A software-defined GPS and Galileo receiver – a single-frequency approach. Fig. 2.2, Ch 2, p 20 and Fig. 3.3, Ch 3, p 38, Birkäuser
18. Giugno L, Luise M (2005) Optimum pulse shaping for delay estimation in satellite positioning. In: 13th European signal processing conference, Antalya, 4–8 Sep 2005
19. Antreich F, Nossek JA, Issler J-L (2008) Systematic approach to optimum chip pulse shape design. In: 3rd CNES/ESA workshop on GNSS signals and signal processing, Toulouse, 21–22 Apr 2008

20. Gold R (1967) Optimal binary sequences for spread spectrum multiplexing. IEEE Trans Inf Theory 13:619–621. doi:10.1109/TIT.1967.1054048
21. Zenick R, Kohlhepp K (2000) GPS Micro navigation and communication system for clusters of micro and nanosatellites. In: 14th AIAA/USU conference on small satellites, Logan, UT, 21–24 Aug 2000
22. Hein G, Avila-Rodriguez J-A, Wallner S (2006) The Galileo code and others. Inside GNSS, Sep 2006, pp 62–74
23. Sun R, Maessen D, Guo J, Gill E (2010) Enabling inter-satellite communication and ranging for small satellites. 4S symposium, Funchal, 30 May–4 June 2010
24. Dunn C, Bertiger W, Franklin G, Harris I, Kruizinga G, Meehan T, Nandi S, Nguyen D, Rogstadt T, Brooks Thomas J, Tien J (2002) Instrument of GRACE: the instrument on NASA's GRACE mission: augmentation of GPS to achieve unprecedented gravity field measurements. In: Proceedings of the ION GPS 15th international technical meeting of the Satellite Division of the Institute of Navigation, Portland, OR, 24–27 Sep 2002
25. Kaplan ED, Hegarty CJ (eds) (2006) Understanding GPS – principles and applications, 2nd edn. Artech House, Norwood
26. Misra P, Enge P (2006) Global positioning system: signals, measurements, and performance, 2nd edn. Ganga-Jamuna Press, Lincoln
27. Vig JR (2008) Quartz crystal resonators and oscillators for frequency control and timing applications – a tutorial. Rev 8.5.3.9. http://www.ieee-uffc.org/frequency_control/teaching/vig/vig3_files/frame.htm. Accessed 26 Oct 2011
28. Hewlett Packard (1997) Fundamentals of quartz oscillators. Application note 200–2, Electronic counters series
29. Vig JR (1992) Introduction to quartz frequency standards. SLCET-TR-92-1 Rev 1. http://www.oscilent.com/esupport/TechSupport/ReviewPapers/IntroQuartz/vigcateg.htm. Accessed 23 Dec 2010
30. Grelier T, Guidotti P-Y, Delpech M, Harr J, Thevenet J-B, Leyre X (2010) Formation flying radio frequency instrument, first flight results from the PRISMA mission. In: 5th ESA workshop on satellite Navigation Technologies (NAVITEC), ESA ESTEC, Noordwijk, 8–10 Dec 2010

Chapter 7
Vision Based Relative Navigation

Domenico Accardo, Giancarmine Fasano, and Michele Grassi

Abstract Vision is a key technology for the relative navigation of formation flying satellites especially when they operate in close proximity. Indeed, with a vision system the relative position and attitude (usually referred to in the machine vision literature as "pose") of co-flying satellites can be extracted in real-time. This information can be used either to maintain or change the formation geometry. A crucial aspect affecting vision system design and development is the operation with largely variable lighting conditions and the interference caused by the presence of other celestial bodies in the sensor field of view. This requires the implementation of effective image processing techniques and algorithms by which robust, accurate and reliable pose estimation can be achieved. This chapter provides an overview of sensors, techniques and algorithms enabling the relative navigation based on vision, with specific reference to space missions which have already tested in flight this technology.

7.1 Introduction

This chapter describes sensors, technologies, techniques and algorithms relevant to vision systems for the relative navigation of Formation Flying (FF) satellites.

Vision is a key technology for the relative navigation of satellites flying in formation, especially when they come into close proximity as a result of the relative orbit selected to achieve a specific mission objective, either scientific or technology demonstration and validation.

The vision system may be passive or active, being equipped in this last case with laser range finders that illuminate reflectors installed on the tracked satellite

D. Accardo (✉) • G. Fasano • M. Grassi
Department of Aerospace Engineering, University of Naples "Federico II",
P.le Tecchio 80, 80125 Naples, Italy
e-mail: domenico.accardo@unina.it; g.fasano@unina.it; michele.grassi@unina.it

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_7,

external surface according to prefixed geometries. In this regard, several solutions have been proposed and some of them already tested in flight. They will be described in this chapter.

From the vision-based navigation standpoint the tracked satellite can be cooperative or not. In the former case it is usually equipped with Light Emitting Diodes (LEDs) or Reflectors which can be easily imaged by a camera on board a chaser satellite independently from the particular lighting conditions or the interference of other celestial bodies (mainly Earth and Sun) in the camera Field of View (FOV). In the latter case, techniques and algorithms capable of extracting features of the tracked satellite, such as binarization, contour mapping, and edge detection, can be used to set up the synthetic information that well be used to determine the relative pose (i.e. position and attitude). These techniques and algorithms will be described in this chapter.

Concerning pose estimation, several techniques and algorithms can be used to compute the pose from the synthetic information extracted from the images. These techniques can be broadly divided in monocular and three dimensional (3D). The former ones perform relative navigation on the basis of angular-only estimates, while the latter methods make use also of range information provided by a stereovision system or by a laser range finder. Moreover, pose estimation can be refined on the basis of relative dynamics modeling and optimal filtering techniques. These techniques will be described in this chapter, as well.

The chapter is organized as follows: in the first section an overview of sensors and technologies for vision-based relative navigation is provided with specific reference to space missions already performed or still in flight. The second section instead describes the image processing techniques and algorithms adopted to extract the information on the position of selected targets in the sensor reference frame. Finally, the third section focuses on the algorithms and methods to estimate the relative pose on the basis of a synthetic information extracted from the vision system.

7.2 Missions, Sensors and Technologies

This section provides an overview of sensors and technologies for vision-based relative navigation with specific reference to space missions which have already validated in flight such technologies.

At the present, no formation flying missions with specific scientific objectives have been performed carrying on board a vision-based system for the relative navigation, although in the open literature vision is reported as a key technology for next-generation formation-flying missions. Thus, with the aim of performing a quantitative analysis of developed sensors and technologies, we consider already flown space missions in which experiments have been executed in flight to test autonomous navigation technologies for rendezvous, docking, proximity flight and formation flying. In particular, in the following we will describe the following missions and programs:

Table 7.1 Flown missions overview

Mission	Goal	Status	Satellites
ETS-7	Demonstrate autonomous rendezvous and docking technologies for on-orbit servicing missions	Launched on November 28 1997. Tests executed in July-August 1998. Mission was successful	Chaser and target launched into a mated configuration on 550-km 35° inclined orbit
XSS-10/ XSS-11	Test GN&C key technologies for autonomous rendezvous and proximity flight	Launched in 2003 and 2005 respectively. Tests were successfully completed	Two satellites: one is the Delta II second stage (XSS-10) and Minotaur I upper stage (XSS-11)
DART	Test autonomous rendezvous technologies for future on-orbit servicing missions	Launch on 15 April 2005. Duration: 24 h. Partial success due to collision between the satellites	The chaser on 650 km, 97.7° orbit. The target is the MULBCOM satellite already in orbit (1999)
OE	Test technologies for autonomous rendezvous and docking for future on-orbit servicing missions	Launched on March 9 2007. Mission was successful. Satellite separation ranges from a few meters to hundreds of kilometers	ASTRO (the chaser) and NEXTSat (the target) on 492-km 46-inclined orbit
PRISMA	Autonomous GN&C demonstration for future FF and rendezvous missions	Successfully Launched on 15 June 2010. Still in operation	Mango (the chaser, 140 kg), and Tango (the target, 40 kg) on 700-km sun-synchronous orbit

- The Micro-Satellite Technology Development Program, led by the Air Force Research Laboratory (AFRL),
- The Engineering Test Satellite, flown by the National Space Development Agency of Japan (NASDA),
- Demonstration of Autonomous Rendezvous Technologies (DART), conducted by NASA,
- Orbital Express (OE), realized by the Defense Advanced Research Projects Agency (DARPA) and NASA,
- PRISMA, led by the Swedish Space Agency with the support of the German Space Agency (DLR), the French Space Agency (CNES) and the Danish Technical University (DTU).

All the mentioned missions aim at testing in-flight autonomous Guidance, Navigation and Control (GN&C) technologies by using two co-flying satellites and monocular vision (with or without laser ranging aiding) to extract the relative navigation state. For reader convenience, Tables 7.1 and 7.2 synthesize the main information relevant to the above mentioned missions, that are described in more details in the following sections. Specifically, the first table provides an overview of

Table 7.2 Synthetic view of flown mission technology aspects and results

Mission	Sensors	Relative navigation observables/techniques	Performance
ETS-7	RVR (Rendezvous Radar), PXS (Proximity camera sensor) on chaser. CCR (Corner Cube Reflectors) and PXS markers on target	RVR measures LOS and range to target during final approach. PXS determines 6-DOF relative motion during docking	*RVR*: 22 cm (Range), 0.02°(LOS) *PXS*: <5 cm (Position), <0.4° (Attitude)
XSS-10 XSS-11	CCD Camera on XSS-10, CCD Camera and LIDAR on XSS-11	Relative navigation based on centroiding information from CCD camera (XSS-10). Range and LOS determination with LIDAR (XSS-11)	–
DART	AVGS (Advanced Video Guidance Sensor) on the chaser including laser diodes and CCD camera. Reflectors on the target satellite with prefixed geometry	Processes reflector positions on the focal plane to provide LOS to target up to 1 km separation (spot mode). 6-DOF relative position and attitude within 300 m in tracking mode	–
OE	Narrow-FOV Camera (long range operations), Wide-FOV Camera (close range operations), Infrared Camera (situation awareness, nighttime ops.), laser range finder (mid-range ops.), AVSG for short range operations	Relative range and LOS (from cameras and laser). Independent computation of attitude in final approach and capture by Boeing Vis-STAR software. AVSG used within 200 m prior to capture by imaging four non-coplanar targets	*Vis-STAR*: sub-degree in attitude *AVGS*: ±7 cm (range), ±0.01 (LOS), ±0.4° (pitch/yaw), ±0.15° (roll)
PRISMA	Sensor based on flight-proven star camera technology μ-ASC (Advanced Stellar Compass). Four camera heads used for long-range and short range operations	LOS determined in far range. Perspective 4-Point Problem solved for relative pose of a cooperative target in close proximity. Pose determination of non-cooperative target by feature tracking and matching with on-board 3D model	*VBS*: 3 arcsec. (LOS), <1° attitude (cooperating mode) <5° attitude (non-cooperative mode)

mission status, goals and satellite/orbit parameters. The second table offers a synthetic view of adopted sensors, measured navigation parameters and reported accuracies.

Some of the missions in table have been specifically executed to test advanced technologies for future on-orbit servicing platforms and missions, whose

importance has been widely recognized [1]. Nevertheless, sensors, technologies, techniques and algorithms tested in these missions are of great relevance also to formation flying missions, since approach/separation and proximity flight are phases relevant to formation flying applications in which the several satellites fly with short separations or formation re-configuration maneuvers are planned.

It is worth noting that, other than the ones reported in Tables 7.1 and 7.2, in the next few years other missions are planned by the European Space Agency (ESA) within the in-orbit technology demonstration program. They are:

- Proba-3, with the aim to demonstrate precision formation flying technologies and techniques as precursor of future ESA missions [2]
- SMART-OLEV, managed by an European Consortium [3], with the aim of developing a servicing system to extend the mission lifetime of GEO communication satellites
- **DE**utsche **O**rbitale **S**ervicing Mission (DEOS), by DLR, with the main objective of testing technologies for the capture of a non-cooperative tumbling satellite [4].

For the sake of conciseness, in the following sections the DART mission is not reported since the sensor embarked on DART is the same sensor (except for some software changes and minor modifications) successfully tested on Orbital Express, which, instead, is described in details. In addition, DART was not successfully completed since the two satellites collided.

7.2.1 Microsatellite Technology Development Program

The Air Force Research Laboratory established the Micro-Satellite Technology Development Program (XSS series of flight demonstrations) to test in space new technologies for future space missions. In this framework, XSS-10 and XSS-11 flights intended to develop and verify key technologies and operational aspects of on-orbit guidance, navigation and control for autonomous rendezvous and proximity flight operations [5, 6]. Specifically, the XSS-10 program, started in 1997 and launched in January 2003 as a successor of the cancelled Clementine II program, consisted of a micro-satellite ejected from the orbiting Delta II second stage and performing a series of semi-autonomous maneuvers around the second stage to demonstrate relative navigation, proximity operations and related real-time communication technologies, and inspection and tracking of a Resident Space Object (RSO). XSS-10 (an illustration of the system can be found at the web site www.aerospace-technology.com/projects/xss-10microsatellite/) was equipped with a visible camera, a GPS receiver, an Inertial Measurement Unit (IMU), a light weight star sensor, a transponder for the ground link and a flight software for relative navigation and proximity operations. A specific objective of the XSS-10 mission was to maneuver around (by using prefixed inspection points) and demonstrate station keeping capabilities relative to a RSO based on visible imaging, relative position and inertial position/attitude knowledge. In particular, a major goal was to

demonstrate the capability of acquiring and tracking a space object with the on-board CCD visible sensor by using a relative navigation scheme based on centroiding information from the camera collecting images of the RSO at prefixed inspection points. Specifically, the GN&C software updated the relative state by combining RSO images from the camera with an on-board relative state propagator relying on the Clohessy-Wiltshire formulation of the relative motion, while tracking the RSO at the designed inspection points. Inter-point maneuvers were performed to place the RSO centroid in the centre of the camera Field of View (FOV): off-sets from the center represented error in the relative state estimate. The GN&C software used this centroid location information to update the relative state at each point. A Delta II second stage image taken from on-board XSS-10 can be found at the web site www.globalsecurity.org/space/systems/xss.htm.

The XSS-11 was launched on 11 April 2005 on a Minotaur expendable launch vehicle. After completing systems checkout of the Minotaur I upper stage, the micro satellite successfully demonstrated rendezvous and proximity operations with the rocket body. It has accomplished more than 75 circumnavigations of the launch vehicle. The XSS-11 spacecraft contained a IMU for angular rate and acceleration sensing, a coarse sun sensor assembly for sun acquisition, a visible camera system for star detection and target imaging, and a scanning light detection and ranging (LIDAR) instrument to determine relative range and angle measurements to the target.

7.2.2 Engineering Test Satellite

In 1997 the National Space Development Agency of Japan (NASDA) performed the Engineering Test Satellite (ETS-VII) flight experiment to validate enabling technologies for on-orbit autonomous coordinated flight, rendezvous and docking, and proximity flight operations of future missions. The mission consisted of two satellites, “Hikoboshi” (the chaser satellite) and “Orihime” (the cooperating target satellite), launched in mated configuration on the H-11 rocket to a 550-km, 35° inclined orbit [7]. During the experiments, the chaser released the target and drifted away to distances of up to 9 km. Subsequently, a rendezvous maneuver was executed to drive the chaser back close to the target and execute the docking [8]. This experiment represents one of the first flights in which vision-based technologies for relative navigation between two satellites was implemented in orbit and carried out autonomously.

In order to perform the relative navigation task with respect to the target, the chaser satellite embarked GPS receivers, a rendezvous laser radar (RVR), a visible Proximity Operation Sensor (PXS), an Inertial Reference Unit (IRU) and an Earth Sensor Assembly (ESA). The target satellite was cooperative in the meaning that it was equipped with RVR Corner-Cube Reflectors (CCRs) and PXS markers for chaser relative navigation. The sensors and the related markers were arranged so that each sensor can distinguish its own marker set. It embarked also a GPS receiver

Table 7.3 ETS-VII vision-based navigation phases and sensors

Phase	Relative range (m)	Main navigation sensor	Observables
Final approach	500–2	RVR	Relative range/LOS
Docking	2–0.5	PXS	Relative pose

Table 7.4 ETS-VII RVR accuracy specification

Measurement ranges	Measurable relative range: 0.3−660 m FOV for range measurement: ±3° Measurable LOS range: ±4° Max. measurable relative attitude angle: 15° Max. measurable relative speed: 1 m/s Max. measurable LOS rate: 0.5°/s
Measurement accuracy	Range: Bias = 10/60 cm (at min.-max. range) Random (3σ) = 0.6/22 cm (at min.-max. range) LOS: 0.05° (Bias), 0.02°(Random, 3σ)

whose data were transmitted to the chaser for the relative navigation by exploiting a direct communication inter-link between the two satellites [9].

The sensors embarked on the two satellites allowed three different relative navigation modes to be implemented depending on the inter-satellite separation [9, 10]. In particular, the GPS was used for separations larger than 500 m. Instead, for shorter separations (see Table 7.3) the relative navigation between the two satellites was performed by exploiting vision-based technologies.

Specifically, the RVR was used as the main sensor for the relative navigation in the final approach phase (i.e. from 500 m to about 2 m). The sensor, whose accuracy specifications are summarized in Table 7.4 [9, 10], measured the relative range and Line-Of-Sight (LOS) to the target (i.e. target azimuth and elevation angles). The RVR optical head contains a near-infrared laser diode which radiates pulsed light at 810 nm wavelength into a 85° cone. The light beam is reflected from the CCRs on the target back to the RVR optical head where it is imaged on a two-dimensional CCD array. Image processing is then implemented to evaluate the coordinates of the reflected beam centroid on the CCD plane from which target azimuth and elevation can be computed. The relative range instead is determined from the data of the Avalanche Photo Diode (APD) by computing the phase difference between the transmitted and received light beams. The RVR in-orbit accuracy was ascertained to be well within the specification and in agreement with the GPS, PXS and ground test data [9, 10]. Specifically, in-orbit data agreed with ground test measurements (with the two satellites in mated configuration) within 2 cm in range and 0.1° in LOS, with random errors (3σ) of about 5 mm in range and less than 0.01° in LOS [11].

During docking (from 2 m to about 0.5 m) the relative navigation was performed with the proximity sensor, which consisted of a CCD camera and an array of LEDs on the chaser, and a 3D marker (micropitch CCR) on the target. The LED array

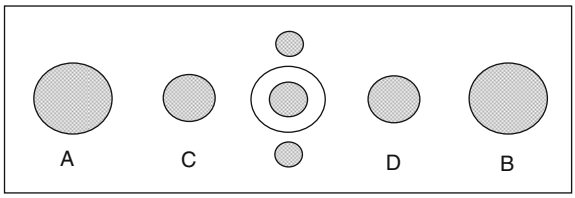


Fig. 7.1 PXS marker configuration schematic

Table 7.5 ETS-VII PXS main specification parameters

Relative position accuracy (at 10-m maximum range)	Bias: <22 cm (X), <50 cm (Y,Z) Random (3σ): <5 cm (X), <1 cm (Y,Z)
Relative attitude accuracy (at 2-m maximum range)	Bias: <1.1° (roll), <0.85° (pitch, yaw) Random (3σ): <0.4° (roll, pitch, yaw)

radiates pulsed visible light (640 nm) within a 30-deg cone. The marker consists of seven round shaped markers, with the central one behind the others that are symmetrically positioned in sets of two (see Fig. 7.1 for a schematic). Relative pose is determined by imaging markers on the camera focal plane. Specifically, the central marker and one of the three sets are used depending on the relative range. Pitch and yaw rotations are detected by exploiting the position of the central marker that is not collinear with the others, roll is computed by detecting the rotation of the baseline vector between the two side markers. The relative position components are estimated by detecting the variation of the baseline length (longitudinal axis) and the displacements on the focal plane of the baseline center (traverse axes). Table 7.5 shows PXS accuracy specifications [9, 10, 12]. The PXS in-orbit accuracy was ascertained to be much better than the specification [9, 10]. Specifically, in-orbit data agreed with ground test measurements (with the two satellites in mated configuration) within 2 cm in range and 0.1° in LOS, with random errors (3 s) of about 5 mm in range and less than 0.01° in LOS [11]. An illustration of the two satellites and pictures of the sensors and reflectors used for the relative navigation can be found in Refs. [13, 14].

7.2.3 Orbital Express

Orbital Express was a cooperative agreement between the Defense Advanced Research Projects Agency (DARPA) and NASA. Mission main goal was to conduct the first autonomous component exchange and first U.S. refueling, as

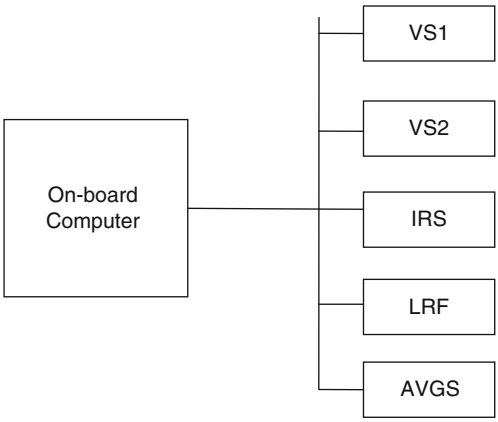


Fig. 7.2 ARCSS composition schematic

well as autonomous rendezvous and docking. It had to demonstrate enabling technologies for autonomous rendezvous, proximity flight, docking and satellite on-orbit servicing. It consisted of two satellites launched in a mated configuration: Astro (Autonomous Space Transport Robotic Operation vehicle, the chaser), developed by Boeing, and NextSat (next-generation serviceable satellite prototype), the target, developed by Ball Aerospace. Launched in May 5 2007 on Atlas 5 rocket within the U.S. Air Force Space Test Program STP-1, after separation the two satellites underwent a series of AR&D (Approach, Rendezvous and Docking) maneuvers experiencing a variety of separation ranges [15]. The flight test lasted about 3 months on a 46°-inclined 492-km orbit. An illustration of the system can be found at the web site boeing.net/bds/phantom_works/orbital/oe_029.html.

Astro embarked the Autonomous Rendezvous and Capture Sensor System (ARCSS), which included optical sensors for the relative navigation with respect to NextSat. Specifically, with reference to Fig. 7.2 there were two sets of optical sensors: one consisting of three imaging sensors and a laser rangefinder (installed on a common optical bench), developed by Boeing, and the other consisting in the AVGS, developed by DARPA for DART mission. The two sets were used by the Astro GN&C depending on the separation between the two satellites. NextSat instead embarked the AVGS and ARCSS targets.

Going into more details, the first set of sensors included: two fixed-focus cameras, one narrow-field (6.5°) camera (VS1), for far range operations, and one wide-field camera (about 40°) (VS2), for mid-to-short ranges operations, and a long-wave infrared imager (IRS), for continuous situational awareness, that could detect the NextSat heat. A spotlight provided target illumination starting at about 24 m. The narrow-field camera could detect and acquire the target satellite from about 200 km, when it was a point source, and determine the bearing to it in order to track it for approach operations. The wide-field camera could supplement the first camera from about 100 km, and the IR camera operated

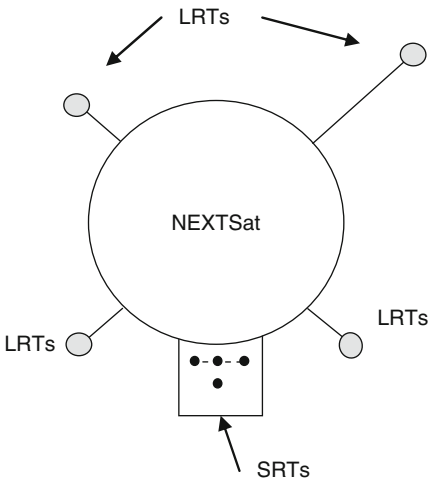


Fig. 7.3 Schematic of AVGS SRT and LRT positioning on NEXTSat [17]

from about 24 km. This last one also provided backup in eclipse and difficult illuminating conditions.

The laser range finder (LRF) started to operate at about 10 km from the target by adding ranging information to the already available ones [16].

Data from the ARCSS available sensors were processed by a vision-based Software for tracking, attitude and ranging (Vis-STAR) to determine bearing, range and relative attitude of the target. Specifically, at large distances, only bearing and range were computed. Instead, at shorter separations (starting from about 500 m) the target image became sufficiently large so that Vis-STAR could start computing independently the attitude, from passive imaging (silhouette tracker), in addition to range (from the laser range finder) and bearing. Attitude determination was performed by correlating acquired target images with pre-loaded pictures of NextSat taken at different angles. Edge tracking techniques were used to discriminate among ambiguous attitude solutions. Sub-degree accuracy was claimed in each rotation axis. Vis-STAR could work with both visual and IR data [16].

The ARCSS included also the AVGS, inherited from the DART program (except for some software changes and minor modifications) [14], developed at Marshall Space Flight Center. It is a near-field proximity navigation sensor ($\pm 8°$FOV) capable of computing 6-DOF solutions (relative position and attitude) for ranges up to 300 m. It represents an advancement of a prior concept, the Video Guidance Sensor (VSG), successfully tested in flight on STS-87 (1997) and STS-95 (1998) missions. Specifically, the AVGS has got improved performance and capabilities for long-range operations [15].

The AVGS incorporates laser range finders emitting IR laser beams illuminating two sets of retro-reflective targets on NextSat (see schematic in Fig. 7.3): 4 Long-Range Targets (LRT), viewed from less than 10 m to about 100 m, and 4 Short-Range Targets (SRT), viewed from docking up to about 30 m, under nominal

Table 7.6 AVGS dual track mode and mated configuration 1σ performance

	Dual track mode	Mated configuration
Parameter	Noise performance	Noise performance
Relative range (m)	0.07	0.0025
Azimuth-elevation (°)	0.01	0.005
Pitch/yaw (°)	0.4	0.12
Roll (°)	0.15	0.04

conditions. In the schematic of Fig. 7.3, the central SRT is not collinear with the others, it is about 20 cm off the figure plane. These targets are tracked over different ranges of operation, and their reflections on the focal plane are processed, instead of imaging the overall shape of the target [17]. In this way it is possible to compute relative attitude, range and bearings at close separations in the approach corridor (last 200 m prior capture). Although the DART mission experienced a failure without the possibility of using the AVGS in closed loop, valuable bearing data were collected with the AVGS operating in spot mode, showing nominal performance [17].

AVGS relies on two sets of laser diodes operating, respectively, at wavelength of 800 and 850 nm, a mirror, a fixed-focus solid-state camera, that images the light reflected by the corner cubes installed into a known pattern on the target satellite, image capture electronics and a digital signal processor to convert the video data into the pose information. The two sets of lasers are fired asynchronously (at 5 Hz) at the two different wavelength with an offset of 0.1 s. One wavelength (800 nm) is absorbed by a filter placed on the retro-reflectors, while the other is passed through the filter and reflected back to the sensor to be imaged on the focal plane. Thus images at 10 Hz of the target are captured: one image contains only the background without the CCRs and the other contains the background and the CCRs. The two images are subtracted from one another leaving the CCRs which can be then easily detected in the resulting image under variable ranges and lighting conditions [15]. Relative range, bearing (azimuth and elevation) and attitude are computed by processing 2D images by inverse perspective algorithms which combine the spot data with the known target pattern.

Flights tests results on board Orbital Express with the AVGS operating in dual track mode (i.e. tracking both LRTs and SRTs), corresponding to ranges varying from about 10 to 30 m, shows that the sensor behaved according to specifications. Flight results in dual track mode were obtained by comparing the two pose computations achieved independently by tracking LRTs and SRTs. It is reported that the range standard deviation was about 0.07 m, the azimuth and elevation standard deviation was less than 0.01°, pitch and yaw noise was under 0.4°, and the roll noise less than 0.15°, according to sensor specifications [15]. For reader convenience, the reported performance is summarized in Table 7.6. In the same Table also the flight results related to the mated configuration (range about 1.22 m) [15] are reported for the sake of completeness.

Table 7.7 Prisma vision-based navigation test phases and main sensor

Range	Test	Mode/Main sensor
10–100 km	Approach and rendezvous	Far range mode: far range camera, LOS and range determination
5–100 m	Proximity operations	Short range mode: short range camera cooperative/non-cooperative target, pose determination
0–5 m	Final approach and recede	Short range mode: cooperative target, pose determination

7.2.4 Prisma

PRISMA is a mission funded by the Swedish National Space Board aiming at testing in space enabling technologies for formation flying and rendezvous. The mission is supported also by CNES, DLR and the Danish Technical University. Mission main goal is demonstrating autonomous GN&C, with specific reference to formation flying, homing and rendezvous, proximity operations and final approach and recede operations [18]. The mission relies on two satellites (an illustration of the system can be found at the web site www.spaceportsweden.com/about-prisma.aspx), MANGO (the main) and TANGO (the target). MANGO weights about 150 kg, it is 3-axis stabilized and has got full 3D delta-V maneuverability in order to perform a series of maneuvers around the target on both long and close range by using a set of sensors installed on board. TANGO is a 40-kg satellite equipped with a simple magnetic 3-axis attitude control system and has got no orbit control and maneuvering capability. It can communicate with the main by means of a 450 Hz interlink channel. The two satellites were launched in June 2010, in a mated configuration, as secondary payload with a Dnepr launcher into a 98° inclined sun-synchronous orbit at an altitude of about 700 km [18–20].

MANGO embarks various sensors to measure the relative position with respect to TANGO. Specifically, depending on the separation between the two satellites the following relative navigation experiments are executed:

- Differential GPS, for separations higher than 30 m
- Vision-based, for separations from 500 km down to about 10 m
- Radio-frequency based, for separations from 30 km to about 3 m [18–20]

We focus on the phases in which the vision sensor was used for relative navigation. The vision-based sensor embarked on MANGO is designed and developed by the Technical University of Denmark [21]. In the PRISMA mission it has to be evaluated as a multi-range range tracking and rendezvous sensor. To this end, it is used during the homing and rendezvous experiment, as well as for proximity operations and final approach/recede maneuvers (see Table 7.7).

The first experiment aims at evaluating the vision-based sensor capability of identifying the target as a non-stellar object at a distance up to 500 km and to track the target down to ranges as short as 10 m by means of a sequence of autonomous

approach maneuvers. The second experiment main goal is evaluating the sensor performance in operations relevant to in-orbit servicing, inspection and assembly missions. In this experiment, sensor performance are tested by configuring the target both as a cooperative one, i.e. equipped with light emitting diodes, and as a non-cooperative one with varying attitude and rates [18].

The vision-based sensor has been developed as an extension module of the μ-Advanced Stellar Compass (μASC), a fully autonomous-high-accuracy CCD-based star tracker platform, which can host from 1 to 4 Camera Head Units [21, 22]. The visual sensor designed for PRISMA is equipped with two cameras units, one for far range operations (down to about 100 m from the target) and the other for short range operations (from about 500 m down to a few centimeters).

This design allows performing the experiments described above. The far range camera is a standard Camera Head Unit, basically a star tracker. Indeed, for large separations the target is imaged as a point source, like the stars.

By image processing techniques, the sensor can discriminate the target from the stars on the background, thus determining the direction to it with about 3 arcsec. accuracy [22]. As the inter-satellite separation reduces (up to about 100 m), the target becomes more extended and luminous, so that the stars cannot be detected anymore. The sensor then provides the direction and the distance to the target. When the target image on the focal plane is large enough (at separations < 100 m), the sensor starts determining also the pose with respect to the target by using the short range camera, that is a camera head with modified focal length, iris and electronic shutter, to allow operations in close proximity.

Regarding the pose determination in short range, as previously outlined the target can act both as a cooperative one and as a non-cooperative one. In the first case, it is expected that the relative pose is computed with centimeter-level accuracy in relative position and better than 1° in relative attitude (in all axes) by imaging a set of coplanar Light Emitting Diodes (LED) installed on the target into a unique, asymmetric non-collinear pattern on each panel [21]. The pose is then determined as the solution of the Perspective 4 Point problem [21, 22], which will be described in the following sections. In the second case, the relative attitude and position are determined by using target feature tracking and searching for the rigid transformation matching the acquired images to a target 3D model stored on board [22]. The relative navigation solution in this case is expected to be less accurate and more dependent on the illumination conditions. Specifically, it is expected that the position accuracy is about ten times worse and the pose accuracy is about five time worse.

7.3 Image Processing Algorithms and Techniques

This section presents the image processing techniques and algorithms that are commonly adopted to estimate the relative pose between two space platforms by exploiting images acquired onboard using an electro-optical camera. In the reference

model, the camera is equipped with an Image Forming System that comprises optics and an image array detector. The detector is formed by several elementary sensing elements called pixels that are arranged in rectangular arrays formed by rows and columns. Each pixel is capable to output a signal that is proportional to the total energy of photons in the camera spectral waveband that impacted its surface during a preset exposure time. Of course, this signal is corrupted by noise that is determined by several sources such as dark current, photon noise and fixed pattern noise [23]. The noise level depends on the quality of the whole Image Forming System.

In Pixel Size limited imaging systems each pixel measures the energy within a specific Istantaneous Field of View (IFOV), i.e. a rectangular solid angle, that is integrated in the whole Field of View (FOV). This is the most diffused camera configuration. In the less common case of Diffraction Limited imaging systems, the IFOV is determined by the size of the optics diffraction blur that is larger than the size of a single pixel. Several technologies have been developed to manufacture array photodetectors such as Charge Couple Device (CCD) and Complementary Metal-Oxide Semiconductor (CMOS) for visible and Near Infra-Red wavebands whereas Micro Bolometers are available for Thermal Infra-Red wavebands [23]. It is worth noting that particular attention must be given to select a proper photodetector for space electro-optical systems, since they are very sensitive also to electromagnetic input in other wavelengths such as X-rays. The effect of this input can determine unwanted degradation of image processing performances because of the presence of blank pixels [24].

In general, the type of image processing techniques described here are independent on camera spectral waveband. Notwithstanding, the camera works as a radiated energy sensor rather than a scattered energy sensor as the waveband moves toward Infra Red. As a consequence, images tend to be thermal maps rather than scattering maps as it is for the visual wavebands.

7.3.1 Assumptions

In this section the type of image considered is a grayscale image, i.e. an image that is given by digitizing the signal output of each pixel so that it is associated to a specific level in a scale of gray tones from white to black. Color images will be not considered in this section since no visible waveband color camera has been found in the references that is adopted for visual based relative navigation in space. Color images are obtained by superimposing Bayer filter mosaic over photodetector array pixels with alternate passing waveband in red, green, and blue light [25]. Full color images are obtained by Bayer pattern images by demosaicing. Indeed, the information about spectral distribution of incoming light can be useful to recognize objects in the image but it does not affect the process of selecting pixel of interest from background that is the main task of image processing. Grayscale images can always be derived by color images by averaging the intensities on three contiguous pixels.

The fact that the image is digitized means that the voltage values measured by each pixel are processed by sampling, quantization, and coding.

For a 2D image, both spatial and temporal sampling must be considered. Spatial sampling means that the voltage measured in each pixel is associated to the position of its center. Temporal sampling means that the measured pixel intensity is associated to a specific time in its exposure interval.

The effect of quantization results in that the continuous value measured in each pixel is set equal to a discrete level chosen in a finite set of values from darkest value, i.e. black, to brightest one, i.e. white. If the distribution of levels is uniform with light intensity then "linear quantization" is applied. In many cases, a logarithmic distribution of levels with light intensity is applied in order to increase the number of levels at darkest luminosity values [26]. The grayscale resolution of an image is given by the number L of different levels available. Usually, it is determined by giving the number N of bits that are needed to store the information in a CPU. The following expression correlates L to N:

$$L = 2^N \tag{7.1}$$

A grayscale image is composed by an array of integers $I_{i,j}$ that represent the digital values associated to the intensity of light that impacts on the pixel surface during the exposure time. As a consequence, the digital image formation process is the result of three integration processes:

1. Spectral integration, i.e. $I_{i,j}$ is proportional to the spectral integral on the spectral waveband of the photodetector;
2. Spatial integration, i.e. $I_{i,j}$ is proportional to the spatial integral of the local intensities measured in each point of its surface;
3. Time integration, i.e. $I_{i,j}$ is proportional to the time integral during the exposure time.

In real systems, the effect of noise and distortions must be estimated. A detailed discussion about these effects can be found in the references [23]. We have environmental distortions and optics distortions. The first ones are determined by environmental conditions such as temperature, wet or foggy weather in the Earth atmosphere but also by X-rays and Cosmic Rays in the Outer Space. In particular, X-rays can determine the unwanted phenomenon of blank pixels [24]. Optics distortions are forms of optical aberrations that are determined by imperfections in manufacturing lenses. The effect of distortions is to transform straight lines into curved lines in the images. The most common types of aberrations are pincushion aberration and barrel aberration [23]. It is worth noting that also lens glass transparency properties degrades with space radiation doses. For this reason, a proper doping of glass is performed to reduce darkening [24].

Finally, photodetector noise must be considered. Two main noise sources are considered [23]:

– A random noise that is called shot noise. This name is determined by the fact that it is generated by random motion of particles such as photons or electrons that

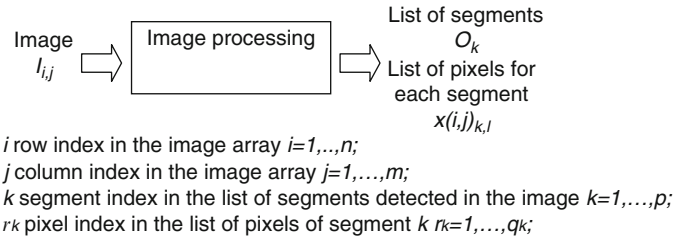


Fig. 7.4 Typical flow chart of image processing algorithm

can be modeled by Poisson or shot distribution, for a single particle. Anyway, the overall effect can be modeled by means of a Gaussian distribution when the number of particles becomes large enough. It can be divided into:

- Photon noise if it is determined by the random motion of photons. This type of noise is due to photons that are not generated by imaged objects or background but from other sources;
- Dark current if it is determined by the random motion of electrons inside photodetector readout electronics.

– The Fixed Pattern Noise, i.e. the bias noise of each pixel that determines a fixed pattern in the whole array.

In general, nominal photodetector noise is handled by keeping its magnitude within the first level during quantization. Indeed, several issues can determine a performance degradation such increasing the level of noise as it happens for thermal effects and electrostatic effects.

7.3.2 General Form of Image Processing Algorithm

Image processing techniques aim at extracting information about the localization in the image of a specific object S that is supposed to be contained in a single image acquired by an electro-optical camera. In particular, Fig. 7.4 reports the typical flow chart of an image processing algorithm. It accepts as input a grayscale image and it returns as output a list of p detected segments. Each segment is formed by an segment identifier O_k with $k = 1,\ldots,p$ and by a list of q_k pixel coordinates $x(i,j)_{k,l}$ for each $r_k = 1,\ldots,q_k$. that form a connected set of pixels. Hopefully, once image processing has been performed at least one segment has been detected in the image. If more than one segment is detected, a proper strategy must be addressed to find out how many segments O_k can be associated to S. In general, a customized strategy must be defined in each case. Some common strategies perform the recognition of S by inference on the measured size of detected segments and the distribution of geometric invariants [27]. Also information about the thermal signature or the color

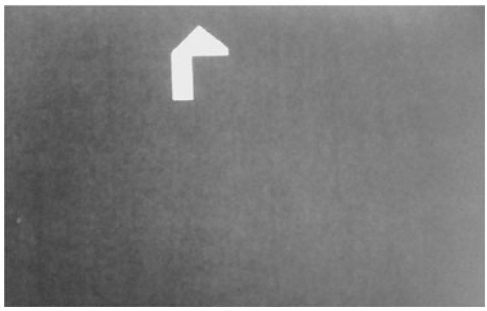


Fig. 7.5 Example image of a pattern acquired by a camera for testing image processing algorithms

of the segment can be used to this aim, if a thermal or a color camera are used. Since none of currently available image processing techniques is perfect in terms of object detection performance, a residual error must always be accepted. In case image processing fails to detect an object that is actually present in the image, then a *missed detection* event is determined. Conversely, a *false match* occurs when S is associated to a segment O_k that is not the true image of S.

Three main image processing categories are available to address the above described task, such as:

1. Image segmentation by binarization and labeling. This algorithm recognizes a multiple set of pixels as segments. Each segment is associated to a selected identifier or label so that it can be effectively analyzed in order to be associated to an object S. Segments are selected by means of image binarization and labeling;
2. Image segmentation by edge detection and labeling. In this case, the algorithm recognizes as "edge" the pixel where a proper term that is related to the slope of the measured digital intensities with surrounding pixels is greater than an assigned threshold. All detected edges are associated to segments;
3. Image segmentation by correlation based template matching. This image processing techniques needs that one or more template images of the searched object S in the same resolution of the camera are stored in the memory. The position of S is located by finding the pixel with maximum of correlation between the template and the acquired image.

Indeed, the first algorithm is less performing than the last one in terms of object recognition but it is much faster in terms of computational load. The edge detection is an intermediate compromise among the others since it has average object recognition performance and computational load. The selection of the proper solution for a specific application will be the result of a tradeoff analysis.

In the following, the picture reported in Fig. 7.5 will be considered as reference for the application examples of image processing techniques. It is a grayscale image of a pattern on a dark background that has been acquired by means of a commercial camera with a resolution of 1280×800 pixels.

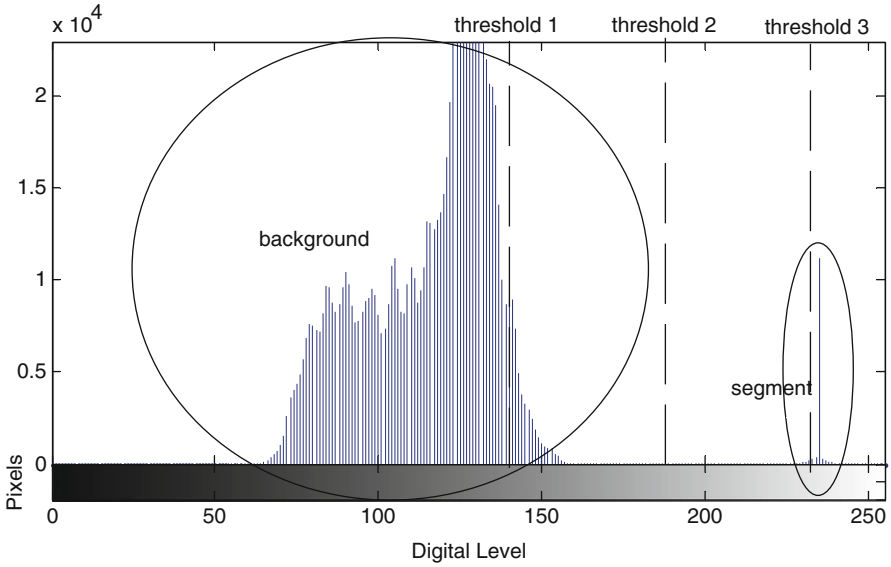


Fig. 7.6 Histogram of digital level distribution in the example image with three thresholds

7.3.3 Binarization

The binarization transforms a grayscale image into a binary image, i.e. an image where each pixel is given by a binary value, by applying the following rules:

1. The binary value of all pixels that are associated to image *background* must be set to 0;
2. The binary value of all pixels that are associated to a *segment* must be set to 1.

For a correct separation of *background* and *segments* a statistical analysis of digital level distribution must be performed. Figure 7.6 reports the histogram of the digital level distribution in the reference image. A large amount of pixels is within the digital levels from 60 to 160. They represent *background* pixels. In this image the background seem to be distributed as a combination of two different Gaussian distributions. The first one has a mean about digital level 80 whereas the second one has a mean about digital level 130. Indeed, this condition is given when background luminosity is different in two image subsections. In this case, the background level is quite high and the quantization was not realized in the best form. The binarization is realized by separating two regions in the histogram such as the region of background pixels and the region of segment pixels. This is performed by determining the mean μ and the standard deviation σ of digital

level distribution. A lower threshold t_1 and an upper threshold t_2 are defined as reported in (7.2):

$$(t_1 = \mu - k\sigma) \text{ and } (t_2 = \mu + k\sigma) \tag{7.2}$$

If the value of t_1 determined by means of (7.2) is less than 0 then t_1 must be set to 0. If the value of t_2 determined by means of (7.2) is greater than the maximum digital level t_{max} then t_2 must be set to t_{max}.

The constant k is a gain that has a critical role in determining the performance of the binarization. Common values adopted are $k = 1.2$ or $k = 2$, but customized values can be derived [28] after proper statistical analysis of images acquired during typical operating conditions. Indeed, a small value of k can determine a threshold that lets some background pixels to be erroneously classified as segment pixels. This is the case of "threshold 1" in Fig. 7.6. At the same time, a large value of k can determine a threshold that lets some segment pixel to be erroneously classified as background pixel. This is the case of "threshold 3" in Fig. 7.6. The proper value of k must identify a threshold that allows for a correct selection of pixel, as it is for "threshold 2". Figure 7.7 represents the effects of binarization on the example image when each one of the above mentioned thresholds are selected.

7.3.4 Labeling

Labeling is an operation that determines the total number of segments detected in the image and it associates the proper set of pixels to each segment. It receives in input the binary image and it returns a matrix of integers with the same size of the image array. In this matrix each pixel is associated to an integer that represents the i-th segment. This operation is performed by scanning all rows and columns in the whole binary image extracted during binarization. Each time a new pixel with a binary value of 1 is encountered, its neighboring pixels are checked in the matrix of integers. If one of them has been previously associated to a segment, then also the selected pixel is associated to the same segment. In case none of the surrounding pixels has been already associated to a segment then a segment counter is increased by one unit and the resulting value is associated to the selected pixel in the integer matrix. A particular attention is needed for "U" shaped segments. Indeed, the two vertical lines of "U" could have been initially associated to different segment indexes if scanning proceeds from the top of the image to its lower part. Anyway, at least a pixel will exist in the bottom of the U-shaped segment that will have two neighboring pixels that have been set to different indexes. When this event happens, one of the two indexes is cancelled and the relevant pixels are set equal to the value of the other one.

When labeling is completed, a list of segments O_k is available and S can be detected, if it is present, by performing an inference on its geometric features, such as its size or the distribution of figure invariants [27].

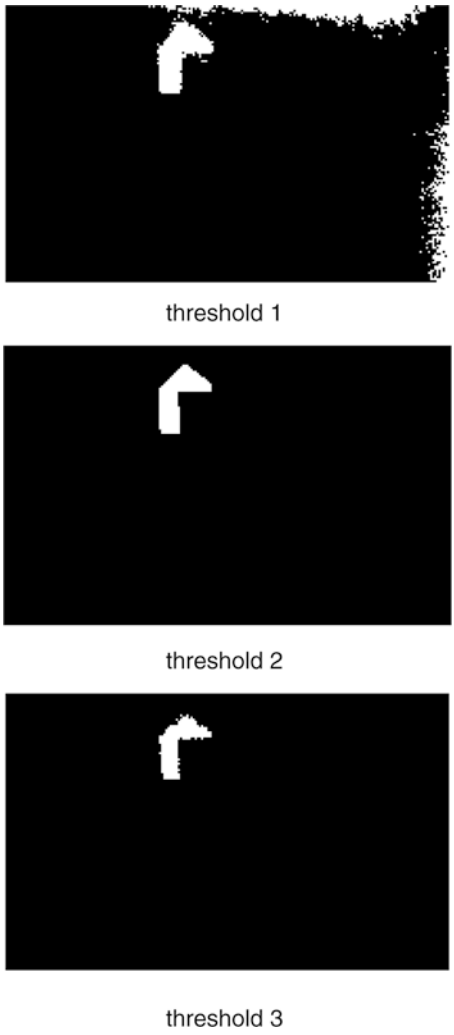


Fig. 7.7 Effects of threshold selection on the binarization of images

7.3.5 Edge Detection

With images with a composite background the performance of standalone binarization can be unsatisfactory in terms of false alarms or missed detections, since a proper strategy for selecting the threshold cannot be found. The "edge detection" technique can be a suitable solution to solve this issue.

The concept behind "edge detection" is to search for pixels where the slope of the digital level with respect to neighboring pixels is larger than a chosen threshold. Two strategies were developed to find pixel with large digital level slope, such as:

1. Finding a local maximum of the numerical first derivative of digital level;
2. Finding a local zero crossing of the numerical second derivative of digital level.

The estimates of numerical derivatives in the images are determined by performing a filtering transformation of the image by means of numerical convolution with a proper mask M, i.e. a square matrix that has an odd order $d > 2$. The resulting 2D array of convolution named R_{ij} is given by (7.3), in the form:

$$R_{ij} = \sum_{a=-c}^{+c} \sum_{b=-c}^{+c} \left(M_{a+c+1,b+c+1} \cdot I_{i+a,j+b} \right) \tag{7.3}$$

where $I_{i+a,j+b} = 0$ if one of its subscripts is less than 1 and $c = \frac{(d-1)}{2}$. A binary array of edges or *edge map* T_{ij} can be derived by applying the following rules:

1. The binary value of elements T_{ij} that are associated to an *edge* must be set to 1;
2. The binary value of all other elements must be set to 0.

Considering (7.3) the *edge map* T_{ij} can be estimated by means of the rule of (7.4), such as:

$$T_{ij} = \begin{cases} \begin{cases} 1 \; if \; R_{ij} \geq d_t \\ 0 \; if \; R_{ij} < d_t \end{cases} \\ case(a) - \text{to detect the maximum of first derivative} \\ \begin{cases} 1 \; if \; \left| R_{ij} \right| \leq d_t \\ 0 \; if \; \left| R_{ij} \right| > d_t \end{cases} \\ case(b) - \text{to detect the zero crossing of second derivative} \end{cases} \tag{7.4}$$

where d_t is a properly selected positive edge detection threshold. Several techniques have been developed to derive the expression of matrix M. When the first derivative must be estimated, such as in case (a), the Prewitt and the Sobel solutions are documented [26]. The Prewitt solution adopts the values reported in (7.5) for M.

$$\begin{aligned} M_{hor,\mathrm{Prewitt}} &= \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad M_{vert,\mathrm{Prewitt}} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \\ M_{diag1,\mathrm{Prewitt}} &= \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad M_{diag2,\mathrm{Prewitt}} = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{aligned} \tag{7.5}$$

Fig. 7.8 Results of edge detection on the example image by means of Sobel mask filter

where $M_{hor,\mathrm{Pr}\,ewitt}$, $M_{vert,\mathrm{Pr}\,ewitt}$, $M_{diag1,\mathrm{Pr}\,ewitt}$, and $M_{diag2,\mathrm{Pr}\,ewitt}$ are the expressions of M matrix used to detect, respectively, horizontal, vertical, left diagonal, and right diagonal edges. Four passes of (7.3) are needed to detect all types of edges.

The Sobel solution adopts the values reported in (7.6) for M.

$$M_{hor,Sobel} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad M_{vert,Sobelt} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$
$$M_{diag1,Sobel} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad M_{diag2,Sobel} = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \tag{7.6}$$

where $M_{hor,Sobel}$, $M_{vert,Sobel}$, $M_{diag1,Sobel}$, and $M_{diag2,Sobel}$ are the expressions of M matrix used to detect respectively horizontal, vertical, left diagonal, and right diagonal edges. The main difference between Previtt and Sobel masks is the presence of the value 2 in the center location. This value allows for image smoothing and provides an increase in the performance of edge detection [26]. For this reason, Sobel is in general preferred to Prewitt, since they have the same computational load. Figure 7.8 represents the result of edge detection by means of a Sobel filter on the example image. In the case of second derivative zero-crossing edge detection, such as in case (b), the Laplacian mask is used that has the expression reported in (7.7).

$$M_{Laplacian} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{7.7}$$

Advanced *edge detection* techniques, such as the one developed by Canny, require increased computational resources in terms of image preprocessing and post-binarization analysis. Conversely, they guarantee very good performances when complex images must be processed [26].

The searched object S can be recognized among the detected segments in the same way as reported for binarization. In this case, a pixel region filling algorithm

can be applied when closed edge perimeters are detected in order to include all the pixels in the internal area before performing recognition [26].

7.3.6 Correlation Based Template Matching

When images are very noisy, even the edge detection algorithm does not ensure a sufficient level of performance. In these cases, correlation based image processing techniques can be applied if a remarkable increase in the computational load can be accepted.

Template matching requires that one or more sample images w_{hk} of S, i.e. templates, are stored in the memory of the image processing CPU. The template images w_{hk} are rectangular 2D arrays that are *hmax* rows and *kmax* columns. Obviously, the size of w_{hk} must be smaller than I_{ij}. In order to find out the position of S in the image the correlation C_{ij} between I_{ij} and w_{hk} must be computed by means of the expression reported in (7.8).

$$C_{ij} = \sum_{h=1}^{h_{\max}} \sum_{k=1}^{k_{\max}} \left(w_{h,k} \cdot I_{i+h,j+k} \right) \tag{7.8}$$

The limits of sums in (7.8) must be contained into the size of I_{ij}. In case the sum exceeds the limits of I_{ij} it can be padded with zeros. Equation 7.8 is similar to (7.3). The main difference is the size of w_{hk} with respect to the size of M that can determine a remarkable increase in the number of computations.

For what concerns object recognition, the position(s) of S are located in the pixels where the local maximum of C_{ij} is larger than a selected threshold t_{corr}. If just one occurrence of S is expected in the image then it can be associated to the global maximum of C_{ij}, in case it exceeds t_{corr}. It is worth noting that the performance of correlation C_{ij} as determined by (7.8) can be sensitive to changes of light intensity. For this reason, the *normalized correlation coefficient* γ_{ij} is commonly used as reported in (7.9).

$$\gamma_{ij} = \frac{\sum_{h=1}^{h_{\max}} \sum_{k=1}^{k_{\max}} \left(w_{h,k} - \overline{w} \right) \sum_{h=1}^{h_{\max}} \sum_{k=1}^{k_{\max}} \left(I_{i+h,j+k} - \overline{I}_{i+h,j+k} \right)}{\sqrt{\sum_{h=1}^{h_{\max}} \sum_{k=1}^{k_{\max}} \left(w_{h,k} - \overline{w} \right)^2 \sum_{h=1}^{h_{\max}} \sum_{k=1}^{k_{\max}} \left(I_{i+h,j+k} - \overline{I}_{i+h,j+k} \right)^2}} \tag{7.9}$$

Where $\overline{w}$ is the average value of the mask and $\overline{I}_{i,j}$ is the average value of $I_{i,j}$ in the same region of w. The term γ_{ij} will be bounded in the interval $[-1, 1]$ and its maximum is on a pixel included in the region of $I_{i,j}$ that is the most similar to the template. As for (7.8), the position(s) of S are located in the pixels where the local maximum of γ_{ij} is larger than a selected threshold t_{γ}.

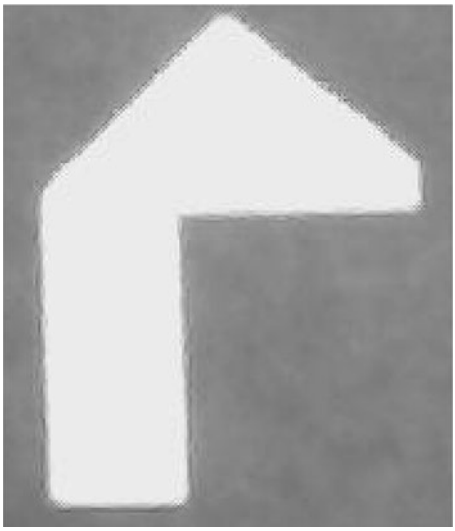


Fig. 7.9 Template image of object *S* for applying correlation processing to the image of Fig. 7.5

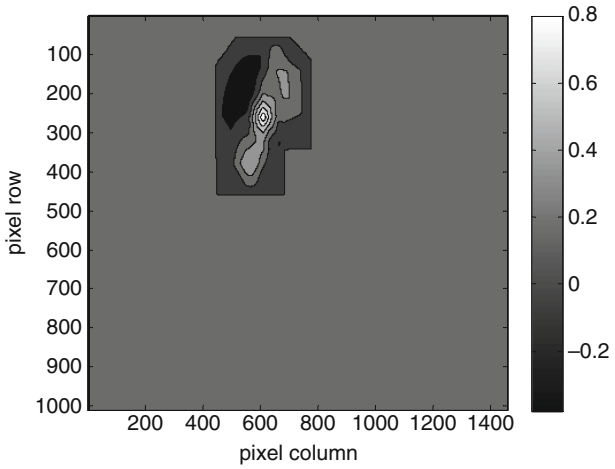


Fig. 7.10 Contour plot of correlation coefficient determined by performing template matching between the example image and the template in Fig. 7.9

Figure 7.9 reports an example of template that has been used for testing template matching algorithm by means of the normalized correlation coefficient on the example image in Fig. 7.5. The resulting contour map has been reported in Fig. 7.10. In this case, the position of the maximum of correlation is exactly on the center pixel of the object.

It is worth noting that the correlation matching is sensitive to object scaling and rotation. If the amount of scaling or rotation is known, than a proper transformation can be applied to the template in order to correct this effect. In particular, the correction of scaling requires a spatial resampling of the template whereas the rotation requires a linear transformation of template by means of a rotation matrix. The correction is much more difficult when the magnitude of scaling and/or rotation are unknown.

An advanced application of template matching can be realized by means of Sequential Template Matching [29]. This technique recognizes the object in the first image acquired using one of the above described techniques. Subsequent frames are processed using as template a rectangular window extracted from the previous frame that includes the largest edge detected in the recognized position of the object. This technique allows tracking objects that perform rotations or scaling at lower rates than the image processing rate used for the developed system.

7.3.7 Notes on Computational Resources

When the proper image processing routine has to be selected, the order of magnitude of elementary computations needed must be known so that the required processing time can be predicted when a specific CPU is used. This is a critical issue for real time applications. A general rule of thumb for the presented algorithms is summarized in the following bullets:

- The number of computations needed for binarization is in the order of the size of I_{ij};
- The number of computations needed for edge detection is in the order of the size of I_{ij} times the size of M;
- The number of computations needed for template matching by correlation is in the order of the size of I_{ij} times the size of γ_{ij}.

7.4 Pose Determination Techniques and Algorithms

As anticipated in the introduction, this section focuses on algorithms and methods to estimate relative position and attitude of a formation flying satellite on the basis of synthetic information extracted from an optical system onboard another platform.

In general, there exists a wide literature about vision-based pose estimation and several techniques have been developed with main applications in the field of ground robotics and virtual reality.

When considering spaceborne applicability of these techniques, the particular features of the on-orbit visual environment play a key role. In fact, the on-orbit environment's lack of atmosphere and rich background, that might diffuse sunlight, creates highly contrasting scenes. When the contrast of the scene exceeds the dynamic range of the camera, then part of the image data is lost. The structures and spacecraft are often covered with loosely attached reflective materials (foil) or featureless thermal blankets. Such surfaces when illuminated with directional sunlight and on-board lights create shadows and specularities that pose difficult problems for imaging cameras and vision systems [30]. Furthermore, the need of near real-time operation with space qualified hardware poses strict requirements on the computational burden of the pose estimation process.

The discussion that follows will briefly review the approaches and algorithms which have been used in space or are under consideration for future spaceborne implementations.

Before going into these details, it is important to note that in general terms, complete (6 degrees of freedom) pose estimation can be carried out by vision-based systems only at short distances, of the order of some tens of meters at most. Thus, most of the literature about autonomous pose estimation deals with rendezvous and docking application. To be consistent with this literature, in this discussion the satellite platforms will be referred to as "chaser" and "target", besides the classical wording of "chief" and "deputy". At larger distances, laser ranging systems can provide range measurements, and optical sensors can be used to provide high accuracy angular information, though it can be impossible to provide an estimate of relative attitude.

Within this framework, it is interesting to discuss potential and limitations of angle-only relative navigation.

Observability criteria for angles-only relative navigation are discussed in Ref. [31] and can be intuitively interpreted. Basically, in close formations which can be described with linear dynamic models such as Hill's equations [32], relative position and velocity cannot be determined with angle measurements alone, since the angular measurement profiles are not unique to a given set of initial conditions. For example, two deputy satellites moving on different radial/along-track ellipses with respect to a chief satellite on circular orbit, will produce the same angular measurements profile regardless of their distance from the chief (Fig. 7.11).

However, as described in Ref. [33], angles-only measurements allow one to determine some geometric aspects of relative motion, i.e., the type of relative trajectory. On the other hand, a calibrated thrust maneuver can be used to achieve range observability [34], if its effect on relative motion is not along the line of sight. Of course, in this case position and velocity estimation are based on dynamic filtering of angular measurements. On the other hand, this strategy is more likely to be adopted in rendezvous and docking phases rather than in formation flying missions.

Several taxonomies can be used to classify pose estimation techniques. First of all, they can be divided into *model-based* and *non-model-based techniques*, depending on whether or not an *a-priori* knowledge of the tracked object is assumed

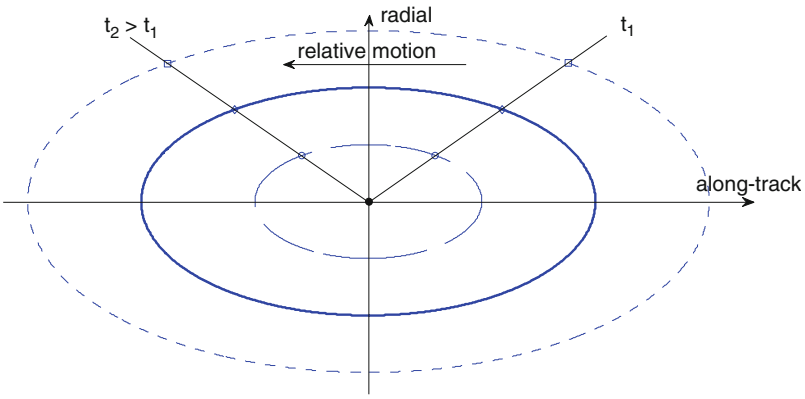


Fig. 7.11 Example of different relative orbits with the same angular measurement profile

[35]. Non-model-based techniques (see for example [36]) are not discussed here because they are of less interest for formation flying missions.

In the case of known objects, artificial targets (markers), natural features, or surface data can be used as input for the pose estimation process.

As a rule of thumb, cooperative techniques based on markers onboard the target satellite usually rely on more direct analytical or semi-analytical approaches, while model-based techniques are usually more complex, are based on iterative methods, and foresee different phases with different computational weight.

Furthermore, the techniques can be broadly divided in monocular and three-dimensional. The former ones perform relative navigation on the basis of angular-only estimates, while the latter methods are based on the use of three-dimensional information such as the one provided by stereovision or laser aided systems. This latter classification will be used as a general scheme for this discussion.

7.4.1 Monocular Techniques

The general theoretical framework for all the monocular techniques based on feature points is the *perspective n-point (PnP) problem*, that is the problem of determining the three-dimensional position and orientation of an object relative to a camera based on a two-dimensional image of n known points located on the object [37].

General conditions for existence and uniqueness of the solution have been widely discussed in the scientific literature and, as intuitive, depend on the number of feature points [37–39].

Both for a single point and for two points there are infinitely many solutions, which are bounded in the latter case. More interestingly to the considered application, for three non collinear points there are finitely many solutions (up to four). In

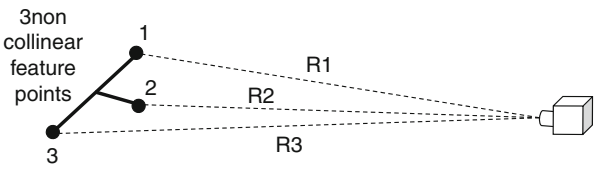


Fig. 7.12 Problem geometry for a monocular camera imaging three non collinear points

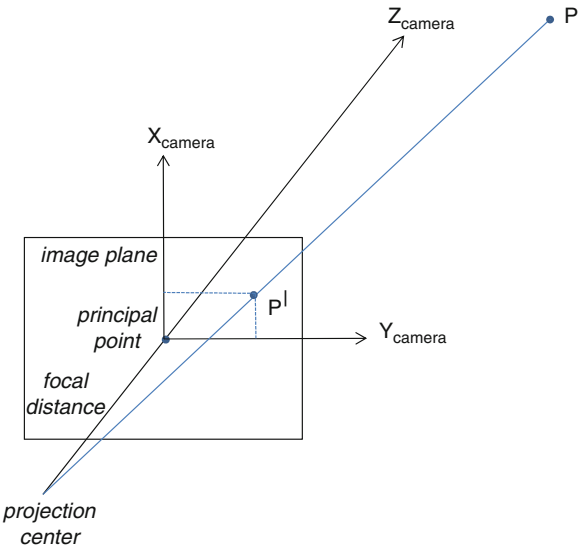


Fig. 7.13 Pinhole camera model (neglecting optical distortions)

the case of four non coplanar (three non collinear) points there are finitely many solutions, while the problem admits a unique solution in the case of four coplanar points (again, three non collinear). Finally, for five points the solution can be ambiguous, while for six points there is a unique solution. The actual number of solutions in practice depends on the location of the feature points and their geometry with respect to the vision sensor. It is also important to note that the solution sensitivity to unavoidable sensor uncertainties, and thus the actual accuracy of the pose estimate, depends on the geometric configuration of the features.

Standing these conditions, there are several mathematical approaches to determine the problem solution (see for example [40]). For the sake of clarity, some details will be provided about the algorithm used in the NASA Video Guidance Sensor [41], whose evolution is the AVGS, and the solution adopted onboard PRISMA for the short range cooperative mode [21].

Let us first consider the case of three non collinear features imaged by a camera as depicted in Fig. 7.12.

After image processing and feature point extraction, pixel coordinates of the centroids can be translated in line-of-sight unit vectors by means of the intrinsic model of the camera, obtained by calibration [42]. The classical pinhole camera model is depicted in Fig. 7.13 (neglecting optical distortions).

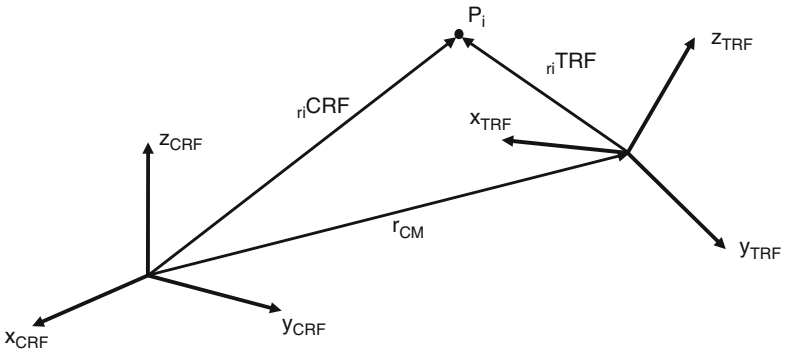


Fig. 7.14 Reference frames for relative position vector computation

First the cosine of the angle between each pair of unit vectors is determined by finding the dot product of the vectors. Then the law of cosines is applied to derive three non-linear equations with the ranges to each spot as unknowns:

$$l_{12}{}^2 = R_1{}^2 + R_2{}^2 - 2R_1R_2\cos\vartheta_{12} \tag{7.10}$$

$$l_{23}{}^2 = R_2{}^2 + R_3{}^2 - 2R_2R_3\cos\vartheta_{23} \tag{7.11}$$

$$l_{13}{}^2 = R_1{}^2 + R_3{}^2 - 2R_1R_3\cos\vartheta_{13} \tag{7.12}$$

Where l_{ij} is the distance from point i to point j, known from the target geometry, R_i is the range from camera to spot i, and θ_{ij} is the angle between unit vectors i and j.

Equations 7.10, 7.11, and 7.12, form a set of three coupled non-linear equations in the unknowns R_1, R_2, and R_3, which can be solved numerically using a non-linear root solver such as the Newton–Raphson method [43].

Once the unknown ranges are computed, they can then be used for relative attitude determination. In particular, the TRIAD algorithm can be adopted [44] using as input the relative vectors $\underline{R}_{12} = \underline{R}_1 - \underline{R}_2$ and $\underline{R}_{23} = \underline{R}_2 - \underline{R}_3$, which are known in the target reference frame (known geometry) and in the camera reference frame (calculated from the range vectors). Once relative attitude is known, the position of the center of mass of the target spacecraft can be calculated in the camera reference frame by a simple algebraic relation, that is

$$\underline{r}_{CM}^{CRF} = \underline{r}_i^{CRF} - M_{TRF\to CRF}\underline{r}_{P_i}^{TRF} \tag{7.13}$$

With reference to Fig. 7.14 *CRF* indicates the camera reference frame, *TRF* is the target reference frame, $M_{TRF\to CRF}$ is the relative attitude matrix, i is the considered feature point, and $\underline{r}_{P_i}^{TRF}$ is the vector from the center of mass to the feature

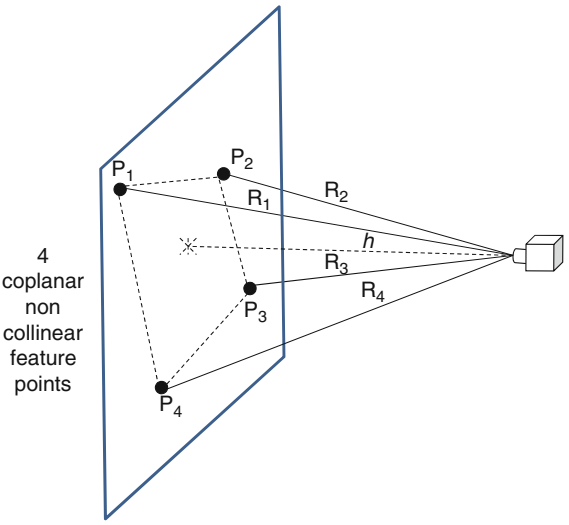


Fig. 7.15 Problem geometry for a monocular camera imaging four coplanar non collinear points

point *i*, calculated in the *TRF*. Precision in the relative position computation can be improved by averaging over the all detected features.

If more than 3 targets are imaged, then the solution is over-determined for both range and attitude. In fact, different solutions for ranges from camera to feature points can be calculated which are then averaged. From the relative attitude point of view, the solution can be optimized in a least squares sense by means of the QUEST algorithm and the elegant solution provided by Wahba [45, 46].

Indeed, in the case of three features, singularities can result when the feature plane is normal to the camera line of sight [34]. Moreover, non coplanar configurations allow increasing sensitivity to small rotations [37, 41]. Thus, at least four features are usually considered in practical applications to enable robust pose estimation.

For example, in the AVGS system, four non-coplanar retro-reflectors are used [47], which means that four independent angular estimates are available.

Again in a monocular configuration, the solution adopted in PRISMA experiment (short range cooperative mode) is based instead on a coplanar geometry with five LEDs placed in a unique non-collinear pattern for each panel of the target satellite [21]. The pose estimation process is based on the four points coplanar problem, which, as stated above, admits a unique solution achievable for example following Abidi and Chandra [47]. Thus, in general multiple pose solutions can be generated which are then averaged. The problem geometry is depicted in Fig. 7.15.

As before, the input data are the components in the camera reference frame of the unit vectors corresponding to the feature points. The solution is based on the fact that the four points and the camera center span four different tetrahedra sharing the same base plane and the same height, *h,* given as the distance from the base plane to the camera center.

Considering the base triangle $\Delta P_1P_2P_3$ (see Fig. 7.15), its area can be calculated by Heron's formula as

$$A_1 = \frac{1}{4}\sqrt{(s_{12}{}^2 + s_{13}{}^2 + s_{23}{}^2)^2 - 2(s_{12}{}^4 + s_{13}{}^4 + s_{23}{}^4)} \tag{7.14}$$

where s_{ij} represents the (known) distance between the two points P_i and P_j.

Then, for every tetrahedron, the volume can be calculated as a third of the product of the base area times the height h, or as a sixth of the parallelepiped span of the vectors forming the tetrahedron.

Thus, the following equalities can be imposed:

$$V_1 = \frac{1}{3}A_1 h = \frac{1}{6}\left|\overrightarrow{P_1}\cdot\left(\overrightarrow{P_2}\times\overrightarrow{P_3}\right)\right| = \frac{1}{6}R_1R_2R_3\left|\overrightarrow{u_1}\cdot\left(\overrightarrow{u_2}\times\overrightarrow{u_3}\right)\right| \tag{7.15}$$

where R_i represent the (unknown) distances and $\overrightarrow{u_i}$ are the relevant unit vectors, computed from images.

Applying these relations to the different tetrahedrons and combining the resulting equalities, it is possible to express the distances R_2, R_3, and R_4, as a function of R_1, as in (7.16)

$$R_2 = \frac{B_3A_3F_2}{A_3B_4F_1}R_1 \tag{7.16}$$

where B_i is twice the area of the projected triangles on the CCD (determined on the basis of the unit vectors $\overrightarrow{u_i}$) and F_i is the distance from the camera center to the projected point on the CCD (evaluated from pixel coordinates corrected for lens distortion and camera focal length).

After some mathematics [21, 47], at the end six different analytical expressions for each distance R_i are obtained, which can be averaged.

It is interesting to note that this solution method assumes that the reference between the projected point and which point is projected is known beforehand. In general, placing the LEDs in unique asymmetric patterns will give only a single solution where solving the six equations for R_i yields similar results.

Once the distances are known, relative attitude and position can be calculated as before (for example, relative attitude by means of the QUEST algorithm and relative position by means of an algebraic relation).

Another general approach to relative pose estimation on the basis of feature points and line of sight information is provided by Junkins et al. in the framework of the studies related to VisNav [48], a relative navigation system based on active infrared beacons and position-sensing diode technology. In the considered system, at least four measurements are used in the collinearity equations by applying a non linear iterative Gaussian least squares processing. The algorithms can accommodate any number of measurements. Relative attitude is modeled by means of modified Rodriguez parameters [49].

The iterative procedure starts by using an initial guess of relative position and attitude. Given this guess, it is possible to predict the angular position of the beacons, and given the real measurements, the residuals can be calculated. The correction is then computed by using a linearization of the measurement model and by imposing the minimization of the quadratic residuals. The procedure is iterated until convergence.

An interesting point of this algorithm is that it provides an estimate of the error covariance matrix in addition to its converged relative navigation estimate, which is very useful when using this technique within a Kalman filter algorithm. In particular, the covariance estimate forms part of the optimal differential correction calculated during each iteration. The procedure is independent from the orientation of the line of sight vectors, but different patterns will of course influence the final measurement covariance.

From an architectural point of view, algorithms for pose estimation based on known feature points usually rely on high contrast fiducial markers installed on the target satellites. Of course, this limits their use to the case of cooperative targets. Furthermore, mis-detection of one or more of these markers can lead to vision system failure.

Using natural features and object models can alleviate these limitations, at the cost of a deep complication of the pose estimation process, especially in the monocular case. In both monocular and 3D vision, model-based pose estimation relies on the idea of detecting natural features, such as lines or surfaces, and matching them with known model features, and is comprised of two phases usually based on different algorithms: pose determination (or acquisition), and pose refinement and tracking. Initial pose acquisition is the heaviest phase from the computational point of view. Within the monocular vision framework, different techniques can be adopted such as active appearance models [50, 51], robust principal component analysis/eigentracking [52] or simultaneous pose and correspondence determination (soft-POSIT) [53].

As an example, the main idea in active appearance models is to learn a parametric function that governs the possible appearances of an object. That is, if we can learn how to change a set of parameters in order to synthesize any desired appearance of the object, then we can find a perfect match between observed appearance (current image) and expected appearance (model). This information can then be used to recognize the object pose [50].

The visual orbit environment, with the possible presence of shadows, specular reflections and background features, can make this process very difficult and sometimes unreliable in the monocular framework. Thus, the models are usually made up of lines [54], which provides a good invariance to pose and illumination changes, and edges are used as the main features extracted in the images.

When an initial pose estimate has been produced, the pose tracking phase (which includes also pose refinement) is entered, and subsequent pose estimates can be based on fastest algorithms. For monocular systems, edges are again used as the main features to track, and non linear optimization processes are adopted which provide the best alignment between edges of the projected model and edges

extracted in the image. Some examples of this process are the virtual visual servoing framework [55] and the iterative reweighted least-squares formulation [35, 56].

7.4.2 Three-Dimensional Techniques

Three-dimensional pose estimation methods take advantage of the range information provided by laser rangefinders and/or stereo vision systems. As in the case of monocular techniques, some methods are based on feature points provided by fiducial markers, while other techniques make use of natural features and known models. The approaches are thus similar, while the main advantage of 3D vision systems is to provide a larger operation range and an increased robustness against the challenges of the visual environment on orbit.

When artificial markers are used, the input information for the algorithm is given directly by their three-dimensional coordinates in the camera reference frame. Thus, as done in Ref. [57] and in a similar way to the previous sub-section, relative vectors $\underline{R}_{ij}$ can be determined in the camera reference frame, and relative attitude can be estimated for example by means of the QUEST method. Afterwards, relative translation is computed algebraically.

Within the model-based approaches, the availability of (dense or sparse) 3D information makes easier to recover the pose. Conceptually, pose estimation is still comprised of an initial pose acquisition phase, followed by pose refinement and tracking. However, different mathematical techniques are used in the two phases compared with monocular systems.

Pose determination can be described as a search within the pose space of the object, and is formulated as a template set matching problem. The target satellite model is represented as a set of volumetric pixel (voxel) templates, one for each possible pose. Within a template, each voxel is assigned a binary value to indicate whether that region is empty or occupied. The resolution of the voxel grid has to be established on the basis of a trade-off between pose determination accuracy and size of the pose search space (i.e., memory occupancy and computational burden of the pose acquisition process). Pose estimation methods aim at maximizing the correlation between 3D sensed data and the voxel templates relevant to the possible object poses. In the geometric probing approach [30, 58], the set of all templates is composed into a binary decision tree. Each leaf node references a small number of templates. Each internal node references a single voxel, and has two branches, “true” and “false”. The sub-tree branching from the “true” branch contains the subset of templates that contain the node voxel. Conversely, the sub-tree branching from “false” branch contains the subset of templates that do not contain the node voxel. Traversing the tree at any image location executes a geometric probing strategy that efficiently determines a good match with the template set.

In order to improve efficiency and reliability, it may be possible to reduce the dimensionality of the search space by resolving some of the degrees of freedom prior to executing the geometric probing method. For example, this can be the case of satellites with very distinct major and minor axis, or with identifiable features which can be reliably extracted and used to resolve some of the positional ambiguity.

Regarding pose tracking, a common technique is the Iterative Closest Point algorithm [59, 60]. The ICP algorithm is an iterative procedure minimizing the distance between points in one set and the closest points, respectively, in the other. ICP implementation can follow the subsequent steps [60]:

1. Selection of the closest points between the model and the data set (the first data set is relevant to the initial pose estimation);
2. Rejection of outliers;
3. Computation of geometrical registration between matched data points and the model. This is done to minimize the mean square error between the two data sets, following for example the approach shown in Ref. [61];
4. Application of geometrical registration to the data;
5. Termination of the iteration after convergence is reached.

7.4.3 Dynamic Filtering Techniques

As for other relative navigation sensors, vision-based measurements can be used in dynamic filtering schemes to filter out sensor noise, increase data rate, improve reliability in case of loss of measurements, and enable multi-sensor data fusion.

Algorithms based on Kalman filtering are commonly used in this framework and can be embedded in pose tracking methods [35]. Due to the non linear relation between Cartesian coordinates and line of sight measurements, sub-optimal approaches such as the Extended Kalman Filter are used in absence of range information.

Especially when considering sensor fusion with other information sources, or there is the necessity to cope with loss of valid measurements, the choice of the dynamic model plays an important role. Regarding position and velocity estimation, Hill's equations can be effectively used for near circular orbits in the case of very short distances and relatively large measurement rate [57], while attitude dynamics equations have to be included in the dynamic model to provide relative attitude filtering [62]. Of course, more complex models of relative motion are more robust in case of loss of measurements.

As a general consideration, since vision-based systems provide measurements in the body reference frame, transformed measurements have to be used. This requires introducing attitude uncertainties in the model, for example by modifying the measurement covariance matrix, as done in Ref. [57].

An example of the potential of the dynamic filtering is shown in Fig. 7.16, which reports results of MonteCarlo simulations (mean plus and minus one standard deviation) relevant to a circular formation with radius 5 m, and compares errors on relative position estimation as derived from an Extended Kalman Filter based on

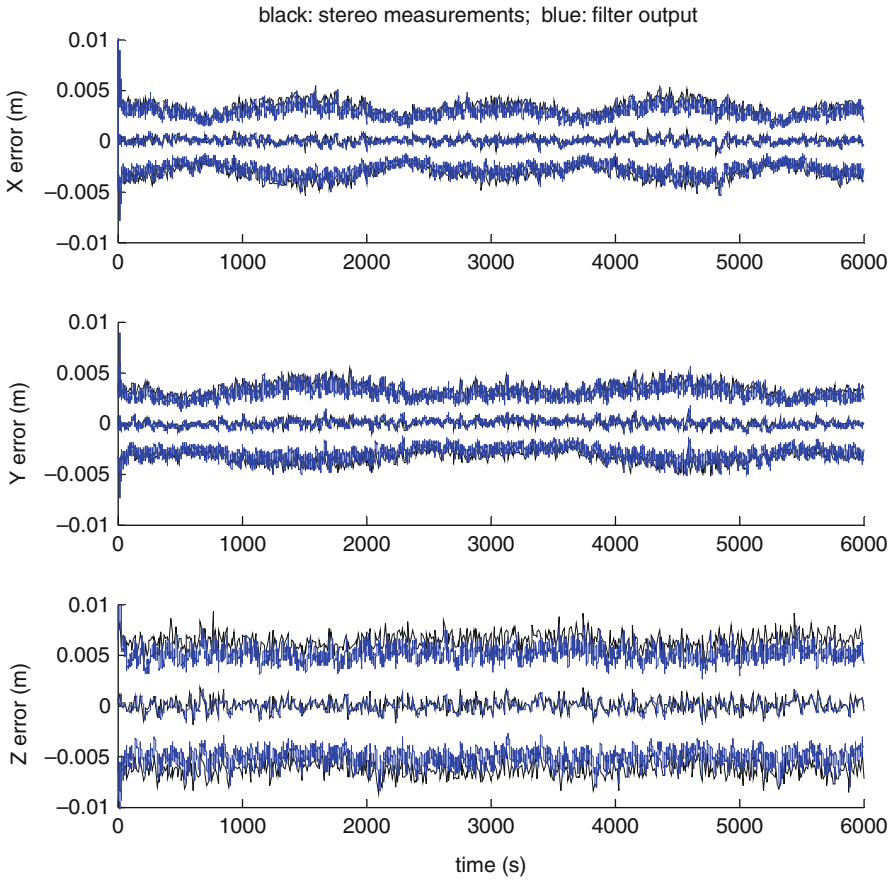


Fig. 7.16 Example of EKF performance (10 Hz) with 0.1 Hz stereo aiding frequency and 0.01° attitude determination error (Reprinted with permission of the American Institute of Aeronautics and Astronautics, published in Ref. [57])

Hill's equations (in blue, at 10 Hz) and from stereo-vision processing (in black, at 0.1 Hz), assuming an error on attitude determination for the chaser platform of 0.01° (1σ), compatible with the use of medium/high accuracy attitude sensors.

Dynamic filtering provides the mathematical framework for fusion of vision-based information with other systems such as differential GPS. Within this context, it is interesting to note that also far range line of sight measurements can be very useful for relative navigation in absence of GPS corrections, thanks to previous range measurements integrated in a non linear filtering scheme such as an Extended Kalman Filter.

7.5 Conclusions and Future Trends

In this chapter sensors, technologies, techniques and algorithms for vision-based relative navigation have been described. Particular emphasis has been given to sensors and technologies validated in flight with past or present space missions. Specifically, all the considered missions rely on two spacecrafts performing formation flying and/or rendezvous and docking maneuvers to test autonomous GN&C technologies. One of the two satellites is equipped with cameras and, in some cases, lasers to determine the relative position and attitude of the other satellite carrying on the external surface reflectors and/or LEDs, which can be imaged by the vision system.

Results of successful missions as ETS-VII and Orbital Express have demonstrated the capability of a vision system to compute the pose in real-time with high accuracy when the two satellites fly in close proximity, which is relevant to the terminal approach and docking phases of on-orbit servicing missions and to formation flying applications in which the satellites fly with short separations. Preliminary flight results from the recent Prisma mission confirm that in the framework of close range relative navigation vision-based systems represent an appealing technology with a great potential in terms of autonomy, reliability, and miniaturization, with their very small power, size, mass and cost requirements.

In the near future, it is likely that robotic on-orbit servicing applications will represent a strong driver towards non cooperative methods which can exploit natural satellite features thus avoiding to impose any requirements on target satellite buses. Then, the evolution of relative navigation techniques is likely to see a tighter integration and a more effective fusion of vision-based measurements with other information sources such as differential GPS. This will also allow to better face the challenges of the on-orbit visual environment.

References

1. On-orbit Satellite Servicing Study, Project Report, NASA, Oct 2010
2. Villien et al (2008) Formation flying guidance navigation and control design for science missions. In: Proceedings of the 17th world congress the international federation of automatic control, Seoul, Korea, July 6–11, 2008
3. Kaiser C et al (2008) Simulation of the docking phase for the SMART-OLEV satellite servicing mission international symposium on artificial intelligence, robotics and automation in space 2008, Hollywood, USA, Feb 26–29, 2008
4. Sellmaier F et al (2010) On-orbit servicing missions: challenges and solutions for spacecraft operations 2010 conference, Huntsville, Alabama, USA, AIAA 2010–2159
5. Davis TM, Melanson D (2005) XSS-10 micro-satellite flight demonstration. First Annual Space Systems Engineering Conference, Atlanta, Georgia (USA), Nov 8-11, 2005 Paper no. GT-SSEC.D.3
6. Woffinden DC, Geller DK (2007) Navigating the road to autonomous orbital rendezvous. AIAA J Spacecraft Rockets 44(4):898–909

7. Kasai T et al (1999) Results of the ETS-7 mission – rendezvous docking and space robotics experiments. In: Proceedings of the 5th international symposium on artificial intelligence, robotics and automation in space, ESTEC, Noordwijk, The Netherlands, 1–3 June 1999 (ESA SP-440), pp 299–306
8. Visentin G, Didot F (1999) Testing space robotics on the Japanese, ETS-VII satellite, in testing space robotics. ESA Bull 99:61–65
9. Mokuno M et al (2004) In-orbit demonstration of rendezvous laser radar for unmanned autonomous rendezvous docking. IEEE Trans Aerosp Electron Syst 40(2):617–626
10. Mokuno M, Kawano I (2011) In-orbit demonstration of an optical navigation system for autonomous rendezvous docking. AIAA J Spacecrafts Rockets 48(6):1046–1054
11. Kawano I et al (2001) Result of autonomous rendezvous docking experiment of engineering test Satellite-VII. AIAA J Spacecraft Rockets 38(1):105–111
12. Oda M, Kine K, Yamagata F, (1995) ETS-VII – a rendezvous docking and space robot technology experiment satellite. In: Proceedings of 46th international astronautical congress, Oct 1995, Norway, pp 1–9
13. Ueno H et al (2002) NASDA perspective on unmanned on-orbit servicing – projects and prospects. 1st bilateral DLR – CSA workshop on on-orbit servicing of space infrastructure elements via automation and robotics technologies, Cologne, Germany, Nov 24–26, 2002
14. http://www.jaxa.jp/projects/sat/ets7/index_e.html
15. Pinson RM et al (2008) Orbital express advanced video guidance sensor: ground testing, flight results and comparisons, AIAA guidance, navigation and control conference and exhibit, 18, Honolulu, United States, 18–21 Aug 2008
16. Weissmuller T Leinz M (2006) GN&C technology demonstrated by the orbital express autonomous rendezvous and capture sensor system. In: 19th annual AAS guidance and control conference, Breckenridge, Colorado (USA), Feb 4–8, 2006
17. Heaton AF et al (2008) Orbital express AVGS validation and calibration for automate rendezvous. In: AIAA guidance, navigation and control conference and exhibit, Honolulu, United States, 18–21 Aug 2008
18. Persson S et al (2005) PRISMA – demonstration mission for advanced rendezvous and formation flying technologies and sensors, Paper IAC-05-B56B07. In: Proceedings of 56th international astronautical congress, Fukuoka, Japan, Oct 17–21, 2005
19. Nilsson F et al (2007) Autonomous rendezvous experiments on the PRISMA in orbit formation flying test bed. In: Proceedings of 3rd international symposium on formation flying, missions and technologies, Noordwijk, The Netherlands, 23–25 Apr 2008 (ESA SP-654, June 2008)
20. Bodin P et al (2009) PRISMA: an in orbit test bed for guidance, navigation, and control experiments. J Spacecraft Rockets 46(3):615–623
21. Benn M Jorgensen JL (2008) Short range pose and position determination of spacecraft using a μ-advanced stellar compass. In: Proceedings of 3rd international symposium on formation flying, missions and technologies, Noordwijk, The Netherlands, 23–25 Apr 2008 (ESA SP-654, June 2008)
22. Jorgensen JL Benn M (2010) VBS – the optical rendezvous and docking sensor for PRISMA. NordicSpace, (ISSN: 0805-7397), pp. 16–19, http://www.nordicspace.net
23. Holst GC (2000) Electro-optical imaging system performance, 2nd edn. JCD Publishing, Winter Park
24. Jørgensen JL (1999) In-orbit performance of a fully autonomous star tracker. SAO/NASA astrophysics data system. http://adsabs.harvard.edu/full/2000ESASP.425.103J. Accessed 15 Dec 2011
25. Bayer BE (1976) Color imaging array. US patent 3,971,065 (1976)
26. Gonzalez RC, Woods RE (2008) Digital image processing, 3rd edn. Pearson International, Upper Saddle River
27. Alferez R, Wang Y-F (1999) Geometric and illumination invariants for object recognition. IEEE Trans Pattern Anal Mach Intell. doi:10.1109/34.771318
28. Pratt WK (1991) Digital image processing, 2nd edn. Wiley Interscience, Mountain View

29. Sako H, Whitehouse M, Smith A, Sutherland A, (1994) Real-time facial-feature tracking based on matching techniques and its applications. Pattern Recognition, 1994, vol 2 – conference B: Proceedings of the 12th IAPR international conference on computer vision and image processing, Jèrusalem, Israel, Oct 1994
30. Jasiobedski P, Greenspan M, Roth G, (2001) Pose determination and tracking for autonomous satellite capture. In: Proceedings of 6th international symposium on artificial intelligence, robotics and automation in space, Montréal, Québec, Canada, June 2001
31. Woffinden DC, Geller DK (2009) Observability criteria for angles-only navigation. IEEE Trans Aerosp Electron Syst 45(3):1194–1208
32. Hill GW (1878) Researches in the lunar theory. Am J Math 1(3):245–260
33. Schmidt J, Lovell T (2008) Estimating geometric aspects of relative satellite motion using angles-only measurements. In: Proceedings of AIAA/AAS astrodynamics specialist conference and exhibit, Honolulu, Hawaii, Aug 18–21, 2008
34. Woffinden DC, Geller DK (2007) Relative angles-only navigation and pose estimation for autonomous orbital rendezvous. J Guid Contr Dyn 30(5):1455–1469
35. Kelsey JM, Byrne J, Cosgrove M, Seereeram S, Mehra RK, (2006) Vision-based relative pose estimation for autonomous rendezvous and docking. In: Proceedings of IEEE aerospace conference 2006, Piscataway
36. Tomasi C, Kanade T (1992) Shape and motion from image streams under orthography: a factorization method. Int J Comput Vision 9(2):137–154
37. Yuan JS-C (1989) A general photogrammetric method for determining object position and orientation. IEEE Trans Robot Autom 5(2):129–142
38. Fishler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
39. Wolfe WJ, Mathis D, Sklair CW, Magee M (1991) The perspective view of three points. IEEE Trans Pattern Anal Mach Intell 13(1):66–73
40. Haralick RM, Lee C-N, Ottenberg K, Nolle M (1994) Review and analysis of solutions of the three point perspective pose estimation problem. Int J Comput Vision 13(3):331–356
41. Calhoun PC, Dabney R (1995) Solution to the problem of determining the relative 6 DOF state for spacecraft automated rendezvous and docking. Proc SPIE 2466:175. doi:10.1117/12.211505
42. Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed 11 Dec 2011)
43. Press WH (1986) Numerical recipes. Cambridge University Press, New York, pp 269–272
44. Lemer GM (1978) Three-Axis Attitude Determination. In: Wertz JR (ed) Spacecraft attitude determination and control. D. Reidel Publishing Co., Dordrecht, pp 420–428
45. Shuster MD, Oh SD (1981) Three-axis attitude determination from vector observations. J Guid Control 4(1):70–77
46. Wahba G (1965) A least squares estimate of satellite attitude. Problem 65.1. SIAM Rev 7(3) Jul 1965, p 409
47. Abidi MA, Chandra T (1990) Pose estimation for camera calibration and landmark tracking. In: Proceedings of IEEE international conference on robotics and automation 1990, vol 1, Cincinnati, 13–18 May 1990, pp 420–426
48. Crassidis JL, Alonso R, Junkins JL (2000) Optimal attitude and position determination from line-of-sight measurements. J Astronaut Sci 48(2–3):391–408
49. Schaub H, Junkins JL (2003) Analytical mechanics of space systems, 1st edn, AIAA Education Series. American Institute of Aeronautics and Astronautics, Washington, DC
50. Mittrapiyanuruk P, DeSouza GN, Kak AC (2004) Calculating the 3D-pose of rigid objects using active appearance models. In: Proceedings of the international conference in robotics and automation, 2004, New Orleans
51. Abbasi S, Mokhtarian F (2001) Affine-similar shape retrieval: application to multiview 3-D object recognition. IEEE Trans Image Process 10(1):131–139

52. Black MJ, Jepson AD (1998) EigenTracking: robust matching and tracking of articulated objects using a view-based representation. Int J Comput Vision 26(1):63–84
53. David P, DeMenthon DF, Duraiswami R, Samet H (2002) Softposit: simultaneous pose and correspondence determination. In: Proceedings of 7th ECCV, vol III, Copenhagen, Denmark, May 2002, pp 698–703
54. Antoine P, Eric M, Keyvan K (2011) Vision-based space autonomous rendezvous: a case study. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems (IROS), 2011, San Francisco, 25–30 Sep 2011, pp 619–624
55. Comport AI, Marchand E, Pressigout M, Chaumette F (2006) Real-time markerless tracking for augmented reality: the virtual visual servoing framework. IEEE Trans Vis Comput Graph 12(4):615–628
56. Drummond T, Cipolla R (2002) Real-time visual tracking of complex structures. IEEE Trans Pattern Anal Mach Intell 24(7):932–946
57. Fasano G, Accardo A, Grassi M (2009) A stereo-vision based system for autonomous navigation of an in-orbit servicing platform. In: AIAA Infotech@Aerospace 2009, Seattle, USA, ISBN-10: 1-56347-971-0, Apr 2009
58. Greenspan MA, Boulanger P (1999) Efficient and reliable template set matching for 3D object recognition, 3DIM99: Proceedings of the 2nd international conference on 3-D digital imaging and modeling, Ottawa, Canada, Oct 4–8, 1999, pp 230–239
59. Besl PJ, McKay HD (1992) A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 14(2):239–256. doi:10.1109/34.121791
60. Abraham M, Jasiobedski P, Umasuthan M (2001) Pose determination and tracking for autonomous satellite capture. In: Proceedings of 6th international symposium on artificial intelligence, robotics and automation in space, Montréal, Québec, Canada, June 2001
61. Horn BKP (1987) Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am 4(4):629–642
62. Aghili F, Kuryllo M, Okouneva G, English C (2010) Robust vision-based pose estimation of moving objects for automated rendezvous and docking. In: Proceedings of international conference on mechatronics and automation (ICMA), Xi'an, China, 2010, 4–7 Aug 2010, pp 305–311

Part III
Technological Challenges

Chapter 8
Autonomy

Claudio Iacopino and Phil Palmer

Abstract Missions involving multiple spacecraft, autonomously working together, have become of great interest in the last decade as they offer a number of scientific and engineering advantages. This trend is responsible for an increasing demand on mission planning and scheduling systems able to coordinate the different spacecraft and to allocate tasks amongst them. New approaches are therefore needed to handle this new level of complexity, combining together autonomous solutions for the ground and space segment. The chapter is organized as follows: after an introduction on the motivations of using autonomy solutions for space missions, sect. 8.2 presents a short survey on the applications of autonomy in space Operations. The section identifies the focus of the chapter in Mission Planning and Scheduling, one the most critical aspect for distributed mission. Section 8.3 gives the state of the art regarding mission planning and scheduling applications first on single platforms and then on distributed platforms. Section 8.4 explains the most popular technology for distributed missions, the multi agent paradigm. The section is not meant to cover all the aspects of this technology. It focuses instead on the challenges related to the new trends of self-organizing systems and natural-inspired approaches, techniques that can offer promising advantages for distributed missions.

8.1 Introduction

In the last decade, the interest for space missions involving multiple spacecraft is rapidly growing. They offer a number of key scientific drivers over single monolithic spacecraft: signal separation (e.g., large synthetic apertures), signal space coverage (e.g., multi-point sensing) and signal combination (e.g., data fusion) [14]. Beside them, distributed missions bring unique advantages also on the engineering level: increase of reliability and extensibility, application of economies of scale and

C. Iacopino (✉) • P. Palmer
Surrey Space Centre, University of Surrey, GU2 7XH, United Kingdom

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_8,

of new computational paradigms – concepts such as grid, cloud computing or self-organization present potentials to be applied to swarm of satellites. Distributed missions are definitely the new trend for the space sector. Multiple platforms are indeed already largely adopted for communication, geo-location (GPS) and meteorology purposes. However, only recently this paradigm is showing its potentials for Earth Observation and Space Exploration.

The Earth Observation field, despite the economic downturn, will grow at a rate of 16% per year over the next decade, as showed in the research report by Market Intel Group (MiG) [34]. This is due to the miniaturisation of satellites and corresponding reduction in cost that has opened the opportunity of many end users to use satellite data. The Global Monitoring for Environment and Security – GMES is a prime example of the global nature of Earth imaging data for international response. In this case, five ESA spacecraft devoted to Earth Observation (Sentinel-1 to Sentinel-5) need to cooperate with other existing and/or planned missions provided by ESA, EUMETSAT, other national agencies or private companies such as Surrey Satellite Technology Ltd (SSTL) with the Disaster Monitoring constellation.

Distributed space missions are usually identified by different names including constellations, clusters and swarms. The first category is the traditional constellation approach where, to have maximum coverage, satellites are sparsely distributed and do not require precise inter-spacecraft positioning. This solution is usually used for Earth Observation, communication, geo-location (GPS) and meteorology purposes. In the cluster scenario, spacecraft are intentionally closely together to increase the sensing capabilities. In such a cluster, formation-flying astrodynamics, combined with precise attitude determination and control systems must be used in order to avoid spacecraft collisions and maintain the configuration. The terminology swarm satellite is broadly similar to a cluster satellite; however, the concept of swarm implies a larger number of spacecraft than a cluster and smaller and cheaper platforms.

For future missions large numbers of spacecraft are envisaged and hence the current Operations concepts designed for individual spacecraft are not necessarily transplantable in the context of missions such as the ANTS framework where thousands of spacecraft are envisaged. As the number of spacecraft in distributed missions increases, it is foreseeable a decrease of the platform size. The mass of a spacecraft is going to be a critical parameter concerning the processing and communication capabilities. Now there is lots of interest in nanosatellites, platforms of the order of few kilograms, as they are not anymore just technology demonstrators but are able to achieve scientific results.

8.1.1 Nanosatellite

A good example of a nanosatellite is the pioneering SNAP-1 that was launched in 2000 by the Surrey Space Centre weighting 6.5 kg [56]. SNAP-1 was able to perform wide area earth imaging and formation flight with a larger microsatellite

Tsinghua-1 launched at the same time. SNAP-1 demonstrated that nanosatellite platforms were useful and achievable already 10 years ago. Importantly the entire mission was possible for a relatively cheap $2 million (at the time) as it was based around commercial off-the-shelf technology (COTS). Other more recent nanosatellites include MicroLink-1 (featuring redundant LEON processors and 20 Gbits of EDAC SDRAM) and the ESA Munin (powered by a Texas Instrument TMS320C50). Even on low power miniature nanosatellites therefore, it is possible to incorporate reasonably powerful on board computers, which can only increase in power in future missions. Unfortunately, the same trend is not shared by the communication capabilities relying on antennas and RF subsystems, critical challenges for nanosatellites.

Finally, it is worth considering the CubeSat platform that has become a standardised spacecraft configuration adopted by space researchers around the world. CubeSats sit on the border between pico and nano classes spacecraft as they are a standardised 10 $\times$10 $\times$10 cm cube with the weight restricted to 1 kg. However, up to three cubes can be connected together for a combined 3 kg weight, in a cuboid configuration. Example of this architecture is the mission STRaND, a fully-functionals satellite that it is going to demonstrate a number of advanced technologies in space, such as smartphones. Researchers around the world have focused on developing compatible COTS technology within the mass and power constraints, so that at this point, the actual platforms and payloads are highly customisable with multiple successful launches to date. It is clear however, that for this type of platform the computational power available and the communications capabilities are challenging constraints that must be carefully evaluated.

8.1.2 Why Autonomy?

Autonomy is a concept of growing interest, particularly relevant for distributed space missions. An increasing number of missions is showing the feasibility and the advantages of autonomy applications, in several aspects of mission Operations. The following list points out the main motivations for such a interest:

- Decrease of Operational costs – A major, often overlooked, cost of space missions is the ongoing cost of ground control. This can account for around 5% of the unit cost per year, even for the relatively simple architecture currently used. This is due to the significant resources needed by the ground station staff for commanding the spacecraft. Consequently, it forms a large part in the overall cost of the Operations.
- Efficiency/Science Return – One of the main goals of mission Operations is to ensure maximum efficiency in the use of the resources and to avoid periods of inactivity. Without some degree of autonomy, this cannot be ensured due to communication delay and possible failures.

- Robustness – Capability of maintaining acceptable performances under non-nominal conditions, (in case of failures).
- Flexibility/Adaptability – Ability to change its behaviour according to changes in the environment. This concept would enable Goal-oriented Operations [20]. The operators have only partial information about the spacecraft state; under this new approach, instead of sending low-level commands for each instant of time, they communicate high-level goals to be achieved and relative deadlines.
- Responsiveness – Desirable property to respond rapidly to environment changes such as emergencies.
- Complexity handling – Autonomy is necessary to address computationally hard problems such as distributed constraint satisfaction and decentralized optimization that are dynamically, evolving and highly complex in terms of interactions and dimensionality.
- Enabling new mission concepts – Future mission scenarios, such as high-responsive Earth Observation constellations or swarms of satellite for exploration mission would not be possible without a high-level autonomy.

While autonomy should enable more complex and robust missions, it will also reduce controller workload allowing controllers to handle more spacecraft and, thanks to Goal-oriented Operations, to avoid information overload. The advantages presented above are already significant for traditional space missions but become necessary for distributed missions where the complexity of performing mission operations increases dramatically with additional platforms.

8.2 Autonomy in Space Missions

Already 10 year ago NASA proposed a roadmap for the gradual inclusion of autonomy solutions in the space missions [56]. This vision sees the following steps to be realized in these decades: isolated on-board autonomy experiments, ground based autonomy, on board autonomy (automated subsystems, AOCS), high-level spacecraft autonomy (flight navigation, collision avoidance, planning and FDIR) and lastly fleet spacecraft autonomy (distributed missions). This section is going to present a short survey on the applications of autonomy in space Operations, identifying the focus of the chapter in Mission Planning and Scheduling, one the most critical aspect for distributed missions.

8.2.1 What is Autonomy?

Autonomy comes from the Greek word *autonomia* that means freedom to live by one's own laws. It is actually commonly accepted as the condition or quality of being autonomous, self-government. This concept is translated in science as the

independence from external control. In Computer Science, the concept of autonomy is strictly connected with the field of Artificial Intelligence where it became a key element in research subfields such as the Agents theory. An agent is indeed defined as an autonomous entity with some capabilities of perceiving the environment and acting on it to achieve its goals. Autonomy is theoretically independent by the concept of intelligence that implies the agent's skills. An agent highly sophisticated might be less autonomous than a simple well-defined agent able to act independently without a human user. In real situations, making a system autonomous is a very challenging task as it means facing the uncertainty of the reality that implies the development of intelligent software. In complex systems such as a spacecraft, the level of autonomy becomes then strictly connected with the level of intelligence of the software. An agent by definition is supposed to be autonomous in all his functionalities but in reality, a system can present different degrees of autonomy/intelligence depending on the activities performed without external control. In the space sector, the field more influenced by autonomous technologies is the Operations because it involves human operators as external control. In general, it is possible to talk about autonomy in several contexts, as the number of systems and functionalities involved is very high. The following list presents a short survey on the applications of autonomy in space Operations: rover navigation, GNC, signal processing, maintenance and monitoring and mission planning.

- **Rover Navigation**. Rover navigation receives lots of attention due to the increasing number of missions sending exploratory rovers to Moon and Mars. In this case, autonomy is involved in tracking and path planning. NASA's Mars rovers Spirit and Opportunity demonstrated autonomous navigation technologies using four pairs of stereo cameras, as well as feedback from inertial wheel movements. The spectacular success of the missions allowed further software modules to be uploaded to demonstrate additional capabilities such as target tracking and global path planning. These modules are part of a the framework CLARAty, a reusable robotic software framework that supports heterogeneous robotic platforms and integrates advanced robotic capabilities [45].
- **GNC**. In the last years, a number of missions are successfully demonstrating autonomy solutions of increasing complexity applied to Guidance Navigation and Control (GNC). These range from orbital maintenance for a single spacecraft, PROBA-1 [54] and PROBA-2 mission [42], to rendezvous (docking) for two spacecraft, PRISMA [8]. Moreover a number of studies is addressing formation flying with multiple missions.
- **Signal Processing**. Space missions, in particular for Earth Observation are constantly increasing the size of the data acquired. The memory and processing on board available follow the same trend, however the downlink bandwidth cannot be easily extended. One of the main strategies used is to compress the data before the downlink [41]. This reduces the bandwidth needed but it can be very computational expensive. Autonomy could decrease directly the data to be downloaded implementing advanced on board functionalities such as target identification, discharge of poor data such as cloudy images and bands selection

in multi-spectral sensors for specific target features. These last two functions have been showed on the mission Earth Observing One respectively [4, 13]. Similar needs concern rover explorations, where the downlink bandwidth is not sufficient for the data collected by the rover. The Onboard Autonomous Science Investigation System (OASIS), a NASA project, aims at carefully selecting the data with the highest science interest for downlink [10].

- **Monitoring and FDIR.** On ground monitoring and Fault Detection Isolation and Recovery (FDIR) are major tasks for Operations that involves thousands of telemetries. Such amount of data is not only critical for the downlink bandwidth but also for the analysis and the responsiveness in case of failures or anomalies. The telemetry analysis is indeed a meticulous and time demanding process and involve the correlation analysis between telemetries of different sub-systems and the correlation analysis on the temporal dimension in case of periodic anomalies [41]. The challenge is from one side to improve the tools for the analysis on ground and on the other side to push most of these elaborations on board, downloading only the telemetries that are involved on a specific anomaly. Just a recent example is the mission PROBA-2 [42].
- **Mission Planning and Scheduling**. Mission planning is a crucial activity because it concerns with planning and scheduling the activities of a spacecraft in the most efficient way possible to increase the science return. It can deal with problems of high complexity (easily NP-Hard). In these cases, human operators can only find acceptable solutions. More general fields such as Operation Research and Automated Planning and Scheduling have developed a number of methods and techniques to handle these problems in a more efficient way.

Some of these approaches are quite mature but the perspective of distributed missions is introducing new challenges, specifically for GNC and mission planning, the two applications more affected by the platform multiplicity. It is therefore the aim of this chapter to focus on these aspects, specifically on the Mission Planning as GNC is going to be deeply explored by other chapters.

8.3 Automated Mission Planning and Scheduling

Planning and scheduling spacecraft operations involves generating a sequence of spacecraft commands from a set of higher-level science and engineering goals without violating any system rules or constraints and optimizing the resources available. Due to the complexity and the computation power required, historically the responsibility for Mission Planning has remained within the Ground Segment. Clear advantages such as reducing Operations cost and increasing the efficiency, are pushing this activity on board of the spacecraft. However, the single spacecraft are not able to deal with the whole complexity of the system, in particular in distributed scenarios. As Funase [23] and Damiani [17] suggested the solution is going to be a separation of responsibility between the higher level task distribution

and allocation mechanism done at ground segment and the lower-level localised re-planning and scheduling performed at the space segment. The following paragraphs present first the solutions developed for single platforms divided in on ground and on board autonomy applications and then is exploring the studies concerning the multiple platforms. A focused survey on the characteristics and differences between single platforms P&S systems has been provided by Moylan and Atkins [43].

8.3.1 Single Platform

Few solutions, highly sophisticated, have been developed for solving Planning and Scheduling problems in the space context. The SPIKE scheduling system was designed for NASA's Hubble Space Telescope and has been used since 1990 [35]. It performs long-term schedules, typically around 1 year in duration. SPIKE has been adopted for a number of missions, mostly astronomy missions and now it is evolving in the system for the future James Watt Space Telescope [26]. A more advanced planner is ASPEN (Autonomous Scheduling and Planning Environment) [11] developed by the Artificial Intelligence group at the NASA's JPL. ASPEN is a reconfigurable framework, composed of a constraint modelling language to define the application domain, a constraint management system, a set of search strategies for plan generation, a language for representing plan preferences and a graphical interface for visualizing plans/schedules. The spacecraft knowledge is encoded under seven classes: activities, parameters, parameter dependencies, temporal constraints, reservations, resources and state variables. ASPEN adopts a local, early-commitment, iterative search to optimize the objective functions. It has been applied to several missions such as Deep Space One (DS-1) and Earth Observing One (EO-1) to generate baseline plans that could then be modified on board. A parallel project is the ESA APSI, aiming to create a software framework to improve the cost-effectiveness and flexibility of development of mission planning support tool using automated planning and scheduling technology. In APSI, a given problem is represented as state variables, consistency features and a domain theory, a collection of synchronisations and rules. The APSI framework follows a timeline-based approach. The core comprises five layers at different reasoning abstractions: user layer, solver layer (decomposing goals), domain management layer (maintaining a decision network), component layer and temporal layer. The last one manages the temporal network. The validly of such approach has been demonstrated in several cases of study from exploration missions such as Mars Express to astronomy missions such as Integral and XMM-Newton [52]. In a different context, Mars Exploration Rover (MER) mission, the system MAPGEN was deployed in the ground segment system [5]. It is the result of integrating an interactive activity plan editor with a constraint-based planning framework, lying on a simple temporal constraint network (STN). Though this system is able to generate valid plans, it is not performing optimization.

Table 8.1 Benefits comparison between ground segment autonomy and on board autonomy

Ground autonomy	On board autonomy
CPU power available	Reactivity to the environment
Software flexibility	Processing data without communication delay
Testing procedure not impacting the mission	Reduced communication to ground
Interactions with operators and experts in short loop	
Lower cost of software development	

On board autonomy for Mission Planning is concentrated on the ability to modify and maintain dynamic plans for individual spacecraft. Promising results have been achieved by missions such as the NASA DS-1 [44], launched in 1998 with the system RAX-PS [37]. This was the first mission functioning for a few days completely autonomously: the planner component was able to translate high-level goals, coming from the ground segment, in plans. These plans were sent to a multithread executioner module in charge of translating them in executable commands for the flight systems. Lastly, a diagnosis and recovery mechanism could monitor the execution and, if necessary, trigger failure recovery procedures. The planner itself consisted in a search engine operating over a constraint-based temporal database. The spacecraft constraints were codified in a domain model formed of state variables and timelines. Domain-dependent heuristics were used to speed-up the search. NASA EO-1, launched in 2000, demonstrated the architecture ASE with a new re-planning system CASPER [38]. This system implemented continuous planning, increasing the reactivity of the system. It was able to repair a plan in case of anomaly in few seconds while the previous system RAX-PS was taking hours. The advantage is due to the technique called *iterative repair*, a planning approach that fixes flaws in an existing plan repeatedly until a good plan is found. Casper tries to address only the problems the anomaly introduced into the current plan. RAX-PS used a backward chaining refinement search technique combined with constraint propagation. In contrast, Casper uses local search to refine incrementally an existing plan. The ESA demonstrator mission PROBA-2 [25] launched in 2009 focused as well on the increasing of autonomy on board, interacting with the ground station only to download data and to upload the observations list. Further, the FDIR component was able to handle a large amount of anomalies. Some additional theoretical studies have been carried out for instance on the UK-DMC mission, the NEAT architecture was able to perform on board continuous planning using an evolutionary algorithm [9]. Table 8.1 summarises the benefits of the two different paradigms.

A key element that needs to be considered is the availability of the information necessary for taking decisions. The paradigm chosen should then depend also on the location of such information. For example if the requirement is to make the system more responsive to failure, than on board autonomy is the obvious direction but if instead the system should be more reactive to user requirements, than the ground segment should handle the complexity. In general both the scenarios are required; hence autonomy needs to be addressed either for the ground either for the

space segment. Solutions developed for individual spacecraft are not necessarily transplantable in distributed mission contexts. In essence, the distributed missions introduce a new element of complexity in the mission planning: the coordination mechanism between the spacecraft. This is strictly dependent on the inter-satellite link and the computation power that are critical resource constraints. Finding the right level of sharing of responsibilities between ground and space segment is one of the main challenges for future missions.

8.3.2 Multiple Platform

Multiple platforms are already largely adopted for communication, geo-location (GPS) and meteorology purposes. In these cases however the satellites are sparsely distributed and do not require coordination. The new challenges come from three main scenarios:

- Cluster or swarm satellites used for remote sensing or astronomy
- Swarm satellites used for space exploration
- Earth Observation constellations

In the first case, signal separation and signal space coverage are the first benefits. However, formation-flying astrodynamics, combined with precise attitude determination and control systems must be used in order to avoid spacecraft collisions and to maintain the configuration. Examples of studies for future astronomy missions are: the 2-satellites optical interferometry mission NASA ST3, the tetrahedral formation ESA LISA for detecting gravitational waves or the five free-flying formation of the ESA DARWIN mission designed to search for Earth-like exoplanets. Other proposals are outlined for studying space weather; more ambitiously is the project QB50, started on July 2011. It is planning to develop a network of 50 double CubeSats, separated by a few hundred kilometres and carrying identical sensors, for the studying key constituents of the thermosphere (90–320 km). In this scenario, the challenges concern mainly relative navigation, therefore it will not further explored in this chapter.

In the space exploration scenario, frequent and timing communications to the ground can be unfeasible. The coordination problem therefore needs to be handled on board. A challenging proposal exploring mission is the Autonomous Nanotechnology Technology Swarm (ANTS) aiming to explore the solar system with swarms of thousands of picosatellites ($<$1 kg) [16, 32]. Nowadays work on board autonomous mission planning and scheduling among multiple spacecraft is still first steps and it is an area of theoretical research. An example is the Shared Activity and Coordination model [15], where spacecraft modelled as agents negotiate with each other directly to partition the work among them. Projects such as the control system D-SpaCpanS [50] and the NASA TechSat21 [6] have shown how a hierarchical planning architecture could be a good trade-off between excessive communications and excessive computation. The drawback in that case is the rigid role structure

between the satellites. A challenging mission was Three Corner Sat (3CS) [12], three university nanosatellites with onboard science data validation, responsive re-planning, robust execution and anomaly detection, but unfortunately it failed due to a problem with the rocket during launch. The recent works are showing lots of interest in modelling the whole system using the multi agent paradigm [2, 3, 30]. In particular the ESA study DAFA (Distributed Agents for Autonomy) investigated the advantages of using distributed agents in a space system [46]. It analysed different scenarios, implementing two demonstrators: one for the Global Monitoring for Environment and Security system (GMES) and one for collaborative rover exploration on Mars [1]. Similar research has been carried on considering the current NASA rovers [21]. Very interesting is the approach of Tripp and Palmer [55] that applied an indirect communication system to reduce the communication overhead. The communication and computation overhead due to the coordination process itself is indeed the main issue that arises from these studies. It is required therefore to consider carefully the coordination paradigm depending on the specific scenario. Van der Host [57] has recently showed the trade-off between centralized and distributed planning in front of the number of spacecraft to coordinate. Concluding, a key element to consider is the cost of communication link. In the case of space exploration missions, this cost is too high or the delays are not acceptable. This is indeed the scenario where on board autonomy is going to be extensively adopted. The main challenge in this case regards the intersatellite communication system, required for coordinating the platform. This aspect will be further explored in other chapters.

In the Earth Observation constellations scenario, the focus of this book, the main goal is to satisfy efficiently the user community. The plan can be produced on ground and uploaded to the whole constellation because the communication link is not a critical resource. The level on uncertainty is quite low and this approach does not exclude re-planning capability on board in case of failure. In this scenario therefore, the need of autonomy is shifted from on board to the ground segment. One of the few autonomous Operations examples that have been demonstrated in space is the tandem mission TerraSAR/TanDEM-X [39], where basic functionalities of automated scheduling have been implemented though without optimizing the resources. The real big challenge is coordination and optimization at the same time. A number of studies have recently shown interest on Earth Observation constellations [18, 29, 48], in particular in case of disaster management [49, 59]. Most of them try to reduce the coordination aspect to an optimization problem and to solve it with classic techniques such as greedy [48, 59], backtracking [29] or simple heuristics [18]. In these cases either they did not achieve efficient solutions either they considered small problems (reduced number of spacecraft). Moreover, these works did not consider the dynamics of the problem itself. This scenario is a dynamic problem. In case of the GMES or of the Charter system, the five ESA spacecraft devoted to Earth Observation (Sentinel-1 to Sentinel-5) need to cooperate with other existing and/or planned missions provided by ESA, EUMETSAT, other national agencies or private companies such as Surrey Satellite Technology Ltd (SSTL) with the Disaster Monitoring constellation or the

RapidEye constellation. Lastly, the solution envisaged needs to be particular responsive to the requests coming from user community. The demonstrator of the DAFA study [46] aims at addressing these issues with a multi agent architecture based on negotiation paradigm and deliberative agents. The main drawback however is still in the lack of scalability and flexibility.

The following section is going to focus on the technologies required to model the coordination and control aspects of such missions, specifically in the EO constellation scenario. As explained above, in this case the goal is to develop a system highly responsive, adaptable to a dynamic environment and scalable to a various number of spacecraft. The technology investigated is the multi agent paradigm, as it seems the most popular approach for meeting these requirements.

8.4 Technologies for Distributed Systems: Multi Agent Systems

Multi Agent Systems (MAS) is a very wide field that brings together techniques and theories from multiple disciplines. It first appeared to solve communication problem in distributed environment such as networks or Internet. Today multi agent systems are commonly used and applied not only for communication and coordination purposes but also for modelling natural systems (agent-based modelling), as computational model and simulation-based design. It seems to be a convenient approach in the following cases [54]:

- Problems too large for a centralized agent.
- Problems naturally regarded as a society of autonomous interacting components.
- Information sources or expertise spatially distributed.
- Performance enhancement (computational efficiency, reliability, extensibility, robustness, maintainability, responsiveness, flexibility).

There are a number of different mechanisms that can be used to coordinate together multiple agents for a common purpose. These approaches are strictly connected with the capabilities of agents that range across the spectrum from reactive to deliberative architecture [22]. Reactive agents have relatively simple rules of interaction, but the overall system behaviour that emerges shows complexity that can be utilized to solve global objectives. Deliberative approaches are characterised by explicitly planning the individual behaviours of the agents in advance. The planning can be centralised or distributed but the agents are confined to operate in a deliberate way in order to meet mission objectives. In essence, these two ends of the spectrum can be characterised as performing a task in a highly planned manner (deliberative), or relying instead on an instantaneous spontaneous manner (reactive).The reactive approach is highly suited with dynamic problems with uncertainty. It is the most suitable for describing natural complex systems with high number of entities interacting with complex dynamics but is generally less efficient than the deliberative approaches and could lead to unpredictable and unstable system behaviours [36]. Deliberative approaches require large quantities

of processing power to resolve all the rules built into the plans and produce in general more inflexible solutions, not able to face dynamic problems. Considering the benefits explained above, MAS seems to be the best choice for the multiple spacecraft scenario. The majority of work on autonomous systems for spacecraft, mentioned in the previous paragraph, focused indeed on this paradigm. However, almost all applied deliberative techniques, exemplified by methods for deliberative planning and re-planning. The reason for this choice is mainly the higher reliability that deliberative architectures can offer. Mission planning is a critical scenario and requires high levels of confidence in the system outputs and performance. On the other side, the limitations of this approach badly fit with the requirements of a distributed platform. Reactive architectures can match these requirements and, thanks to new communication paradigms and control techniques, might achieve a suitable level of reliability to be applied to real scenarios.

A discussion on the different dimensions of analysis and design of MAS is out of the scope of this chapter. The section focuses on the challenges regarding reactive architectures and how they can be applied to distributed platforms.

8.4.1 Self-organization and Emergence

The key element behind the reactive multi agent systems is the self-organizing and emergent behaviours offered by these systems. In a swarm consisting of a large number of entities, the result of combining simple behaviours at local level can end in an emergent complex behaviour at the system level able to achieve significant results. Galaxy structures, cells or organisms are just few examples. The traditional reductionist view trying to explain everything, referencing to the law governing the constituent elements, lacks in understanding of how these phenomena arise [24]. This lack is becoming crucial in fields such as engineering that is facing problems of increasing complexity. Holland claimed that a collection of dumb agents is often better suited than a single smarter one [33]. As an agent get smarter, its functionality and flexibility reduces.

The literature does not offer a unique definition of self-organization and emergence and the distinction between them is indeed quite blurred. A number of authors tried to clarify the differences and provide a precise classification, such as Prigogine, identifying emergence in dissipative structure [27], Gabbai or Müller-Schloer more focused on engineering aspects. Hills [31] summarized the notion of emergence as “the whole is greater than the sum of the parts”. This idea is resumed by De Wolf and Holvoet [19] that defined a system as self-organizing when it shows structures or patterns at system level without a central or external authority (flocking patterns, snow flakes, so on). The concept of emergence is instead attributed to a system property that arises out of the multiplicity of relatively simple interactions and that cannot be reduced as a sum of such interactions (weather, life, thought). Self-organization and emergence are indeed concepts strictly interconnected

though in some cases they might appear as separate phenomena. In the following, these two terms are going to be used without distinctions, except if required.

As presented, self-organization is a desirable characteristic that need to be imported in artificial systems that cope with high uncertainty and dynamic environments, such as distributed space missions.

8.4.1.1 Design Emergence

The challenge in designing a self-organizing system is that there is no systematic way to formulate required micro-level behaviours given desired top-level macro behaviours. Nowadays a theory of emergence is still far to be achieved. Researchers have been experimenting with several mechanisms leading to emergent phenomenon. The different approaches, as presented by Serugendo [51], can be divided in four categories:

- **Cooperation**, collective behaviour emerges as result of local cooperation and local handling of non-cooperative situations. In the AMAS theory [7] the agents are able to recognise cooperation failure (NCS). The local handling of NCS aims at increase the adaptability. The difficulty lies in the exhaustive list of all the NCS that might result unfeasible for a real system.
- **Reinforcement**, dynamically modification of the agent behaviour by reinforcement, such as reward or punishment. This approach is focused on the adaptability skills of the single agent more than of the whole system.
- **Direct interactions**, basic principles such as broadcast and localization coupled with local interactions [40, 58]. The advantage here is the exactly known outcome of the emergent behaviour. However, it is suitable only for a limited number of tasks such as spatial pattern formation, topological agent placement where simple global equilibrium states (patterns) can be modelled in linear term.
- **Indirect interactions**, interactions only with the environment. Stigmergy is the most popular paradigm that has been successful applied to a number of complex tasks such as emerging pattern formation, coordination and management. It is robust, effective and simple to implement though, because of the nonlinearities involved, it is not possible to have direct control of the system behaviour, but only a statistical confidence on the results.

The research for self-organizing systems is shifting the focus of the classic AI. Traditionally, the focus has been on increasing the capabilities of one single agent while now the key element to have emerging properties is the communication mechanism between the agents. Considering complex tasks such as planning and requirements such as scalability, robustness and adaptability, stigmergy looks the most promising though, to be applied to the conservative world of Space Operations, it needs to demonstrate reliability and efficiency. A clear methodology for analysis and control of the emerging behaviours is therefore required.

8.4.2 Stigmergy

The term stigmergy has been introduced in the 1950s by the French biologist Grassé [28]. It comes from the Greek words "stigma" (sign) and "ergon" (work) indicating how the communication mechanism is based on traces left in the environment. This information stored in the environment forms a field that supports agent coordination stimulating their actions. Such techniques are common in biological distributed decentralized systems such as insect colonies where the information assumes usually the shape of pheromones. It is possible to identify two major type of stigmergy concerning whether the signs consist of special markers (e.g., pheromone) that agents deposit in the environment – this is called marker-based stigmergy – or whether agents base their actions on the current state of the solution – sematectonic stigmergy [47]. The most popular example is the ant foraging process, a marker-based stigmergy. The ants looking for food deposit pheromones along the path and they keep depositing pheromones on the way back, once they found food. These pheromones influence the following ants to get the same path. However, only the shortest path will end having the strongest pheromone distribution because it is the one requiring the minimum travelling time. Such a simple heuristic allows the ants to converge most of the time to the shortest path. This is an autocatalytic phenomenon based on positive feedback. It represents indeed a collective spontaneous problem solving strategy.

From the engineering point of view, stigmergy presents a number of attractive benefits:

- **Simplicity**, the system is formed by simple reactive agents with limited cognition capabilities.
- **Scalability**, it allows coordination of large numbers of simple agents without direct communication.
- **Robustness**, the system's performance is robust against the loss of a few individuals.
- **Environmental integration**, the environmental dynamics is at the same time affected by and affecting the system due to the explicit use of the environment in agent interactions.

Thanks to these advantages, stigmergy has become the core of two parallel research fields: Swarm Intelligence and System Ecosystem. Swarm Intelligence is specifically focused on decentralized self-organizing collective behaviours natural or artificial [3]. From one side, it studies the natural models such as ant foraging, wasp differentiation, termite building nest and so on. On the other side, it applies these models to a number of tasks: optimization, clustering, task allocation, network routing and so on. The result is a wide spectrum of metaheuristics that aim at reusing the key features of the natural models in the artificial systems. The term metaheuristic indicates a set of algorithmic concepts that form a general algorithm framework.

A parallel approach is to applies swarm intelligence directly to multi agent system, developing what Parunak and Brückner called Synthetic Ecosystem [7]. The aim is to provide practical engineering solutions of industrial strength. The synthetic ecosystems approach is not applying the actual social animal coordination mechanisms; rather it seeks to capture the underlying logic of the biological systems. Brückner showed how to develop a manufacturing system based on pheromone field where the industrial machines and workpieces are agents that communicate using pheromones such as ants. Inspired by digital pheromones, Tripp and Palmer developed an on board coordination system for a cluster of satellites [55]. He showed how stigmergy can efficiently reduce the computational and communication overhead. In the architecture adopted, the ground segment creates the environment broadcasting all the tasks to the spacecraft with some additional high-level aggregated information, feedback/feedforward, previously received from the spacecraft. The spacecraft communicates to the ground segment the tasks executed, feedback, and the future intentions, feedforward. Such a system using the broadcast is inherently scalable and resistant to failure, although the major disadvantage is the lack of guarantee on task completion or conflicts.

Concluding, these research fields are based both on constructive paradigms where the agents share experience by means of pheromones. The main difference is the scope: Swarm Intelligence is focused on pure optimization while Synthetic Ecosystem on coordination. In distributed systems these concepts are strictly interconnected, both their experiences therefore need to be combined.

8.4.3 Technological Challenges

The previous section introduced a new concept in Mission Planning: self-organizing multi agent systems. Though several applications have already been developed, mainly using the stigmergy approach, many issues are still open. This section is going to discuss the main technological challenges in the application of these technologies in real critical scenarios, such distributed space missions.

8.4.3.1 Methodology Principles

The traditional practices for MAS are not suitable for designing emergent systems because they do not include adaptation mechanisms. A clear design methodology is indeed required but so far the literature shows only a number of attempts. Some authors suggest the use of natural models as paradigms for the agents coordination. Those models have been successfully applied to a number of problems but in general, it is difficult to shape them to the specific problem requirements. Another direction is the extrapolation from these natural models to design patters. A design pattern is a reusable solution to a commonly occurring problem. They help to understand the basic bricks able to form self-organization but they do not explain

how to put them together. A different way is instead the evolutionary design where the main driver of the design process is the natural selection. Obviously, the evaluation of the global behaviour requires specific technique such as simulation-based design.

8.4.3.2 Methodology Techniques

In the last years, the progress in computer technologies has given the possibility to test extensively the system variables by means of simulations. Simulation-based Design is part of the wider field of Agent-based Modelling aiming at describing MAS as dynamical systems. Simulation-based Design is gaining lots of attention and the number of tools and methods for MAS simulation is rapidly increasing. Such tools offer lots of flexibility in the system design, graphical interfaces and statistical analysis support. Interesting is the combination of Simulation-based Design with Evolutionary Computation. The last is a technique that has been successfully demonstrated in a wide number of applications including design. In the case of MASs, each individual of the population represents a specific MAS. The difficulty here lies in assessing the fitness of each individual. Simulating the specific MAS against a set of problems can give a score for selecting the best individuals. A parallel field facing similar issues is Evolutionary Swarm Robotics. Here the robot controllers, representing the robot behaviours, evolve to solve a cooperative task. Interesting works see evolving artificial neural networks robot controller for competitive team-game evolving behaviours such as synchronization, coordination and obstacle avoidance. This field put in evidence the importance of the evolutionary pressure represented by the fitness function that needs to drive the evolution towards better solutions. In case of complex problems, defining the evolutionary pressure is extremely challenging. It seems indeed premature to talk about emergence design using exclusively evolutionary computation. The last is surely adequate for micro-evolution but not yet for macro-evolution that are the more remarkable in this contest. Concluding, the design process of self-organizing systems still lacks of clear guidelines, leaving the designers with a wide number of decisions to take. The simulation-based design and its evolutionary extension are powerful tools but are still trial and error processes, requiring a high amount of time and computational resources. Even on the ground segment, the computational resources are not so prominent. Interesting results might come from the description of these systems as dynamical systems. This type of description might be able to model the nonlinear dynamics that characterize the system emergence and to identify the relative equilibrium points. However, they are still premature studies.

8.4.3.3 Controlled Emergence

The goal of self-organizing systems is not only to organize themselves but to yield structures that solve engineering problems. The idea of controlling something such

as emergence seems nonsense and in fact, it is still in its infancies. However, for engineering applications this is a priority and for space Operations a critical issue. Parunak was the first rising the problem and suggesting two possible methods: specialised agents or shaping the system envelope [51]. Concerning the first, a number of projects have showed how to increase the reliability of these systems. The Autonomic Computing field for example has been promoting the Manager architecture where a specific component is in charge to monitor the global dynamic and to intervene by global parameters. Analogous is the approach of Organic Computer, the Observer/Controller Architecture, where this special role is played by a component called Observer. Both these two approaches have been criticized as they reintroduce the usual centralized control loop in an architecture that is decentralized by definition. Other researches have shown how it is possible to model the global dynamics using a number of agents with different behaviours, called antagonist agents, that using negative feedback are able to inhibit the action of the canonical agents.

The method of shaping the system envelope recalls the approach of modelling the system as a nonlinear dynamical system and describing its behaviours using chaotic dynamics. In this way, it would be possible to directly control and shape the equilibrium points. This approach could be very powerful but might result difficult to apply to real system. So far, the only works published in this direction regards natural systems such as swarms, ant colonies that can be described as chaotic systems.

8.4.3.4 Hybrid Architecture

Another promising technology is the combination of deliberative and reactive agents in the same population. Several hybrid approaches have been presented in literature but this field is premature. The challenge here is to exploit properly the advantages that the two paradigms can offer: the deliberative agent is appropriate for exploiting a well-defined problem in efficient way while the reactive agent because of his fast reasoning and planning is more suitable for exploration tasks in dynamic environments. These two capabilities if well combined could offer incredible flexible and powerful solutions. Such a system could be used for monitoring transient unpredictable events (e.g. fire outbursts detection) where a great reactivity is required either at local level (spacecraft) either at global level (constellation). This duality is reflected in the wider trade-off optimization versus adaptation. The adaptation properties required in distributed space applications are not only confined to re-planning but they regard a more general ability to change the global behaviour according to the environment. The issue is that normally a software more flexible and adaptable is inevitably less efficient and optimized. Independently from the specific technology chosen (multi agent, etc.), it is necessary to evaluate this trade-off: high-level of adaptation whilst ensuring that the schedule is always valid and optimized, even in dynamic and uncertain environment.

8.5 Conclusions

The distributed mission scenario presents new challenges for the autonomy solutions. They need to be highly responsive, adaptable to face dynamic environments and scalable to a various number of spacecraft. Today a number of advanced technologies are available for meeting these requirements but they need further development to achieve a suitable reliability. The progress in computer technologies is opening the possibilities of these advanced technologies even to multiple platforms formed by nanosatellite.

The solution identified is going to be the sharing of mission planning activities between the ground segment and the satellites. While each spacecraft is in charge of planning its own image acquisition schedule, the ground is in charge of coordinating the workload of the whole constellation [30]. Ground segment autonomy is necessary to avoid conflicts at global level, to monitor and support the on board autonomous process while on board autonomy is necessary to increase the reactivity of the spacecraft toward transient events sensed by the spacecraft itself. Finding the right level of sharing of responsibilities between ground and space segment is one of the main challenges for future missions.

Despite the technological challenges, the biggest challenges come from the human side. Autonomy and specifically goal-oriented operations reduces the low-level control of the human operators giving a higher view of the spacecraft status. One challenge is making this transition acceptable. This can be achieved offering a number of tools able to display goals, plans and plan execution status. They must indicate deviations from the plan and any corrective actions the system takes. They have to cluster the low-level commands and summarize them as a single event, allowing data mining on the same telemetry. The human operator needs to become a supervisor and needs powerful tools for this task.

A second big challenge is the mindset shit. The space sector and in particular the Operations is quite resilient to new approaches because is strongly adverse to the risk of failure. In this case, the issue is to shift from a deterministic to a nondeterministic engineer design, but if distributed missions is the future it must be accepted and embraced. Of course, it may take some time for the collective mindset to shift from the current rigid and inflexible guaranteed approach towards a more reactive and probabilistic system and one of the challenges is indeed the creation of a roadmap to make this process feasible.

References

1. Amigoni F, Brambilla A, Lavagna M, Blake R, le Duc I, Page J, Page O, de la Rosa Steinz S, Steel R, Wijnands Q (2010) Agent technologies for space applications: the DAFA experience. In: International conference on Web intelligence and intelligent agent technology (WI-IAT), 2010 IEEE/WIC/ACM, Toronto, Canada, vol 2. IEEE Computer Society Washington, DC, pp 483–489

2. Barreiro J, Jones G, Schaffer S (2009) Peer-to-peer planning for space mission control. In: In Aerospace conference, 2009 IEEE, Big Sky, Montana, USA, pp 1–9
3. Blum C (2008) Swarm intelligence: introduction and applications. Springer, Berlin
4. Bonnet G, Tessier C (2007) Collaboration among a satellite swarm. In: Proceedings of the 6th international joint conference on autonomous agents and multiagent systems, Honolulu, HI, USA, p 54
5. Bornstein B, Thompson D, Chien S, Bue B, Tran D, Castano R (2011) Efficient spectral endmember detection and scene summarization onboard the EO-1 spacecraft. In: AI in Space workshop, interantional joint conference in artificial intelligence, JICAI-11, Barcelona
6. Bresina JL, Morris PH (2007) Mixed-Initiative planning in space mission operations. AI Mag 28(2):23
7. Brueckner S (2000) Return from the ant. synthetic ecosystems for manufacturing control. PhD thesis, Humboldt-Universität zu Berlin
8. Campbell M, Schetter T (2002) Comparison of multiple agent-based organizations for satellite constellations (TechSat21). J Spacecr Rockets 39:274–283
9. Capera D, George J, Gleizes MP, Glize P (2003) The AMAS theory for complex problem solving based on self-organizing cooperative agents. In: Proceedings of the twelfth IEEE international workshops on enabling technologies: infrastructure for collaborative enterprises, 2003. WET ICE 2003, Linz, Austria, p 383
10. Carlsson A, Noteborn R, Larsson R, Bodin P (2010) PRISMA operational concept: servicing a variety of experimental teams for teh flight demonstration of formation flying technologies. In: 61st international astronautical congress, small satellite missions symposium, Prague, Czech Republic
11. Carrel AR, Palmer PL (2005) An evolutionary algorithm for near-optimal autonomous resource management. In: 'i-SAIRAS 2005'-the 8th international symposium on artificial intelligence, robotics and automation in space, München, Germany, vol 603, p 25
12. Castano R, Estlin T, Gaines D, Castano A, Chouinard C, Bornstein B, Anderson RC, Chien S, Fukunaga A, Judd M (2006) Opportunistic rover science: finding and reacting to rocks, clouds and dust devils. In: Proceedings of 2006 IEEE aerospace conference, Big Sky, Montana, USA
13. Chien S, Rabideau G, Knight R, Sherwood R, Engelhardt B, Mutz D, Estlin T, Smith B, Fisher F, Barrett T, Stebbins G, Tran D (2000) ASPEN – automated planning and scheduling for space mission operations. In: IN SpaceOPS, Toulouse, France
14. Chien S, Engelhardt B, Knight R, Rabideau G, Sherwood R, Tran D, Hansen E, Ortiviz A, Wilklow C, Wichman S (2002) Onboard autonomy software on the three corner sat mission. In: Proceedings of the SpaceOps 2002 conference, Houston, Texas, USA
15. Chien S, Sherwood R, Tran D, Cichy B, Rabideau G, Castano R, Davis A, Boyer D (2005) Using autonomy flight software to improve science return on earth observing one. J Areospace Comput Inf Commun 2:196–216
16. Clement B, Barrett A (2004) Coordination challenges for autonomous spacecraft. Appl Sci Multiagent Syst 10:7–26
17. Clement BJ, Barrett AC, Schaffer SR (2004) Argumentation for coordinating shared activities. In: 4th international workshop on planning and scheduling for space, Darmstadt, Germany
18. Curtis SA, Rilee ML, Clark PE, Marr GC (2003) Use of swarm intelligence in spacecraft constellations for the resource exploration of the asteroid belt. In: Proceedings of the 3rd international workshop on satellite constellations and formation flying, Pisa, Italy
19. Damiani S, Verfaillie G, Charmeau MC (2005) An earth watching satellite constellation: how to manage a team of watching agents with limited communications. In: Proceedings of the fourth international joint conference on autonomous agents and multiagent systems, Utrecht, The Netherlands, pp 455–462
20. De Florio S (2006) Performances optimization of remote sensing satellite constellations: a heuristic method. In: International Workshop on Planning and Scheduling for Space, Baltimore, MD, USA
21. De Wolf T, Holvoet T (2005) Emergence versus self-organisation: different concepts but promising when combined. Eng Selforgan Syst 3464:1–15

22. Dvorak D, Ingham M, Morris JR, Gersh J (2007) Goal-based operations: an overview. In: AIAA infotech, Rohnert Park, CA, USA
23. Estlin T, Chien S, Castano R, Doubleday J, Gaines D, Anderson RC, de Granville C, Knight R, Rabideau G, Tang B (2010) Coordinating multiple spacecraft in joint science campaigns. In: International symposium on space artificial intelligence, robotics, and automation for space (i-SAIRAS 2010), Sapporo, Japan
24. Ferber J (1998) Multi-agent systems: an introduction to distributed artificial intelligence. Addison-Wesley, Harlow
25. Funase R, Nakasuka S (2005) Cooperation between onboard executive and planner on ground and its demonstration on remote-sensing nano-satellite PRISM. In: 8th international symposium on artificial intelligence, robotics and automation in space, Munich, Germany
26. Gabbai J, Yin H, Wright W, Allinson N (2005) Self-organization, emergence and multi-agent systems. In: International Conference on neural networks and brain, 2005. ICNN&B '05, Beijing, China, vol 3, pp 24–1863
27. Gantois K, Santandrea S, Teston F, Strauch K, Zender J, Tilmans E, Gerrits D (2010) ESA's second in-orbit technology demonstration mission: Proba-2. Esa Bull Eur Space Agency 144:22–33
28. Giuliano ME, Hawkins R, Rager R (2011) A status report on the development of the JWST long range planning system. In: IWPSS – international workshop on planning and scheduling for space, ESOC, Darmstadt, Germany
29. Glansdorff P, Prigogine I (1971) Thermodynamic theory of structure, stability and fluctuations. Wiley-Blackwell, New York
30. Grandjean P, Pesquet T, Muxi A, Charmeau M (2004) What on-board autonomy means for ground operations: an autonomy demonstrator conceptual design. In: SpaceOps, Montreal, Canada
31. Grasse P (1960) The automatic regulations of collective behavior of social insect and "stigmergy". J Psychol Norm Pathol 57:1–10
32. Grasset-Bourdel R, Verfaillie G, Flipo A (2011) Building a really executable plan for a constellation of agile earth observation satellites. In: IWPSS – international workshop on planning and scheduling for space, ESOC, Darmstadt, Germany
33. Grey S, Radice G (2010) Design and testing of an autonomous multi-agent based spacecraft controller. In: 61st international astronautical congress (IAC 2010), Prague, Czech Republic
34. Hillis WD (1988) Intelligence as an emergent behavior; or, the songs of eden. Daedalus 117 (1):175–189
35. Hinchey MG, Rash JL, Truszkowski WF, Rouff CA, Sterritt R (2005) Challenges of developing new classes of NASA self-managing missions. In: Proceedings of the 11th international conference on parallel and distributed systems, 2005, Phoenix, AZ, USA, vol 2
36. Holland JH (1999) Emergence: from chaos to order. Basic Books, New York
37. Hollister GS (2011) Military space requirements markets and technologies forecast - 2012-2017. Tech. rep., Market Info Group LLC
38. Van der Horst J, Noble J (2010) Distributed and centralized task allocation: when and where to use them. In: Self-adaptive networks (SAN) workshop, IEEE international conference on self-adaptive and self-organising systems (SASO2010), Budapest, Hungary
39. Johnston MD, Miller GE (1994) Spike: intelligent scheduling of hubble space telescope observations. Intell Sched pp 391–422
40. Jones M (2009) Artificial intelligence : a systems approach. Jones and Bartlett, Sudbury
41. Jonsson AK, Morris PH, Muscettola N, Rajan K, Smith B (2000) Planning in interplanetary space: theory and practice. Artif Intell Plan Syst pp 177–186
42. Knight S, Rabideau G, Chien S, Engelhardt B, Sherwood R (2001) Casper: space exploration through continuous planning. Intell Syst IEEE 16(5):70–75
43. Lenzen C, Worle M, Mrowka F, Geyer M (2011) Automated scheduling for TerraSAR-X/TanDEM-X. In: IWPSS – international workshop on planning and scheduling for space, ESOC, Darmstadt, Germany

44. Mamei M, Vasirani M, Zambonelli F (2005) Self-organizing spatial shapes in mobile particles: the TOTA approach. Engineering Self-organising systems. Springer, Berlin/Heidelberg, pp 138–153
45. Martinez-Heras JA, Donati A (2011) Making sense of housekeeping telemetry time series. In: AI in space: intelligence beyond planet earth – interantional joint conference in artificial intelligence, IJCAI-11. Barcelona, Spain
46. Montenbruck O, Markgraf M, Naudet J, Santandrea S, Gantois K, Vuilleumeier P (2008) Autonomous and precise navigation of the PROBA-2 spacecraft. In: AIAA/AAS astrodynamics specialist conference and exhibit, Honolulu, HI, USA
47. Moylan G, Atkins E (2006) Research trends in autonomous Space-Based planning and scheduling. In: International workshop on planning and Scheduling for space – IWPSS, Baltimore, MD, USA
48. Muscettola N, Nayak PP, Pell B, Williams BC (1998) Remote agent: to boldly go where no AI system has gone before. Artif Intell 103(1–2):5–47
49. Nesnas I (2007) Claraty: A collaborative software for advancing robotic technologies. In: Proceedings of NASA science and technology conference, Adelphi, MD, USA
50. Ocon J, Rivero E, Sanchez Montero A, Cesta A, Rasconi R (2008) Multi-agent frameworks for space applications. In: SpaceOps 2010, Huntsville, AL, USA
51. Parunak HV (2003) Making swarming happen. In: Proceedings of swarming and network-enabled C4ISR, Tysons Corner, VA, USA
52. Parunak HV, Brueckner SA (2004) Engineering swarming systems. Methodol Softw Eng Agent Syst 11:341–376
53. Pralet C, Verfaillie G, Olive X (2011) Planning for an ocean global surveillance mission. In: International workshop on planning and scheduling for space (IWPSS) 2011, ESOC, Darmstadt, Germany
54. Raghava Murthy DA, Kesava Raju V, Srikanth M, Ramanujappa T (2010) Small satellite constellation planning for disaster management. In: International astronautical federation, Prague, Czech Republic
55. Richards R, Houlette R, Mohammed J (2001) Distributed satellite constellation planning and scheduling. In: Proceedings of the fourteenth international florida artificial intelligence research society conference, Key West, FL, pp 68–72
56. Schoeberl M, Bristow J, Raymond C (2001) Intelligent distributed spacecraft infrastructure. In: Earth science enterprise technology planning workshop, Pasadena, CA, USA, pp 23–24
57. Serugendo GD, Gleizes MP, Karageorgos A (2006) Self-organisation and emergence in MAS: an overview. Informatica 30(1):45–54
58. Steel R, Niezette M, Cesta A, Fratini S, Oddi A, Cortellessa G, Rasconi R, Verfaillie G, Pralet C, Lavagna M, et al (2009) Advanced planning and scheduling initiative: MrSPOCK AIMS for XMAS in the space domain. In: The 6th international workshop on planning and scheduling for space, IWPSS-09, Pasadena, CA, USA
59. Sycara K (1998) Multiagent systems. AI Mag 19(2):79
60. Teston F, Barnsley M, Settle J, Vuilleumier P, Santandrea S (2003) PROBA: an ESA technology demonstration mission with earth imaging payload. first year of in orbit results. In: 4th IAA symposium on small satellites for earth observation, Berlin, Germany, vol 11
61. Tripp H, Palmer P (2010) Stigmergy based behavioural coordination for satellite clusters. Acta Astronaut 66(7–8):1052–1071
62. Underwood CI, Crawford M, Ward JW (2001) A low-cost modular nanosatellite based on commercial technology. In: Proceedings of the 12th annual AIAA/USU conference on small satellites. SSC98, Logan, UT, USA, vol 4
63. Viroli M, Casadei M, Omicini A (2009) A framework for modelling and implementing self-organising coordination. In: Proceedings of the 2009 ACM symposium on applied computing, Honolulu, HI, USA, pp 1353–1360
64. Wang P, Tan Y (2008) Joint scheduling of heterogeneous earth observing satellites for different stakeholders. In: SpaceOps 2008, Heidelberg, Germany

Chapter 9
Relative Navigation

Nadjim Horri and Phil Palmer

Abstract Satellite formations are being considered for a large variety of current and future space missions including in-orbit inspection, SAR interferometry, magnetospheric observation and gravimetry. In the case of cooperative satellite formations, differential GPS, radiofrequency and optical navigation techniques have been demonstrated as viable approaches for relative navigation on a number of recent space missions. Future challenges include accurate relative navigation and positioning in six degrees of freedom, with the limited power and computational resources of small satellites. This article explains the relative navigation requirements and their dependency on the space applications. The software and hardware challenges on relative navigation for future satellite formations are also described.

9.1 Introduction

There is an increased interest in the space community for space missions requiring relative navigation for various applications such as docking, in-orbit servicing, the inspection of non cooperative satellites and more generally formation flying. Satellite formations are being proposed for an increasingly wide range of space missions including in-orbit inspection, SAR interferometry, gravimetry, and various science applications. Small satellites are being increasingly proposed to carry out such formation flying missions, because of the increase in their capabilities in recent years as well as their lower cost of development and launch.

However, the requirements on these satellite formations are rapidly evolving and increasingly challenging. The relative navigation challenges vary with the type of mission.

N. Horri, Ph.D. (✉) • P. Palmer, Ph.D.
Surrey Space Centre, University of Surrey, GU2 7XH Surrey, UK
e-mail: n.horri@surrey.ac.uk; p.palmer@surrey.ac.uk

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_9,

In the case of SAR interferometry, the main challenges are the achievement of very accurate positioning and pointing with conventional Attitude and Orbit Control System (AOCS) hardware and dual frequency Global Positioning System (GPS) receivers [1]. In the case of inspection missions, the accuracy requirement is still significant but less stringent. The main challenges for current inspection missions [2–5] are the provision of continuous and reliable relative navigation information with the limited power and computational resources onboard small satellites. In close satellite formations, the emphasis is generally on relative position knowledge rather than absolute position knowledge, which sometimes implies the need for specific relative navigation software. Relative position determination for such missions requires specific relative navigation sensors [3].

In future satellite formations, relative attitude control will also be needed to meet system level requirements. More generally, satellite formations will require autonomous guidance, navigation and control [5–7]. This implies less reliance on communications with the ground and more emphasis on inter-satellite links. In particular, satellites will be required to autonomously avoid collisions, which are likely in the case of close formations.

The main technologies of relative navigation are GPS/GNSS based receivers, radiofrequency (RF) transceivers and optical navigation sensors. The PRISMA mission developed by the Swedish Space Corporation is of particular interest because it flew all three types of sensors [2, 3, 8]. A technology based on the use of pulsars has also been recently proposed but has not yet been flight proven [9]. Other types of sensors, which have been used for interplanetary landers, such as lidar, can also be envisioned for certain relative navigation phases, from medium to large separation distances. Future formation flying missions may require the combined use of several types of relative navigation sensors. There is therefore a need for algorithms allowing for the fusion of data from different types of sensors.

In the case of formations of multiple satellites, there is also an increasing requirement for strategies of fault detection, isolation and recovery under various scenarios such as gradual degradation of performance or even the failure of one of multiple satellites in a formation, which may have consequences when such satellites are automatically tracked and can eventually lead to unforeseen collisions. Collision avoidance techniques are part of an autonomous orbit control software, which is discussed in Sect. 9.5.3.

This article is organised as follows: Sect. 9.2 introduces the relative navigation technologies, the relative navigation accuracy requirements are then discussed in Sect. 9.3, the importance of attitude control is discussed in Sect. 9.4, while Sects. 9.5 and 9.6 respectively introduce the software challenges and hardware limitations of relative navigation.

Table 9.1 Relative navigation technologies for various GNC modes of a satellite formation

Relative GNC mode	Description	Separation	Key sensors
Autonomous formation flying	Satellites out of sight, approaching target orbit	Tens of meters to few kilometres and beyond	GPS
Homing	Satellites approaching and in sight	Few kilometres to few meters	ONS/GPS
Precision 3D operations	Relative orbit and/or attitude manoeuvres	100 m to few meters	GPS/ONS
Collision avoidance/ transient stages	Autonomous collision avoidance, safe mode	Few kilometres to 100 meters	RF/GPS
Final approach	Rendezvous and docking case	From zero to few meters	ONS

9.2 Technologies of Relative Navigation Sensors

Various types of absolute and relative navigation sensors can be required depending on the main objectives of the mission, but also depending on the mission phases.

While several tests or relative navigation technologies were conducted on satellites that started on the same orbit before separation, future missions may involve satellites separately injected on different orbits. In this scenario, the first orbital manoeuvre consists of assigning target orbits to the satellites, which can be achieved by applying velocity increments to move satellites on a drift orbits [10]. GPS sensors are the obvious solution in earth orbit. This phase does not rely on relative navigation sensors. Relative navigation sensors are activated when the satellites are in sight and/or in communication range. From that moment, various relative Guidance Navigation and Control (GNC) modes can be entered depending on the mission phases [3, 10].

For autonomous formation flying at distances ranging from tens of meters to a few kilometres, GPS is the primary source of relative orbit measurements. Optical navigation with long range cameras can also complement GPS receivers in this range for objectives such as autonomous approaches.

In the case of precision proximity operations at a close range, starting from 100 m, GPS is generally complemented by short range optical navigation sensors, when the primary mission objective is in-orbit inspection or docking. At very close range (meters), for rendezvous operations, short range optical navigation sensors (ONS) are generally the only source of measurements [4]. RF transceivers are of particular interest for collision avoidance and other orbital manoeuvres [3]. These various scenarios are summarised in Table 9.1.

9.2.1 GNSS Technology

GNSS navigation technology is probably the most commonly used for both absolute and relative navigation purposes. The first missions to use GNSS for relative navigation were the Japanese ETS-7 formation of two satellites, followed by the

formation keeping experiment between Snap-1 and Tsinghua-1, the DART mission, EO-1. More recently, the Orbital Express formation of two satellites [11] and the ATV ESA mission have demonstrated in-orbit docking. The Japanese satellite Ikaros conducted a time limited in-orbit inspection of a solar sail deployment and PRISMA has demonstrated various relative orbit control scenarios, with autonomous collision avoidance [3].

While dual frequency receivers allow for the highest accuracy, the use of single frequency GPS receivers on each satellite is sufficient for most relative navigation applications, apart from SAR interferometry. The positions of satellites can be communicated between them. In the case of a two satellite formation, the target typically communicates its GPS measurement to the chaser satellite, which accordingly corrects its relative motion. The use of Differential GPS, a technology initially developed for ground receivers, has recently been proven to be a viable approach, especially in terms of relative positioning accuracy. GNSS technology is however obviously only available for earth orbiting spacecraft.

9.2.2 RF Technology

RF technology presents a relative navigation solution that can be of interest to both earth orbiting satellites and interplanetary spacecrafts, although coarse GPS measurements are typically used by the RF transceiver in earth orbits. The FFRF sensor on the PRISMA mission has been a successful test of the technology [4]. This transceiver uses GPS like signals at two frequencies in the S frequency band. It is deemed particularly appropriate for intermediate phases such as deployment, orbital manoeuvres and especially for collision avoidance. It is also used for the transition to optical metrology. It delivers a coarse but unambiguous pseudo-range (ten times less accurate than GPS in normal operation mode) and an accurate but ambiguous carrier phase. It presents limitations in terms of cross track accuracy. Another issue is the need for calibration to correct for biased measurements. On PRISMA, these biases were corrected by using the GPS receiver. More generally, the RF transceiver should be used in combination with other sensors.

9.2.3 Optical Navigation Sensors

Optical navigation sensors are primarily used for in-orbit inspection missions and for docking. The long range (more than 10 m) and short range imaging (below 10 m) are generally performed by different dedicated cameras. The long range cameras are generally based on star sensor technology. In this case, there is a need for use of optical navigation software, which is discussed in Sect. 9.5.1. There are a

number of practical considerations to account for in the case of proximity operations or of a final approach, including plume impingement following firings from thrusters, which may considerably affect the accuracy of pose estimation algorithms.

9.3 Relative Navigation Accuracy Requirements

The technical specifications on relative navigation accuracy are particularly important to the overall mission design of a satellite formation, with direct implications on the satellite design at system and subsystem level. These requirements depend primarily on the application.

9.3.1 Missions Requiring Very Accurate Relative Positioning

The TanDEM-X bi-static SAR interferometry mission, comprising of two satellites launched in June 2007 and June 2010 is due to produce SAR digital elevation maps with an unprecedented vertical accuracy of 2 m. This corresponds to a 1 mm accuracy requirement on the baseline between the two satellites, which is also unprecedented. This mission has so far passed a 6 month test phase according to plan and has started delivering SAR images from both satellites, with a 20 km separation. The next milestone, which is particularly interesting from a relative navigation perspective, is generation of the digital elevation maps by keeping the satellites in close proximity, typically a few hundred metres. This phase is due to start later this year.

To meet this very high accuracy requirement, precise baseline determination software was developed for TanDEM-X, which is based on a forward-backward filter smoother. This filter handles both ionosphere free pseudo-range and carrier phase measurements with 1 m accuracy, requiring a sufficient number of visible GPS satellites and single difference carrier phase measurements, requiring commonly observed GPS satellites allowing for a much higher accuracy. Relative attitude and orbit design will therefore be of paramount importance to increase the availability of commonly observed GPS satellites when these become available.

Certain science missions also require high relative precision the maintenance of very large baselines. One example of such a mission is the Laser Interferometer Space antenna (LISA) satellite formation, composed of three satellites on an equilateral triangle in the same orbit as the earth to detect ripples in space-time. This will require the use of precise relative navigation to maintain a small offset angle between the laser telescopes of all three satellites.

9.3.2 Missions Requiring Moderate Position Accuracy

Accuracy requirements can be moderate for certain applications but their maintenance remains a challenge under specific technological constraints.

In the case of inspection missions, the required position knowledge is generally moderate but careful planning of relative orbits is required to make sure that the satellites maintains positioning requirements during inspection or relative manoeuvring. In the case of the PRISMA mission, absolute and relative position accuracies were respectively of 2 and 0.5 m. Maintaining these accuracies in real time is however challenging for a microsatellite with limited onboard computational resources. The onboard computer on PRISMA is radiation tolerant but has limited processing power. Orbit determination data for example can only be updated every 30 s on that satellite and sub-metric real time relative positioning accuracies were made possible thanks to a precise orbit model.

Gravimetry missions are another example of satellite formations that do not require highly accurate relative position knowledge. There attitude control requirement is of the same order as that of medium resolution earth observation satellites. The attitude profiles are however required to be smooth and stable. The challenge in this case is to keep the satellites in a typical loose formation with a separation of 20 km $\pm$ 500 m, while keeping the satellite under the sole effects of gravity and disturbance torques for thousands of seconds [12].

9.4 Relative Attitude Control Requirements

While the relative orbit control for such systems has been widely investigated, the relative attitude control of multiple satellites in a formation has comparatively received less attention. In the following, we consider two examples of mission scenarios that may require special attention in terms of relative attitude control.

9.4.1 Relative Attitude Control of Multiple Satellites

A number of future applications will require formations of multiple satellites. With the increase in the number of satellites in a formation, more frequent attitude and orbital manoeuvres will be required, including relative orbit reconfiguration manoeuvres but also coordinated attitude control of multiple satellites that will have to synchronise their attitudes to ensure formation keeping by maintaining themselves within their respective fields of view. This will be necessary to carry out inter-satellite communications. To respond to this requirement, satellites will have to be more responsive and more power efficient at the same time.

The attitude control requirements will therefore have to increase in terms of agility, pointing accuracy without increasing the Attitude Determination and Control Subsystem (ADCS) share of the power budget. One way of meeting this challenge is the development of more efficient small actuators in terms of output torque, given a level of the energy consumption. In this context, control moment gyros appear to be worthy of being part of future technology demonstration missions on small satellites. Another approach will be the development of advanced control methods, such as optimal control, within the available computational resources.

The approaches to coordinated attitude control can either be centralised, based on nearest neighbour tracking and following a centralised hierarchy, or decentralised. The current trend is towards a decentralised control approach [13], whereby each satellite in the formation typically interacts with neighbouring satellites to adjust its relative pointing, without hierarchy or with multiple team hierarchy and without the need for global information about the whole formation. One of the future challenges of decentralised control techniques for coordinated attitude control will be the maintenance of simultaneous station keeping and formation keeping objectives, within energy limitations. This suggests the need for optimal control techniques of sufficiently low implementation complexity.

9.4.2 Relative Attitude Control for In-Orbit Inspection

The attitude control software will be particularly important to future inspection missions. In particular, to enhance the quality of relative optical navigation, relative navigation modes such as those developed for the PRISMA mission should be complemented by relative attitude control modes.

In the case of an inspection mission, relative attitude control modes may in the simplest cases consist of assigning inertial or nadir pointing objectives to maintain the target within the field of view of the chaser, assuming the maintenance of a conventional 2:1 elliptic relative orbit. In the nadir pointing case, the chaser may for example be equipped with a camera pointing in the inverse tangential direction, with 90° off-pointing with respect to the Local Vertical Local Horizontal frame (LVLH) frame [5].

However, if certain mission phases require the chaser to point precisely towards the target, then a relative attitude control mode would have to be included [5] to ensure such relative pointing objectives. This capability was not available on the PRISMA mission, which relied on careful selection of relative orbits, while maintaining conventional attitude control modes such as sun and zenith pointing.

While formations of multiple satellites should only be requested to maintain a minimum field of view between satellites, the lower level of complexity of an inspection mission with two satellites should allow for the objective of pointing precisely in the target's direction.

9.4.3 *Effect of Attitude Control on Relative Orbit Determination*

The last two subsections addressed the importance of relative pointing to optical navigation or to maintain inter-satellite links for formation keeping. It is sometimes overlooked that attitude control also has a direct impact on orbit determination accuracy, particularly in the case of GNSS based navigation. Indeed, the number of visible GPS satellites, which directly affects GPS positioning accuracy, is dependent on the satellite's attitude. The best visibility on orbit is generally achieved with an antenna pointing to the local Zenith direction, which is generally achievable in nadir pointing mode. This requirement is however in contradiction with other objectives such as sun tracking, target tracking and other science modes.

When a GPS antenna deviates from Zenith, the receiver's automatic channel allocation algorithms still assume a zenith orientation for the computation of visible satellites. The attitude deviation from Zenith will therefore cause the receiver to neglect visible satellites and occupy free channels in an attempt to acquire out of view satellites. This problem constitutes a challenge when high position accuracy has to be maintained for large durations. This challenge can be met by careful mission design. A channel allocation scheme can also be useful to account better for the actual antenna bore-sight direction [14]. A technique was also proposed for the Can-X nanosatellite mission, which consisted of saturating the roll spin rate of the chaser, as it was found that the number of visible GPS satellites dropped most significantly when a maximum spin rate was reached [15].

9.5 Software Challenges

9.5.1 *Relative Pose Estimation from Optical Imagery*

The optical navigation sensors require dedicated software for the estimation or relative positions (pose) from the available images of the target. The current approach consists of storing features of the target satellite and comparing them with the features extracted from images to determine the pose estimate. This approach requires prior knowledge of the spacecraft to be inspected and presents challenges in terms of accuracy for a given computational demand [4].

One alternative has been considered at the Surrey Space Centre, which consists of modelling the target as an ellipsoid or spheroid, rather than storing any features of it. On the image plane, a minimum bounding ellipse, which is the projection of the ellipsoid on the 2D image plane, is drawn around the object. The parameters of the ellipse can then be used to estimate the pose and even the relative orientation [16].

9.5.2 Relative Orbit Determination Algorithms from GPS or RF Sensors

The relative positioning using differential GPS is generally based on the so called Lease Squares Ambiguity Decorrelation (Lambda) method, which had proven effective for most mission scenarios. However, the PRISMA mission tested a number of challenging scenarios, including large attitude manoeuvres causing rapid decrease of the numbers of commonly tracked GPS satellites as well as frequent relative orbit manoeuvres. The Lambda orbit determination method (with ambiguity resolution) failed for these scenarios. As a remedy, a forward-backward Extended Kalman filter was used in the absence of orbit manoeuvres, while ambiguities were incorporated into the state vector of the Kalman filter during orbital manoeuvres. This method was robust but came at the cost of degraded accuracy and is only used when the ambiguity resolution does not work. This illustrates that more advanced orbit determination algorithms are required for missions requiring frequent orbit and attitude manoeuvring. The RF technique is similar to the GPS orbit determination, except that each RF unit can also act as a transmitter. In this case, a coarse pseudo-range measurement is obtained from the GPS coarse acquisition code, modulated in the transmitter's frequencies, generally in S-band, before finer carrier phase measurements can be made. The use of a dual frequency RF transceiver makes ambiguity resolution possible.

One area for research and development in the context of relative orbit determination is sensor fusion. Future missions requiring relative navigation are very likely to rely on more than just one source of relative position measurement. Sensor fusion potentially allows for more accurate relative motion estimation, especially when the information from different types of sensors is complementary. The estimation algorithms based on Extended Kalman Filtering (EKF) can however face challenges in the case of multi-rate measurements, especially when the most accurate sensors have significant second order terms. This might lead to misleading estimations and even divergence, making common alternatives to the EKF such as the unscented KF worth considering [17]. Note that sensor fusion can also consist of combining relative motion measurements with absolute inertial measurements.

9.5.3 Autonomous Relative Navigation

There classical approach to satellite navigation and orbit control has traditionally been centralised at the ground segments, from which commands were sent by operators to initiate orbital manoeuvres and other mission critical tasks. In the case of future satellite formations, the tasks of coordinated relative navigation can rapidly become complex and risky to fully centralise at a ground station. As an

alternative, satellites will be required to be more autonomous and correct their positions automatically using their onboard relative navigation software.

9.5.3.1 Enabling Technologies for Autonomous Orbit Control

The solutions to onboard autonomy can either rely on inter-satellite communications or on the tracking of predefined trajectories. A number of satellite missions have demonstrated various enabling technologies for onboard autonomy. For example, Uosat-12 from the Surrey Space Centre allowed, in collaboration with Microcosm, for the autonomous tracking of predefined trajectories. The Proba-2 satellite and the BIRD microsatellite from the German space Agency (DLR) have demonstrated enabling technologies for the onboard management of computer system failures and of attitude anomalies and other limit exceeding parameters [18]. The DEMETER microsatellite from the French Space Agency (CNES) has demonstrated a number of autonomous orbital manoeuvring scenarios, while PRISMA also demonstrated autonomous close proximity operations between the Mango and Tango satellites. These technologies will have to be tested on formations of multiple satellites, by taking into consideration the mission level requirements of satellite formations.

9.5.3.2 Autonomous FDIR and Collision Avoidance

While Failure Detection, Isolation and Recovery (FDIR) software was generally dedicated to the management of subsystem failures, collision avoidance should also be part of any future FDIR system in satellite formations. Autonomous collision avoidance was also demonstrated onboard PRISMA using the RF relative navigation sensor.

Collision avoidance is of particular importance to any mission involving proximity operations or close formations. Successful autonomous docking was demonstrated on Orbital express [11] and on the ATV mission [19]. However, a number of anomalies occurred on other missions demonstrating rendezvous capabilities, including the DART mission and the ETS-7 mission. In the case of the ETS-7 mission, such anomalies caused a number of entries into a safe mode including a pre-programmed manoeuvre to move the chaser 2.5 km away from the target. In the case of the DART mission, an anomaly caused excess fuel expenditure and was suspected to have caused a collision. These examples suggest that the inclusion of additional safeguards for collision avoidance and more generally for FDIR software can be of great benefit to future missions involving proximity operations or in-orbit inspection [20]. Agent based techniques have been proposed to autonomously perform FDIR [6]. A number of approaches have been recently developed for collision avoidance based on passive safe trajectories (without thrusting) and active safe trajectories, which is less restrictive but requires failure detection in real time. Current research in this

field aims to the development of fuel optimised safe trajectories [7, 21]. There are however also computational constraints to account for. The approach proposed in reference [21] for a multiple satellite formation is sequential and based on the computation of piecewise optimal solutions, using the fact that reconfiguration manoeuvres often do not affect other satellites in the formation. The computation time using this approach was shown to be considerably reduced compared to a conventional finite element method that solves the global optimal solution.

9.6 Hardware Limitations

The decision of choosing between Commercial Off The Shelf (COTS) components or radiation tolerant space qualified components is one of the choices to be made at the system design level. There is generally a compromise to be found between the need for high performance at system and subsystem levels and the reliability of the satellite. The enabling technologies generally have to be reliable and power efficient. In the following, we consider some key equipment which affects the reliability and performance of relative navigation onboard small satellites.

9.6.1 GPS Receiver Selection

GPS receivers are the most commonly used sensors for orbit determination in earth orbit. Other sensors such as those based on RF technology would generally be used in conjunction with GPS receivers, which offer a way of calibrating them in-orbit. Most small satellites fly single frequency receivers. For example, GPS receivers on the PRISMA satellite are known not to be sufficiently radiation tolerant. This intolerance to radiations is in this case responsible for the absence of measurements during all passes over the South Atlantic anomaly, when the GPS receiver is turned off. Latch ups were also reported to happen once every 15 days per GPS receiver on the PRISMA mission. These relatively frequent data gaps will not be admissible for all space missions, which will require the development of radiation tolerant GPS receivers.

9.6.2 Processing Power Options

The choice of the processor is another important element in the trade-off between reliability and performance. The choice made on the Dutch-Chinese Fast formation of two satellites FAST-D and FAST-T was one of performance, to meet the high manoeuvrability demands of relative navigation. The powerful (2GH) Intel atom

processor, with relatively low power requirements (2.5 W) was indeed the choice made by the FAST team. This processor largely outperforms the space certified LEON3FT for example, which is limited to 4 MFLOPS (million floating point operations per second), in terms of processing power [22]. The fact that the processor is not fault tolerant is addressed through the application of fault tolerant software, which requires an overhead. This does not completely equal the reliability of space qualified equipment but significantly reduces the cost and performance loss incurred by using space qualified hardware.

The PRISMA team on the other hand opted for higher reliability with a space qualified LEON 3 processor allowing for up to 20 MFLOPS and the onboard software was designed to work within this relatively limited processing power. The optical navigation sensor (fully autonomous), the GPS receiver (using an ARM7 microcontroller) and even the radiofrequency FFRF sensor were however developed with their own processing units.

The challenges ahead will include the development of more reliable COTS processors and data handling subsystems and/or the enhancement of the performance of flight proven processors.

9.6.3 Innovative Propulsion Technologies

A MEMS micro-propulsion subsystem was meant to provide low but highly accurate and continuous thrust on PRISMA, which could have added a significant capability in terms orbit control accuracy. This micro-propulsion system failed due to a leak in the fuel tank. More conventional hydrazine thrusters were however available to continue the relative orbit control experiments, but micro-propulsion technology will have to evolve in terms of reliability to meet the fine relative orbit control requirements of future missions. One of the challenges will be the development of power efficient micro-propulsion technologies. In the case of microsatellites, conventional prolusion will still be required to provide sufficient level of thrust, but micro-propulsion will be of particular interest to relative GNC modes requiring highly accurate orbit control. The ability to produce variable thrust will also be a desirable feature and will enable the application of feedback control strategies.

9.6.4 Limitations on Nano-Satellites

A number of the future planned satellite formations, such as CanX, QB-50 will be composed of nano-satellites or even pico-satellites. Based on the current technology, these satellites have limited Attitude Determination and Control System (ADCS) Capabilities, making the prospect efficient coordinated relative attitude control for example very challenging. A number of enabling technologies, such as

micro-wheels for three axis control are currently being developed to increase the limited ADCS capability. Another challenge is the development of micro-propulsion based on MEMS technology for this class of satellites. There are also limitations on the optical navigation capabilities at such small sizes, where the use of GNSS should be the main source of relative navigation data. In the case of nano-satellite formations, less conventional mission concepts such as that of tethered spacecrafts can be of interest to certain applications, particularly when the satellites are required to maintain a close formation.

9.7 Conclusions

A number of technologies have been validated for relative navigation applications, including differential GPS, radiofrequency and optical navigation. However, due to increasing autonomy, relative positioning and relative attitude control requirements, future satellite formations will require further development of current relative guidance, navigation and control techniques. These missions will require higher measurement reliability, the fusion of data from different types of sensors and higher autonomy, while meeting system level constraints. In the case of formations of multiple satellites, there will also be a need for higher processing power, which is currently available from COTS technologies and higher attitude control capability.

Acknowledgments The authors would like to thank Luke Sauter from the Astrodynamics group at the Surrey Space Centre for his useful advice, particularly with regards to collision avoidance.

References

1. D'Amico S, Montenbruck O (2010) Differential GPS: an enabling technology for formation flying satellites. In: Sandau R, Röser H-P, Valenzuela A (eds) Small satellite missions for earth observation: new developments and trends. Springer, Berlin
2. De Florio S, D'Amico S, Ardaens J-S (2010) Flight results from the autonomous navigation and control of formation flying spacecraft on the prisma mission. In: Proceedings of the 61st international astronautical congress, Prague, Czech Republic
3. Larsson R, Noteborn R, Chasset C, Bodin P, Karlsson T, Carlsson A, Persson S,(2010) Flight results from SSC's GNC experiments within the prisma formation flying mission. In: Proceedings of the 61st international astronautical congress, Prague, Czech Republic
4. Terui F, Ogawa N, Oda K, Uo M (2010) Image based navigation and guidance for approach phase to the asteroid utilizing captured images at the rehearsal approach manuscript template and style guide. In: Proceedings of the 61st international astronautical congress, Prague, Czech Republic
5. Horri NM, Kristiansen K, Palmer PL, Roberts RM (2012) Relative attitude dynamics and control for a satellite inspection mission, Acta Astronautica 71:109–118 March 2012

6. Mueller JB, Surka DM (2001) Agent-based control of multiple satellite formation flying. In: Proceedings of the 6th international symposium on artificial intelligence & robotics & automation in space, Canadian Space Agency, St-Hubert, Canada, 18–22 June 2001
7. Breger L, How JP (2008) Safe trajectories for autonomous rendezvous of spacecraft. J Guid Contr Dyn 31(5):1478–1489
8. D'Amico S, Montenbruck O, Larsson R, Chasset C (2008) GPS-based relative navigation during the separation sequence of the prisma formation. In: Proceedings of the AIAA guidance, navigation and control conference and exhibit, Honolulu, Hawai, 18–21 Aug 2008
9. Lan S, Chen X, Zhang JZ, Shi X, Xu G (2009) Intersatellite range determination using multi-detectors observation of pulsars. In: Proceedings of the 7th IAA symposium on small satellites for earth observation, Berlin, Germany, May 2009
10. Wertz JR, Bell R (2003) Autonomous rendezvous and docking technologies – status and prospects. In: Proceedings of the SPIE AeroSense symposium, space systems technology and operations conference, Orlando, FL, 21–25 April 2003
11. Ogilvie A, Allport J, Hannah M, Lymer J (2008) Autonomous satellite servicing using the orbital express demonstration manipulator system. In: Proceedings of the 9th international symposium on artificial intelligence, robotics and automation in space, Hollywood, USA, 26–29 Feb 2008
12. Cesare S, Parisch M, Sechi G, Canuto E, Aguirre M, Massotti L, Silverstrin P (2009) Satellite formation for a next generation gravimetry mission. In: Proceedings of the 7th IAA symposium on small satellites for earth observation, 4–8 May 2009
13. VanDyke MC, Hall CD (2006) Decentralized coordinated attitude control within a formation of spacecraft. J Guid Contr Dyn 29(5):1101–1109
14. Hauschild A, Markgraph M (2009) Accuracy dependency of the GPS navigation solution on the attitude of LEO satellites. In: Proceedings of the 7th IAA symposium on small satellites for earth observation, Berlin, Germany, 4–8 May 2009
15. Johnston-Lemke B, Zee RE (2010) Attitude manoeuvring under dynamic path and time constraints for improved GPS coverage of formation flying nanosatellites. In: Proceedings of the 61st international astronautical congress, Prague, Czech Republic
16. Wokes S, Palmer PL (2010) Perspective reconstruction of a spheroid from an image plane ellipse. Int J Comp Vision 3(5):369–379
17. Virgili LP (2006) Design and evaluation of navigation and control algorithms for spacecraft formation flying. Ph.D. thesis, University of Barcelona, Spain
18. Kayal H, Barwald W, BrieR K, Gill E, Halle W, Montenbruck O, Montenegro S, Sandau R, Terzibaschian T (2003) Onboard autonomy and fault protection concept of the bird satellite. In: Proceedings of the international conference on recent advances in space technology, Istanbul, Turkey, 20–22 Nov 2003
19. Prat G, Maisonobe L (2009) Monitoring the ATV rendez-vous and docking with ISS: a challenge in term of frame transformations dealing. In: Proceedings of the 21st international symposium on spaceflight dynamics, Toulouse, France, Oct 2009
20. Hoffmann GM, Gorinevski D, Mah RW, Tomlin CJ, Mitchell JD (2007) Fault tolerant relative navigation using inertial and relative sensors. In: Proceedings of the AIAA guidance, navigation and control conference and exhibit, Hilton Head, South Carolina, 20–23 Aug 2007
21. Sauter L, Palmer P (2011) An onboard semi-analytic approach to collision-free formation reconfiguration. IEEE Trans Aerosp Electron Syst, pending publication
22. Maessen D, Guo J, Gill E, Gunter B, Chu QP, Bakker G, Laan E, Moon S, Kruijff M, Zheng GT (2010) Conceptual design of the FAST-D formation flying spacecraft. In: Small satellite missions for Earth observation: new developments and trends. Springer, Berlin

Chapter 10
Communication in Distributed Satellite Systems

Klaus Schilling and Marco Schmidt

Abstract The information flow between the different components of a distributed mobile sensor system is crucial in order to enable coordination for an efficient overall performance. This section provides further details for the special situation of network nodes consisting of several satellites and ground stations. Via the communication system partially autonomous functions at each satellite are to be coordinated to enable joint observations. Such self-organized activities of the space segment have to integrate with teleoperations based on supervisory control interaction from ground stations. Related suitable communication design approaches are the central topic of this chapter.

10.1 Introduction

Distributed satellite systems have to exchange data in the networked space segment to enable appropriate coordination for observation, also when no ground station contact is available. By example, when the distributed instruments are to be pointed to the same target observation areas for subsequent data fusion. Further related navigation sensor data are to be exchanged to enable coordinated observations despite disturbances. Real-time corrective reaction capabilities are to be enabled by data acquisition via typically low bandwidth communication links, in order to increase the quality of the obtained observation data for this satellite based sensor network.

Also in the context of provision of telecommunication services, networked satellite systems in low Earth orbits offer interesting capabilities for telecommunication links at a minimum use of resources due to shorter ground contact distances. Challenges are related to significant orbit dynamics affecting the link topology and

K. Schilling (✉) • M. Schmidt
Julius Maximilians University, Am Hubland, D-97074 Würzburg, Germany
e-mail: schi@informatik.uni-wuerzburg.de; schmidt.marco@informatik.uni-wuerzburg.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_10,

thus complicating establishment of reliable intersatellite communications for data transfer. Corrections of Doppler effects or even relativistic effects might be necessary for a sufficient communication performance.

Combination of data from various networked, coordinated satellites enables provision of high performance results despite the limitations in resources of each individual satellite, in particular if distributed systems of small satellites are employed. Technology challenges to implement such innovative distributed spacecraft system concepts require efficient and robust telecommunication in the space segment for a cooperative coordination, but also between satellites and ground segment (cf. Chap. 11). It is the emphasis of the following chapters to discuss related data exchange methods.

10.2 Requirements for Typical Scenarios

Interesting applications of distributed satellite systems are anticipated for robust low-bandwidth communication systems, as well as for space weather and Earth observations [1]. Interesting examples related to SAR antennae are provided in Chaps. 1 and 2 of this book. In order to illustrate the broad diversity of potential distributed space missions, here in-situ multi-point observation of space weather phenomena in the upper atmosphere is addressed by the QB50 mission [2], composed of 50 small satellites. Upper atmospheric layers can hardly be observed from ground, as the accumulated noise effects from dense lower layers limit measurements, while from space only the low density layers above are to be penetrated by the instrumentation. Due to the limited lifetime of satellites in the poorly observed range between 90 and 300 km (corresponding to a survival period of hours to a month due to deceleration by atmospheric particles) expensive traditional satellites are rarely placed at such altitudes. This opens interesting opportunities for distributed small satellites to fill these gaps of knowledge.

For multi-point observations, it is always essential to have a baseline distance between the different distributed sensors and to exchange data related to navigation, observation and satellite status in order to enable appropriate reconfiguration. In constellations all communication and coordination is channeled via a ground station. For avoidance of degradation of observation data due to noise or drift effects between ground contact periods, direct intersatellite links are demanded to enable self-reconfiguration in orbit.

10.3 Communication Concepts for Distributed Systems

The communication and tele-operation infrastructure provides a key element in establishing distributed satellite systems. Approaches for efficient organization of the data flow in the networked system, consisting of several satellites and ground

stations, are addressed in this section. In distributed applications on Earth the internet protocols TCP/IP became the established standard and attracts significant development efforts for further improvements. To benefit from terrestrial activities, transfer of these technologies to the space environment is attractive. Space related network standards are promoted by the Consultative Committee for Space Data Systems (CCSDS) (http://www.ccsds.org/). In particular adaptations to significant delays and to higher noise levels have been included. A significant advantage for this IP-based approach is the easy transfer of satellite data to terrestrial networks without any protocol conversion needs.

10.3.1 Centralized/Decentralized Network Architectures

The network configurations can be based in a centralized way on "servers" [3], or on a distributed approach based on ad-hoc networks [4]. Solutions between those extremes are by example "embedded star topology" [5], suitable for highly deterministic behavior with limited node-mobility. Typical network architectures used in terrestrial context are:

- Purely centralized: many wireless products support this kind of architecture (IEEE 802.16, IEEE 802.11a/b/g). In order to exchange data between each other, two nodes always need to relay over another central node (e.g. base station, access point).
- Cellular systems: can be seen as a generalization of centralized systems. Interconnected base stations form the cell center points and soft hand-overs are to be provided, when a mobile terminal changes from one cell to another. Typical products are GPRS components.
- Ad-hoc networks: here every node has similar rights and does not rely on a special infrastructure.

In a centralized network architecture for formations, one of the satellites would serve as an access point (AP) while the others act as clients. This well accepted architecture is sensitive to failure of the AP. Therefore functions of the AP have to be continuously monitored and in case of a failure transfer of AP-functions to another satellite is to be organized. This requires the AP to frequently broadcast an "alive" message. All communication in the satellite network is relayed via the AP, causing inefficient paths by example for neighboring clients. As the power consumption increases exponentially for a linear extension of the communication distance, the energy efficiency in such centralized systems is low.

Cellular systems exhibit similar problems and are designed for a high number of clients being far beyond the size of present distributed satellite scenarios.

In the decentralized approach by ad-hoc networks a coordinating instance as an AP is missing for controlling medium access or for ensuring that data frames arrive at their destination. This leads to reduced hardware requirements, but causes an increased software development effort. Terrestrial solutions including MANETs

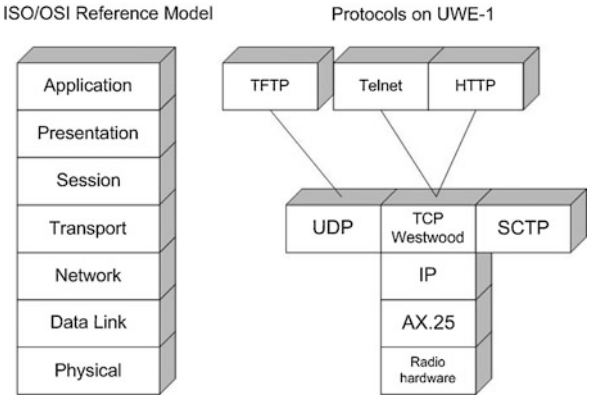


Fig. 10.1 The specific implementation of ISO/OSI reference model layers on-board of UWE-1. Here for comparison reasons several transport layer alternatives were realized

(Mobile ad-hoc Networks), VANETs (Vehicular Ad-hoc Networks) or WSNs (Wireless Sensor Networks) are well established, but still have to be adapted to the use in space. The multi-hop feature reduces communication distances and related energy demands.

10.3.2 Internet Protocols (IP)

First experiments related to Internet Protocols (IP) in space were performed 1999 by NASA during the UoSat-12 mission. One of the first missions, totally operated only over the TCP/IP protocol stack, was the CHIPsat mission launched in 2003 from NASA and the Space Science Laboratory in Berkley. In 2005 the pico-satellite UWE-1 (University Würzburg's Experimental satellite) was launched with the main scientific objective to optimize IP parameters in adaptation to the measured space environment [6, 7]. UWE-1 carried the on-board data handling system µ-Linux, implemented on a microcontroller. Thus advantage could be taken from an integrated, appropriate IP-stack for related telecommunication experiments. The advantages of IP and its higher layer protocols (e.g. TCP, UDP) are the world wide usage, resulting in a fully tested reliable protocol stack and a broad spectrum of available applications using the IP interface. UWE-1 communication was based on a commercial transceiver, normally used by radio amateurs for data transmission via packet radio. The main experiments were related to cross layer optimizations between AX.25 and higher protocol layers (i.e. IP) and to application layer protocols like HTTP and TFTP (cf. Fig. 10.1).

A major disadvantage of the TCP/IP protocol stack is the performance problem of the TCP protocol in space conditions. As the TCP protocol was intended for usage in the terrestrial internet, a congestion avoidance algorithm decreases the transmission rate, if congestion occurs. This behavior is an essential feature of TCP in the terrestrial internet, when the network is overloaded by traffic. A congestion situation in the terrestrial internet is indicated by the loss of data packets. In satellite

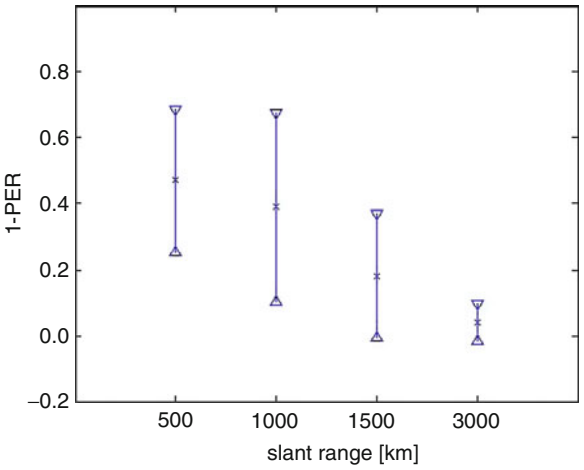


Fig. 10.2 UWE-1 packet error rate (PER) determination for the AX.25 radio link

communication the situation is totally different: loss of packets is normally caused by transmission errors, nevertheless TCP reacts in this situation with decreasing the transmission rate. Therefore it is important to choose very carefully the communication protocols. An alternative is the usage of UDP instead of TCP, a connectionless transport protocol. In this case the application layer has to provide mechanisms to guarantee the correct reception of data packets. Another possibility is to use a TCP extension protocol, which overcomes typical problems of TCP.

The results of the UWE-1 experiments displayed, that IP communication is feasible, but different optimizations are necessary to enable an efficient telecommunication between satellite and ground stations. Especially the high Packet Error Rate (PER) observed on the communication link with UWE-1 affects the performance of the AX.25 protocol. The measured PER values are presented in Fig. 10.2. The values are expressed in terms of confidence intervals, the variance of these intervals reveal the necessity to improve the combination between AX.25 and IP with additional redundancy for the communication link. Further redundancy for the telecommunication can be generated by hardware or by software algorithms to solve the problems of error rates due to significant noise effects from the space environment.

10.3.3 Mobile Ad-hoc Networks (MANet)

Realization of robust network communications attracts significant research efforts in terrestrial applications. A mobile ad-hoc network (MANet) combines several agents to a self-organizing telecommunication network. It includes in particular initialisation and reconfiguration capabilities for the case of defects or of changes in the topology. Therefore in formations of satellites, exhibiting high dynamics and link interruptions, a reliable reconfiguration capability of the communication path,

Fig. 10.3 Network of mobile sensor systems with heterogeneous dynamics

employing several agents of the space and ground segment, will increase robustness. Related routing methods are analyzed in the following section.

For characterizing MANET performance in mobile systems at the University Würzburg's demonstrator and test facility based on WLAN (IEEE 802.11) has been installed, consisting of a system of several mobile robots and fixed stations as nodes (cf. Fig. 10.3) [8].

In this test facility experiments to prepare future MANET applications in space have been performed with respect to re-routing performance. Typical ad-hoc routing protocols developed for mobile systems were compared in teleoperation scenarios for mobile vehicles, including:

- Reactive protocols, such as "Ad-Hoc On-demand Distance Vector (AODV)" or "Dynamic Source Routing (DSR)",
- Proactive protocols, such as "Optimized Link State Routing (OLSR)",
- Hybrid protocols, such as "Better Approach To Mobile Ad-hoc Networking (BATMAN)".

Pro-active protocols expect nodes to frequently send messages, which increases power consumption, while re-active protocols require significant start up time before establishing data streams.

A testbed was realized to record during test runs the crucial data about neighbors, route requests, potential routers, link costs and hop counts (a typical example of round trip times related to changing transmission topology is provided in Fig. 10.4). Thus resulting characteristics of the packet stream like packet loss rates, time needed for route reestablishment, packet inter-arrival time, network topology and bandwidth can be evaluated. Files from the different nodes are to be synchronized (with respect to time or to events). Typically default parameter settings need to be adapted to the specific scenario to exhibit reasonable performance.

The performance measurements turned out to be very sensitive to noise effects, thus a careful setup preparation is mandatory to generate comparable results.

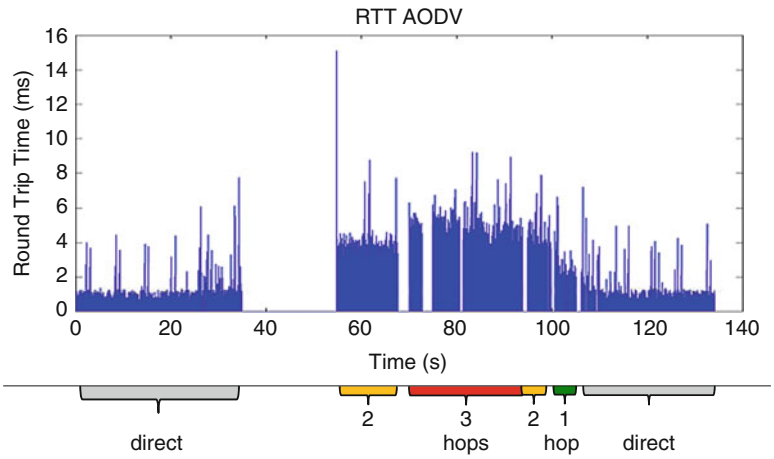


Fig. 10.4 Typical round trip time behaviors for a changing transmission topology, displaying in particular the significant transmission interrupts due to route reestablishment

Table 10.1 Performance comparison for test runs with tuned parameter settings in the protocols

Protocol	Packet loss (%)	Min. time for rerouting (s)	Max. time for rerouting (s)
OLSR	32.6	5.0	<21.6
DSR	28.8	2.0	<40.4
BATMAN	16.0	0.8	<26.2

For future use of MANETs in space also adaptation procedures of protocols to the specifics of the encountered space environment are to be investigated.

Independent from the used specific protocol, Table 10.1 displays a large amount of packet losses as well as a significant maximum time for rerouting in case the network topology changed during test runs. Such characteristics are to be considered in the design of the communication link and possibly parameter are to be recalibrated due to the encountered specific space environment.

10.3.4 Delay Tolerant Networks

Delay or disruption tolerant networks (DTN) were designed for heterogeneous nodes without continuous network connectivity [9, 10]. This is typical for space applications, where link interruptions due to limited radio range or due to occultation effects, as well as dynamically changing link topologies and distances are typical for satellites in different orbits. Therefore DTNs have attracted much interest in the space community [11, 12]. In such situations standard MANET fail, as they first attempt to establish a complete route and then transfer the data. In case of the DTN a "store and forward" concept is applied, where data are passed step by step to available nodes in the hope to finally reach the target node. In order

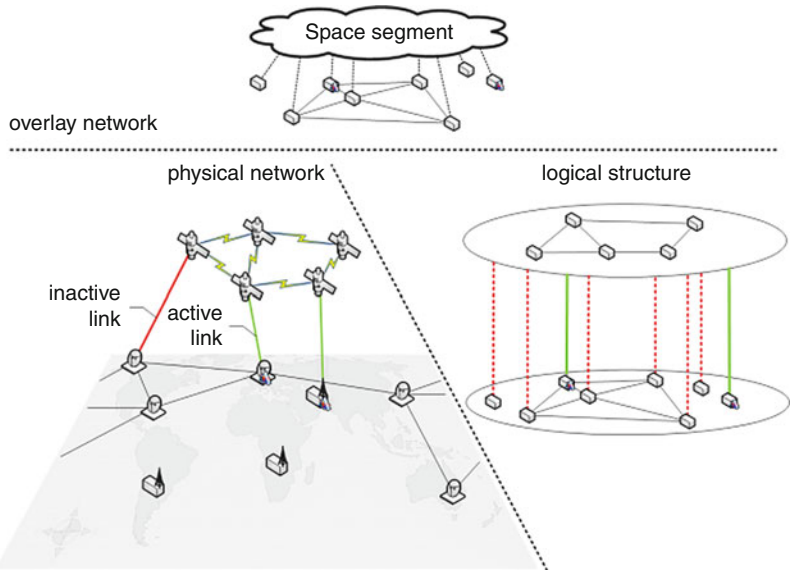


Fig. 10.5 Schematic of an overlay network approaches for an integrated space and ground segment taking into account the available physical network structure and the abstracted logical structure

to increase chances for success, messages are replicated. Thus a huge amount of traffic and of local storage needs is generated. Nevertheless also "hybrid" approaches trying to combine advantages from MANET with DTNs are investigated [13].

Standards are provided by the bundle protocol (e.g. RFC 4838 and RFC 5050), defining a series of contiguous data blocks as a bundle, and operating as an overlay network (cf. Fig. 10.5) with a naming approach based on Endpoint Identifiers [12]. In the extreme case, even concepts for interplanetary internet have been addressed in this context [14].

10.3.5 Distributed Satellite Control via Networks

For precision observations, network controls have to be established in order to coordinate in real-time pointing of the instruments on the different satellites towards joint observation targets. Thus, via the intersatellite links related control loops are to be closed. Challenging problems result as information in a communication link is transferred in digital and packetized form in an event oriented manner, while control is usually based on continuous real-time information inputs, or at least on inputs acquired on fixed sampling intervals. In non time-critical situations the telemetry and telecommand data can be delayed to a specified level and then be executed, while in real-time context it is still a topic of intensive research.

10.4 Conclusions and Future Trends

The paradigm shift from large spacecrafts incorporating multiple payload capabilities to decentralized, distributed small satellite systems challenges communication system technology in order to coordinate such a highly dynamic sensor network. While in terrestrial data networks a huge amount of effort has been invested, ranging from sensor networks to cloud computer networks, in the space environment related technologies are still in the initial stage. While basic concepts can be transferred, specific space characteristics, like interruptions, variations in delays due to high relative velocities of nodes require adaptations. To realize the vision of future self-organizing sensor networks in space in particular combined communication and control approaches need to be further developed, in particular for time-critical situations. Breakthroughs in this field will further improve the basis for sensor data fusion in order to achieve higher resolution and precision in observation data.

Acknowledgments The authors acknowledge the contributions from many collaborators in computer Science of University Würzburg and in the company Center for Telematics, where for years this topic was investigated and implemented for the satellite missions UWE-1 and -2.

References

1. Schilling K, Garcia-Sanz M, Twiggs B, Sandau R (2009) Small satellite formations for distributed surveillance: system design and optimal control considerations. NATO RTO Lecture Series SCI-209
2. Muylaert J (2009) An international network of 50 double cubesatsfor multi-point, in-situ, long duration measurements in the lower thermosphere and for re-entry research. QB 50 workshop, Brussels. https://www.qb50.eu/project.php
3. Knoblock EJ, Wallett TM, Konangi VK, Bhasin KB (2001) Network configuration analysis for formation flying satellites. In: Proceedings of IEEE aerospace conference, Big Sky, Montana, pp 2/991–2/1000
4. Sidibeh K, Vladimirova T (2006) IEEE 802.11 Optimisation techniques for inter-satellite links in LEO networks. In: Proceedings of 8th international conference advanced communication technology, Phoenix Park, Korea, pp 1177–1182
5. Clare LP, Gao JL, Jennings EH, Okino C (2005) A network architecture for precision formation flying using the IEEE 802.11 MAC protocol. In: Proceedings of IEEE aerospace conference, Big Sky, Montana, pp 1335–1347
6. Schilling K (2006) Design of pico-satellites for education in system engineering. IEEE Aerosp Electron Syst Mag 21:9–14
7. Schmidt M, Zeiger F, Schilling K (2006) Design and implementation of in-orbit experiments on the pico-satellite UWE-1. In: Proceedings of the 57th international astronautical congress, Valencia, Spain, IAC-06-E2.1.07
8. Zeiger F, Krämer N, Schilling K (2008) Parameter tuning of routing protocols to improve the performance of mobile robot teleoperation via wireless ad-hoc networks. In: Proceedings of the 5th international conference on informatics, automation and robotics (ICINCO 2008), Funchal, Madeira

9. Fall K (2003) A delay-tolerant network architecture for challenged internets. In: Proceedings of ACM SIGCOMM 2003, Karlsruhe
10. DTN Ref. Implementation. http://www.dtnrg.org/wiki/Code
11. Jenkins A, Kuzminsky S, Gifford KK, Holbrook M, Nichols, K, Pitts L (2010) Delay/disruption-tolerant networking: flight test results from the international space station. In: IEEE aerospace conference 2010, Big Sky, Montana
12. Wood L et al (2008) Use of the delay-tolerant networking bundle protocol from space, Conference paper IAC-08-B2.3.10. 59th international astronautical congress, Glasgow
13. Ott J, Kutscher D, Dwertmann C (2006) Integrating DTN and MANET routing, SIGCOMM'06 workshop 2006, Pisa
14. Krupiarz CJ, Jennings EH, Pang JN, Schoolcraft JB, Seguí JS, Torgerson JL (2006) Spacecraft data and relay management using delay tolerant networking. In: SpaceOps 2006 conference, Rome, Italy, AIAA 2006-5754

Chapter 11
Ground Station Networks for Distributed Satellite Systems

Marco Schmidt and Klaus Schilling

Abstract A space mission is typically divided in space and ground segment, the focus of this chapter lies on ground station networks. Especially highly distributed ground station networks offer new opportunities for the operation of distributed satellite systems. Actual networking concepts are presented and research challenges in the field of ground station networking are discussed.

11.1 Introduction

Distributed satellite systems can be used in wide range of application fields, for example to increase the spatial or temporal resolution in an earth observation mission. Benefits of multi-satellite systems are discussed extensively within this book. However, an important aspect which needs to be addressed is the operations phase of such a distributed space mission. It is obvious that the operation of several space vehicles in a single mission requires a more sophisticated operations concept. This involves a number of additional tasks, like planning and scheduling to coordinate science observations or contact windows. Depending on the number of satellites, this can be a very time consuming and extensive task.

The current state of art procedure for multi satellite systems is to control each satellite independently from ground, defined as constellation. Examples are manifold, for example the Cluster 2 mission consists of four satellites each one controlled individually from ESOC mission control. In communications Iridium satellites and in navigation GPS satellites form constellations, too. For future missions, containing a vast number of space vehicles, this would result in an enormous effort for mission operators. There are two main approaches pursued to overcome such issues: One possibility is to place more autonomy in the space

M. Schmidt (✉) • K. Schilling
Julius Maximilians University, Am Hubland, D-97074 Würzburg, Germany
e-mail: schmidt.marco@informatik.uni-wuerzburg.de; schi@informatik.uni-wuerzburg.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_11,

segment, hence not all satellites need to be operated at the same time. A different strategy is, to use a network of highly distributed ground stations to efficiently operate a distributed space segment. A new concept of low-cost ground station networks evolved in the last years, mirroring the distributed satellite approaches in the ground segment, too. It is especially well suited for the operation of huge number of satellites. The advantage of these low-cost receiving stations is, that they resemble in architecture and are therefore compatible with the most small satellite platforms. Also for larger satellites promise highly distributed ground systems a new step in satellite operations. The antenna systems of low-cost stations were designed for specific frequency bands, the signal processing can be performed with COTS components. This is a very interesting aspect, in this way it is possible to operate many satellites at the same time with a ground network, which was originally not established for multi satellite operation. But due to the interoperability of satellite and ground systems, the existing resources can be combined with simple means to take advantage of the distributed nature. The main challenge is an intelligent network concept for a distributed space segment in combination with loosely coupled ground stations.

In contrast, the classic ground station approach contains typically a broad spectrum of specialized hardware, which limits the utilization to a specific kind of satellite mission or type. These dedicated ground stations can't be grouped together arbitrarily for the operations phase of a distributed system. Especially interesting are concepts from the field of distributed intelligence and control for distributed space missions, unfortunately only some of them are transferable to the traditional ground station concept.

In this section the differences between the traditional ground station approach and low-cost ground system architectures are explained. These differ on the one side in topology, which influences the operations phase in many ways. On the other side they differ significantly in the type of components employed. Finally, the technological challenges for ground systems in future distributed space missions are pointed out. A very interesting aspect is the establishment of highly distributed ground station networks, which can be utilized for the operation of a huge number of satellites. Section 11.2 handles in detail current ground station network projects. Furthermore, the resulting requirements for multi satellite operation are elaborated, like coordination and scheduling. The chapter finalizes with an outline of future developments in the field of highly distributed ground systems.

11.1.1 Traditional Ground Station Approach

Since the first space missions the ground segment consisted of several entities, exchanging information between these parts can be considered already as networking. The aggregation of different stations was already performed in the beginning of the space era to achieve better coverage or to increase redundancy. A good example for this "traditional" or "classic" approach is the ESA ESTRACK system,

consisting of nine ground stations combined to a network, it was initiated already in the early 1970s. Also NASA started quite early with the combination of different stations to networks, the Deep Space Network (DSN) was established in 1958 [1].

Before addressing the traditional ground station approach, an important aspect related to the taxonomy of the term ground station needs to be clarified: In literature is the ground segment typically divided into mission control and ground station network [2, 3]. The mission control is aggregated in control center (Mission Control Center (MCC), Spacecraft Operations Control Center (SOCC)) and is responsible for mission planning and mission operations, for example monitoring and commanding the spacecraft. The ground station network, composed of different ground stations, is on the other side dealing with signal reception and transmission, orbit tracking etc. The mission control and the ground station network are not only logically divided, also often geographically separated from each other. A typical ground station contains in this context several receiving stations including different types of antennas as well as corresponding hardware equipment. Exemplary, the ground station in Weilheim (part of the ESTRACK system) contains six different antennas ranging from 6 to 30 m to support deep space missions as well as near Earth missions. So, a ground station describes in a traditional sense a sophisticated system containing a larger number of facilities for satellite communication, acting as the interface between satellite and mission control center. The traditional ground station network contains only a few, but therefore highly specialized stations, to support a broad spectrum of space missions. Contrary, in the context of small satellites projects, a ground station is considered rather as a single entity used to access one single satellite, a ground station consists here typically only of one antenna system and corresponding equipment for communication with a single satellite in LEO (compare Sect. 11.1.2).

The ESTRACK system is used to explain the overall organization of a large ground station network. More specifically the architecture of a typical ground station, i.e. a satellite receiving station will be explained. Issues related to ground station networks itself are discussed in Sect. 11.2. The ESTRACK core ground network consists currently of nine ground stations located in Australia, Africa, Europe and South America. Since 1968, ESTRACK supported more than 60 missions. The ground stations of the ESTRACK system contain many highly specialized stations. Thus, dedicated stations for deep space missions and geostationary satellites exist and can in general not be replaced with each other. The ESTRACK system has for example dedicated 35 m diameter antennas for deep space missions, for near Earth missions 15 m diameter antennas are provided. The frequency bands range from S-Band to Ka-Band [4], thus all kind of missions can be supported by ESTRACK, but dedicated stations for specific type of missions need to be requested.

An exemplary ground station for communication with a LEO satellite consists typically of the following units: Antenna system, transmitting and receiving equipment as well as telemetry, tracking and control (TT&C) equipment: The antenna system contains different sized apertures, mainly large parabolic dish antennas for signal reception (the size depends on several factors, important is to provide the

necessary gain for the link budget of a space mission). The equipment for receiving and transmitting the electromagnetic waves comprises low noise amplifiers (LNA), high power amplifiers (HPA), up- and down-converter. For telemetry and telecommand are baseband devices used, its tasks are carrier acquisition, tracking and demodulation, processing of data, and ranging. Each station is composed of a broad spectrum of devices and components and constitutes a complex facility. As mentioned above, an individual ground station is often connected to a larger network of ground stations, which are then connected to several control centers (MCC and SOCC) and data networks.

The devices and components in these ground stations are usually not commercial of-the-shelf (COTS), they have to fulfill high quality requirements and satisfy common space related standards. Thus, specialized hardware equipment can be very expensive. Additionally, it has to be considered, that a single ground station provides only limited coverage for a LEO satellite. To increase the coverage for a low altitude satellite mission, other ground stations need to be incorporated in a network. However, even for a single ground station are extensive redundancy concepts required to achieve high degree of availability for a mission. Thus, redundant equipment (receivers, transmitters, TT&C equipment, computers etc.) is essential to avoid the loss of data. In many cases it is mandatory to have a second ground station available which could take over in the case of any failure. Hence, the traditional ground station is a complex system, containing a variety of facilities, devices and components. The effort to operate and maintain such a system is enormous; the fraction of the mission budget dedicated to operation costs should not be underestimated. Of course, not each satellite mission needs a dedicated ground station network, the established networks from the space agencies are utilized for several missions simultaneously. So, traditional ground station networks are open to paying customers. To run a huge ground station network a large part of the available budget comes from facility fees. Thus, the typical average cost of a tracking hour ranges between 300 and 450 Euros [4].

On the first glance the “traditional” and the “low-cost” ground station approach seem fairly similar, both contain equipment for satellite tracking and reception, nevertheless significant differences exist, which are pointed out in the next section.

11.1.2 Low Cost Receiving Stations

The term low-cost ground station might be a bit misleading, but it is important to point out that the receiving stations in small satellite projects have differences in capabilities and architecture compared to traditional ground station systems. Thus, the term “low-cost ground station” is used to emphasize the difference to the traditional approach. Low-cost means here, that these stations are not composed of high quality standard components, they are rather composed of commercial of-the-shelf components (COTS), i.e. radio equipment, modems etc. Other authors refer to that kind of station as radio stations or academic ground stations. Hence,

one has to be careful when the term ground station is used: While the classic term ground station describes a complex system containing a broad spectrum of expert systems, the same term means in the context of small satellites projects a single entity used to access one single satellite. A low-cost ground station is in contrast to the classical definition a standalone system, comprising all components to communicate with the satellite, i.e. antenna, transceiver, tracking hardware, data distribution system, etc. There is no clear distinction any more between mission control and ground stations. Furthermore, a ground station consists here typically only of one antenna system and corresponding equipment for communication with a single satellite in LEO. All these stations are very similar in architecture and contain similar type of antennas. They utilize comparable hardware components for signal processing as well.

The reason why academic ground stations are so similar in architecture is that they are designed for the communication links typically used in small satellites. Due to the limited mass and power budget and the restricted pointing capabilities are primarily UHF and VHF transceivers used for communication. The utilized frequency bands, 70 cm and 2 m, are part of the amateur radio bands and are under the supervision of the International Amateur Radio Union (IARU). Higher frequency bands, for example S-Band, will be the next step in the evolution process of small satellites, but are still rarely used.

This affects of course the architecture of low-cost ground stations: As hardware components for UHF and VHF are commercially available, many ground stations are built up from low-cost commercial of-the-shelf components. Typically are standard radio transceiver and Terminal Node Controller (TNC) components used, which are connected to simple desktop computers (example architectures are described in Refs. [5, 6]). That standard computer is normally connected to the Internet for data exchange. A variety of antennas and suitable tracking systems are offered. To control the antenna and radio equipment several software solutions are available (open source as well as proprietary). A schematic diagram of the ground station at the University of Würzburg is shown in Fig. 11.1.

The ground stations of an academic network are connected through the Internet, i.e. the Internet Protocol (IP) is used on the network layer, on the transport layer the Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) is used, depending on the application on top. Each ground station in the network can be seen as an access node to a satellite, which is in contact range. Typically only one communication link to a satellite can be established at the same time. Currently, IP is only for the data exchange between the ground station computers used, there is no real end-to-end communication between satellites and distant ground stations on basis of IP realized.

The primarily used protocol for data exchange between ground station and satellite is AX.25 [7], which is conform to HDLC ISO standard 3309. AX.25 originates from the X.25 protocol and was adapted for the special needs of the radio amateur community. It is used as data link layer protocol in packet radio mode. Error detection is possible due to a 2 byte checksum attached to each frame, corrupted packets are discarded by default from a Terminal Node Controller (TNC) device.

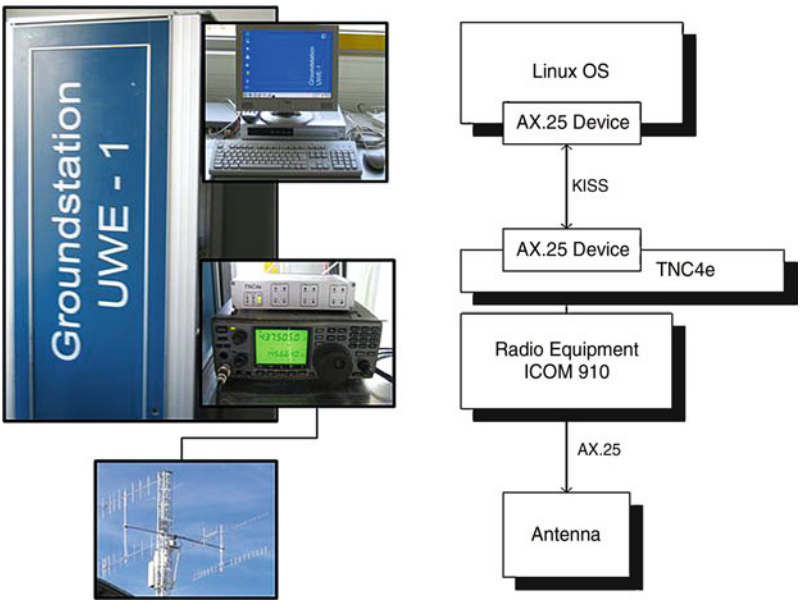


Fig. 11.1 Low-cost ground station at the University of Würzburg

In many ground stations is the TNC replaced by software modem. This enables more control over the radio link and delivers new opportunities for post processing of corrupted data. AX.25 is accepted in the small satellite community as standard communication protocol, i.e. all major ground station network projects support AX.25. Nevertheless, a migration to CCSDS protocols might be on a long term view reasonable for compatibility reasons.

11.1.3 Technological Challenges

In the last years the tremendous progress in computational power brought a number of innovations to the ground segment. This implies as well technological challenges for ground stations and offers new approaches due to recent advances in digital signal processing or network control. Hence, this section points out the specific technological challenges and implications for the ground segment to operate multi satellite systems.

In this context the terms multi-satellite systems, distributed satellite systems and satellite networks are used interchangeable, and describe in general a space segment with more than one space vehicle, which are used to achieve a common goal. The definitions of a formation, constellation or cluster are used to distinguish between different topologies or satellite control strategies. Beside the space segment also the ground segment is integral part of each space mission. In a simple

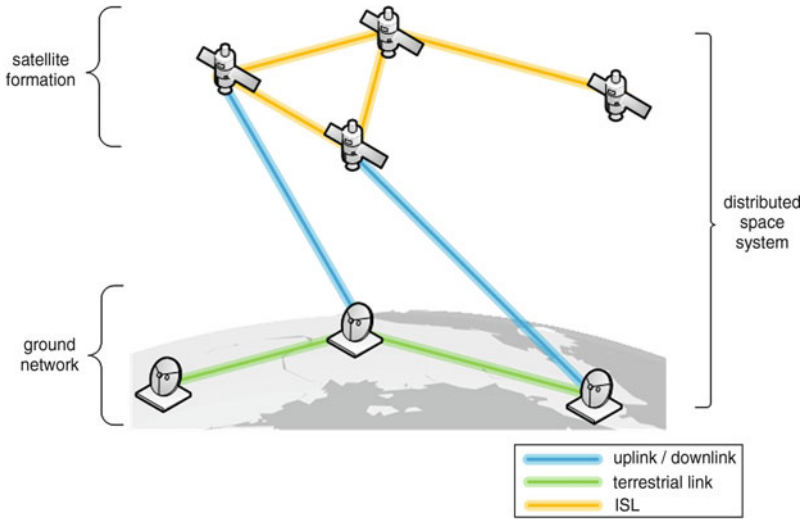


Fig. 11.2 Distributed space systems

view, one could already see one satellite and one ground station as a distributed system (consisting of two nodes communicating with each other during contact). But considering a more general case, having a multi satellite system connected to a ground station network, is more suitable. Therefore, one can identify two different distributed systems, one in space and another one on ground, both in close relationship with each other (see Fig. 11.2). This viewing angle is sometimes neglected when multi-satellite systems are discussed, but of course the topology of the ground segment plays an important role for telemetry and telecommand. Due to the movement of the satellites on their orbits, the communication links between ground stations and satellites change frequently. That adds new challenges to the operation of a satellite network, careful planning and scheduling is necessary.

This section emphasizes the challenges for low-cost ground station networks, i.e. loosely coupled radio stations with a large number of network nodes. The corresponding infrastructure is already available; many ground stations were built in the scope of small satellite projects. First steps in the direction of networking these stations are already underway (GENSO, GSN—compare Sect. 11.2). The described challenges need to be faced to enable an efficient and intelligent operation of multi satellite systems.

The distinction in traditional and low-cost ground station network approach is necessary, due to differences in topology, architecture and requirements. With respect to satellite swarms and formations is especially the low-cost approach promising, as the large number of network nodes can be used to realize new operation concepts. Future mission scenarios including more than tens of space vehicles in a single mission need an efficient operation concept based on recent advances from distributed system control.

One of the major challenges is to coordinate and control the existing, but very heterogeneous, ground station networks. As mentioned above, current ground station network implementations consist of loosely coupled radio stations, which are composed from a variety of COTS components. Even if the overall architecture is similar, i.e. each station contains the same functional entities like radio equipment and modem, the individual realization of each entity might be very different: For example, the modulation of a satellite signal is often performed from a hardware device called Terminal Node Controller (TNC), meanwhile software modems became popular and can be used instead of a TNC. Especially with respect to the software setup are low-cost stations very heterogeneous and utilize a number of commercial, but also self developed applications. The challenge is to combine the individual stations in such a way that the overall network can be used to efficiently operate a multi satellite system. Nevertheless, each station belongs to an independent research institute which restricts their gearing to a certain depth. In particular on hardware level is the coordination of such a loosely coupled system very difficult. Hence, it has to be ensured that the stations are treated in an abstract way to fully benefit from the distributed nature.

A very crucial point is the establishment of communication resp. protocol standards for distributed space missions. Nowadays mainly conservative point-to-point links between satellites and ground stations are employed. But the success of the terrestrial internet demonstrated, how powerful a self configuring and reliable communication network can be utilized for a broad spectrum of applications. Though there is neither a satisfying solution nor an agreement on interfaces and protocols for space communication in multi hop networks. While the IP protocol stack proved its reliability and performance in terrestrial applications, in space missions several drawbacks were identified (especially related to the TCP protocol). Therefore, many researchers tend to utilize dedicated protocol implementations, like the recommendations of the CCSDS committee, defining protocols optimized for the space environment. A major problem is then the interoperability between the distributed system in space (relying on CCSDS) and the distributed system on ground (mainly based on IP). So, communication standards for the ground segment are one of the major research challenges which need to be faced for future space mission. Which alternative will prevail in the coming years is hard to predict, the long term objective could be the establishment of a “spacenet” comparable to the terrestrial internet.

In current low-cost ground station networks the availability of individual stations is relatively often affected from failures. The reason is the structure of such networks, composed of loosely coupled stations, which might not be available due to trivial reasons. In terrestrial communication networks are failures handled autonomously, for example if an internet router fails, other routers automatically reconfigure routing tables to take over its responsibilities. In space applications this self-configuring capabilities are so far only rarely utilized. Thus, flexibility becomes on the one side an important role, on the other side mechanisms to compensate the unavailability of a single station need to be provided. As the individual stations in low-cost ground networks are compatible with each other, it

is in general possible to implement to certain degree graceful degradation. The main challenge is here to adopt recent advances from self-configuring and autonomous networks to provide a high degree of flexibility and to take advantage of the distributed nature of such networks.

The usage of low-cost ground station networks is especially promising to realize new operation concepts. While the traditional approach uses only dedicated links to connect one satellite to one ground station, new approaches exist which propose to use a ground network as a whole to operate a satellite. The ground network can be seen as a sensor network, which is able to determine the orbit or link quality based on a multi-point measurement from ground. Such new operation concepts will enter a new era of satellite operations, but placing a number of additional challenges: Beside the technological challenges like communication and failsafe capabilities, further issues need to be addressed: While in the traditional ground station approach the management of the network is carried out from a single institution, the organization in low-cost networks is fully distributed. There is no central entity with full access to all nodes in the network. From terrestrial applications the benefits and drawbacks of decentralized control mechanisms are well known, the challenge remains to transfer that knowledge to ground networks considering their particularities. This addresses as well new requirements in cooperation between research institutes, from administrative and legal point of view.

11.2 Ground Station Networks

11.2.1 Infrastructure

Already from the very beginning of the space era, different facilities for satellite operation were grouped together as the ground segment. The ground station network is here often associated with the receiving stations. Nevertheless, there is not always a clear distinction made between the ground network and appending facilities for mission control. These facilities are the Mission Control Centers (MCC), which are responsible for telemetry and telecommand of a spacecraft, and the Science Control Centers (SCC), which are dedicated to the science payload of a mission. These centers are not only logically separated; often they are also geographically distributed in different locations. All facilities are typically connected through a data network to collect, exchange and archive data and to grant access to end users.

In the scope of many space missions a broad spectrum of ground infrastructure was established. The available resources were steadily upgraded and emerged over time to support different satellite missions. Therefore, those networks and appended facilities have grown to a complex system containing a variety of network architectures and technologies. Hence, it is hard to derive a generic architecture of a traditional ground station network. Some examples of currently utilized ground systems are shortly introduced to give an overview: The ESTRACK system

operated by ESA was already mentioned in Sect. 11.1, it has a core network of nine stations and cooperates with other organizations to increase the coverage. The main control center is located in Germany (ESOC in Darmstadt). A huge network which provides services to more than 100 satellites is the Air Force Satellite Control Network (AFSCN) in the US, which has 10 stations worldwide with 22 antennas. Beside the operation of military satellites they are utilized as well to provide launch and early orbit tracking. Two control centers are in charge of command of the assigned satellites. Another interesting ground station network is the deep space network (DSN), which is dedicated to deep space missions and contains three facilities for communication with deep space orbiters, separated approximately by 120° in longitude (US, Spain and Australia). Each receiving location contains at least four large parabolic dish antennas (ranging from 26 to 70 m).

In the context of low-cost ground stations, a dedicated mission and science control center on a distant location is mostly not set up. Each individual ground station originates from a small satellite project, the ground station itself performs all necessary tasks for satellite operation and can be seen as a standalone system. When these stations are combined to a network, the task of the operation centers are performed from software applications, using the internet for data exchange.

In the following are different aspects addressed which play a major role for the ground segment when distributed space missions are considered. Especially a large number of network nodes in a distributed mission have impact on operation concepts.

11.2.2 Highly Distributed Ground Networks

The trend to distributed space missions puts certain requirements on the ground segment. Especially future missions, containing a large number of space assets within a single mission demand for a corresponding counterpart on ground. Thus, highly distributed ground station networks are needed to cope with highly distributed space segments. The term highly distributed ground network describes in this context a network of ground stations with a large number of network nodes, where more than one station can communicate with a distributed space mission at the same time. Hence, several communication links can be established between space and ground in parallel.

To emphasize the needs of highly distributed ground networks, the QB50 mission is introduced. The QB50 project is a very promising mission relying on a multi satellite system [8], initiated from the Von Karman Institute (VKI) and supported from many European experts. The scientific objective of QB50 is the in situ measurement of spatial and temporal variations of key parameters in the lower thermosphere. The measurements are performed from 50 double CubeSats, launched together and separated consecutively from the launch vehicle. All satellites are equipped with identical sensors, which obtain data approximately 100 km separated from each other. Furthermore, the re-entry process of the distributed satellite system is studied. Interesting is the fact that the 50 satellites are provided

from 50 different institutes, so the satellite swarm itself is very heterogeneous. For operation is the ground station network GENSO foreseen (detailed description of GENSO follows below). The project is a promising mission with scientific value; nevertheless suitable strategies for efficient operation of such a satellite network are not available yet. One key aspect for the proper operation of 50 satellites flying near to each other in nearly the same orbit is the availability of appropriate ground resources, as well as scheduling and data management (compare Sect. 11.2.3).

The QB50 mission concept is used to illustrate the specific requirements placed on the ground segment in such a distributed mission including a large number of vehicles. Beside the large number of satellites, the short contact times in a LEO orbit (5–6 contacts between 5 and 15 min a day) complicate the operation. Another issue is the restricted lifetime of the distributed space mission, due to the low altitude of the satellite orbits. Of course the low altitude orbit is intended, larger satellites are normally not placed in such low orbits because of the remaining atmosphere which restricts the lifetime dramatically. Hence, the restricted lifetime of the mission prevents the possibility to use a single ground station to communicate sequentially with each satellite in the network in a large time horizon. Thus, the question needs to be faced how to gather the data from the space vehicles efficiently down to earth. The most convenient way is to collect the data from all the 50 satellites by the utilization of a highly distributed ground network. QB50 intends to use the GENSO network for this reason, which is due to the large number of participating stations an ideal platform to operate many satellites in parallel.

In the following are two implementations of highly distributed ground station networks described, mainly composed from low-cost receiving stations. They provide the capability to track a distributed space missions with a large number of receiving stations and are therefore well suited for operation of satellite swarms or formations.

11.2.2.1 Global Education Network for Satellite Operations (GENSO)

The Global Education Network for Satellite Operations (GENSO) was started in 2006 from the International Space Education Board (ISEB), which consists of the educational departments of CSA, JAXA, NASA and ESA. Objectives of the GENSO project are to allow remote access for operators to their satellites over the network, provide remote control of the participating ground stations and to define and implement a global standard for educational ground stations. The GENSO software provides a distributed platform based on a client and server architecture. A server application, the so called Ground Station Server (GSS), is installed on each ground station participating in the network. Mission Control Clients (MCC) are used from the satellite operators to bring the data from the satellite back to the owner. Secure access is granted from Authentication Servers [9]. From a technical point of view is the GSS software responsible for handling the interfaces to the hardware devices, i.e. rotor and radio equipment. Software libraries like the Ham Radio Control Libraries were integrated to support a broad spectrum of COTS components.

The MCC software is used from an operator to control the configuration of a connected GSS server. It contains a graphical user interface, but also provides interfaces for customized software solutions for satellite operation. The public response to GENSO is very remarkable and many small satellite developers announced their interest to join the GENSO network. More than 30 universities and radio amateurs are currently involved in the early operational phase [10]. The project issued already a first version of its software, but is still under testing phase.

11.2.2.2 Ground Station Network (GSN) of the UNISEC Group in Japan

The ground station Network project of the UNISEC group was already initiated in 1998. As JAXA also takes part in the ISEB, there is close cooperation between developers from the GSN and the GENSO projects. The GSN network was established on a national level in Japan. Objective of GSN is to construct a network-based ground station system with functionalities for remote control via Internet. More than ten ground stations in Japan are participating in the project [11]. Conducted experiments demonstrated already successful how the operation of pico-satellites can be improved by the usage of several ground stations. The GSN system bases on the Ground Station Management Server (GMS), a package of software functions which are necessary to remotely control ground station hardware. As each ground station has different hardware architecture, specific device drivers need to be implemented to interface the GMS client software. The data exchange between client and local ground station hardware uses the Web Services technology, a W3C term for communication between devices using web standards, in this way is a platform independent remote control concept realized. An operator using GMS can directly control hardware devices like TNC or radio equipment.

11.2.3 Planning and Scheduling

One of the most limited resources of a satellite in a low Earth orbit is for sure the communication time. The contact window between a satellite and ground station can be determined from the orbit elements and the location of the ground station. For a LEO satellite are typically 5–6 contact windows between 5 and 15 min available each day. This implies two major drawbacks: First, the satellite is only visible for a few minutes each day, which limits the amount of transferable data dramatically. Moreover, the ground station is not utilized for a large fraction of each day. Second, the contact windows have fixed start and end times due to orbit geometry and a ground station can serve only a single satellite at a time. Therefore, overlapping contact windows of individual satellites lead to the question which of these satellites should be operated first. To handle conflicts respectively overlapping contact windows, scheduling comes into play. The term scheduling describes in general the assignment of scarce resources to activities, with the objective to optimize one or

more performance measures [12]. Planning is a very general term and involves formulating a sequence of actions to achieve a desired objective, for example path or task planning. According to [13], scheduling is a special case of planning, where the actions are already chosen, leaving only the problem of determining a feasible order. In certain environments is the process of planning and scheduling still performed manually and is therefore very time consuming. It takes for example already 5 h to create a preliminary schedule for a small number of tracking stations of the Air Force [14]. In larger networks, the creation and optimization of suitable schedules can take weeks. A variety of systems for automatic scheduling in space applications were already investigated in the past, this section focuses on the satellite range scheduling domain.

The term scheduling is used for a wide field of problems, in space context the Satellite Range Scheduling (SRS) problem was investigated in deep. It is defined as the generic problem of scheduling task requests for communication antennas, introduced from Schalck [15]. A satellite can only communicate with a ground station when it is within the transmitting horizon of the ground station, depending on the orbit geometry this occurs periodically. Creating an efficient schedule is very sophisticated due to the huge number of combinatorial possibilities of satellites and ground stations. The SRS domain can be further divided into the Single Resource Range Scheduling Problem (SiRRSP) and Multi Resource Range Scheduling Problem (MuRRSP), where SiRRSP describes the case of a single ground station (respectively antenna) allocation to several satellites. The general problem of assigning a ground network to several satellites is contained in the MuRRSP domain. It was shown from Burrowbridge, that the SiRRSP (i.e. only one ground station) can be solved in polynomial time when only LEO satellites are considered [16]. In contrary, in the MuRRSP domain (i.e. more than one ground station needs to be assigned), the contained problems are NP complete [17].

SRS scheduling was investigated in deep for a number of different applications, each having special constraints or requirements, e.g. related to scheduling time horizon and type of satellite missions. Comparing the results from these different applications shows, that the performance of an algorithm depends strongly on the problem formulation and structure. While for example a genetic algorithm proved to obtain good results for the satellite range scheduling problem in the AFSCN network, genetic algorithms failed to get comparable results in similar scheduling problems. So, heuristics and strategies can't be transferred easily from one application to a slightly modified version of the same problem. There is no state of the art algorithm which delivers good average result on all scheduling problems in space context.

In classic satellite range scheduling (SRS) is the objective to increase the utilization of the ground network, this holds not necessarily for low-cost ground station networks, here the objective is rather to extend the contact time with the satellites by sharing resources. The participants of these ground station networks are faced with the following problem: Very often the ground stations were built up in the scope of small satellite projects, but as these satellites are normally launched

to low Earth orbit (LEO), the available communication time is restricted to a few minutes a day. By sharing the ground stations between two institutions, the contact time with the satellites can be doubled on both sides. That is an important difference to traditional ground station networks, which are sharing the ground stations with commercial interest, increasing the utilization is therefore necessary for economic reasons. Major emphasis in low-cost ground station networks is placed on flexibility, i.e. a flexible scheduling process is desired. Traditional ground station networks on the other side have to create schedules for long time horizons.

First approaches tailored for the special particularities of low-cost ground station networks were proposed recently: These consider also the case that several ground stations are used to track a single satellite. An approach for scheduling the requests in the GENSO network was presented from Preindl et al. [18], it is based on the booking request structure used in the GENSO network. A similar approach dedicated to flexible scheduling in highly distributed ground station networks is discussed in Ref. [19]. Main focus lays on the ability to create schedules for short time horizons to cope with the requirements of flexibility in low-cost ground station networks.

11.2.4 Coordination and Administration of Ground Station Networks

Any ground station network consists of a huge variety of components, devices and facilities, which are networked together to a complex topology. Both types of ground station networks, traditional and low-cost approach, have to place a considerable amount of effort to control and administrate the overall network (control is in this context related to the operations of the ground station network itself, not a satellite which is controlled resp. operated from the ground station network). Basically there are two choices how the administration can be organized: A centralized administration from a single organization or a distributed solution where several entities coordinate and manage the system in a distributed or decentralized manner.

The centralized approach as well as the decentralized approach has benefits and drawbacks. While central services or functions are easier to implement, they might be affected by single point of failures (which has to be compensated with proper redundancy). On the other side are decentralized control strategies under certain conditions very sophisticated to implement, as appropriate ways of coordination and synchronization need to be provided. For example arises in a decentralized organization the question who has the permission to take decisions, assign resources or change the behavior of the network. Especially when the participants share their resources on a voluntary basis, a centralized coordination can be difficult to realize. The advantage of decentralized control is the possibility of graceful degradation and failsafe capability in the case of failures.

Traditional ground station networks are mainly centrally organized, i.e. there is typically one controlling organization on top, for example ESOC in the case of the ESA ESTRACK system, which is able to manage or control the complete network. For example is maintenance or migration to new software or hardware managed centrally by ESOC [4]. But also the ESTRACK system has interfaces to other ground station networks, which support the ESTRACK core network with stations from other organizations. This implies corresponding interfaces for remote operation and introduces a certain degree of decentralization.

Low-cost ground networks are mainly decentralized organized. The individual stations are loosely coupled through the internet; there is no full access to the ground station hardware itself. Remote operation is possible through various software interfaces provided from projects like GENSO or GSN (see above). Nevertheless, there is no central entity which takes care about collective hardware updates or maintenance actions. Each station is to a large fraction self-reliant, the associated institution decides whether a station is running or idle. This requires a more abstract view on a ground station, as a stations should be "replaced" when it drops out with another station (this is in general possible due to the similar architecture of low-cost stations). This is a bit comparable with the routing process in the internet, where a router could at any time neglect its service, which can be compensated easily from the neighboring nodes by updating the routing tables. Similar concepts are foreseen for highly distributed ground networks, where any ground station can take over the signal reception when a communication window exists for the desired satellite. An approach in that direction was proposed from Cutler [20], who uses a virtualization concept for ground stations to simulate a similar network structure like observed in computer networks.

11.3 Future Developments

While the traditional satellite operation scheme assigns one ground station for operation to one satellite, new concepts using a network of ground stations to track a single satellite were brought up recently. Hence, the conventional one-to-one link from satellite to receiving station is replaced by a one-to-many link (one-to-many links are of course only possible for reception of satellite signals, for telecommand still one-to-one links need to be employed). Nevertheless, these redundant links are not only useful for backup purposes, they can be used as well to realize distributed measurements or link quality improvements. Two exemplary future applications taking advantage of that principle are shortly described in the following:

A proposed idea is to determine orbit elements with Doppler shift measurements from ground. The ground station network can be seen in this scenario as a sensor network. The determination of orbit elements from simultaneous measurements is state of the art; the challenging aspect is to use a highly distributed ground network to determine the orbit of several space vehicles flying in close orbits or formation.

Especially during the Launch and Early Orbit Phase (LEOP) can an orbit determination system based on the measurements in a ground station network like GENSO support operators. This would complement the orbit data issued from other organization like NORAD to achieve more accurate tracking.

A second application scenario is the usage of simultaneous downlinks in a ground network to detect transmission errors. This principle was initially proposed from Stolarski [21], the parallel received bit streams from a transmitter are compared with each other. When transmission errors are detected a majority voting is used to determine the correct bit. The algorithm was demonstrated with a transmitter on a weather balloon and three receiving stations on earth, it was proven that this mechanism can be used to decrease the bit error rate. A step further is proposed in Ref. [22], the described system synchronizes automatically data frames in low-cost ground station networks. Frames which were received simultaneously from several stations can after the synchronization process used to detect and correct transmission errors. In small satellite missions the primarily used communication protocol is AX.25, it does not support forward error correction, but it contains a frame checking sequence which can be used to check if the data correction was successful.

It can be expected that future distributed systems in the space and in the ground segment will complement each other in an integrated distributed efficient infrastructure. This way fault tolerant, robust systems with high efficiency are expected, taking advantage from the terrestrial trends and advancements towards distributed networked computing.

References

1. Fisher F, Mutz D, Estlin T, Paal L, Law E, Golshan N, Chien S (1999) The past, present, and future of ground station automation within the DSN. In: Proceedings of the 1999 I.E. aerospace conference, Aspen, CO, pp 315–324
2. Wertz J, Larson W (1999) Space mission analysis and design. Microcosm Press, Torrance
3. Wittmann K, Hanowski N (2008) Handbuch der Raumfahrttechnik, chapter Raumfahrtmissionen, pp 41–67, no 1. Hanser
4. Maldari P, Bobrinskiy N (2008) Cost efficient evolution of the ESA network in the space era. Space Oper Commun 5:10–18
5. Hsiao F-B, Liu H-P, Chen C-C (2000) The development of a low-cost amateur microsatellite ground station for space engineering education. Global J Eng Educ 4:83–88
6. Tuli T, Orr N, Zee R (2006) Low cost ground station design for nanosatellite missions. In: AMSAT-NA space symposium, San Francisco
7. Beech W, Nielsen D, Taylor, J. (1997) Ax.25 link access protocol for amateur packet radio, version 2.2
8. Muylaert J (2009) An international network of 50 double cubesats for multi-point, in-situ, long-duration measurements in the lower thermosphere (90–320 km) and for re-entry research. 7th CubeSat Developers Workshop, April 2010
9. Page H, Preindl B, Nikolaidis V (2008) Genso: the global educational network for satellite operations. In: International astronautical congress 2008, vol B4.3
10. Page H, Biraud M, Beavis P, Aguado F (2010) Genso: a report on the early operational phase. In: International astronautical congress 2010, Prague, Czech Republic, vol B4.3.8

11. Nakamura Y, Nakasuka S (2006) Ground station networks to improve operations efficiency of small satellites and its operation scheduling method. In: AIAA Aerospace Conference Proceedings
12. Leung J (2004) Handbook of scheduling. Chapman & Hall/CRC, Boca Raton
13. Smith D, Frank J, Jonsson A (2000) Bridging the gap between planning and scheduling. Knowl Eng Rev 15:47–83
14. Howe A, Whitley D, Barbulescu L, Watson L (2000) A study of air force satellite access scheduling. In: World automation kongress, Maui, Hawaii
15. Schalck M (1993) Automating satellite range scheduling. Graduate School of Engineering of the Air Force Institute of Technology, Ohio
16. Burrowbridge S (1999) Optimal allocation of satellite network resource. Master's thesis, Virginia Polytechnic Institute and State University
17. Barbulescu L, Whitley D, How A (2004c) Leap before you look: an effective strategy in an oversubscribed scheduling problem. In: Proceedings of National Conference on Artificial Intelligence (AAAI-04) , San Jose, CA, July 2004, pp 143–148
18. Preindl B, Seidl M, Mehnen L, Krinninger S, Stuglik S, Machnicki D (2010) A performance comparison of different satellite range scheduling algorithms for global ground station networks. In: International astronautical congress 2010, Prague, Czech Republic, vol B4.7.1
19. Schmidt M, Schilling K (2009) A scheduling system with redundant scheduling capabilities. In: International workshop for planning and scheduling in space, Pasadena, USA
20. Cutler J (2003) Ground station virtualization. In: The fifth international symposium on reducing the cost of spacecraft ground systems and operations, Pasadena
21. Stolarski M (2009) Distributed ground station system experimental theory confirmation. In: The European ground system architecture workshop (ESAW), Darmstadt, Germany, pp 1–12
22. Schmidt M, Schilling K (2010b) Ground station majority voting for communication improvement in ground station networks. In: SpaceOps, Huntsville

Part IV
Studies and Missions

Chapter 12
Overview of Distributed Missions

Maria Daniela Graziano

Abstract Former studies of missions to perform SAR interferometry and multistatic SAR applications are presented and detailed. Missions and studies to perform Earth observations by optical remote sensing are also reported.

12.1 Introduction

Distributed Space Systems (DSS) involve two or more space vehicles and a cooperative infrastructure to improve resolution and coverage achievable from scientific remote sensing platform [1]. One of the main technology challenges of DSS is the precise control of the relative positions of satellites, which separation distance affects the sensing performance. To this end, starting from 1965 a number of formation flying experiments have been conducted. The first one was performed when Gemini 6A stayed close to Gemini 7 for about 5 h [2]. The experiment foresaw that only one spacecraft was actively controlled while the other one stayed passive. In 1997, there was the first "autonomous" formation flying mission, the Japanese ETS-7. It was made up by two satellites, named Chaser and Target, developed to acquire the basic technologies of rendezvous docking and space robotics. Specifically, Target was placed 200 mm away from Chaser and the robotic arm was used to hold it in place and three rendezvous and docking operations was successful completed [3]. Since then, a lot of formation flying missions or demonstration initiatives have been developed.

There is no unique way to classify them but a quite large accepted rule foresees the difference between trailing and cluster formations. In cluster formation, the satellites don't follow the same path but their orbits are chosen to the end to keep

M.D. Graziano (✉)
Department of Aerospace and Mechanical Engineering, Second University of Naples, Via Roma 29, 81031 Aversa (CE), Italy
e-mail: mariadaniela.graziano@unina2.it

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_12,

constant (or to change if required) the relative distances. The first complete study about cluster formation is Techsat-21 mission (see Sect. 12.3), started in 1997 when Air Force Research Laboratory presented the innovative mission concept based on reconfigurable formation of microsatellites. Typically, a cluster is composed of identical satellites, as happens for Techsat-21, CanX-4–5, 3CSat or TICS (see Sect. 12.5 and 12.7), but it has been also studied the possibility to have different payloads on the flight platforms, as foreseen for F6 or FASTRAC (see Sect. 12.5 and 12.7).

In trailing formation, the satellites have one or more payloads in common but the spacecrafts are designed to reach different goals. The satellites follow the same path to the end to improve the resolution and coverage performance. As an example, even if Terra aims to study Earth system and climate and Aqua is designed to measure water in all its forms, they have two common payloads, which data can be synergically used to improve resolution and coverage. Moreover, LandSat-7 crosses the equator 1 min before EO-1 (see Sect. 12.4), creating a historical image archive of Earth surface.

It is worthy of mention, the ESA Swarm [4] mission, composed of three identical spacecraft in near polar orbit with altitude ranging from 400 to 550 km. The mission cannot be defined a cluster or a trailing formation, because two spacecraft fly side-by-side at lower altitude measuring the East–west gradient of the magnetic field while the third one fly at higher altitude. The mission, which launch is planned for before mid-2012, will map the magnetic field of the Earth with unprecedented spatial and temporal accuracy.

In the following sections, more details are given for relevant missions and mission studies related to distributed systems and Earth observations. The following chapters will then complete the figure of the past and current activities.

12.2 TOPSAT

The Global Topography Mission, whose name was successively changed into TOPSAT [5, 6], was proposed as a result of Topographic Science Working Group [7] results about the usefulness of high resolution Digital Elevation Models (DEMs) for many scientific applications. SAR interferometry was identified as the key technique to attain DEMs on a global scale, whereas a multibeam laser altimeter would have been used to derive very accurate measurements over limited spots on the Earth surface. The mission was planned as a joint USA/Italy effort, with Italy providing the satellites and USA the payloads.

The main objective of the interferometric SAR on board TOPSAT would have consisted in global elevation measurements with horizontal resolution of 30 m, horizontal accuracy of 10 m, and vertical accuracy of 1–3 m over 90% of Earth land and ice sheets, in less than 6 months. In addition, high vertical accuracy (20 cm–1 m) data would have been collected over selected areas of Earth lands and ice sheets by laser altimetry.

Different solutions were considered to implement the SAR interferometer:

- Ka band system relying on a rigid truss
- L-band system achieving a vertical baseline thanks to a tether
- L-band system with horizontal baseline achieved with two free-flying satellites

The first option was abandoned due the difficulties in both the deployment of the interferometer and the control of its vibrations. The tethered option was afterwards discarded as a consequence of the partial failure of the TSS-1 mission. Therefore, the formation option was selected with the relative orbit reeling on ascending node and anomaly separations, which would nowadays called a pendulum configuration with no difference in orbital inclination. Baseline reconstruction accuracy was considered one of the most severe technical challenges. It was in fact estimated that 3 cm (with early 1990s technology) could have been achieved by post processing of GPS measurements.

The mission was cancelled due to budget restrictions and NASA replaced it with the Shuttle Radar Topography Mission on board the Space Shuttle.

12.3 Techsat-21

Techsat-21 [8, 9] is a flight experiment developed by Air Force Research Laboratory (AFRL) to assess the utility of the space–based sparse-array aperture formed by the satellite cluster.

The formation is made up by three identical microsatellites in LEO near circular orbit, flying in various configurations with relative distance ranges from 100 m to 5 km, equipped with an X-band 2-D electronically steered antenna.

The choice of microsatellites formation instead of one large satellite increase the launch flexibility, the system reliability and allows unlimited aperture size and geometry.

The primary radar mode is the Moving Target Indicator (MTI), working so that each satellite transmits a different signal and it is able to receive also the signals for the other two satellites [10]. To detect a moving target, which appears as a spark in the angle-Doppler domain, a narrow antenna main beam is required, implying a too large number of samples with a consequently unacceptable increase of processing time. To overcome this limitation, a spatial undersampling has been considered and an analysis of the clutter associated with the lost side lobe control has been developed, showing that a better target detectability has been possible using periodically thinned array. Therefore, the basic configuration for Techsat-21 consists of planar 19-element array, vertically displaced to provide the highest gain at the maximum range. The array geometry aims at creation of triangular grid, composed of one satellite in the center and six satellite distributed on the first ellipse and other 12 satellites on the largest ellipse. Each satellites makes one cycle along its ellipse during one Earth orbit, so that the triangular grid is maintained. The rotation of the array leads to a continuous change of the grating lobe structure.

The problem has been solved selecting a pulse repetition frequency (PRF) according to grating lobe structure to the end of maximizing the signal to clutter factor.

The processing, communications, control functions as well as payload functions are distributed among the three satellites, which are able to operate as a single "virtual" satellite. In order to have three microsatellites working as one "virtual" satellite, the formation configuration and evolution must be controlled. Specifically, the nominal separation between the satellites has to be maintained within 10% of their values. Numerical results have been shown meter level tracking error in the formation by using variable, continuous low level trusts. In addition, the spacecraft relative positions accuracy is achieved at millimeter level through the use of carrier phase differential GPS and the active control of the attitude is required to avoid errors in the control force directions. In such close formation, each satellite is able to execute also a fail-safe separation algorithms when the relative distance is below a defined threshold.

The launch was foreseen in 2004 but the program was at first restructured because it reveals "far more challenging than originally thought" [11] and then it was cancelled in 2003 due to continuous cost overruns.

12.4 A-Train and Morning Constellation

In the last two decades, two large satellite formations were put in orbit, "Morning constellation" and "Afternoon Train", mainly dedicated to atmospheric and clouds properties studies.

Figure 12.1 shows the time distance between the satellites, whereas Table 12.1 contains the launch date and equator crossing time.

"Morning Constellation" is made up by Landsat-7, EO-1, SAC-C and Terra, which have been working since March 2001 as a single mission, with satellites crossing the equator from 10 am to 10:30 am. The constellation provides: (1) quasi-simultaneous acquisition of images from the four satellite, (2) carrying out autonomous navigation experiment, (3) testing GPS satellite constellation for atmospheric studies, navigation and attitude and orbit control, (4) demonstrate various formation flying technologies. On October 2000, CONAE presented the Announcement of Opportunity to investigate possible applications of constellation in the fields of hydrology, desertification, urban planning, precision farming, forestry, ecology atmospheric and ionospheric studies and clouds properties, with more than 200 projects approved.

The first satellite of the formation is **Landsat-7** [12], which is part of NASA Landsat program, started in 1975 to collect spectral information from Earth's surface, creating historical archive. It is equipped with a multispectral sensor known as the Enhanced Thematic Mapper Plus (ETM+), which has a high-resolution panchromatic wide band. Landsat-7 orbit is slightly to the west of Earth Observing One (EO-1) satellite and it crosses the equator 1 min before EO-1 satellite.

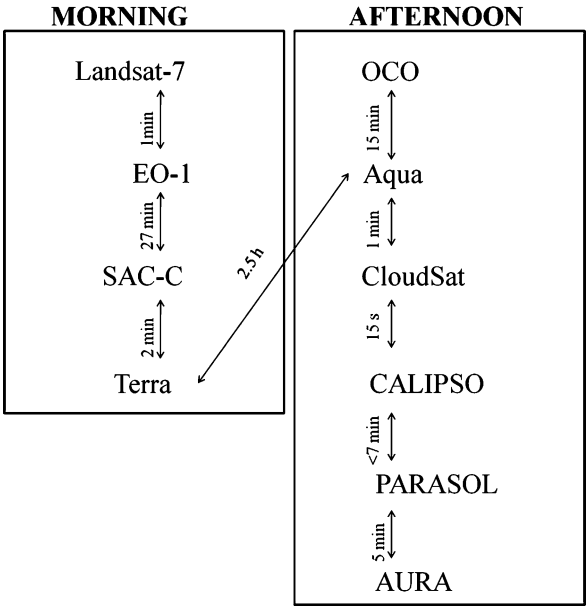


Fig. 12.1 Equator crossing time difference for "Morning" and "Afternoon" satellite formations

Table 12.1 Trailing formation satellites: launch date and equator crossing time

Satellite name	Launch date	Equator crossing time
Landsat-7	April 15, 1999	10:01
Terra	December 18, 1999	10:31
EO-1	November 20, 2000	10:02
SAC-C	November 20, 2000	10:29
Aqua	May 4, 2002	1:30
Aura	July 15, 2004	1:38
Parasol	December 18, 2004	1:33
CloudSat	April 28, 2006	1:31
CALIPSO	April 28, 2006	1:31:15
OCO	Planned for 2013	1:15

EO-1 [13] is equipped with three observing instrument: an Advanced Land Imager (ALI), the first high spatial resolution spectrometer (Hyperion) to look at the Earth and an high spectral resolution wedge imaging spectrometer (LAC) to estimate the content of water vapor in the atmosphere. Given the very short distance between EO-1 and Landsat-7, the 7.7 km-swath of Hyperion and 37 km-swath of ALI overlap the 185 km-swath of ETM + on the equator, allowing a multiple measurement on the same area with several instrument at the same time. EO-1 and Landsat-7 are not only in formation between them, but they work as a single mission with SAC-C and Terra.

SAC-C [14] is the Argentine Earth observing satellite, equipped with ten different instruments for evaluation of desertification processes and their evolution in time to identify and predict agriculture production, to monitor flood areas and to make studies in coastal and fluvial areas.

Terra [15] is the first NASA EOS satellite, called also EOS AM-1, to observe and measure the state of the Earth system and to monitor global environmental changes over time. The satellite is able to provide information about the physical and radiative properties of clouds, measurement of trace gases and vulcanology, improving the ability to detect human impacts on the Earth system and climate.

Terra moves south across the equator at 10:30–10:45 am and north across the equator at 10:30–10:45 pm, followed by Aqua satellite, crossing the equator at 1:30 pm northward and 1:30 am southward. Because of the respective late morning and early afternoon data collection times, Terra and Aqua were formerly named EOS AM and EOS PM, respectively. With Terra and Aqua both flying, the two instruments common to both spacecraft, MODIS and CERES, double the amount of potential daily data obtainable from either satellite alone. MODIS is a Moderate-Resolution Imaging Spectroradiometer operating from visible to thermal infrared to provide measurements of cloud and aerosol properties and their effects on the solar radiation budget and global climate whereas CERES, standing for Clouds and the Earth's Radiant Energy System, aims to study Earth's radiation budget and atmospheric radiation from the top of the atmosphere to the surface, operating in the ultraviolet through thermal infrared in three wide bands. Thus, the synergic utilization of this two instrument has a significant value in providing information on the diurnal cycle of some of the rapidly changing variables, such as clouds and aerosols, and in providing improved daily and longer term statistics on those variables.

Aqua [16] was launched by NASA in 2002. It has the earliest equator crossing time focuses on studies of water in the Earth/atmosphere system. Aqua carries a synergistic instrument payload that measures water in its gaseous, liquid and solid forms, plus atmospheric and surface temperature, land and ocean vegetation.

Aqua satellite flies in formation with other five satellites [16], AURA, PARASOL, CloudSat, CALIPSO and OCO, composing the so called Afternoon Train (A-train), because the equator crossing time is around the 1 pm.

AURA and PARASOL were launched in 2004. AURA [16] aims at study air quality, stratospheric ozone and climate change. Specifically, the satellite is able to observe global distributions of temperature by HIgh Resolution Dynamics Limb Sounder (HIRDLS) scanning the horizon as opposed to looking down on the surface. In addition, a Microwave Limb Sounder (MLS) can measure the concentrations of ozone-destroying chemicals throughout the stratosphere and upper troposphere as well as water vapor greenhouse gas in the upper atmosphere.

PARASOL [16] is a French mission able to make measurements of the total and polarized light in several wavelengths and at several different viewing angles. These data can be combined to study the state of the light reflected by the coupled surface-atmosphere system, characterize clouds and aerosols in the atmosphere, the influence of the human activity on the aerosols percentage in the atmosphere and its influence on the climate change. It is equipped with a POLarization and Directionally of the Earth's Reflectance (POLDER) able to sample three distinct polarizations.

Other two components of A-Train were launched in 2006, CloudSat and CALIPSO [17, 18].

CloudSat is a U.S./Canadian missions to study clouds and how they affect the Earth climate. It carries a powerful Cloud Profiling Radar (CPR) with 500-m vertical resolution in microwave region where the signal is not significatively attenuated by clouds, the radar is able to detect 90% of all ice cloud and 80% of water clouds. CloudSat was the first satellite to use a 94 GHz radar for atmospheric observation.

CALIPSO is a U.S./French collaboration to improve our understanding of the role of aerosols and clouds in the Earth's climate system. CALIPSO's main instrument is called Cloud Aerosol LIdar with Orthogonal Polarization (CALIOP) designed to measure the vertical profile of aerosols and clouds with 30-cm vertical resolution and 333-m horizontal resolution.

The last satellite included in A-Train is called OCO.

OCO-1 [19] satellite was lost during launch in 2009 and it will be replaced by OCO-2, planned for 2013. This satellite will make global, space-based observations of carbon dioxide CO_2, being equipped with three grating spectrometer to provide independent approaches to validate the data and ensure high accuracy.

Also **Glory** [20] satellite should have been included in A-Train formation, but the launch failed in 2011. It should have provided a greater understanding of the seasonal variability of the aerosol properties and understanding whether the temperature increase and the influence of the man-made source on the climate change.

The A-Train satellites are in close proximity to gain a better understanding of important parameters related to climate change and the formation wants to provide synergic measurements on the same areas. The difference in the resolution and the size of the footprint make difficult the synergic use of the data and the formation will need to be precisely aligned.

12.5 F6

F6 (Future, Fast, Flexible, Fractionated, Free-Flying) Program's main goal is an on-orbit demonstration of "fractionated" spacecraft [21–25]. The initiative, developed by the Defense Advanced Research Projects Agency (DARPA), aims to develop the necessary technology to build a single large satellite with free-flying satellites, wirelessly connected, showing the feasibility and the benefits of a system in which the module for power generation and storage, communication and payload are allocated on different flight platforms. Several benefits have been identified as driver factors for the program: (1) Flexibility, thanks to the possibility to add modules for specific purpose, (2) Reduced Risk, since the modules will be put in orbit with multiple launches, (3) Production Learning, based on module industrial production on large scale, (4) Adaptability, which implies the ability of the system to change configuration on demand, (5) Survivability, because the mission can continue even if one or more modules fail, (6) Payload Isolation, due to the distribution of functionalities on several platforms.

As starting points, program constraints have established that each spacecraft module shall be less than 300 kg wet mass and have a lifetime of 1 year after the launch of the last module, the first launch shall be within 4 years of the program start and the launch vehicle, manufactured in US, shall have performed at least one successful launch.

DARPA have identified several technological challenges: networking, to demonstrate that the spacecraft are able to change autonomously the network configuration, considering the ground element as a network node, wireless communications between each node, cluster flight, to demonstrate not only that the modules can displaced within a specific radius but also in a defensive cluster geometry, distributed computing to prove the flexibility and adaptability and wireless power transfer.

The enable technology are currently under development, solicited by multiple Broad Agency Announcement (BAA) and prize award. The program should culminate with an on-orbit demonstration 2013, to test the maintenance of the clusters, removal of a spacecraft from the cluster, the real-time resources sharing, the autonomous reconfiguration of the formation and the emergency maneuvers to debris avoidance.

Whereas F6 spacecraft is fractionated by function, DARPA explored also the potentialities of a system composed by identical satellites working together to achieve a common task. The program, called **TICS** (Tiny, Independent, Coordinating Spacecraft), foresaw swarms of small satellites (1–4 kg), launched by a common aircraft, to enable sensing and servicing missions. As an example, one of the TICS program application foresaw a protection formation to defense large satellites from direct attack. The program was cancelled in 2009.

12.6 Distributed Optical Payloads

The concept of distributed system has been applied also to optical missions.

The general goal of filled large aperture is simple in principle [26]: bringing antenna apertures form the order of magnitude of few to tens square meters to hundreds square meters or square kilometers and optical apertures from few square meters to hundreds of kilometers. The concept of such systems does not differ from standard system since there is a primary mirror (microwave or optical) and a feed (microwave) of a focal plane (optical).

The concept of large filled apertures was as an example exploited at JPL [26]. The innovative concept is that the system, named GEOTEL, is made up by six separated free flying optical modules, to allow for high resolution Earth observation. The primary mirror, also called Primary Electrostatic Membrane Reflector (25-m diameter), is a gossamer concept, thus it is based on a future technology. Severe constraints exist on real time relative motion knowledge (nanometer range and sub-microradian bearing) and control (sub-mm) and many challenges are in the

fields of centralized and decentralized sensing, estimation and control. Apart from the sunshade for the primary mirror, the other elements have a weight of 100 kg (primary mirror), 25 and 50 kg.

A similar concept is proposed by Wertz [27]. In this case, the primary mirror is made up by 96 2-m diameter close free-floating optical modules which all move on near GEO orbit. The primary mirror is protected from Sun by an orbiting Sunshade, which is 170 m wide and 70 m high, and which is made by kapton. It keeps the primary mirror at a temperature of 40 K. Wertz included in the system several external mirror element control units, which are used both for positioning knowledge and as control devices.

A widely explored application for optical distributed missions is the interferometry. The basic configuration foresees a certain number of spacecraft collectors, which reflect light towards the combiner satellite, able to combine the beams and generate an interference pattern.

Several studies have been conducted in the last decades.

Darwin [4] is an ESA study, consisting of four or five satellites with aim to search for Earth-like planets around other stars and analyze their atmospheres for chemical signatures of life. Three or four satellites work as collectors and reflect light to the central spacecraft, by using a technique called "nulling interferometry". This technique implies that light coming from a planet is delayed and then combined to "null" the starlight. Furthermore, Darwin system works also in "imaging mode" as a large telescope, with 100-m diameter, and the satellites flying in formation. In this case, a very high accuracy is required for relative positioning. The study is actually completed and no future developments are foreseen.

Terrestrial Planet Finder (TPF) [28] is a NASA project, to develop a system of small high sensitivity telescope, able to measure temperature, size and orbital parameters of Earth-like planets. The proposed configuration implies four large 3.5-m telescopes using Michelson interferometry principles to cancel starlight and detect only the light coming from the planet. The program has been postponed indefinitely.

The Stellar Imager (SI) [29] is another NASA project, to increase the knowledge of solar/stellar magnetic activity, and magnetic process, focusing on their role in the origin of life and on the evolution of the stars. The system is a UV-optical interferometer for 0.1 milli-arcsecond spectral imaging of stellar surface, based on 20–30 1-m primary mirrors distributed over a parabolic virtual surface. The focal length varies from 1 km with 100 m-diameter array to 10 km with 1,000-diameter array. Two combiners have been suggested to increase the system redundancy and also a "reference craft" to perform metrology on the formation. The goal is to enable the mission in the last 2020s.

Symbol-X [30] mission is a hard X-ray pointed telescope, based on two formation-flying platforms, the mirror and the detector, with a 44,000 km perigee altitude and 253,000 km apogee altitude to optimize global resources of both spacecrafts. The instrument, which foresees a 20–30 m focal length optics able to focus energy up to at least 80 keV, can observe a large variety of astrophysical sources, depending on the observation time. Clearly, the formation flying requires a

precise control of the relative positions. To this end a laser source is put onboard the "detector" spacecraft to measure the line of sight of the beam reflected by a corner cube. Furthermore, also the formation guidance is located on the "detector", whereas during not critical mission phases each satellite can communicate directly with the ground. Unfortunately, due to CNES budget restriction, the program was cancelled in 2009.

12.7 Technological Demonstration Missions

Several missions and initiatives have been proposed to demonstrate the enabling formation flying technologies. In the following, some of more representative demonstration programs are introduced, showing the main mission objectives.

3CSat (Three Corner Satellite) [31, 32] is a project developed by Arizona State University, University of Colorado and New Mexico State University. The three nanosatellites, called *Sparky*, *Ralphie* and *Petey*, are not attitude stabilized or controlled, so that the formation configuration degrades as mission evolves.

The mission main objectives are to take steroscopic clouds images, demonstrate formation flying and end-to-end command and data handling and it was the first to use cellular phone technology for satellites communication. Because of weight and space limitation of Boeing Delta IV Heavy rocket, only two of them were launched in 2004.

FASTRAC (Formation Autonomy Spacecraft with Thrust, Relnav, Attitude and Crosslink) [33, 34] project was developed by University of Texas, and it was sponsored by Air Force Research Laboratory (AFRL). The mission is composed of two 30 kg satellites: *Sara Lily* and *Emma*, launched in 2010 to investigate enabling technology for satellite formation. The only difference between the two satellites is a micro-discharge plasma thruster located in Sara Lily where Emma hosts a MicroAerospace Solution Inertial Measurement Unit (IMU) to measure the separation of the two satellites. The main technical mission objectives include the verification of satellite communication data exchange, the demonstration of autonomous thruster operation and of on-orbit real time GPS relative navigation. In addition, a distributed ground station network has been used to communicate with the satellites.

CanX-4 and CanX-5 [35, 36] nanosatellites will perform an autonomous formation flying demonstration mission, developed by Canadian Advanced Nanospace eXperiment (CanX). The two nanosatellites will fly with separation distance from 50 to 1,000 m to demonstrate precision formation flying technologies, which include autonomous achievement and maintenance of different formation configurations, high accuracy relative position measurement by DGPS, submeter position control, fuel efficient algorithms and intersatellites communication system.

During mission phases, one satellite, "deputy", will perform all necessary maneuvers to maintain the formation configuration, whereas the other satellite,

"chief", will adjust its attitude to be in the FOV of the same GPS satellites viewing the deputy. CanX-4 and CanX-5 deployment, planned for the end of 2012, will be performed by using the first XPOD Duo separation system, a "jack-in-the-box" container able to eject nanosatellites from any launch vehicles by means of opening-door release mechanism.

References

1. Leitner J (2002) Distributed space systems: mission concepts, systems engineering, and technology development, systems engineering seminar, April 2, 2002, NASA-GSFC (http://ses.gsfc.nasa.gov/ses_data_2002/020402_Leitner_DSS.pdf)
2. On-line source (2012) http://nssdc.gsfc.nasa.gov
3. On-line source (2012) http://robotics.jaxa.jp
4. On-line source (2012) http://www.esa.int
5. Zebker HA, Farr TG, Salazar RP, Dixon TH (1994) Mapping the world's topography using radar interferometry: the TOPSAT mission. Proc IEEE 82:1774–1786
6. D'Errico M, Moccia A, Vetrella S (1994) Attitude requirements of a twin satellite System for the global topography mission. 45th Congress of the international astronautical federation, Jerusalem, Israel, 9–14 Oct 1994, p 10
7. Topographic Science Working Group (1988) Topographic science working group report to the land processing branch, Earth science and application division, NASA headquarters. Lunar and Planetary Institute, Houston, p 64
8. Martin M, Kilberg S (2001) Techsat 21 and revolutionizing space missions using microsatellite. Fifteenth USU/AIAA, Logan, UT, Aug 2001
9. Burns R, McLaughlin CA, Leitner J et al (2000) TechSat 21: formation design, control, and simulation. In: IEEE Proceedings of aerospace conference, March 2000, Big Sky, MT, vol 7, pp 19–25
10. Steyskal S, Franchi M (2001) Pattern synthesis for TechSat21 – a distributed space-based radar system. IEEE aerospace conference, March 2001, Big Sky, MT, vol 2, pp 725–732
11. On-line source (2012) http://space.skyrocket.de/doc_sdat/techsat-21.htm
12. Bryant R, Moran MS, McElroy SA, Holifield C, Thome KJ (2003) Data continuity of earth observing 1 (EO-1) advanced land imager (ALI) and Landsat TM and ETM+. IEEE Trans Geosci Remote Sens 41(6):1204–1214
13. Ungar SG, Pearlman JS, Mendenhall JA, Reuter D (2003) Overview of the earth observing one (EO-1) mission. IEEE Trans Geosci Remote Sens 41(6):1149–1159
14. Colomb FR, Varotto CF (2003) SAC-C and the AM constellation: three years of achievements. Recent advances in space technologies, Proceedings of International Conference on Recent Advanced in Space Technologies, 2003, Nov 2003, Instanbul, Turkey, pp 20–22
15. On-line source (2012) http://terra.nasa.gov/
16. NASA facts, Formation Flying: The Afternoon "A-Train" Satellite Constellation, March 2003, FS-2003-1-053-GSFC
17. On-line source (2012) http://cloudsat.atmos.colostate.edu
18. Stephens GL (2003) The CloudSat mission, geoscience and remote sensing symposium. IGARSS '03. In: Proceedings of 2003 I.E. International
19. On-line source (2012) http://space.skyrocket.de/doc_sdat/oco.htm
20. On-line source (2012) http://glory.gsfc.nasa.gov
21. On-line source (2012) http://www.darpa.mil
22. On-line source (2012) http://defensesystems.com

23. O'Neill MG, Yue H, Nag S, Grogan P, de Weck OL (2010) Comparing and optimizing the DARPA system F6 program value-centric design methodologies. In: AIAA SPACE 2010 conference & exposition, Anaheim, CA, 30 Aug–2 Sep 2010
24. Brow O, Eremenko P (2006) Fractionated space architectures: a vision for responsive space. In: Fourth responsive space conference, Los Angeles, CA, 24–27 Apr 2006
25. On line source: Space and Security, 2001. http://www.spacesecurity.org
26. Mettler E, Breckenridge W, Quadrelli MB (2005) Large aperture space telescope in formation: modeling, metrology and control. J Astronaut Sci 53(4):391–412
27. Wertz JR (2004) High resolution structureless telescope. Final report of contract number NAS5-03110, Subcontract number 07605-003-016
28. On-line source (2012) http://www.terrestrial-planet-finder.com/
29. SI-The Stellar Imager, NASA Vision Mission Study Report, 15 Sep 2005. http://hires.gsfc.nasa.gov
30. Ferrando P et al (2006) Simbol-X: mission overview, space telescopes and instrumentation II: ultraviolet to gamma ray. In: Turner, Martin JL, Günther H. Proceedings of the SPIE, vol 6266, p 62660F
31. On-line. source (2012) http://space.skyrocket.de/doc_sdat/3csat.htm
32. Chien S, Engelhardt B, Knight R, Rabideau G, Sherwood R, Hansen E, Ortiviz A, Wilklow C, Wichman S (2001) Onboard autonomy on the three corner sat mission. In: International symposium on artificial intelligence, robotics and automation for space (i-SAIRAS 2001), Montreal, CA, June 2001
33. On-line source (2012) http://space.skyrocket.de/doc_sdat/fastrac-1.htm
34. On-line source (2012) http://fastrac.ae.utexas.edu
35. On line source (2012) http://www.utias-sfl.net
36. Orr NG, Eyer JK, Larouche BP, Zee RE (2007) Precision formation flight: the CanX-4 and CanX-5 dual nanosatellite mission. In: Proceedings of 21st annual AIAA/USU conference on small satellites, Logan, UT, Aug 2007

Chapter 13
TanDEM-X

G. Krieger, M. Zink, M. Bachmann, B. Bräutigam, H. Breit, H. Fiedler, T. Fritz, I. Hajnsek, J. Hueso Gonzalez, R. Kahle, R. König, B. Schättler, D. Schulze, D. Ulrich, M. Wermuth, B. Wessel, and A. Moreira

Abstract TanDEM-X (*T*erraSAR-X *a*dd-o*n* for *D*igital *E*levation *M*easurement) is a highly innovative Earth observation mission that opens a new era in remote sensing. TanDEM-X comprises two formation flying satellites, each equipped with a synthetic aperture radar (SAR) to map the Earth's surface with high spatial resolution. Together, the two satellites form a unique single-pass SAR

G. Krieger, Ph.D. (✉) • M. Zink, Ph.D. • M. Bachmann • B. Bräutigam • I. Hajnsek, Ph.D.
J.H. Gonzalez • D. Schulze • A. Moreira, Ph.D.
Microwaves and Radar Institute, German Aerospace Center (DLR),
Münchner Straße 20, 82234 Wessling, Germany
e-mail: Gerhard.Krieger@dlr.de; manfred.zink@dlr.de; markus.bachmann@dlr.de; benjamin.braeutigam@dlr.de; irena.hajnsek@dlr.de; jaime.hueso@dlr.de; daniel.schulze@dlr.de; alberto.moreira@dlr.de

H. Breit • T. Fritz, Ph.D. • B. Schättler
Remote Sensing Technology Institute, German Aerospace Center (DLR),
Münchner Straße 20, 82234 Wessling, Germany
e-mail: helko.breit@dlr.de; thomas.fritz@dlr.de; birgit.schaettler@dlr.de

H. Fiedler, Ph.D. • R. Kahle, Ph.D. • M. Wermuth, Ph.D.
German Space Operations Center (GSOC), German Aerospace Center (DLR),
Münchner Straße 20, 82234 Wessling, Germany
e-mail: hauke.fiedler@dlr.de; ralph.kahle@dlr.de; martin.wermuth@dlr.de

R. König
Dept 1, GFZ German Research Center for Geoscience,
c/o DLR Oberpfaffenhofen, 82234 Wessling, Germany
e-mail: rolf.koenig@gfz-potsdam.de

D. Ulrich
AED 71, Astrium GmbH, 88039 Friedrichshafen, Germany
e-mail: dieter.ulrich@astrium.eads.net

B. Wessel, Ph.D.
German Remote Sensing Data Center, German Aerospace Center (DLR),
Münchner Straße 20, 82234 Wessling, Germany
e-mail: birgit.wessel@dlr.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_13,

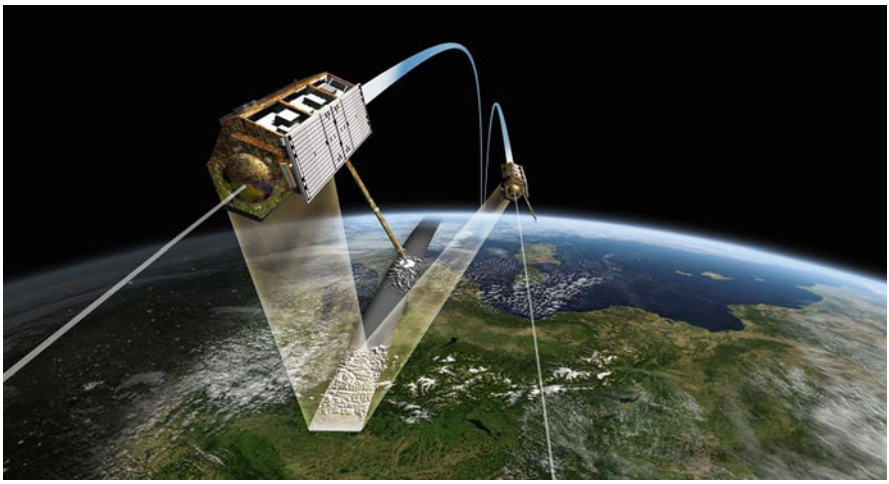


Fig. 13.1 TDX and TSX flying in close formation over Europe

interferometer, offering the opportunity for flexible baseline selection. Primary objective of TanDEM-X is the acquisition of a global digital elevation model (DEM) with unprecedented accuracy and resolution (12 m horizontal and 2 m vertical resolution). Besides the primary mission goal, several secondary objectives based on along-track interferometry and new bistatic SAR techniques have been defined, representing a further important asset of the mission. TanDEM-X was successfully launched in June 2010 and started operational data acquisition in December 2010. This chapter outlines the TanDEM-X mission concept and its implementation, summarizes the main data processing and calibration steps, and provides an overview of the actual performance and mission status. Furthermore, results from several scientific experiments are presented, showing the great potential of future formation flying interferometric SAR missions to serve a wide spectrum of novel applications.

13.1 Introduction

TanDEM-X (*T*erraSAR-X *a*dd-o*n* for *D*igital *E*levation *M*easurement) opens a new era in spaceborne radar remote sensing. A single-pass SAR interferometer with adjustable baselines in cross and along-track directions is formed by adding a second, almost identical spacecraft (TDX) to TerraSAR-X (TSX) and flying the two satellites in a closely controlled formation (see Fig. 13.1). Three years after TSX, TDX was launched in June 2010 and since December 2010 the two satellites are orbiting at typical cross-track baselines between 200 and 400 m acquiring data for a global Digital Elevation Model (DEM) with 2 m relative height accuracy at a 12 m horizontal posting. Beyond that primary mission objective, TanDEM-X provides a configurable SAR interferometry test bed for demonstrating new SAR techniques and applications.

Table 13.1 Comparison of DTED-2 and TanDEM-X DEM specifications

Requirement	Specification	DTED-2	**TanDEM-X**
Relative vertical accuracy	90% linear point-to-point error over a $1^\circ \times 1^\circ$ cell	12 m (slope $< 20\%$) 15 m (slope $> 20\%$)	2 m (slope $< 20\%$) 4 m (slope $> 20\%$)
Absolute vertical accuracy	90% linear error	18 m	10 m
Relative horizontal accuracy	90% circular error	15 m	3 m
Absolute horizontal accuracy	90% circular error	23 m	10 m
Spatial resolution	Independent pixels	30 m (1 arc sec @ equator)	12 m (0.4 arc sec @ equator)

Digital elevation models (DEMs) are of fundamental importance for a broad range of commercial and scientific applications [1–3]. For example, many geoscience areas like hydrology, glaciology, forestry, and geology require precise and up-to-date information about the Earth's surface and its topography. Digital maps are also a prerequisite for reliable navigation, and improvements in their accuracy need to keep step with the advances in global positioning systems, like GPS and Galileo. In principle, DEMs can be derived from a variety of airborne and spaceborne sensors [4, 5]. However, the resulting mosaic of data from different sources with a multitude of horizontal and vertical data, accuracies, formats, map projections, time differences and resolutions is hardly a uniform and reliable data set. The Shuttle Radar Topography Mission (SRTM, [6–8]) had hence the challenging goal to meet the requirements for a homogeneous and reliable DEM fulfilling the DTED-2-specification. The coverage of this DEM is, however, principally limited to a latitude range from 56°S to 60°N due to the inclined orbit of the space shuttle and its mapping geometry. Further restrictions apply to the X-band DEM with its wide gaps at lower latitudes and the C-band DEM where the data are available to the public only at an artificially impaired spatial resolution. A user survey among a wide range of scientists and potential customers has clearly shown that many applications require both an extended latitudinal coverage and an improved accuracy comparable to DEMs generated by high-resolution airborne SAR systems [3]. Table 13.1 compares the specifications of the DTED-2 and TanDEM-X DEM standards.

Beyond the generation of a global TanDEM-X DEM as the primary mission goal, local DEMs of even higher accuracy level (posting of 6 m and relative vertical accuracy of 0.8 m) and applications based on Along-Track Interferometry (ATI) like measurements of ocean currents are important secondary mission objectives. Along-track interferometry will also allow for innovative applications to be explored and can be performed by the so-called dual-receive antenna mode on each of the two satellites and/or by adjusting the along-track distance between TSX and TDX to the desired value. Combining both modes will provide a highly capable along-track interferometer with four phase centres. The different ATI modes will e.g. be used for improved detection, localization and ambiguity resolution in ground moving target indication and traffic monitoring applications. Furthermore,

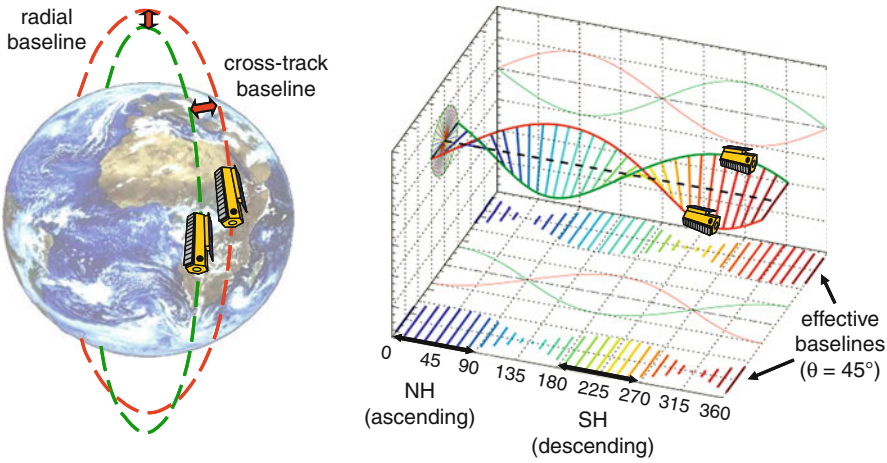


Fig. 13.2 Helix satellite formation for TanDEM-X. *Left*: illustration of the orbits. *Right*: cross-track and radial baselines as a function of the argument of latitude. The latitude positions correspond to one complete orbit (from 0° to 360°). The *dashed line* represents a virtual reference orbit chosen by the midpoint between the two satellites

TanDEM-X supports the demonstration and application of new SAR techniques, with focus on multistatic SAR, polarimetric SAR interferometry, digital beamforming and super resolution.

TanDEM-X has an ambitious time schedule to reach the main mission goal. After the commissioning phase, the first 2 years are dedicated to the global DEM acquisitions, followed by 6 months of additional acquisitions to cover difficult terrain. Only 1 year later, 90% of the TanDEM-X DEM shall be available. The baseline geometry in these first years is optimized for DEM performance. If the baselines are suitable, a limited number of scientific acquisitions can be included already during this phase. After the DEM acquisitions even larger baselines can be adjusted for higher accuracy DEMs on local scales and for the exploration and demonstration of scientific experiments.

13.2 Mission Concept

13.2.1 Helix Satellite Formation

The acquisition of the TanDEM-X DEM requires the coordinated operation of two satellites flying in close formation. Several formation flying options have been investigated during the conceptual development phase [9–13], and the Helix satellite formation shown in Fig. 13.2 has finally been selected for operational DEM acquisition [1]. This formation geometry implies maximum out-of-plane (cross-track) orbit

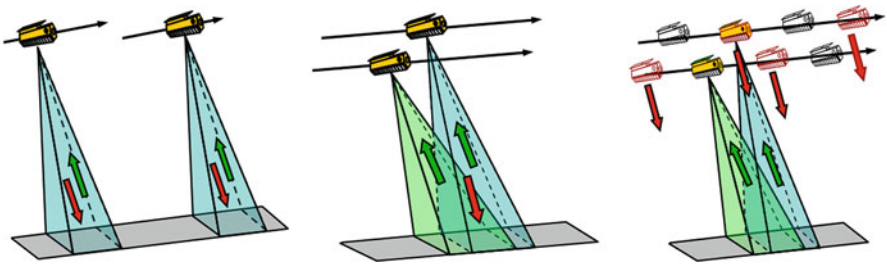


Fig. 13.3 Data acquisition modes for TanDEM-X: Pursuit monostatic mode (*left*), bistatic mode (*middle*), and alternating bistatic mode (*right*)

separation at the equator crossings by small ascending node differences and maximum radial separation at the poles by slightly different eccentricity vectors. This concept of relative eccentricity/inclination vector separation results in a helix-like relative movement of the satellites along the orbit and provides a maximum level of passive safety in case of vanishing along-track separation. The Helix formation enables therefore an inherently safe spacecraft operation without the necessity for autonomous control. It is furthermore possible to optimize the along-track displacement at predefined latitudes for different applications: cross-track interferometry aims at along-track baselines which are as short as possible to ensure an optimum overlap of the Doppler spectra and to avoid temporal decorrelation in vegetated areas. Other applications like along-track interferometry or super resolution require selectable along-track baselines in the range from hundred meters up to several kilometres. A fine tuning and maintenance of the satellite formation is performed via a dedicated cold-gas propulsion system on TDX (cf. Sect. 13.3).

The Helix formation enables a complete mapping of the Earth with a stable height of ambiguity by using a small number of formation settings [1]. Southern and northern latitudes can be mapped with the same formation by using ascending orbits for one and descending orbits for the other hemisphere, as illustrated in Fig. 13.2 on the right. To adjust the baselines, formation reconfiguration manoeuvres can be performed by TDX employing hydrazine propulsion. Smaller, but more frequent formation maintenance manoeuvres are performed by the TDX cold gas propulsion system. These manoeuvres are required to counterbalance the natural drift of the relative eccentricity vector, which would otherwise result in an additional change of the cross-track baselines (cf. Sect. 13.6).

13.2.2 Interferometric Modes

Interferometric data acquisition with the TanDEM-X satellite formation can be achieved in different cooperative modes: Examples are bistatic, monostatic, and alternating bistatic operation which are illustrated in Fig. 13.3. The three interferometric configurations may further be combined with different TSX and TDX SAR

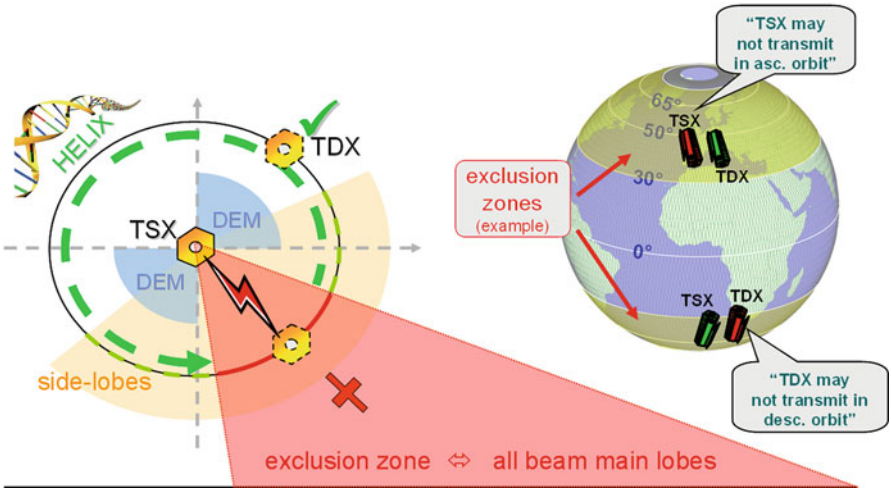


Fig. 13.4 TanDEM-X exclusion zones. Note that in the figure on the left TSX has been chosen as a fixed reference instead of the midpoint between TSX and TDX as in Fig. 13.2

imaging modes like Stripmap, ScanSAR, and Spotlight (Sect. 13.3). Operational DEM generation is performed using the bistatic Stripmap mode shown in the middle of Fig. 13.3. This mode uses either TSX or TDX as a transmitter to illuminate a common radar footprint on the Earth's surface. The scattered signal is then simultaneously recorded by both satellites. This synchronised data acquisition makes dual use of the available transmit power and is mandatory to avoid possible errors from temporal decorrelation and atmospheric disturbances.

13.2.3 Exclusion Zones

For DEM generation, TanDEM-X combines one monostatic and one bistatic radar image in a joint SAR interferogram. To ensure a sufficient overlap of the Doppler spectra, this requires a short along-track distance of typically less than 1 km between the two satellites. The radial and cross-track baselines depend on the argument of latitude and vary between zero and a few hundred meters. As a result, there is the danger that one satellite illuminates the other satellite by its transmitted radar pulses. This could cause interference, or, in the worst case, damage of sensitive electronic equipment. To avoid this risk, the transmission of radar signals has to be suppressed for one satellite at specific arguments of latitude, which are known as exclusion zones (cf. Fig. 13.4).

TanDEM-X ensures exclusion zone compliance by a double fail-safe approach including both a check on ground before command uploading and an additional real-time check on the satellite which suppresses signal transmission within predefined latitude windows (Sect. 13.3).

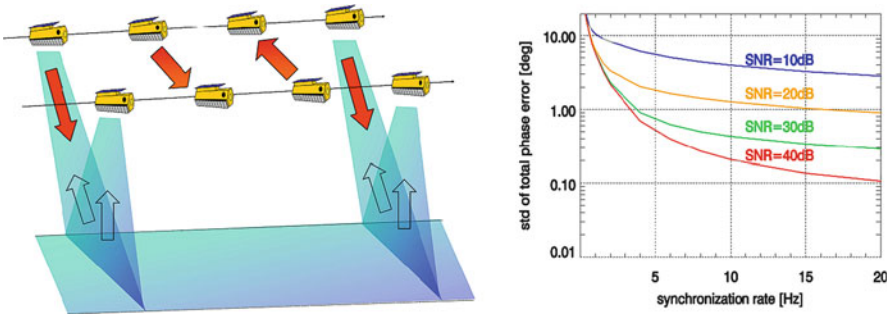


Fig. 13.5 Synchronisation of TanDEM-X satellites by exchange or radar pulses (*left*) and predicted performance (*right*). The performance is shown in terms of the standard deviation of the total synchronisation link phase error as a function of the synchronisation frequency with signal-to-noise ratio as a parameter

13.2.4 Radar Synchronisation

A peculiarity of the bistatic data acquisition is the use of independent oscillators for modulation and demodulation of the radar pulses. Any deviation between the two oscillators causes a residual modulation of the recorded radar echoes. The impact of oscillator phase noise in bistatic SAR has been analyzed in Ref. [14] where it is shown that oscillator noise may cause significant errors in both the interferometric phase and SAR focusing. The stringent requirements for interferometric phase stability in the bistatic mode will hence require an appropriate relative phase referencing between the two SAR instruments or an operation in the alternating bistatic mode. For TanDEM-X, a dedicated inter-satellite X-band synchronisation link has been established via mutual exchange of radar pulses between the two satellites (Sect. 13.3). For this, the nominal bistatic SAR data acquisition is interrupted for some transmit pulses, and the radar signal is redirected from the main SAR antenna to one of six dedicated synchronisation horn antennas mounted on each spacecraft. The pulse is then recorded by the other satellite which in turn transmits a synchronisation pulse (cf. Fig. 13.5, left). This technique allows for mutual phase referencing without exact knowledge of the actual distance between the satellites. On ground, a correction signal can then be derived from the recorded synchronisation pulses. This compensates the oscillator induced phase errors in the bistatic SAR signal. The performance of such a synchronisation link has been investigated in Ref. [15]. The right hand side of Fig. 13.5 shows the predicted standard deviation of the residual phase errors after synchronisation as a function of the repetition frequency of the synchronisation events for different signal-to-noise ratios (SNR) of the bidirectional link. The actual SNR varies with the distance between the satellites as well as with their relative attitude. For the typical DEM

data acquisition mode with baselines below 1 km, the SNR will be in the order of 30–40 dB, and it becomes clear that a phase error below 1° can be achieved for synchronisation frequencies above 5 Hz.

13.2.5 Interferometric Performance and Data Acquisition Plan

Radar interferometry is based on the evaluation of the phase difference between two coherent radar signals acquired from slightly different spatial and/or temporal positions. By this, TanDEM-X is able to measure the range difference from the two satellites to a given scatterer with millimetric accuracy. The height of the scatterer is inferred from this range difference by geometric triangulation. The sensitivity of the phase-to-height scaling depends on the distance between the two satellites, where a larger baseline increases the sensitivity of the radar interferometer to small height variations. However, the conversion from phase to range and hence the conversion of phase differences to height is not unique, since the range difference measurement via phases is ambiguous with the wavelength. Radar interferometry expresses this ambiguity by the so called height of ambiguity

$$h_{amb} = \frac{\lambda r_0 \sin(\theta_i)}{B_\perp} \tag{13.1}$$

where λ is the wavelength, r_0 the slant range from the satellites to the scatterer under consideration, θ_i the local incident angle of the electromagnetic wave, and $B_\perp$ is the perpendicular baseline. The value of $B_\perp$ can be approximately obtained by projecting the vector connecting both satellites onto a plane normal to the master satellite's orbit and then again onto a plane perpendicular to the line of sight.

Figure 13.6 shows the predicted height accuracy as a function of ground range position and the height of ambiguity (cf. [1] for details). It becomes clear that a lower height of ambiguity (i.e. larger baseline $B_\perp$) improves the height accuracy. However, a lower height of ambiguity also increases the difficulties in selecting the correct ambiguity interval during DEM generation (phase unwrapping). To minimize such problems and to ensure a homogeneous performance, TanDEM-X combines acquisitions with different heights of ambiguity. This requires in turn frequent adjustments of the Helix formation parameters which are selected according to an optimized global data acquisition plan (Sect. 13.5). Important constraints in this challenging optimization procedure are besides the interferometric performance the available amount of fuel and thruster cycles, limitations in the onboard storage and downlink capacity in combination with the finite time period for global DEM acquisition, as well as power and thermal constraints [16–18]. Further challenges arise from the interleaved usage of both satellites for the undisturbed continuation of the TerraSAR-X mission.

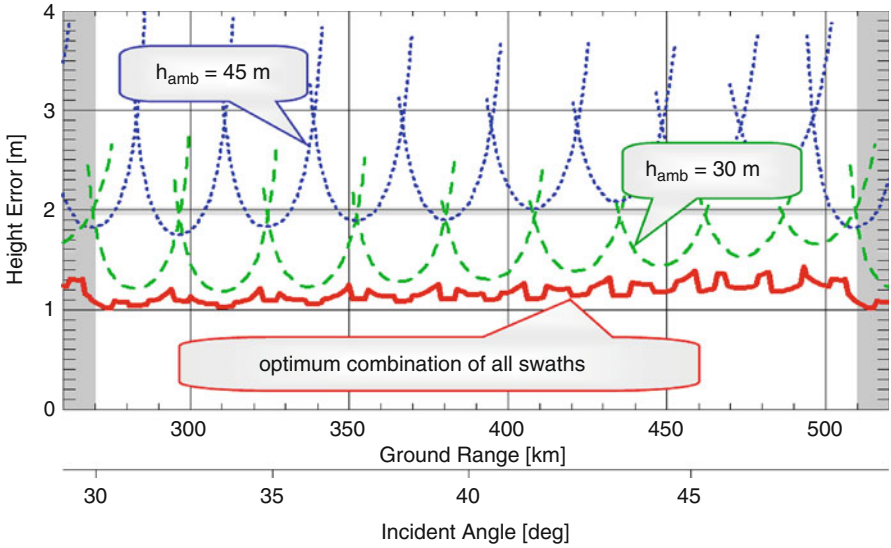


Fig. 13.6 Predicted relative height accuracy for a height of ambiguity of 45 m (*dotted*) and 30 m (*dashed*). The lower *solid curve* shows the error resulting from the combination of multiple swaths. All errors are point-to-point height errors for a 90% confidence interval assuming scattering from bare rock and soil surfaces. First results from the TanDEM-X commissioning phase show a good agreement with the performance model (cf. Sect. 13.9)

13.3 Space Segment

The TanDEM-X (TDX) satellite is essentially a rebuild of the TerraSAR-X (TSX) satellite [19]. The 5 m long satellite structure has a hexagonal cross-section. One side carries the 4.8 $\times$ 0.8 m body-mounted SAR antenna which points to the Earth on the right side of the satellite subtrack. The sun-synchronous dusk-dawn orbit allows the use of a solar array which is fixed on the sun-pointing side of the satellite. SAR measurement data are transmitted to ground stations using an X-band antenna mounted at the tip of a 3.3 m long boom (Fig. 13.7). In addition to the four 1-Newton engines based hydrazine propulsion system for orbit maintenance which is available on both satellites, a high-pressure Nitrogen cold gas system is accommodated on TDX, providing the necessary small impulses based on 40 mN nozzles to support the fine relative orbit control of TDX for the close formation flight.

The X-band (9.65 GHz) radar is based on an active antenna using an array of transmit-receive modules which provides rapid electronic beam steering and programmable antenna patterns (Fig. 13.8). Each module feeds a sub-array comprising two slotted-waveguides, one for vertical and the other for horizontal polarization. The waveguides are based on metalized carbon fibre reinforced plastic to give high inherent stability under all operating conditions.

Nominally, three basic modes of instrument operation are used:

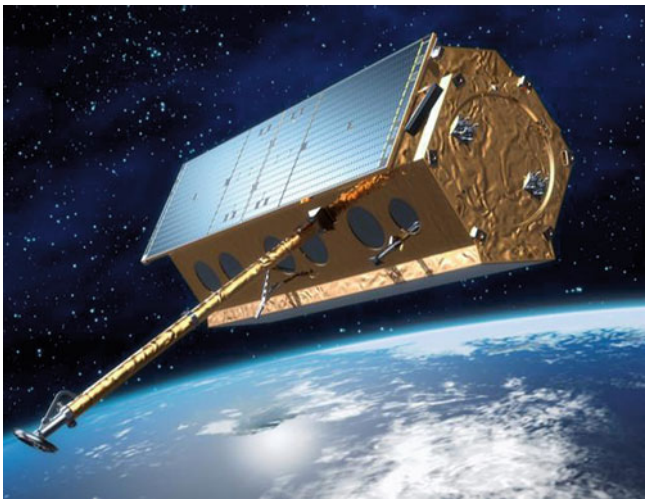


Fig. 13.7 Artist view of the TerraSAR-X satellite. Note the solar panel (*upper left*), the boom (*lower left*) with the X-band downlink antenna, and the X-band radar antenna (*lower right*)

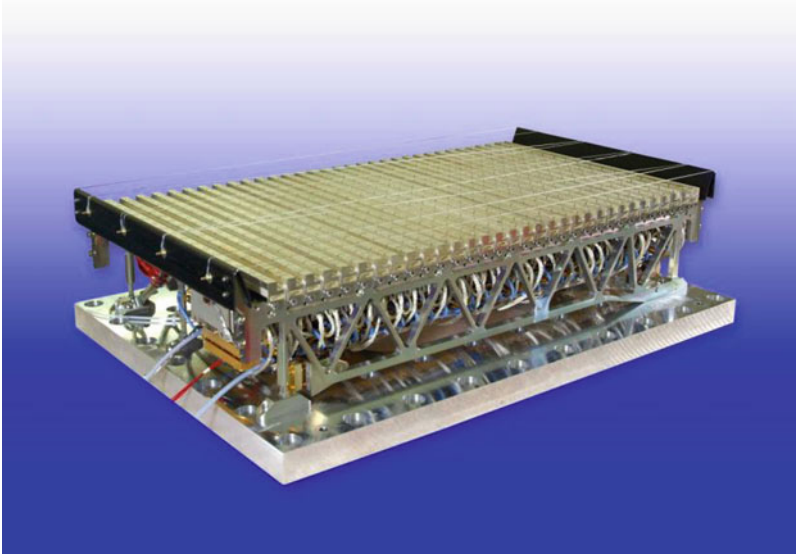


Fig. 13.8 TanDEM-X antenna panel with 32 slotted waveguide subarrays. Each subarray is fed by an individual transmit/receive (T/R) module. The full SAR antenna consists of 12 panels

- Stripmap is the standard mode of operation, providing a resolution of 3 m and a swath width of 30 km on ground.
- In Spotlight mode, azimuth (along-track) scanning of the radar beam allows to improve the resolution to 1 m for the measurement of scene 'patches' of about 10×10 km.
- In ScanSAR mode, elevation (cross-track) scanning of the radar beam allows to increase the swath width to 100 km with a resolution of better than 16 m.

The SAR transmit/receive signal routing to the antenna front-end and simultaneous use of the redundant Central Electronics allows separate radar reception for each along-track half of an antenna. This Dual Receive Antenna (DRA) mode enables fully polarimetric measurements and along-track interferometry. The architecture also provides a routing for calibration signals through the whole chain of hardware involved in transmission and reception with the exception of the radiating waveguides. On-ground processing of the calibration signals derives replicas for optimized pulse compression and amplitude correction factors enabling high measurement stability.

The TDX Solid State Mass Memory capacity is 768 Gbit which is doubled compared with TSX to support the collection of the enormous amount of DEM data. In preparation of the TanDEM-X mission the TSX radar antenna design was extended to allow exchange of so-called "Sync Pulses" to support coherent operation of the two SAR instruments during bi-static operation. Six synchronisation horn antennas on each satellite, fed by transmit-receive modules of the intermediate amplifiers, provide a quasi-omnidirectional coverage.

TDX and TSX carry the Tracking, Occultation and Ranging (TOR) equipment provided by the German Research Centre for Geosciences (GFZ) in Potsdam. TOR consists of the dual frequency Integrated GPS Occultation Receiver (IGOR) together with occultation antennas and a laser reflector. Ground processing (see Sects. 13.7 and 13.10) of the IGOR data from both satellites allows the accurate determination of the interferometric baseline as needed for the DEM generation.

For DEM data collection, formation flight is necessary with satellite separations down to 150 m. This infers risk of collision of the satellites and risk of mutual illumination by the main beams of the radar antennas. A number of safety mechanisms have been introduced in the satellites design to safeguard against these risks:

- Beyond the original hydrazine thrusters based safe mode an additional safe mode based on use of the magnet-torquers is introduced on both satellites. This avoids change in altitude in case of a safe mode drop which could occur infrequently due to single event upsets caused e.g. by radiation and which increases the risk of collision.
- DEM data collection is performed in bistatic operation where one radar transmits and both radars receive the radar signals. Radar transmission is allocated operationally to that satellite in the deterministic Helix formation having no risk to illuminate the partner satellite. Radar transmission by the other satellite is prevented within this part of the orbit by an onboard 'exclusion zone' logic using the actual orbit position determined on board. A similar logic is used in the Ground Segment to avoid planning of such dangerous operations.

- Each SAR transmit-receive chain is diverted to form a communication link using a pair of synchronisation horns, one of six on each satellite. Detection of a poor receive signal level is used by a so-called sync warning logic to immediately suppress radar transmission since the cause could be an unexpected orbit position (e.g. caused by a safe mode drop or a failure in an orbit manoeuvre execution) of the partner satellite which could result in radar illumination.
- An Inter-Satellite Link Receiver & Decoder is accommodated as additional hardware on TDX. This allows TDX to 'listen' to the TSX low rate S-band telemetry. In case TSX is reporting to be in safe mode state, TDX suppresses its radar transmission and as such preventing the risk of radar illumination of TSX.

13.4 Ground Segment Architecture

The TanDEM-X ground segment is built on top of the existing TerraSAR-X ground segment [20] thus providing an integrated solution for two very different system tasks:

- The systematic acquisition and generation of a global DEM over several years and
- The provision of high-resolution single scene SAR products from various imaging and polarisation modes to commercial and science users on a daily basis.

Figure 13.9 provides an overview of the joint TerraSAR-X and TanDEM-X ground segment which consists of three major sub-segments:

– Mission Operations Segment (MOS),
– Instrument Operations and Calibration Segment (IOCS),
– Payload Ground Segment (PGS).

The challenging TanDEM-X mission concept as described in Sect. 13.2 required the introduction of several new and highly innovative elements in all three sub-segments [21]:

- A TanDEM-X Acquisition Planner was added to IOCS. This element provides the long-term acquisition plan for the global DEM and the TanDEM-X scientific data takes (cf. Sect. 13.5). To support the planning, a new functionality was added which systematically predicts and monitors the interferometric performance (cf. Sect. 13.9).
- The IOCS Radar Parameter and Instrument Command Generators were extended to provide satellite specific instrument command sets for bistatic and interferometric data takes. This includes the generation and monitoring of synchronisation pulses that are exchanged between the two satellites.
- The IOCS Calibration element was extended to calibrate and deal with the peculiarities of the bistatic and interferometric modes (cf. Sect. 13.10).
- The MOS Mission Planning System was enhanced to generate combined TerraSAR-X and TanDEM-X mission timelines using both satellites.

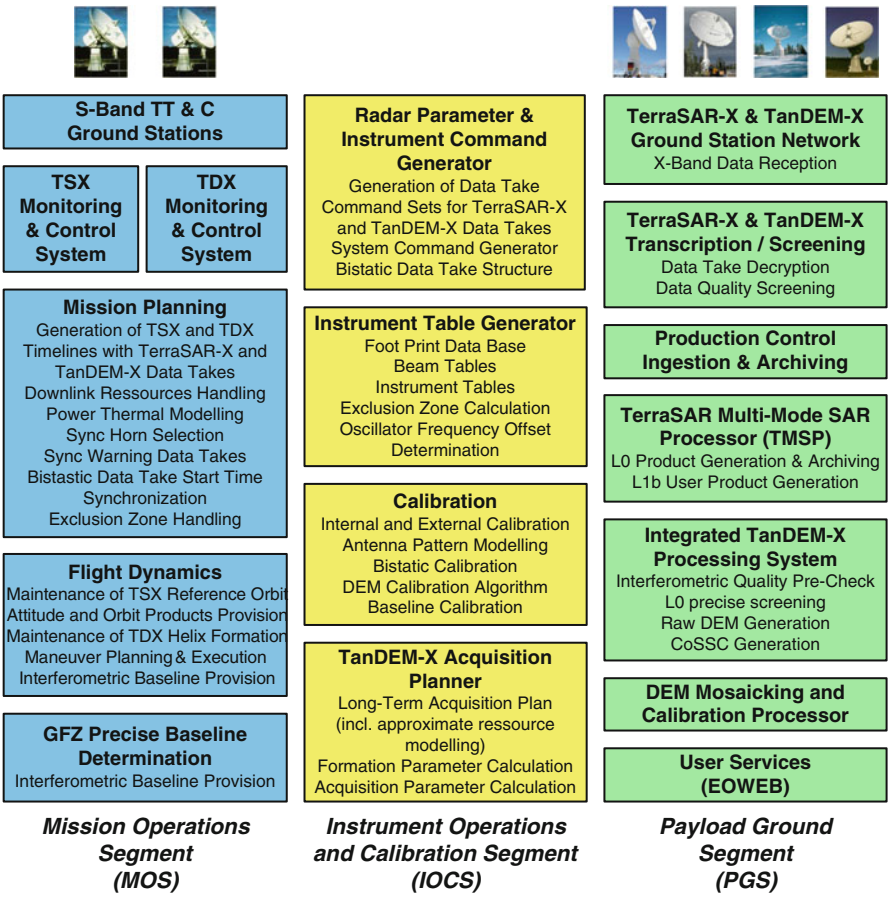


Fig. 13.9 Sub-segments and elements of the joint TerraSAR-X and TanDEM-X ground segment

- The MOS Flight Dynamics System was extended to support the TanDEM-X formation flight which poses new needs for spacecraft navigation and control (Sect. 13.6).
- A new element for precise baseline estimation has been added to MOS. The baseline products are provided by two independent entities (Sect. 13.7).
- A network of ground stations located at appropriate geographical locations (DLR's German Antarctica Receiving Station at O'Higgins, DLR's Inuvik Satellite Station in Canada, DLR's ERIS Chetumal Station in Mexico, and the partner ground station SSC ESRANGE Kiruna in Sweden) was added to downlink a TanDEM-X data volume of about 350 TByte in 3 years of operation.
- An automated chain for the systematic reception, archiving and processing of the TanDEM-X SAR acquisitions into individual "raw" DEMs and their final mosaicking into the global DEM product was added to PGS. For this, an Integrated TanDEM-X Processor (ITP) was developed to process the TanDEM-X SAR raw data into "raw" DEMs in an automated way (Sect. 13.8). The ITP

generated raw DEMs are then further processed by the DEM Mosaicking and Calibration Processing System MCP to the final TanDEM-X DEM product (Sect. 13.11).

Two classes of TanDEM-X products are offered to external users via DLR's EOWEB gateway: MCP generated DEM tiles and ITP generated so-called CoSSC (co-registered single-look slant-range) data products, the latter specifically in support of the secondary mission goals. Ordering a TanDEM-X product means requesting a copy of the systematically processed and archived data in form of a catalogue product. Commissioning of the challenging TanDEM-X mission on top of TerraSAR-X and thus the release of the joint TerraSAR-X/TanDEM-X ground segment for routine operation was done in parallel to the technical system verification. Since the end of 2010, the innovative ground segment is fully operational and successfully supports both missions.

13.5 Global DEM Acquisition Plan

13.5.1 Introduction to DEM Acquisition Planning

In order to generate a global DEM with the specification as described in Sect. 13.1 more than 10,000 interferometric image pairs need to be acquired. Each of these data takes has to fulfill its performance requirements. The TanDEM-X satellites allow interferometric Stripmap acquisitions with a swath width of around 30 km. Considering the 11-day repeat orbit this requires nine adjacent swaths in order to map the complete ground area at the equator. Further north and south the number of required adjacent swaths reduces since the ground distance from one orbit to the next gets smaller. Figure 13.10 shows the acquired area for the TanDEM-X global DEM until August 2011. The composition of the individual acquisitions with a typical data take length of 1,000–1,500 km is visible. This plot reveals in addition that the southern area is acquired earlier which is the effect of height of ambiguity asymmetry between ascending and descending orbits. This asymmetry is due to the non-vanishing along-track baseline at the equator in combination with Earth rotation.

13.5.2 Height of Ambiguity as Key Planning Parameter

The primary performance parameter for the acquisition planning is the Height of Ambiguity (HoA), see Sect. 13.2.5. The smaller the HoA the better is the relative height error but the more difficult is the phase unwrapping. In addition, smaller HoAs cause a loss of coherence over areas which are subject to volume decorrelation. Therefore, the selection of the optimal HoA for the TanDEM-X DEM acquisitions is a balancing act.

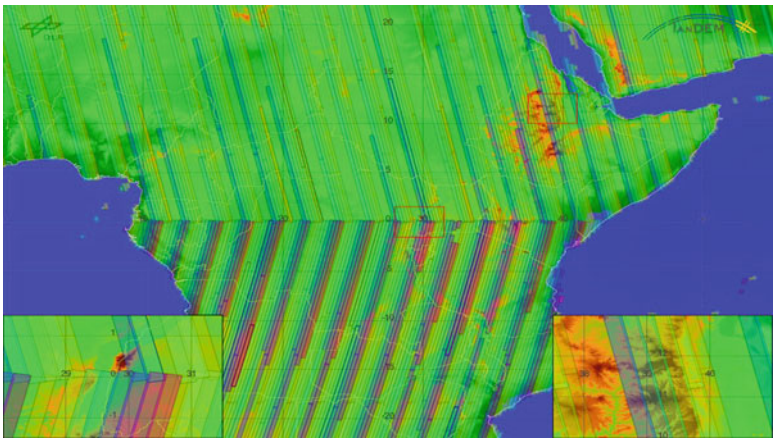


Fig. 13.10 TanDEM-X DEM acquisitions. The colour coded polygons show the DEM acquisitions until August 26, 2011. The colours indicate the different swaths. The *zoom* in the lower left corner shows the connection area between ascending and descending acquisitions at the equator. The *zoom* in lower right corner shows the overlap between adjacent acquisitions. In total, nine different swaths are required in each orbit to map the entire area at the equator

Since the overall mission time allows for at least two global acquisitions, it is foreseen to acquire each area with two different HoA values:

- A first acquisition has been performed in 2011 with a HoA between 45 and 60 m depending on the terrain type. This ensures a robust phase unwrapping.
- A second acquisition with a smaller HoA of around 35–45 m is planned to be finished by the end of 2012 in order to improve the relative height error.

In the planning of the data takes the acquisition geometry has to be considered to achieve the required HoA. The different swaths will be acquired using satellite look angles between 28° and 44° (corresponding to incident angles between 30° and 48°). Depending on the look angle the HoA changes since the effective baseline, the slant range and the incident angle are different (cf. (13.1)). In addition, the acquisition geometry changes constantly during the orbit (argument of latitude) as the satellites are flying in a Helix formation (see Sect. 13.2). This means that the resulting height of ambiguity depends on the used beam and the orbit position. Figure 13.11 shows the variation of the HoA for a typical TanDEM-X Helix formation.

13.5.3 Acquisition Planning

In order to generate the DEM acquisition plan the limited satellite resources have to be taken into account. The driving resource parameters are raw data downlink capacity, available onboard memory and the power/thermal constraints of the satellites.

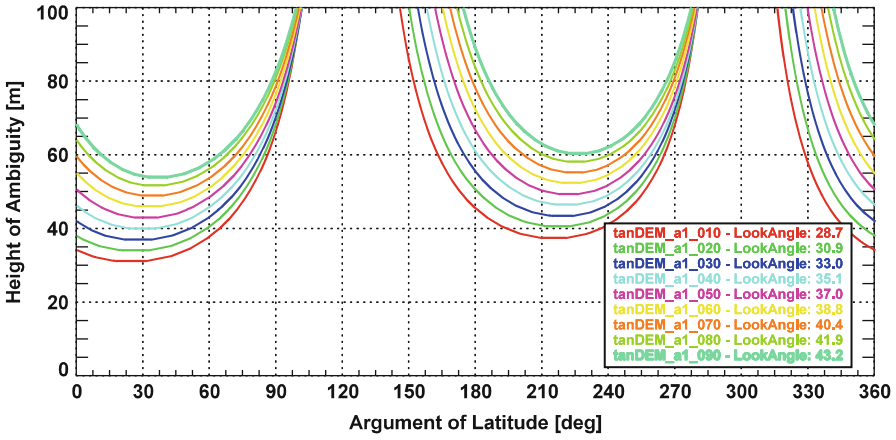


Fig. 13.11 Height of ambiguity as a function of look angle and orbit position (example for a formation with the following parameters: radial: 300 m, horizontal cross-track: 250 m, libration angle: 190°)

These were modelled and considered during the generation of the TanDEM-X DEM acquisition plan.

The generation of the DEM acquisition plan comprises the following steps:

- Step 1: **Calculation of orbit/swath combinations**: During this process the desired ground area (e.g. complete Earth surface) is transformed into a list of orbit positions and corresponding swaths which need to be acquired.
- Step 2: **Formation selection**: The formation parameters are derived by calculating the HoA for each possible acquisition and comparing it against the target HoA of this acquisition. This is performed for several different formations iteratively until an optimum formation is found. Figure 13.12 shows the height of ambiguity together with the formation parameters during the first global acquisition in 2011. It can be seen that while the formation changes the HoA remains approximately constant since the acquisitions are performed with different swaths and from different orbit positions. In October 2011, a very small formation was entered to allow the acquisition of difficult terrain which requires a higher HoA.
- Step 3: **Determination of exact acquisition times**: In the determination of the acquisition times data takes are sequentially selected. All possible acquisitions are evaluated w.r.t. the available resources using a point model. The point model considers performance aspects like the HoA as well as operational constraints like the maximum acquisition duration or a possibility for a later execution. Steps 2 and 3 are performed for each repeat orbit cycle separately. The result is the exact timeline of all acquisitions.
- Step 4: **Detailed acquisition parameter calculation**: With the information of the acquisition times and the flown Helix formation it is now possible to calculate the detailed acquisition parameters. This results in multiple sets of acquisition parameters like e.g. different pulse repetition frequencies and

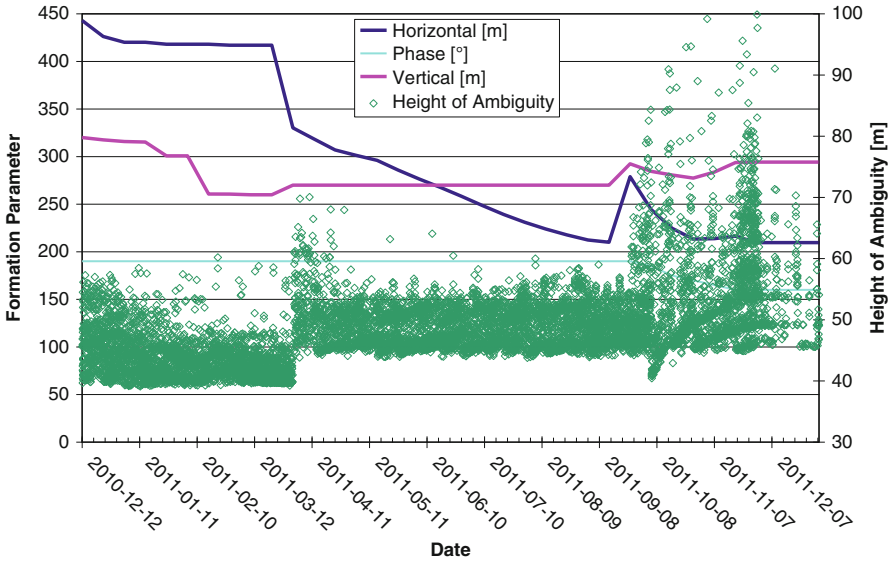


Fig. 13.12 Height of ambiguity and formation parameters (maximum horizontal and vertical separations, libration phase angle) of the first global DEM acquisition

different data compression settings, leading to different interferometric performance and satellite/downlink resources consumption.

- Step 5: **Height error prediction**: For each acquisition possibility of step 4 a performance prediction is evaluated using the performance model described in Sect. 13.2 as well as empirical data retrieved during the TanDEM-X commissioning phase.
- Step 6: **Height error optimization**: The last step of the DEM acquisition planning is the final selection of the acquisition considering the available satellite and downlink resources and the height error prediction. Figure 13.13 shows the height error reduction due to the optimization in step 6.

13.6 Formation Control

13.6.1 Orbit Control Concept

Since June 2007 the TSX satellite is operated in a sun-synchronous dusk-dawn orbit (97.44° inclination, 514 km altitude, frozen eccentricity) with a 167 orbits or 11-day repeat cycle. The TSX osculating orbit is maintained within a maximum absolute cross-track distance of 250 m from a target Earth-fixed reference trajectory, which comprises of exactly matching states at beginning and end of the 11-day cycle enabling highly repeatable data take conditions [22, 23]. Orbit

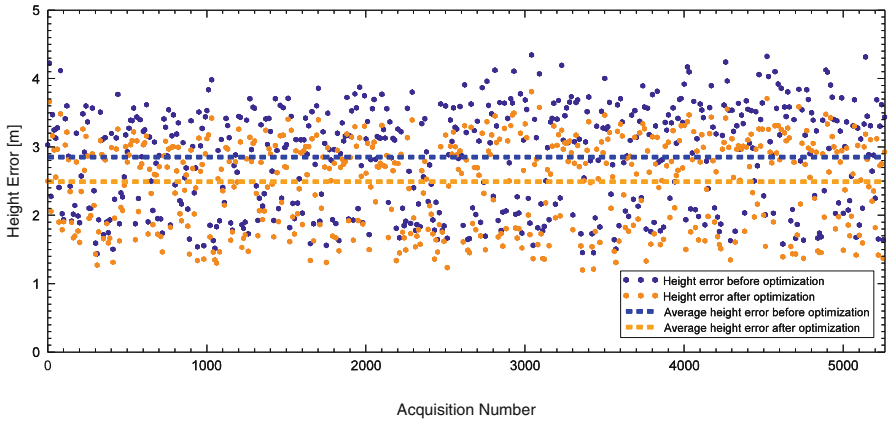


Fig. 13.13 Comparison of the predicted height errors before and after optimization for full exploitation of the available satellite and downlink resources for the second half of the first DEM acquisition year

manoeuvres to counteract luni-solar perturbations and to compensate atmospheric drag are performed 3–5 times per year (out-of-plane) and up to 3 times per week (in-plane) during periods of high solar activity, respectively. This ground-control strategy is foreseen for the entire TSX lifetime including the period of formation flying with TDX, which was launched in June 2010. Thus formation acquisition and maintenance manoeuvres are exclusively performed by the TDX satellite.

Because of the near-circular low-Earth orbit and the mutual distance on the order of 1 km, linearized equations can be used to describe the TDX-TSX relative motion. The model employed builds upon the well-known Clohessy-Wiltshire-Hill equations [24] but employs relative orbital elements as independent parameters and accounts for J2 perturbations. The use of relative eccentricity and inclination vector separation for formation design, the so-called Helix formation, enables safe proximity operations and a flexible adjustment of the interferometric baseline [13].

The starting point of TSX-TDX formation control operations is the TanDEM-X acquisition plan (TAP, cf. Sect. 13.5.3), which contains the timeline of formation geometries and corresponding planned bistatic SAR acquisitions. The target formation parameters are provided to Flight Dynamics (FD), where they are converted into relative orbital elements before applying them within the formation control process. All formation geometries flown within the first year of routine DEM acquisition are depicted in Fig. 13.14.

For the purpose of relative navigation within FD, the filtering of GPS navigation solution data is preferred to the processing of raw pseudorange and/or carrier phase data (cf. Sect. 13.7). In that way the amount of auxiliary information is significantly reduced leading to higher robustness with sufficient relative orbit determination accuracy of typically <0.5 m in cross-track (2D, RMS) and <1 m (RMS) in along-track direction [25].

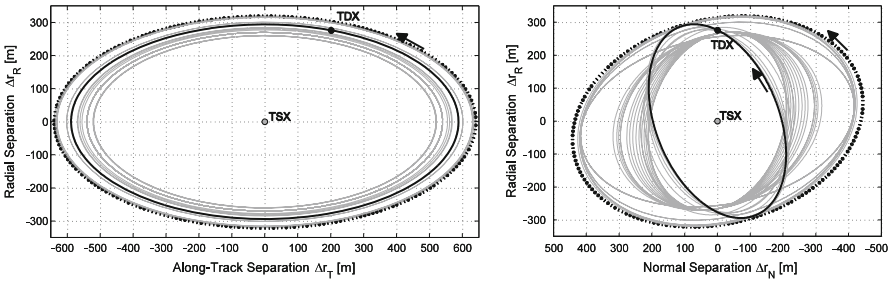


Fig. 13.14 Evolution of TDX-TSX relative motion in the Hill frame: radial-/along-track separation (*left*) and radial/normal separation (*right*). The formation geometries were flown during the period from December 2010 (*dotted ellipse*) to December 2011 (*bold black ellipse*). *Arrows* indicate the relative motion direction. The *filled black circle* indicates the TDX relative position at 90° argument of latitude

Table 13.2 Required and achieved formation control performance in different operation scenarios: mean/standard deviation [m] in radial, normal and along-track direction

	Radial	Normal	Along-Track
Requirement	0.0/20.0	0.0/20.0	0.0/200.0
Coarse ground-control with hydrazine, *3 days man. cycle*, Aug.–Sept. 2010	0.1/13.8	0.0/2.3	34.3/140.6
Routine ground-control with cold-gas, *24 h man. cycle*, since Oct. 2010	0.0/5.2	0.0/0.9	−0.2/25.4
First TAFF test campaign with cold-gas, *8 h man. cycle*, Mar. 2011	0/2.8	(0.0/2.9)	−4.7/13.9

Note: TAFF is not able to perform out-of-plane control, so only radial and along-track accuracies reflect the TAFF performance

In order to meet the required formation control accuracy (Table 13.2) TDX must both replicate the TSX orbit keeping manoeuvres and compensate the natural deviation of the relative eccentricity and inclination vectors. While the cm/s-level absolute orbit maintenance is performed with four 1-Newton hydrazine thrusters, two pairs of 40 mN cold gas thrusters are used exclusively onboard TDX for in-plane formation control including counterbalancing of along-track drifts, which result from differential hydrazine manoeuvre execution errors. The less frequent but larger out-of-plane manoeuvres are performed with hydrazine propellant. Precautions are implemented within FD planning process to ensure that the failure of any TSX or TDX manoeuvres cannot result in a dangerous relative motion [26].

13.6.2 Formation Control Performance

After acquisition of the 20 km wide formation in July 2010 (cf. [27] for details on the acquisition process) the commissioning of the FD formation maintenance function was successfully performed. Because of the achieved excellent control

accuracy and in order to safe precious cold-gas for a likely mission extension, it was decided to make use of the TDX hydrazine propulsion system to control the 20 km along-track formation during the remaining time of the mono-static commissioning phase. The hereby achieved performance (Table 13.2) was fully sufficient for the purpose of calibrating the TDX radar instrument for mono-static operation and surprisingly accurate considering the fact that only two formation maintenance manoeuvre pairs were performed per week.

Shortly after acquiring the tight formation with zero mean along-track separation in October 2010 the ground-based formation control process became fully automated. Since then daily formation maintenance manoeuvre pairs with cold-gas propulsion have been performed. Although the control concept primarily aims on the maintenance of the relative eccentricity vector, the precise control of the along-track separation is fundamental for the purpose of SAR cross-track interferometry, too. The remarkable control performance achieved is summarized in Table 13.2. Because of the J2-induced drift of the relative eccentricity vector (i.e. tilt of the relative motion ellipse), a perfect alignment with the target relative argument of perigee is only achieved in the middle of a manoeuvre cycle. A better radial control performance can therefore only be achieved when further decreasing the manoeuvre cycle as foreseen with the TanDEM-X Autonomous Formation Flying System (TAFF, cf. [28] for more details), which makes use of nearly permanent availability of real-time relative navigation provided through an inter-satellite link. The results of the very first TAFF campaign performed in March 2011, where three formation maintenance manoeuvre pairs per day were autonomously planned and executed onboard TDX, impressively complete the relation between manoeuvre-cycle and achievable control accuracy (Table 13.2).

13.7 Precise Baseline Estimation

13.7.1 DGPS Approaches for Precise Relative 3-D Navigation

The methods for determining precise baselines have been explained in detail in Sect. 5.4.2 above. In the TanDEM-X ground segment three approaches have been implemented to operationally provide baseline products by the two institutes German Research Centre for Geosciences (GFZ) and the German Aerospace Center (DLR):

- The GFZ solution called GFZ based on the EPOS-OC software [29],
- The GFZ solution called GFZB based on the BERNESE software [30],
- The DLR solution called MOS based on the FRNS software [31].

By having three independent software solutions it is possible to control the quality of the individual solutions, to monitor the long-term stability of the biases, and to eventually merge the solutions into a single one for improved accuracy.

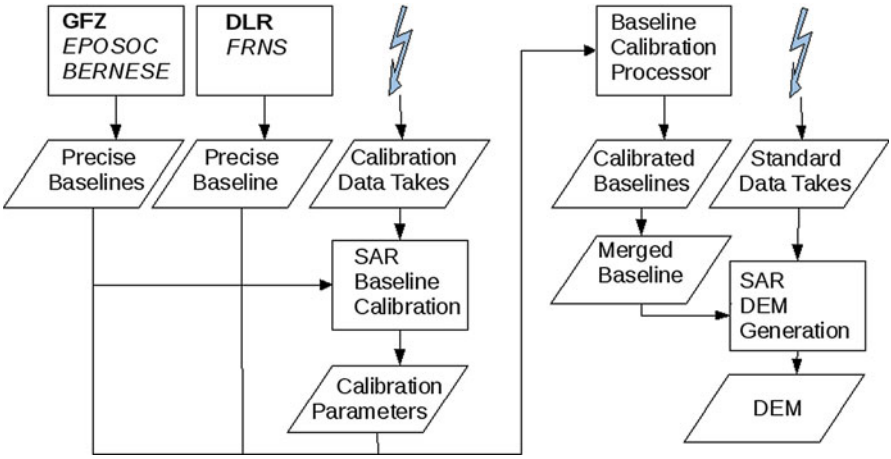


Fig. 13.15 Baseline generation, calibration, and merging chain

13.7.2 Baseline Combination

For the TanDEM-X mission three independent baseline solutions are generated as discussed above. All three solutions are subject to calibration and merging as depicted in Fig. 13.15.

Firstly, all different solutions are calibrated based on bias values inferred from dedicated SAR data takes over well-known test areas. From this the biases of the radial and normal component of the baseline vector can be determined and applied.

Secondly, outlier detection is performed by comparison of the three precise baselines. The outlying solution can then be corrected for or excluded from the next step.

Thirdly and finally, the different solutions are combined in order to minimize the stochastic errors inherent to all individual solutions. Because of some non-recoverable systematic content in the error spectrum, the error of the combined baseline cannot be fully decreased by the square root of 3 as expected for independent error propagation. Experience from the GRACE mission baseline combination allows to assess the improvement at approximately 20%.

The results of the comparisons of the MOS and GFZ baseline solutions for a period of 1 year are presented in Fig. 13.16 for the radial direction and in Fig. 13.17 for the normal direction. In both figures solely the differences of the solutions that went into the merging process are displayed. Each dot in the figures shows the mean, standard deviation or RMS value of the differences per product, i.e. for half a day or approximately 44,000 differences. The very stable radial bias of the solutions becomes obvious in the mean values of the differences over time which happen to be close to zero here. The bias in the normal direction is less stable than the radial one but still better than a millimetre in terms of variability of the mean values.

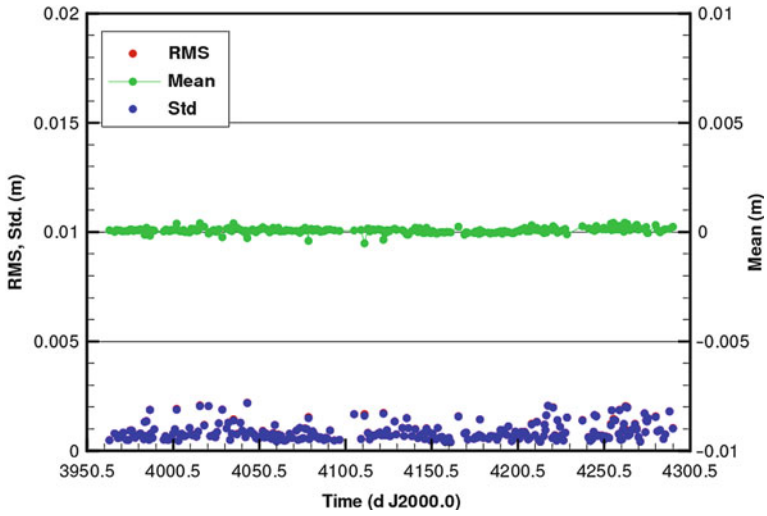


Fig. 13.16 Comparison of the MOS and GFZ baseline products in radial direction. Due to the low mean difference, the values for RMS and standard deviation are almost coincident

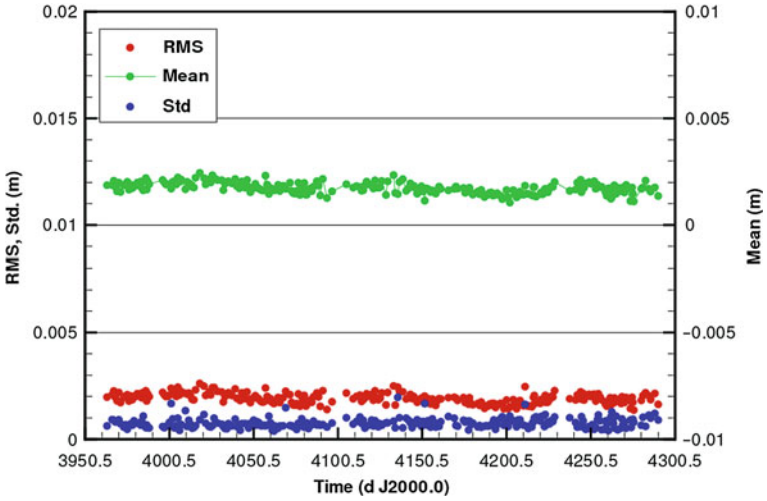


Fig. 13.17 Comparison of the MOS and GFZ baseline products in normal direction

The standard deviations depicted in Figs. 13.16 and 13.17 are measures of short term precision, the RMS values include biases as well.

Table 13.3 gives an overview of the stability of the biases of the different solutions and global standard deviations as a measure of precision. Concluding from Table 13.3 the merged baselines exhibit an accuracy of 1 mm RMS (if the biases are properly taken care of) and therefore comply with the mission requirements.

Table 13.3 Stability and precision of all baseline products

Solutions	Direction	Mean (mm)	Standard deviation (mm)
GFZ–MOS	Radial	0.1 ± 0.1	0.9
	Normal	1.7 ± 0.3	0.8
GFZB–MOS	Radial	-1.3 ± 0.4	0.8
	Normal	-1.0 ± 0.4	0.5

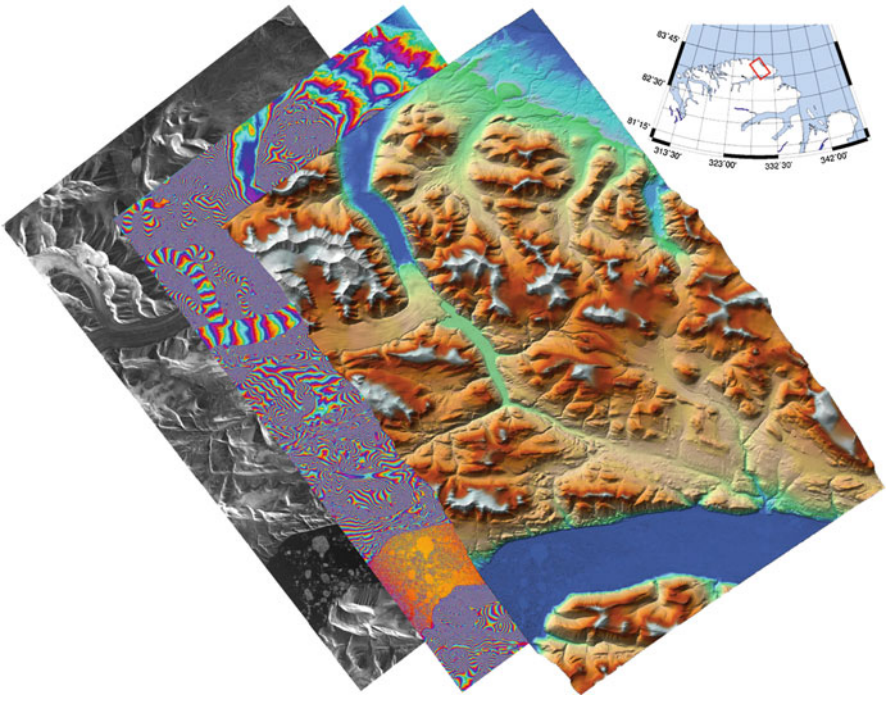


Fig. 13.18 SAR amplitude, interferogram and shaded colour coded raw DEM scene acquired over North Greenland

13.8 Data Processing

The Integrated TanDEM-X Processor (ITP) is the heart of the SAR data workflow in the payload data ground segment, responsible for screening, focusing, interferometric processing and finally the generation of "raw" DEMs from the bistatic raw data of the two instruments [32–34]. It also processes experimental acquisitions in different modes to user products for the scientific community. A raw DEM processing example is shown in Fig. 13.18 together with the interim SAR image and interferogram. The data have been acquired in bistatic mode in December 2010 and cover a 30 km by 50 km part of North Greenland. The terrain height ranges

from sea level up to 1,750 m. Each fringe of the flattened interferogram corresponds to a 58 m height change.

13.8.1 Processing Challenges

More than 500,000 km^2 of the Earth surface acquired at 3 m resolution have to be processed each day in order to be in line with the new acquisitions. Thus, the main requirements on the processing systems are to achieve high throughputs while maintaining the highest possible accuracy at a fully autonomous processing and self-calibration approach without any dependencies on external reference data or human interaction. The very demanding height accuracy targeted for the final DEM, which implicates the use of very small heights of ambiguity, causes as a draw back a significant amount of raw DEMs (10–15%) of complicated terrain to exhibit areas with so called phase unwrapping errors. These systematic local height discrepancies are solved for by an iterative processing loop using additional acquisitions of the same area in later mission phases. Thus depending on the accuracy and stability of the overall system calibration, the final DEMs provided by the ITP are expected to exhibit only very small residual height offsets in the range of cm to meters which are later measured and compensated by the mosaicking and calibration processor (Sect. 13.11) with the help of reference height information, operator interactions and consistency checks. For each processing run extensive quality analysis results are provided for system performance monitoring. The ITP also performs a quality screening of each downlinked DEM data take at the receiving stations to provide a fast feedback for acquisition planning, archival and system monitoring.

13.8.2 Principle of DEM Generation

The approach of the ITP to use the self-consistency of the data for pre-calibration of the raw DEMs can be explained best when recalling the basic principle of DEM generation: measuring the distances to a point on ground from two different and precisely known positions in space. In optical data processing, one can exploit the parallactic angles between such two stereo images. This is not possible in radar imaging, based on ranging from wave front signal travel times without sufficient angular resolution. Thus the triangulation is made from the different delays in the receive times of the radar echoes visible as shifts in the images. This method is called radargrammetry and is well established for very large baselines with corresponding macroscopic image shifts. If the cross-track baselines are small enough, the two instruments get the same spectral signal response with very small relative shifts and the more precise interferometric approach can be taken. Since SAR systems are coherent systems, the delays can be measured very accurately for each matching pixel from the phase differences between the two

instrument data sets. As already pointed out in Sect. 13.2.5 such interferometric phase differences are ambiguous and provide only relative range differences, cyclic in units of the wavelength.

13.8.3 Co-Registration of SAR Images

In order to correctly exploit the two different viewing geometries for height measurements, their differential effects on the two images have to be compensated. Only matching images yield the highest possible coherence—the minimal phase and height noises—from the inherent signal-to-noise ratio of the two SAR data sets. First of all, the images have to be resampled, correcting for the relative shifts and co-registered onto a common grid to overlay matching pairs of pixel in the master satellite's data and the slave's one which image the same ground point. Besides large scale offsets from the different orbits (mostly compensated during bistatic focusing), the precise delays are already related to the local topography due to the radargrammetric parallactic shifts. Thus they are unknown and have to be measured from the data itself. Unlike the relative phases, the residual shifts are much more difficult to determine since they are in the order of fractions of the resolution cell only. The ITP uses coherent cross correlation of equidistantly sampled patches of 32×32 samples in the two images to derive precise shifts in the order of 1/100th of the resolution (hence in the order of the wavelength) and interpolates in between. In cases of incoherent data, fallback solutions based on incoherent cross correlation and orbit geometry are implemented. But not only the spatial mismatch has to be compensated; different viewing geometries and azimuth beam pointings ("Doppler centroids") also cause mutual spectral shifts (in range and azimuth) of the complex data. Spectral components at the edge of the observed bandwidth visible from one point are not present in the other data set. In order not to introduce additional noise (decrease the coherence), the data have to be spectrally filtered for the matching spectral parts only.

13.8.4 Bistatic SAR Processing

Even with spatially and spectrally perfectly matching data, the main problem for interferometry is that both instruments need a common and stable absolute phase reference to correctly relate phase differences to distances or heights. In monostatic SAR, the own instrument oscillator is such a reference but in bistatic acquisitions, the transmitted and received signals are (de-)modulated with two different oscillators; yielding offsets and—worse even—strong variations and drifts of the phases with time. As introduced in Sect. 13.2.4, these relative offsets are measured by exchanging synchronisation pulses and compensated during processing as detailed below. However, the phases of the sync pulses are again

ambiguous by 2π and the derived differential phase offset is ambiguous by π. Hence the bistatic interferometric height measurements of TanDEM-X are determined for each pixel only within a height of ambiguity range with a global offset of half this ambiguity. This will be denoted as π-ambiguity in the following.

The prerequisite for interferometric and radargrammetric exploitation of a SAR image pair are first accurate signal corrections at raw data level and second highly accurate focusing with respect to phase and pixel time registration. The most important signal correction is the compensation of the phase and time drifts in the bistatic channel caused by the two independently drifting oscillators which provide the timing references for pixel registration as well as the carrier frequency signals for modulation and demodulation. The determination of the compensation function is based on differential phase and shift measurements of the bi-directionally exchanged synchronisation pulses. Timing and phase evaluation of both the SAR signal and the synchronisation signal, require precise determination of signal propagation delays taking into account the bi-dynamic geometric constellation of transmitter and receiver. The effective geometric distance $\Delta x(t)$ between the transmitter, and a given point on ground in case of the SAR echo, and the receiver is given by the recursive formulation

$$\Delta x(t) = \left\| \mathbf{x}_{RX}(t) - \mathbf{x}_{TX}\left(t - \frac{\Delta x(t)}{c}\right) \right\| \tag{13.2}$$

and is resolved by an iterative computing scheme involving interpolation of the orbit state vectors $\boldsymbol{x}(t)$. The differential phase compensation is applied as a contiguous correction to the bistatic channel SAR raw data. Differential time drift compensation is sufficiently corrected by piecewise constant shifts updated per SAR echo line. The moderate bistatic acquisition geometry of TanDEM-X permits the application of the TerraSAR-X chirp scaling focusing algorithm extended by a bistatic computation of an equivalent monostatic range history. Therefore, the true bistatic range history $R(t) = \frac{1}{2}(R_{Tx}(t - \tau(t)) + R_{Rx}(t))$ is numerically computed following the bi-dynamic formulation of (13.2) and the descriptive parameters $\tilde{R}_0$, $\tilde{V}^2$, and $\tilde{t}_0$ of the monostatic equivalent range history

$$\tilde{R}(t) = \sqrt{\tilde{R}_0{}^2 + \tilde{V}^2 \cdot (t - \tilde{t}_0)^2} \tag{13.3}$$

are derived by a function fit. Instead of fitting the range history (13.3) directly, the square of it, i.e. a parabola, is optimized. Once the monostatic equivalent range $\tilde{R}(t)$ is fitted to the true range $R(t)$, not only a more accurate azimuth focusing parameter

$$FM(R) = 2 \cdot \tilde{V}^2 / \lambda \cdot R \tag{13.4}$$

is obtained, but also the along-track shift $\tilde{t}_0$ and the across shift $2 \cdot (R_{TX} - \tilde{R}_0)/c$, which are the radargrammetric reference points in the time domain.

13.8.5 Phase Unwrapping

The main task of generating accurate DEMs from bistatic interferometric phases is to assign each of these phase values its absolute phase offset by solving for the ambiguities described above. The first step is to "unwrap" the phases by following phase gradients on consistent paths within the image and thus counting the ambiguity cycles. That is only possible if the phase information is connected for larger areas within the image without too much phase noise in between and phase slopes smaller than π/sample. The ITP uses for that a minimum cost flow algorithm where a function of the coherence represents the "costs" of the paths to follow. Pixel-to-pixel jumps in topographic height larger than the ambiguity or areas disconnected from the rest by random phases (e.g. from the incoherent water bodies or shadow/layovers) may result in wrongly assigned ambiguity offsets—the phase unwrapping errors where segments of the DEM have erroneous offsets of exact multiples of the height of ambiguity. These are unavoidable for complicated terrain and/or incoherent data if the only available information for unwrapping are the phases of one single bistatic acquisition.

13.8.6 Determination of Absolute Phase Offset

The second step is to derive the absolute distance to the arbitrary starting point of the unwrapping process to solve for the global offset. The latter is usually done by ingesting external reference information on tie points into processing. The ITP however uses the novel approach to derive a coarse radargrammetric absolute distance map (actually a DEM which is not geocoded) from the co-registration cross-correlation patch measurements for determining the absolute phase offset of the entire scene. Each individual patch yields only for very coherent data a distance resolution below the wavelength. But for assigning an offset for the entire scene, the difference of the (correctly) unwrapped phase and all coherent radargrametric estimates is averaged, resulting in sub-millimetric offset accuracy [35]. This is more than sufficient for assigning the right global π-ambiguity band to the interferometric phase if the delays of both instruments are correctly calibrated. Finally the absolute unwrapped interferometric phases are geocoded into heights onto geographic coordinates using iterative steps with interpolated grids of phase-to-height conversion functions taking the spatially varying geometry parameters into account—especially sensitive to baseline data accuracy. Additionally, any influences of signal propagation on the pixel localization, phases and heights have to be accounted for. Besides the standard correction of the large tropospheric range delay of

2–4 m valid for both instruments and shifting the interferogram only, also a phase distortion from the very small discrepancy in path length from the angular separation of the two instruments is observed. It is causing a signal delay (and phase) difference in the millimetre range with corresponding height errors in the meter range.

In TanDEM-X, solving for the π-ambiguities in the order of 10–30 m, a relative instrument ranging accuracy in the order of millimetres is required. This outstanding precision is only possible with highly accurate signal and geometric processing algorithms and result of an extensive bistatic system calibration campaign (cf. Sect. 13.10). In the latter, the two measurements of radargrammetric shifts and interferometric phases where compared for thousands of scenes in different instrument settings to a reference phase derived from SRTM DEMs inside the ITP and aligned with delay and phase corrections provided for the processor. This had to be done in accordance with the simultaneous baseline product calibration which influences both. The self-contained processing approach of the ITP thereby enables TanDEM-X to act for nearly 90% of the land surface as a very precise and independent absolute height measurement instrument and thus allows to derive changes in heights (e.g. on glaciers or ice fields) in the meter ranges without being biased by outdated reference heights.

13.8.7 Quality Control

The remaining problem is to determine the correctness of the unwrapped phases for offset determination and for DEM quality control. Also for this task, the radargrammetric estimates are used. An iterative filtering allows to locally compare the phases and the distances to detect unwrapping errors for areas larger than about 1 km^2. Deviating patches are then excluded from the phase offset determination. If the image contains a significant amount (about 3%) of discrepant regions, the resulting DEM is flagged as problematic by the ITP.

13.8.8 Multiple Baseline Phase Unwrapping

The entire mission concept is designed to achieve a high accuracy from large baseline acquisitions while taking into account unwrapping problems from the very small height of ambiguities. To mitigate the latter, the dual baseline phase unwrapping concept has been developed for the processing chain [36, 37]. The Earth is mapped at least twice with different heights of ambiguity. The ITP processes every large baseline acquisition and every one with unwrapping problems with the help of the previously processed acquisitions. The intermediate products from the first processing run are the co-registered complex products (CoSSCs). Overlapping data sets of new raw data and the old CoSSCs are unwrapped and the differential phase is derived. The differential interferogram has a much larger

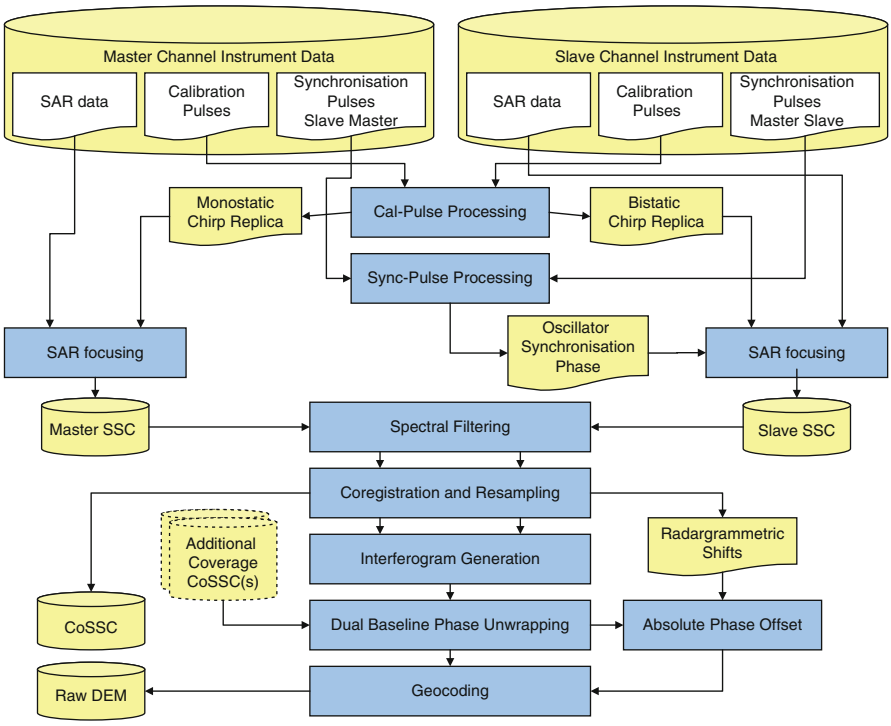


Fig. 13.19 Processing workflow of the Integrated TanDEM-X Processor ITP

height of ambiguity than the two single acquisitions ($h^{-1} = {h_1}^{-1} - {h_2}^{-1}$). It is easier to unwrap but also noisier. From this interferogram and measured unwrapped phase discrepancies between first and second acquisitions, erroneous regions are identified in the unwrapped phases. Since even the differential interferogram may have unwrapping errors for problematic scenes, the two radargrammetric estimates are used to correct the ambiguity band of the differential phases. These modified phases are then used to correct (only) the ambiguity cycle of the large baseline interferogram where necessary. Filtering and region detection algorithms are implemented to maintain the full resolution and quality of the interferogram. The raw DEM derived from this iteratively improved interferogram replaces any erroneous single acquisition based raw DEM prior to the mosaicking and calibration process. The major steps of the entire ITP processing workflow are summarized in Fig. 13.19.

13.9 Interferometric and DEM Performance Monitoring

The excellent radar instrument performance and stability of both satellites is the basis for high-quality SAR imaging and has been demonstrated in the monostatic commissioning phases of both TSX in 2007 [38, 39] and TDX in 2010 [40]. In the

bistatic TanDEM-X configuration two complex SAR images are co-registered and interferometrically processed. The interferometric phase information is translated into elevation values which are characterised by their respective relative height accuracies. The instrument settings and satellite formation parameters have been optimised for global DEM acquisition using a dedicated height error simulator and are now systematically evaluated during the operational mission.

Major factors which affect the relative height accuracy are the radiometric sensitivity of each SAR instrument, range and azimuth ambiguities, quantisation noise, processing and co-registration errors as well as surface and volume decorrelation, scaled by the baseline length [1]. A key quantity in estimating the interferometric performance is the coherence which has been computed for TanDEM-X by the product

$$\gamma_{Tot} = \gamma_{SNR} \cdot \gamma_{Quant} \cdot \gamma_{Amb} \cdot \gamma_{Coreg} \cdot \gamma_{Geo} \cdot \gamma_{Az} \cdot \gamma_{Vol} \cdot \gamma_{Temp} \tag{13.5}$$

where the right hand side describes the different error contributions due to the limited SNR (γ_{SNR}), radar signal quantisation (γ_{Quant}), ambiguities (γ_{Amb}), limited coregistration accuracy (γ_{Coreg}), baseline decorrelation (γ_{Geo}), relative shift of Doppler spectra (γ_{Az}), volume decorrelation (γ_{Vol}), and temporal decorrelation (γ_{Temp}). Different scattering characteristics as well as changing baseline and instrument configurations influence the coherence. An important contribution is the limited signal-to-noise ratio (SNR) which depends on both the sensitivity of the radar instruments and the power of the backscattered signal. For TanDEM-X, the two instruments are almost identical and the mean SNR of both interferometric channels yields a theoretical coherence [41]

$$\gamma_{SNR} = \frac{1}{1 + SNR_{mean}^{-1}} \tag{13.6}$$

which is plotted as the red line in Fig. 13.20. In the same figure the measured coherence of TanDEM-X interferograms is shown. The measurements can be divided into four main land classes: boreal forest and soil and rock test sites (blue triangles and red diamonds) have on average better performance ($\gamma_{Tot} > 0.7$) than areas characterized by tropical or rain forest (green dots). In the latter, coherence losses are mainly due to the existence of volume decorrelation over dense forest. Sandy deserts have in addition very low SNR (red squares).

In vegetated areas, volume scattering is the primary reason for coherence loss. Rain forests with tree heights of over 40 m have been measured in varying TanDEM-X bistatic acquisition configurations. Typical heights of ambiguity for the global DEM acquisition are in the same range. With this knowledge, the baseline geometry can be optimised for vegetated areas to achieve heights of ambiguity greater than the forest heights. Figure 13.21 is an example where the coherence improvement is particularly visible over Amazon rain forest (lower part of the images), whereas in the clear-cuts almost the same high coherence is obtained.

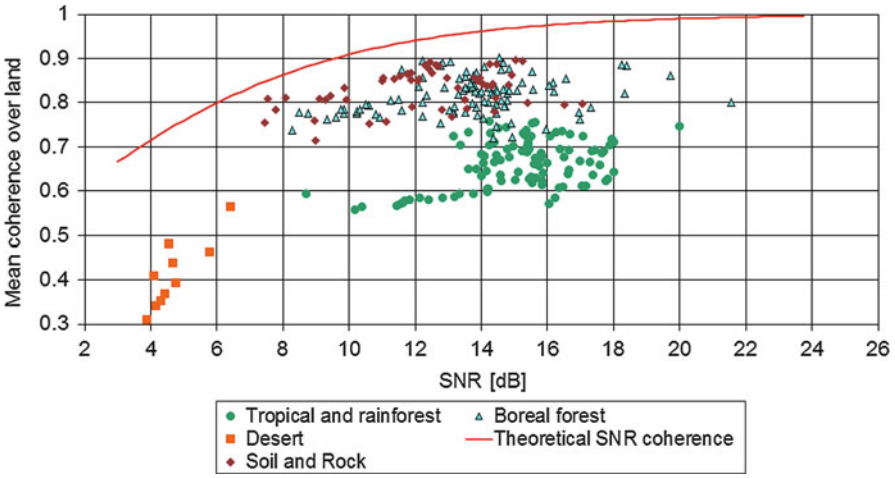


Fig. 13.20 Interferometric coherence of land area over signal-to-noise ratio (SNR). *Red line*: theoretical coherence prediction from SNR only. Single values indicate the coherence of TanDEM-X interferograms

The bistatic image coherence is directly correlated with the relative height accuracy Δh of the corresponding DEM, where high coherence values translate into a low interferometric phase error $\Delta\varphi$ and vice versa:

$$\Delta h = h_{amb} \cdot (\Delta\phi / 2\pi) \tag{13.7}$$

The height error of TanDEM-X elevation data has been determined from the difference between two repeated DEM data takes over the same ground area acquired using the same configuration parameters. Two sources of errors mainly contribute to the difference and can be characterized in the frequency domain: low frequency errors and systematic offsets such as orbit errors and height shifts can be filtered out; and high frequency errors, characterizing the random error in relative vertical height, due to coherence losses and phase unwrapping problems. The relative vertical height error is derived from the 90% distribution interval of the high-frequency errors.[1] Figure 13.22 presents the results obtained from the analysis of several test sites, acquired with different geometries and classified as soil and rock terrain. Even if the final DEM will be obtained by the combination of at least two global acquisitions, for some test sites one acquisition is already sufficient to meet the required specification. A linear fit through the height error estimates shows a good qualitative agreement with the DEM performance predictions from Fig. 13.6: a 90% point-to-point height error of approx. 1.7 and

[1] Since the relative vertical height accuracy depends on the terrain slope, a proper mask has to be applied, in order to discriminate flat and mountainous terrain.

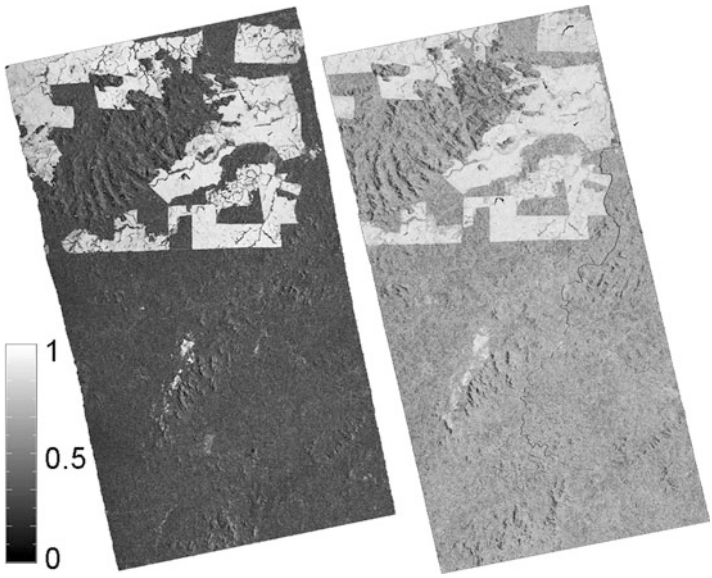


Fig. 13.21 Coherence of Amazon rain forest test site acquired at different acquisition geometries. Low coherence is visualized as dark colour. *Left*/*Right*: height of ambiguity: 25 m/52 m, cross-track baseline 182 m/162 m, incidence angle: 29.9°/47.7°

2.5 m is predicted for a height of ambiguity of 30 and 45 m, respectively (yellow bars in Fig. 13.22).

Figure 13.23 presents a global map of coherence for all processed DEM data takes between December 2010 and December 2011. The total coherence is derived by averaging over land surfaces only. Over 90% of the processed interferograms are characterized by a mean coherence higher than 0.6, which typically assures a sufficient quality of the final DEM. Areas suffering from low interferometric performance are identified by low coherence and phase unwrapping problems in the processing. Difficult terrain—mainly forested or mountainous regions—will be re-acquired in dedicated campaigns with larger heights of ambiguity and other viewing angles.

13.10 Bistatic and Interferometric System Calibration

13.10.1 Expected Height Error

The height accuracy of an InSAR DEM is affected by different error mechanisms that can be classified in noise-like degradation and more systematic contributions. Performance losses due to volume, temporal or SNR decorrelation (Sect. 13.9) cause noise-like height errors that can only be partly mitigated by spectral

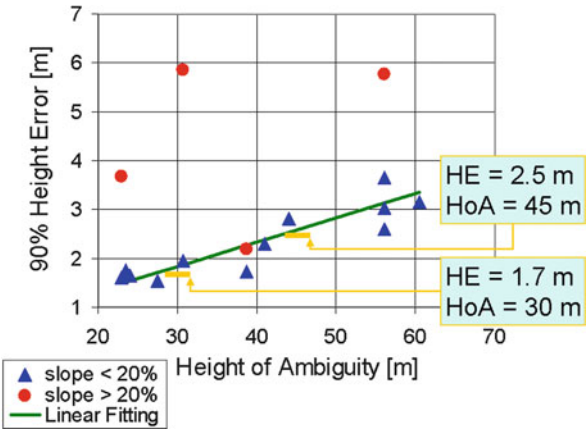


Fig. 13.22 Summary of the relative height error (HE) performance for data takes over soil & rock surfaces. In *green* is the linear fitting through the measured values for slopes <20%, *yellow bars* indicate the predicted performance for specific heights of ambiguity (HoA)

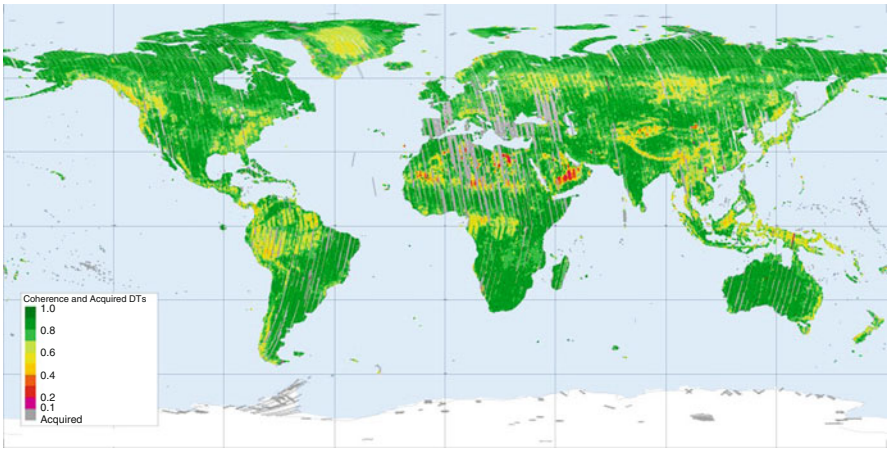


Fig. 13.23 Global coherence map of TanDEM-X data takes acquired between December 2010 and December 2011. High coherence is coloured in *green*, medium coherence in *yellow* and low coherence in *red* and *purple*. *Grey* areas were acquired but still in the processing queue during the time that this coherence map was generated

multilooking or spatial averaging procedures. Other SAR geometry dependent processing effects, like phase unwrapping errors, shadow or layover, can be effectively reduced by appropriately combining several DEMs acquired with different viewing geometries [1, 42]. The TanDEM-X error budget assumes a height error (90% confidence level) of 1.8 m for the noise-like contributions

remaining after the averaging process. On the other side, there are the systematic, but nondeterministic errors. Before launch, a detailed survey on the different components of the TanDEM-X system was carried out. Two dominant systematic error sources were identified: inaccuracies in the baseline determination and drifts in the SAR instrument. The baseline determination error is driven by the residual errors of the relative GPS position determination between both satellites and the determination accuracy of the locations of the SAR antenna phase centres by way of the attitude information provided by the star sensors. The SAR instrument is mainly affected by temperature drifts due to the warming of the transmit/receive modules over an acquisition and by initial offsets due to slight differences in the interferometic behaviour of the two (independent) satellites. These sources cause additional height errors that are expected to reach several meters, which would exceed the relative height error requirement of 2 m (90% confidence level). To quantify the error magnitudes, the influence of the sources was estimated and fed into a simulation tool. With the simulation results, a polynomial error model was established in order to estimate the DEM height error g over a data take

$$g_n(x, y) = a_n + b_n x + c_n y + d_n xy + e_n y^2 + f_n y^3 \tag{13.8}$$

where $a_n . . . f_n$ are the unknown parameters, x are the coordinates of image points in ground range and y in azimuth direction and n is the index of data take. The first term a describes a constant offset of the DEM, b is a tilt in range, c, e and f stand for ramps and gradients in azimuth and d describes a torsion over range and azimuth.

To minimize these height errors, the DEM calibration effort is focused on the initial reduction of the baseline error by means of the baseline calibration [43] (cf. Sects. 13.7 and 13.10.2), the reduction of systematic phase errors and the estimation of the polynomial coefficients in the DEM calibration procedure [44] (cf. Sect. 13.11). These corrections, which use globally distributed test sites and height references [45] and the information of adjacent acquisitions, shall ensure the 2 m relative height accuracy of the final TanDEM-X DEM product (see Fig. 13.24).

13.10.2 Baseline Calibration

The baseline is determined from GPS data provided by dual frequency IGOR GPS receivers on both satellites followed by a double-differential processing approach (cf. Sect. 13.7). Relative accuracies are within the 1 mm (1σ) requirement, but offsets of a few millimetres may not be excluded from this approach. By monitoring the average height error in DEMs over dedicated global test sites, the potential baseline bias perpendicular to flight direction can be characterized [43]. This

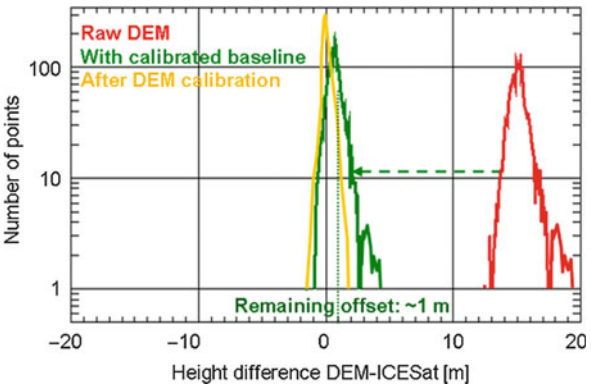


Fig. 13.24 Example of the height error histograms of a TanDEM-X DEM in consecutive calibration stages (ICESat heights are used as references)

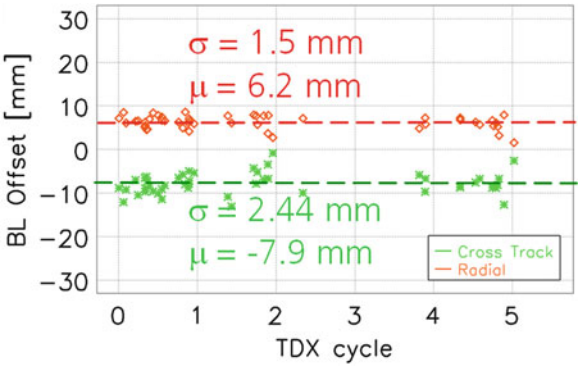


Fig. 13.25 Baseline offsets estimated during the monostatic commissioning phase

procedure has been successfully applied during the TanDEM-X pursuit monostatic commissioning phase, achieving a correction accuracy of around 2 mm (1σ) (see Fig. 13.25). The 2 mm variability around the mean value includes besides the residual baseline errors several other error contributions from the SAR instrument and the limited accuracy of the reference heights, which are considered in the overall DEM error budget.

13.10.3 Calibration of Instrument and External Delays

The interferometrically derived DEMs are initially ambiguous by the height of ambiguity. To correctly resolve this ambiguity, radargrammetry is used which estimates the absolute location and height of large patches in the acquisition (Sect. 13.8). The estimation is based on the travel times of the radar signals. Hence, the technique is very sensitive to differential instrument delays. Such differential delays are caused by different receiver gain settings, different

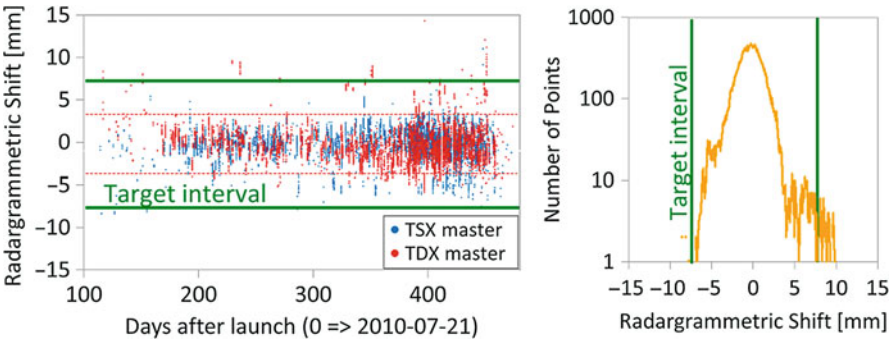


Fig. 13.26 Estimated radargrammetric shifts of the TanDEM-X DEMs with respect to the ones derived with the help of SRTM, after applying the correction parameters. *Left*: shifts over time. *Right*: corresponding histogram. The standard deviation of the shifts is 1.8 mm

bandwidths, different synchronisation horn combinations, or are dependent on the transmitting satellite [43]. The dependencies were derived by statistical evaluations and dedicated analyses of the acquired DEM data. Moreover, also relativistic and differential tropospheric effects have been observed to cause systematic shifts of several meters in the DEMs [46]. All these contributions are compensated during the interferometric processing. To resolve the π-ambiguity, the measured radargrammetric shifts shall be lower than the target interval of ± 7.5 mm (around $\pm\lambda/4$), as shown Fig. 13.26 for 99% of the acquired scenes.

13.10.4 DEM Stability in the Running Mission

After applying the corrections mentioned in the previous section, about 90% of the DEMs show absolute height errors better than ± 10 m (Fig. 13.27) [47]. Most outliers are caused by the π-ambiguity (Sect. 13.8) and will be corrected by reprocessing. During the final DEM calibration process (Sect. 13.11), the DEM acquisitions are calibrated relative to each other and against globally distributed ICESat reference points.

13.11 DEM Calibration and Mosaicking

In this section the last two steps of the DEM processing, namely DEM calibration and DEM mosaicking are outlined. The DEM calibration estimates for each data take its offset, tilt in range and slope in the azimuth direction. In a weighted least-squares block adjustment several neighbouring data takes (up to continental scale) are simultaneously calibrated. This approach guarantees smooth height transitions at data take borders. The DEM mosaicking merges different, corrected DEM

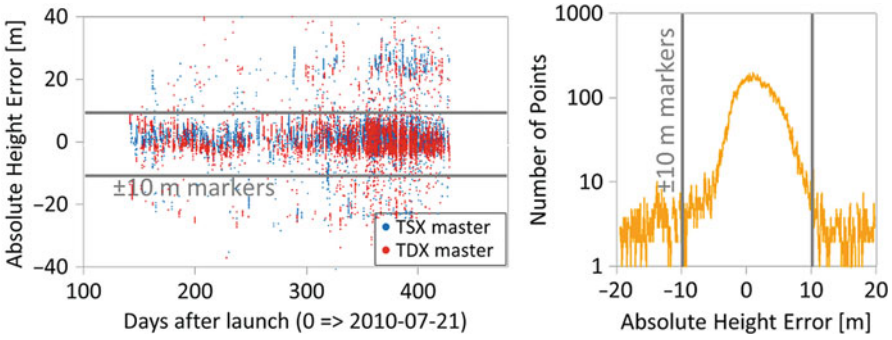


Fig. 13.27 Absolute height difference between the TanDEM-X DEMs and SRTM (and other references) after baseline calibration in bistatic operation. *Left*: height difference over time. *Right*: corresponding histogram. The height differences have a standard deviation of 8.3 m

acquisitions and tiles the resulting DEM mosaic and additional product layers (e.g. amplitude mosaic and height error mask) into one-by-one degree geocells.

13.11.1 DEM Calibration Approach

The DEM calibration is part of the operationally implemented "DEM Mosaicking and Calibration Processor" within the ground segment (Fig. 13.28) [48]. In a first data-driven processing step (DEM preparation) the DEM scenes generated by the ITP (Sect. 13.8) are successively processed and some measures for height discrepancies, a water body detection mask, and a list of calibration (tie) points are produced. After interactive quality control the results of the DEM preparation processor are stored in the archive system (PL).

For TanDEM-X, ICESat (Ice, Cloud and land Elevation Satellite) data are used as height references (ground control points—GCPs). The ICESat spaceborne laser altimeter data provide a very good global coverage for referencing the DEM with a foot print distance of 170 m in along-track and a maximum ground-track separation of 80 km at the equator [49]. According to an accuracy study [50] the standard deviation for selected ICESat points in flat bare surfaces is below 2 m. Therefore, several selection criteria are considered, like number and width of peaks in the waveforms (indicator for vegetation) as well as received signal levels. In order to combine ICESat points and the corresponding TanDEM-X pixels, all elevation values within the ICESat footprint are averaged. Furthermore, equidistant tie-points in the at least 3 km wide overlap between neighbouring DEMs are used as input for the DEM calibration (Fig. 13.29). At tie-points image chips in the dimension of about one by one kilometre are extracted from the raw DEM, the amplitude and the Height Error Mask (HEM). Inside the chips unsuitable areas are excluded with the help of previously generated height discrepancy

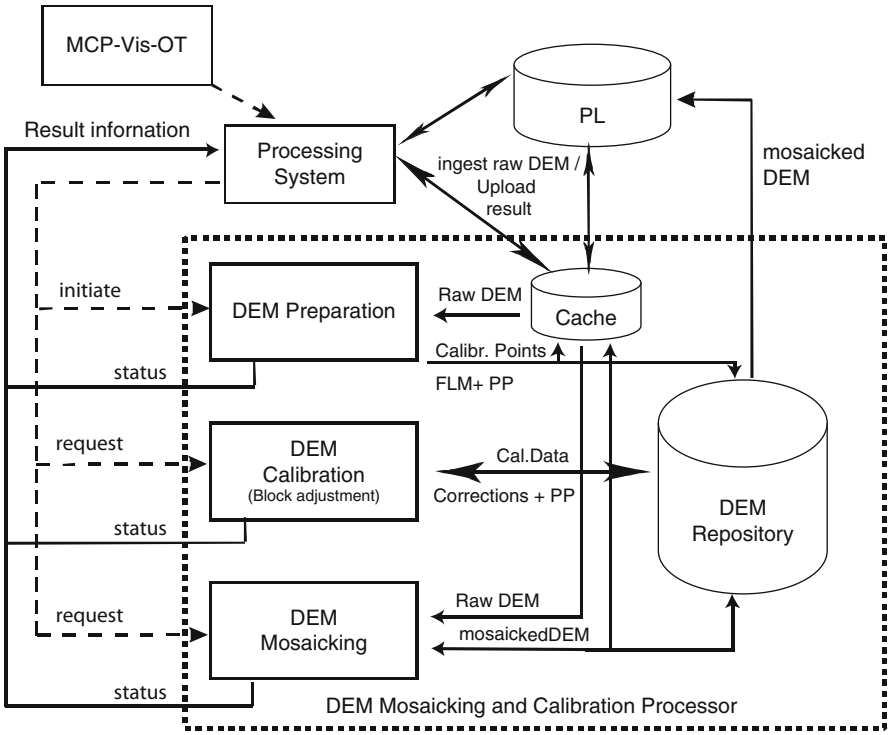


Fig. 13.28 Block diagram of the DEM calibration and mosaicking processor [48] (*MCP-Vis-OT*: visualization and operating tool, *PL*: product library/robot archive system, *FLM*: flag mask, *PP*: parameter pool)

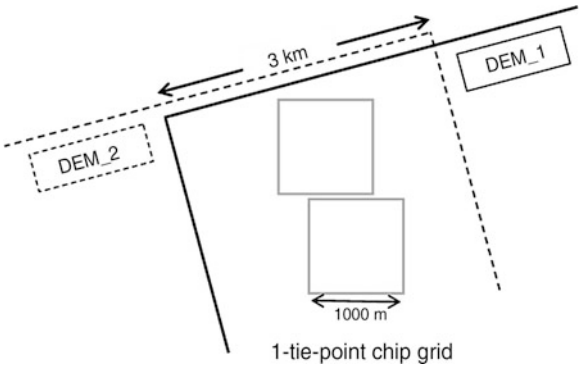


Fig. 13.29 Tie-point chips in the centre of the 3 km cross-track overlap between neighbouring DEMs

masks, water masks and shadow/layover masks. From the remaining 'valid' pixels the median height value is assigned to the tie-point instead of the mean, reducing the impact of outliers. Over the whole mission approx. 40 Mio. chips corresponding to 12 TB will be extracted globally. All calibration points are stored in lists to be used as observations during the DEM calibration. In that way, it is possible to calibrate larger regions without reading huge image data volumes from the archive.

DEM calibration is performed on blocks of larger connected DEM data takes. The polynomial correction parameters (a_n–f_n) of (13.8) will be estimated for each data take within the DEM Calibration Processor by means of a least-squares adjustment [51, 52]. The approach assumes that after application of the error function the heights in overlapping areas should be identical, apart from random noise. The advantage of this method is that the correction parameters can be found independent from terrain types. The observables are averaged heights at each tie-point or GCP. For all observations accuracies resp. weights are set up for the stochastic model. The unknown parameters are estimated iteratively by a weighted least-squares adjustment [53]. Those correction parameters that could not significantly be estimated are excluded from further iterations. As result we also get quality parameters from the adjustment. The DEM calibration is initiated by an operator who also monitors the progress using a Visualization and Operating tool (MCP-Vis-OT, see top left of Fig. 13.28). As the DEMs are acquired and processed according to the globally oriented data acquisition plan (Sect. 13.5) the DEM calibration of larger blocks starts after finishing the first global acquisition.

13.11.2 Mosaicking—Weighted Combination of DEMs

Purpose of the DEM Mosaicking Processor is the fusion of all DEM information layers [54]: DEMs, height error maps, and amplitude data as well as the generation of a fused water indication mask. The calibrated corrections are applied to each DEM and different coverages will be mosaicked by a weighted average of the elevation values [1]

$$h = \frac{\sum_{k=1}^{K} \frac{1}{\sigma_{HEM,k}^2} h_k}{\sum_{k=1}^{K} \frac{1}{\sigma_{HEM,k}^2}} \tag{13.9}$$

with σ_{HEM} as the height error of one elevation. In case of increased HEM values, a special border handling is applied. Amplitude and HEM values are also fused with the help of the height error map. After the DEM mosaicking processor an interactive quality control is performed on the resulting geocells. A by-product is the so-called

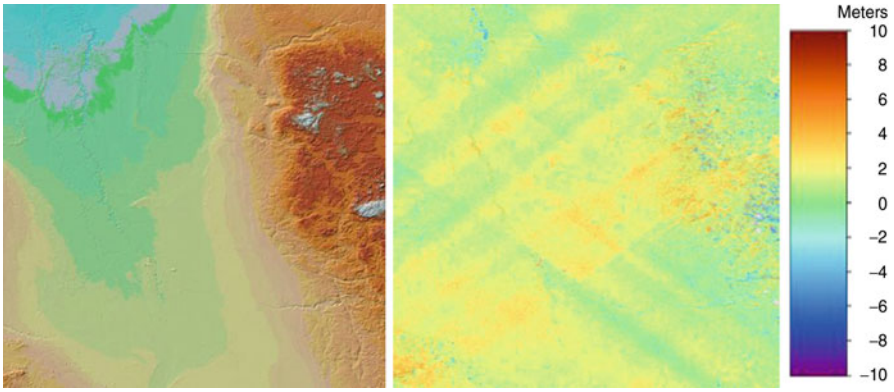


Fig. 13.30 *Left*: TanDEM-X DEM of geocell W97 N47 of Minnesota test site, *right*: difference to SRTM after DEM calibration and mosaicking of three data takes. The *colourbar* on the right refers to the height difference between SRTM and TanDEM-X

Water Indication Mask (WAM). The intention of this WAM is not to map a global water mask, but to have a supplementary information layer to support the DEM editing process. As the TanDEM-X DEM is an unedited, interferometric product, water surfaces will appear noisy because of low coherence and/or backscatter.

The DEM calibration and mosaicking approach was applied to early commissioning phase data takes of a test site in Minnesota (USA). The mosaick (see Fig. 13.30) is compared to SRTM and shows a mean offset of 2 m to SRTM, but just 3 cm offset to ICESat verification points. One can clearly note the stripe pattern from SRTM that is not present in the TanDEM-X DEM. Figure 13.31 shows a mosaicked DEM of Iceland that has been derived from the TanDEM-X data takes of the first acquisition year.

13.12 Scientific Exploitation and Experimental Results

Beyond its primary goal of acquiring a global DEM, TanDEM-X provides the remote sensing scientific community with a unique data set to demonstrate new bistatic and multistatic radar techniques for enhanced bio- and geophysical parameter retrieval. The following subsections summarize some of the advanced capabilities of TanDEM-X which can be operated in a multitude of modes and configurations [1]. First exciting results, which were obtained during the TanDEM-X commissioning phase, already demonstrate the great potential of bistatic and multistatic SAR to serve novel and extremely powerful remote sensing applications [55–57]. The intention of this section is to provide only a succinct overview of the wide range of possible applications. A complete description of the experiments and a detailed discussion of their results can be found in the provided references. Many of the experiments may form the basis for future formation flying SAR missions.

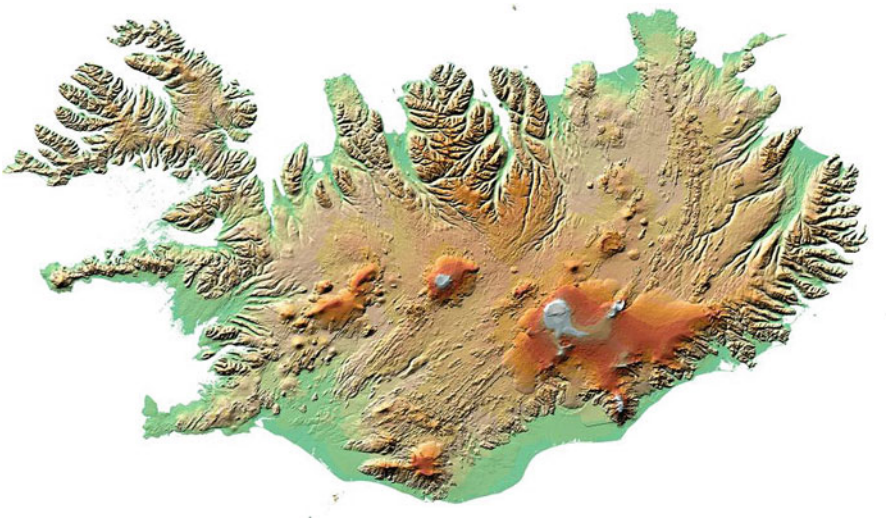


Fig. 13.31 Mosaick of Iceland DEM acquired by TanDEM-X

13.12.1 Velocity Measurements from Space

TanDEM-X has the capability to provide highly accurate velocity measurements of moving objects within a large coverage area. This can be achieved by comparing the amplitude and phase of two SAR images acquired at slightly different times (Fig. 13.32). By adjusting the along-track displacement between the TDX and TSX satellites from almost zero to several tens of kilometres, TanDEM-X can adapt its sensitivity to a broad spectrum of velocities ranging from less than a millimetre per second to more than hundred kilometres per hour. The Helix satellite formation employed by TanDEM-X enables even a minimization of the effective cross-track baseline for a given latitude and incident angle, thereby reducing the complexity in the velocity estimation process. Along-track interferometry can furthermore be enhanced by the so-called dual receive antenna mode (cf. Sect. 13.3) in each of the two tandem satellites, which provides additional phase centres separated by a short along-track baseline of 2.4 m. The combination of short and long baseline SAR data acquisitions improves both the detection and localization of moving objects and resolves phase ambiguities in case of fast moving scatterers. TanDEM-X provides hence a unique SAR system with four phase centres separated in the along-track direction. Potential applications are Ground Moving Target Indication (GMTI), the measurement of ocean currents, as well as the monitoring of sea ice drift and rotation.

Figure 13.33 shows as a first example the observation of ship movements in the Strait of Gibraltar [56]. The data were acquired during the monostatic commissioning phase where the satellites had an along-track separation of 20 km. This separation corresponds to a time lag of 2.6 s. The 2-D velocity vector could be measured with an

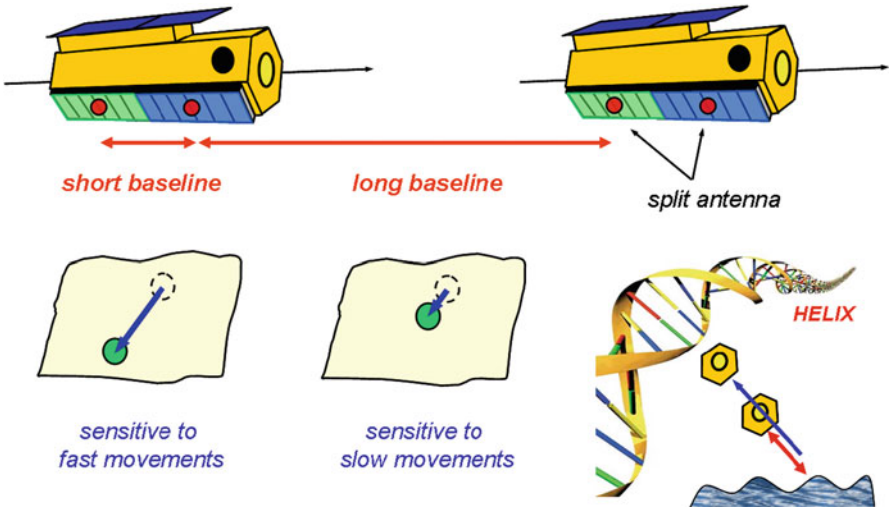


Fig. 13.32 Velocity measurements with TanDEM-X. The Helix satellite formation allows a flexible adjustment of the desired along-track separation between the satellites. In addition, a short along-track baseline is provided by each satellite via the split antenna

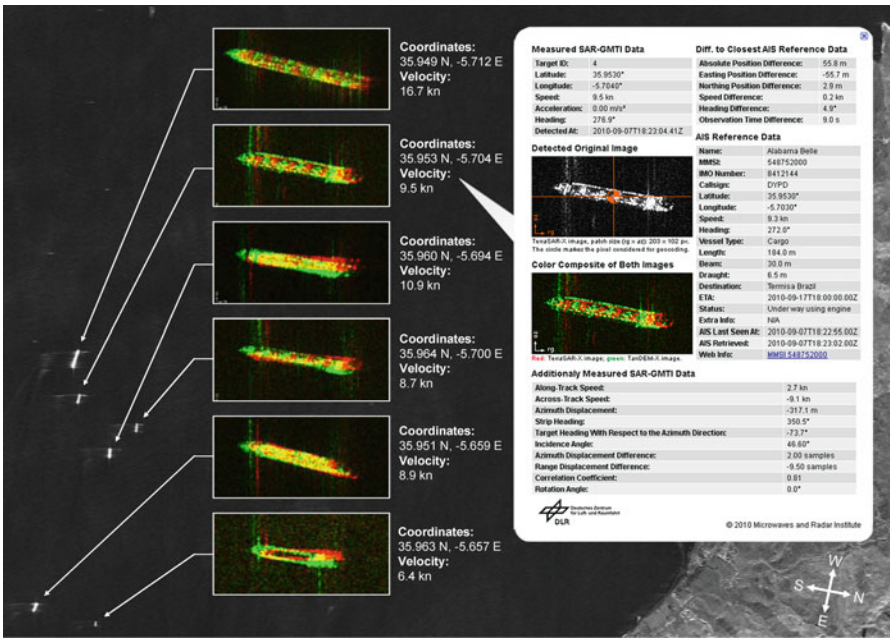


Fig. 13.33 Ship movements observed with TanDEM-X during the monostatic commissioning phase. The ship displacements can be seen from the insets showing TSX and TDX image patches overlaid in different colours. The estimated velocities are in excellent agreement with AIS reference data (*right*) (cf. [56] for details)

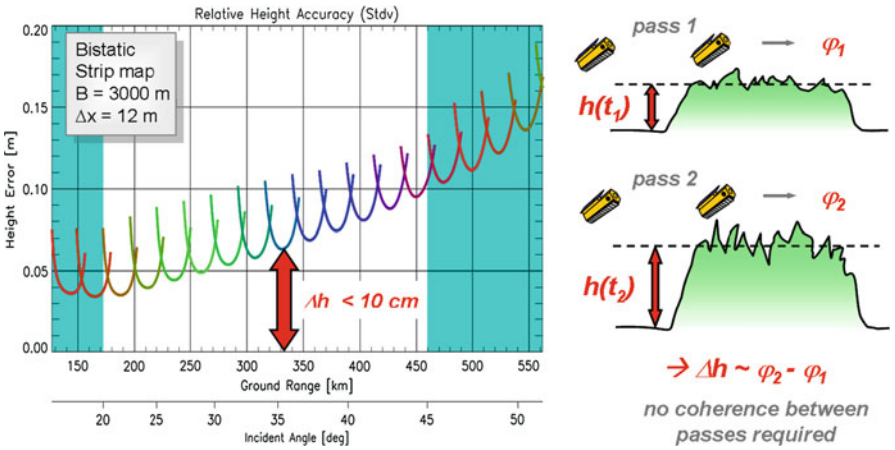


Fig. 13.34 Performance example for large baseline DEM acquisitions with TanDEM-X (cross-track baseline = 3,000 m, posting = 12 m). A relative height accuracy (single point standard deviation) better than 10 cm is predicted

accuracy of 1 km/h by comparing the ship positions in the TSX and TDX SAR images (cf. [56] for details). The velocity measurements have been validated with independent data obtained from the Automatic Identification System (AIS).

13.12.2 Large Baseline Cross-Track Interferometry

Large baseline interferometry takes advantage of the high bandwidth of the TSX and TDX radar instruments, allowing for coherent data acquisitions with cross-track baselines of up to 5 km and more. Note that less than 5% of the maximum possible (critical) baseline length is used during nominal DEM data acquisition. Large baseline interferograms can hence significantly improve the height accuracy beyond the standard TanDEM-X DEM quality, but the associated low height of ambiguity requires typically a combination of multiple interferograms with different baseline lengths to resolve phase ambiguities, especially in hilly and mountainous terrain. Further opportunities arise from a comparison of multiple large baseline TanDEM-X interferograms acquired during different passes of the satellite formation (Fig. 13.34). This provides a sensitive measure for vertical scene and structure changes. Potential applications are a detection of the grounding line which separates the shelf from the inland ice in polar regions, monitoring of vegetation growth (e.g. agricultural fields), mapping of atmospheric water vapour with high spatial resolution, measurement of snow accumulation or the detection of anthropogenic changes of the environment, e.g. due to deforestation. Note that most of these combinations rely on a comparison of two or more single-pass (large baseline)

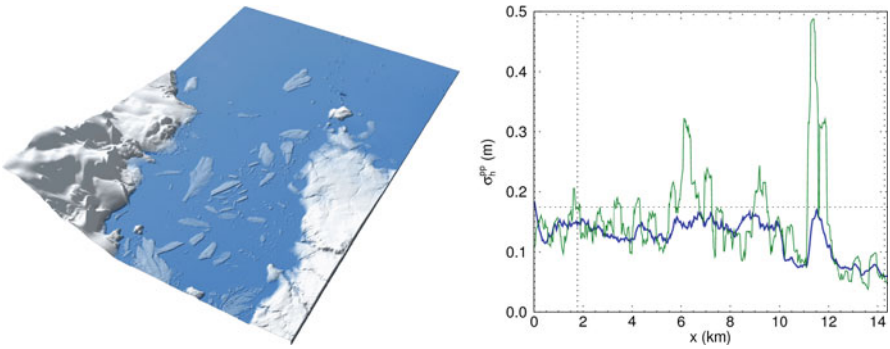


Fig. 13.35 Large baseline TanDEM-X DEM of the edge of October Revolution island (*left*) and predicted (*blue*) versus estimated (*green*) point-to-point height accuracy along a DEM slice (cf. [57] for details)

cross-track interferograms and do hence not necessarily require coherence between the different passes. Further information can be gained from an evaluation of coherence changes, potentially augmented by polarimetric information. This is for instance well suited to reveal even slight changes in the soil and vegetation structure reflecting vegetation growth and loss, freezing and thawing, fire destruction, human activities, and so on. The combination of repeated TanDEM-X single-pass interferograms enables hence the entry into a new era of interferometric and tomographic 3-D and 4-D SAR imaging as it was ERS-1/2 for the development of classical repeat-pass SAR interferometry.

Figure 13.35 shows as a first example a large baseline DEM which was acquired by TanDEM-X on July 16, 2010 in the Russian Arctic (October Revolution Island) [57]. The DEM was part of a longer data take that used a sophisticated commanding to obtain a large baseline interferogram while TDX was still drifting towards TSX from its initial along-track separation of 15,700 km (Sect. 13.6). At the time of data acquisition, the two satellites were 380 km apart from each other, resulting in a temporal separation of 50 s. Earth rotation caused a cross-track baseline of 2 km which corresponds to a height of ambiguity of only 3.8 m. A squinted operation was necessary to provide a sufficient overlap of the Doppler spectra. The right hand side of Fig. 13.35 shows the predicted (black curve) and estimated (gray curve) standard deviation of the point-to-point relative height error for a linear slice through the DEM. The predicted error was calculated from the coherence measurements and the estimated error was obtained by high-pass filtering the DEM slice [57]. Both results show that the relative height accuracy is in the order of 20 cm. This demonstrates the great potential of formation flying SAR missions to obtain high resolution elevation information with decimetre accuracy, thereby enabling new remote sensing applications. An example is the monitoring of height changes over glaciers, ice caps or ice sheets to quantify their ice mass balance and a dedicated formation flying SAR mission has already been proposed for this purpose [58].

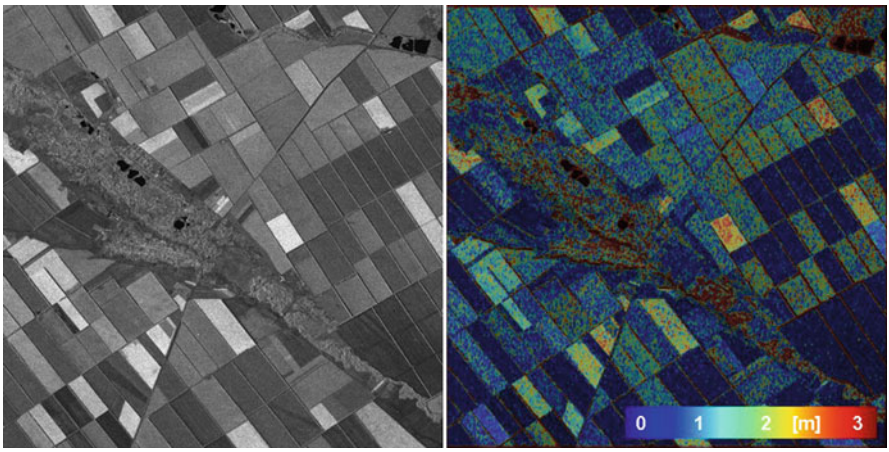


Fig. 13.36 Polarimetric SAR interferometry with TanDEM-X. *Left*: amplitude of SAR image. *Right*: interferometric height difference between HH and VV channels

13.12.3 Polarimetric SAR Interferometry

Polarimetric SAR interferometry combines interferometric with polarimetric measurements to gain 3-D structure information from semi-transparent volume scatterers in a single pass [59, 60]. A prominent example is the measurement of vegetation height and density which forms also the basis of future formation flying SAR missions dedicated to global environmental monitoring. Figure 13.36 shows as an example the interferometric height differences obtained for a dual-polarized TanDEM-X spotlight acquisition of an agricultural field in Russia. The height difference is a result of a different path length of the polarization depending on the volume structure. The data were acquired during the monostatic commissioning phase with a perpendicular baseline of 275 m, demonstrating the potential of crop height estimation.

Not only agricultural vegetation structure but also very sparse forest like in boreal regions can be potentially inverted to forest height. Figure 13.37 shows an example of an inversion of forest height. The RMSE is around 2 m when compared to Lidar (H100) measurements. The forest height inversion has been obtained from data acquired in the pursuit monostatic mode using dual polarimetric observations [61]. Future experiments will also employ fully polarimetric acquisitions.

13.13 Conclusions

TanDEM-X is a first demonstration of a “Distributed Space Mission for Earth System Monitoring”. Key technologies like close formation flying, bistatic SAR operation and synchronisation, precise baseline determination and calibration, as well as highly complex processing chains have been implemented. Sophisticated

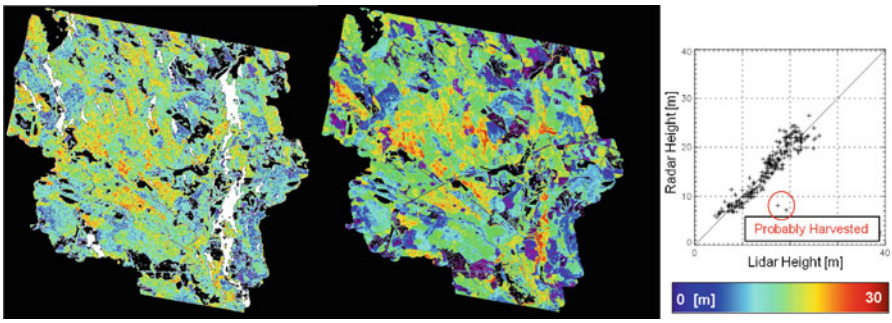


Fig. 13.37 Polarimetric SAR interferometry over forested areas. *Left*: inverted forest height over a boreal forest in North Sweden. *Middle*: lidar forest height measurements. *Right*: validation plot

safety mechanisms allow for safe operation at typical distances between 150 and 1,000 m. The complete system is fully operational since December 2010. Both satellites and the ground segment perform remarkably well. Current fuel consumption and battery degradation on the TerraSAR-X satellite is well below specification and allows predicting life time extensions of 2–3 years, i.e. close formation flying until 2015 seems feasible.

The TanDEM-X system is used to generate a global DEM of unprecedented accuracy. Achieving such a quality level requires at least two global acquisitions at increasing baseline distance. Remaining gaps over difficult terrain will be closed by adjusting the formation. In total, the acquisition for the global DEM is expected to be completed by mid 2013. In parallel to the global DEM acquisition scientific experiments are being served if the DEM-optimised formation geometry is adequate. After the global DEM phase the flexibility of the Helix formation will be fully exploited to allow demonstrating promising future SAR techniques like along-track interferometry, ground moving target identification and digital beamforming from space. Larger cross-track baselines will also enable the generation of regional DEMs with improved height accuracy. These unique demonstrations and experiments will also provide important information for the development and design of future formation flying SAR missions [62].

Acknowledgment The authors represent a large team that was involved in the development of TanDEM-X. We highly acknowledge the great effort and enthusiasm of all our colleagues at DLR and EADS Astrium who made this ambitious mission a reality. The TanDEM-X project is partly funded by the German Federal Ministry for Economics and Technology (Förderkennzeichen 50 EE 1035).

References

1. Krieger G, Moreira A, Fiedler H, Hajnsek I, Werner M, Younis M, Zink M (2009) TanDEM-X: a satellite formation for high resolution SAR interferometry. IEEE Trans Geosci Remote Sens 45:3317–3341

2. Zebker HA, Farr TG, Salazar RP, Dixon TH (1994) Mapping the world's topography using radar interferometry: the TOPSAT mission. Proc IEEE 82:1774–1786
3. Hajnsek I, Moreira A (2006) TanDEM-X: mission and science exploration. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Dresden, Germany
4. El-Sheimy N, Valeo C, Habib A (2005) Digital terrain modeling: acquisition, manipulation, and applications. Artech House, London
5. Li Z, Zhu Q, Gold C (2005) Digital terrain modelling. CRC Press, New York
6. Rosen PA, Hensley S, Joughin IR, Li FK, Madsen SN, Rodriguez E, Goldstein R (2000) Synthetic aperture radar interferometry. Proc IEEE 88:333–382
7. Werner M (2001) Shuttle radar topography mission (SRTM): mission overview. J Telecom (Frequenz) 55:75–79
8. Rabus B, Eineder M, Roth A, Bamler R (2003) The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. ISPRS J Photogramm Remote Sens 57:241–262
9. Moreira A, Krieger G, Mittermayer J (2004) Satellite configuration for interferometric and/or tomographic remote sensing by means of synthetic aperture radar (SAR). US patent 6,677,884, application from 5 Jul 2001
10. Moreira A, Krieger G, Hajnsek I, Hounam D, Werner M, Riegger S, Settelmeyer E (2004) TanDEM-X: a TerraSAR-X add-on satellite for single-pass SAR interferometry. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Anchorage, Alaska
11. Fiedler H, Krieger G (2004) Close formation flight of passive receiving micro-satellites. In: Proceedings of the 18th international symposium on space flight dynamics, Munich, Germany
12. Fiedler H, Krieger G, Zink M (2006) Constellation and formation flying concepts for radar remote sensing. In: Proceedings of advanced RF sensors for Earth observation, ESA/ESTEC, The Netherlands
13. D'Amico S, Montenbruck O (2006) Proximity operations of formation flying spacecraft using an eccentricity/inclination vector separation. J Guid Contr Dyn 29:554–563
14. Krieger G, Younis M (2006) Impact of oscillator noise in bistatic and multistatic SAR. IEEE Geosci Remote Sens Lett 3:424–428
15. Younis M, Metzig R, Krieger G (2006) Performance prediction of a phase synchronization link for bistatic SAR. IEEE Geosci Remote Sens Lett 3:429–433
16. Fiedler H, Krieger G, Zink M, Geyer M, Jäger J (2008) The TanDEM-X acquisition timeline and mission plan. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Friedrichshafen, Germany
17. Fiedler H, Krieger G, Werner M, Reiniger K, Eineder M, D'Amico S, Diedrich E, Wickler M (2006) The TanDEM-X mission design and data acquisition plan. In: Proceedings of European conference on synthetic aperture radar (EUSAR), Dresden, Germany
18. Krieger G, Fiedler H (2007) TanDEM-X mission analysis report. TD-PD-RP-0012, Oberpfaffenhofen, Germany
19. Pitz W, Miller D (2010) The TerraSAR-X satellite. IEEE Trans Geosci Remote Sens 48:615–622
20. Buckreuss S, Schättler B (2010) The TerraSAR-X ground segment. IEEE Trans Geosci Remote Sens 48:623–632
21. Schättler B, Kahle R, Metzig R, Steinbrecher U (2011) The joint TerraSAR-X/TanDEM-X ground segment. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Vancouver, Canada
22. Arbinger Ch, D'Amico S, Eineder M (2004) Precise ground-in-the-loop orbit control for low earth observation satellites. In: Proceedings of the 18th international symposium on spaceflight dynamics, Munich, Germany
23. D'Amico S, Arbinger Ch, Kirschner M, Campagnola S (2004) Generation of an optimum target trajectory for the TerraSAR-X repeat observation satellite. In: Proceedings of the 18th international symposium on space flight dynamics (ISSFD), Munich, Germany
24. Clohessy WH, Wiltshire RS (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 270:653–658

25. Kahle R, Schlepp B, Kirschner M (2011) TerraSAR-X/TanDEM-X formation control—first results from commissioning and routine operations. In: Proceedings of the 22nd international symposium on spaceflight dynamics, Sao Jose dos Campos, Brazil
26. Kahle R, Schlepp B (2010) Extending the TerraSAR-X flight dynamics system for TanDEM-X. In: Proceedings of the 4th international conference on astrodynamics tools and techniques, Madrid, Spain
27. Kahle R, Schlepp B, Meissner F, Kirschner M, Kiehling R (2012) The TerraSAR-X/TanDEM-X formation acquisition—from planning to realization. J Astron Sci (accepted)
28. Ardeans JS, D'Amico S, Fischer D (2011) Early flight results from the TanDEM-X autonomous formation flying system. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Canada
29. Zhu S, Reigber Ch, Koenig R (2004) Integrated adjustment of CHAMP, GRACE, and GPS Data. J Geodesy 78:103–108
30. Dach R, Hugentobler U, Fridez P, Meindl M (2007) Bernese GPS software version 5.0. Astronomical Institute, University of Bern
31. Kroes R (2006) Precise relative positioning of formation flying spacecraft using GPS. Ph. D. thesis, Technical University Delft
32. Breit H, Younis M, Balss U, Niedermeier A, Grigorov C, Hueso Gonzalez J, Krieger G, Eineder M, Fritz T (2011) Bistatic synchronisation and processing of TanDEM-X data. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Vancouver, Canada
33. Breit H, Fritz T, Balss U, Niedermeier A, Eineder M, Yague Martinez N, Rossi C (2010) Processing of bistatic TanDEM-X data. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Honolulu, USA
34. Fritz T, Rossi C, Yague-Martinez N, Rodriguez-Gonzalez F, Lachaise M, Breit H (2011) Interferometric processing of TanDEM-X data. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Vancouver, Canada
35. Rossi C, Eineder M, Fritz T, Breit H (2010) TanDEM-X mission: raw DEM generation. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Aachen, Germany
36. Lachaise M, Bamler R, Rodriguez-Gonzalez F (2010) Multibaseline gradient ambiguity resolution to support minimum cost flow phase unwrapping. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Honolulu, USA
37. Lachaise M, Fritz T, Eineder M (2008) A new dual baseline phase unwrapping algorithm for the TanDEM-X mission. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Friedrichshafen, Germany
38. Mittermayer J, Schättler B, Younis M (2010) TerraSAR-X commissioning phase execution summary. IEEE Trans Geosci Remote Sens 48:649–659
39. Mittermayer J, Younis M, Metzig R, Wollstadt S, Marquez J, Meta A (2010) TerraSAR-X system performance characterization and verification. IEEE Trans Geosci Remote Sens 48:660–676
40. Kraus T, Schrank D, Rizzoli P, Bräutigam B (2011) In-orbit SAR performance of TerraSAR-X and TanDEM-X satellites. In: Proceedings of the URSI commission F—open symposium on propagation and remote sensing, Garmisch-Partenkirchen, Germany
41. Zebker HA, Villasenor J (1992) Decorrelation in interferometric radar echoes. IEEE Trans Geosci Remote Sens 30:950–959
42. Eineder M, Adam N (2005) A maximum-likelihood estimator to simultaneously unwrap, geocode, and fuse SAR interferograms from different viewing geometries into one digital elevation model. IEEE Trans Geosci Remote Sens 43:24–36
43. Hueso Gonzalez J, Walter Antony J, Bachmann M, Krieger G, Zink M, Schwerdt M (2012) Baseline calibration in TanDEM-X to ensure the global digital elevation model quality. ISPRS Journal of Photogrammetry and Remote Sensing, Theme issue "Innovative applications of SAR interferometry from modern satellite sensors", doi: http://dx.doi.org/10.1016/j.isprsjprs.2012.05.008

44. Hueso González J, Bachmann M, Krieger G, Fiedler H (2010) Development of the TanDEM-X calibration concept: analysis of systematic errors. IEEE Trans Geosci Remote Sens 48:716–726
45. Hueso González J, Bachmann M, Scheiber R, Krieger G (2010) Definition of ICESat selection criteria for their use as height references for TanDEM-X. IEEE Trans Geosci Remote Sens 48:2750–2757
46. Krieger G, De Zan F, Hueso Gonzalez J, Bachmann M, Rodriguez Cassola M, Zink M (2012) Unexpected height offsets in TanDEM-X: explanation and correction. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Munich, Germany
47. Bachmann M, Hueso Gonzalez J, Krieger G, Schwerdt M, Walter Antony J, De Zan F (2012) Calibration of the bistatic TanDEM-X interferometer. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Nuremberg, Germany
48. Wessel B, Marschalk U, Gruber A, Huber M, Hahmann T, Roth A, Habermeyer M, Kosmann D (2008) Design of the DEM mosaicking and calibration processor for TanDEM-X. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Friedrichshafen, Germany
49. Zwally J (2002) ICESat's laser measurements of polar, ice, atmosphere, ocean, and land. J Geodyn 34:405–445
50. Huber M, Wessel B, Kosmann D, Felbier A, Schwieger V, Habermeyer M, Wendleder A, Roth A (2009) Ensuring globally the TanDEM-X height accuracy: analysis of the reference data sets ICESat, SRTM, and KGPS-tracks. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Cape Town, South Africa
51. Gruber A, Wessel B, Huber M, Roth A (2012) Operational TanDEM-X DEM calibration and first validation results. ISPRS Journal of Photogrammetry and Remote Sensing, Theme issue "Innovative applications of SAR interferometry from modern satellite sensors", DOI:http://dx.doi.org/10.1016/j.isprsjprs.2012.06.002
52. Wessel B, Gruber A, Huber M, Roth A (2009) TanDEM-X: block adjustment of interferometric height models. In: Proceedings of ISPRS Hannover workshop "High-resolution earth imaging for geospatial information", Hannover, Germany
53. Mikhail EM (1976) Observations and least squares. IEP, New York
54. TD-GS-PS-0021, TanDEM-X ground segment DEM products specification document, Issue 1.7, https://tandemx-science.dlr.de/. Accessed 9 Mar 2012
55. Rodriguez-Cassola M, Prats P, Schulze D, Tous-Ramon N, Steinbrecher U, Marotti L, Nannini M, Younis M, Lopez-Dekker P, Zink M, Reigber A, Krieger G, Moreira A (2012) First bistatic spaceborne SAR experiments with TanDEM-X. IEEE Geosci Remote Sens Lett 9:33–37
56. Baumgartner S, Krieger G (2011) Large along-track baseline SAR-GMTI: first results with the TerraSAR-X/TanDEM-X satellite constellation. In: Proceedings of the geoscience and remote sensing symposium (IGARSS), Vancouver, Canada
57. Lopez-Dekker P, Prats P, De Zan F, Schulze D, Krieger G, Moreira A (2011) TanDEM-X first DEM acquisition: a crossing orbit experiment. IEEE Geosci Remote Sens Lett 8:943–947
58. Börner T, de Zan F, López-Dekker P, Krieger G, Hajnsek I, Papathanassiou K, Villano M, Younis M, Danklmayer A, Dierking W, Nagler T, Rott H, Lehner S, Fügen T, Moreira A (2010) SIGNAL: SAR for ice, glacier and global dynamics. IGARSS 2010, Honolulu, Hawaii
59. Cloude SR, Papathanassiou KP (1998) Polarimetric SAR interferometry. IEEE Trans Geosci Remote Sens 36:1551–1565
60. Papathanassiou K, Cloude S (2001) Single-baseline polarimetric SAR interferometry. IEEE Trans Geosci Remote Sens 39:2352–2363
61. Kugler F, Hajnsek I, Papathanassiou K (2011) Forest parameter characterisation by means of TerraSAR-X and TanDEM-X (polarimetric and) interferometric data. In: Proceedings of PolInSAR, Frascati, Italy
62. Krieger G, Hajnsek I, Papathanassiou K, Younis M, Moreira A (2010) Interferometric synthetic aperture radar (SAR) missions employing formation flying. Proc IEEE 98:816–843

Chapter 14
Cartwheel

Didier Massonnet

Abstract Radar interferometry gained a lot of interest in very few years in the 1990s. Its main product: the interferogram, is a map of the difference of the phases of two radar images acquired on the same site with a time elapsed between the takes that can range from zero to several years. The phase maps are ambiguous, like contour lines that would not carry any number. They need to be "unwrapped" (i.e. assigned a number) which is generally done by continuity. Several SAR system concepts have been dedicated to interferometry. The Cartwheel concept aims at maximizing the interferometric return of a conventional satellite by adding a cheap constellation of receive-only micro-satellites in a special, and very efficient, orbital configuration, which disturbs the least, or not at all, the Transmitter. Here we describe the system from its design to some specificities of the processing of the products, some of them offering unique new capabilities. Finally, we suggest some more advanced uses of the design.

14.1 Rationale

Interferometric techniques were very successful as soon as space borne radar data became available [1–3]. At that time, these radar systems had not been specifically designed for it and it was quite natural to think of optimizing systems for interferometry.

Repeat pass interferometry demands a strict repetition of the orbital path, typically to within 1 km, and a repetition of the mean Döppler to a small fraction of the antenna pattern or its equivalent in the frequency domain. These conditions were met by the standard orbital and attitude correction laws applied to the radar satellites so that the only improvement would be to have appropriate orbital cycles

D. Massonnet, Ph.D.hab (✉)
CNES, Toulouse Space Center, 18 Av. E. Belin, 31401 Toulouse, France
e-mail: didier.massonnet@cnes.fr

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_14,

depending on the nature of the targets. The interferometric products then record parallax form the topography, proportional to the orbital separation, changes in the atmosphere radio electric depth, attached to a given time of observation, and displacements of the terrain occurring in the time elapsed between the recording of the interferometric pair of images. In addition, the uncertainty of the orbital parameters could create "orbital fringes" that can easily be removed by a simple post-processing.

For topographic applications, the best is to remove as completely as possible the contributions changing with time: simultaneous interferometry is the solution. The SRTM mission [4] implemented two radar antennae separated by a 60-m long mast. Despite its success in global topography, the mission was costly, limited in duration due to Space Shuttle constraints, and the products difficult to process due to mast movement artefacts. One can think of using two identical satellites in close formation but, given the cost of a complete satellite, it could be preferable to launch the second later (to increase mission duration) or to launch it in another local hour (to increase the revisit capabilities of the mission). Besides, they are the risk of collision and EM pollution with costly assets. Finally, the relative position of two free flying satellites is appropriate only on parts of the orbit and the time delay, while small, is not zero if both satellites are active. They could even cause ground station bottleneck if both send telemetry at the same time.

A cheaper solution is to use a second, passive satellite. It could be almost identical (having just its transmitter shut down) which has the same drawbacks but the EM pollution and potential time delay artefacts. It could also be radically simplified: but the simplification is likely to adopt an antenna much smaller than the one of the Transmitter, especially in azimuth, otherwise the receiver would be as bulky and difficult to launch as the Transmitter, and its cost might follow.

If now the receiver features a smaller antenna, we face the difficulty of having to combine two radar images with different parameters, in particular a much higher azimuth ambiguity ratio on the receiver's side. It is a rule of thumb in interferometry that it is better to combine two equally "bad" images than a "good" and a "bad" image. In addition we still have the risk of collision: this time between a high cost active satellite and a cheap bug flying around. The constraints of the relative positions of two satellites also remain.

The novelty of the Cartwheel [5] is to have several passive satellites and to combine only their images. This introduces a radical change in the problem. The main satellite (which we will call the Transmitter) is used only as an illuminator. It can fly far ahead or behind the set of receivers and the collision risk is removed, at least with the Transmitter. Interferometry is done with lower grade images, but sharing identical features. The set of receivers can be disconnected from the transmitter in terms of project. It can be launched later when we are sure the Transmitter works reliably. The relation between the Transmitter and the receivers can be very loose: it suffices that the receivers are informed of the Transmitter's program in advance is order to known where to aim their antennae. To the extreme, the Transmitter might not be aware of the existence of the receivers which, by the way, are very discrete. They could still operate if the

Transmitter is jammed by an ill-intentioned ground signal. If we accept the principle of a "freely designed", reflector based antenna, the receiver can be generic to a high degree: one can imagine changing a few sub-system to adapt from C-band to L-band from a generic design that makes up more than 90% of the cost.

14.2 Orbital Configuration

Once we accept the fact that the interferometric products will be obtained from comparison between the records of the passive companions, we must choose their orbital configuration. Working with only two receivers reduces the operational value of the concept. The two receivers will keep changing their relative baseline (both horizontally and vertically) so that a global topography mission would have to operate by processing latitudes bands that should be changed during the mission. Therefore we propose using at least three receivers (remember they are built to be cheap) in the following way. We assume that the receivers start at the same position as the Transmitter. Giving an impulse to one of them along the track will give it a higher apogee half an orbit later. A second impulse can then recover the initial semi-major axis of the Transmitter orbit. In these conditions, the receiver is still synchronous with the Transmitter (same orbital period), but seems to describe a vertical ellipse centred on the Transmitter, with a horizontal axis twice the vertical axis. This ellipse is described during an orbital period with a constant angular velocity as seen from the Transmitter. We can do the same with the two remaining receivers, but after one third (resp. two third) orbital period delay. If the ellipse were a circle, then the three receivers would draw an equilateral triangle in it. Finally, rather than starting from the Transmitter's position, we start a few tens of kilometres ahead of behind it. Figure 14.1 gives an idea of the configuration. Figure 14.2 gives a sketch of the orbital configuration where the summits of one of the triangle are the receivers. Figure 14.2 can be drawn first as a circle with an equilateral triangle inside it, then be extended by a factor of two in the horizontal direction. Figure 14.2 sketches the two extreme vertical baselines (thicker line) corresponding to the height of the triangle aligned vertically or the side of the triangle aligned horizontally. If r is the vertical half axis of the ellipse, the two vertical baselines are respectively $1.5r$ and $r\sqrt{3}$. It is striking to observe that these values never differ by more than $\pm 7.5\%$ from the value of their geometric average. We created a virtual mast with no displacements other than the smooth and small changes easily predicted by orbital mechanics!

The orbital configuration does not cost much. A 1 km vertical half axis requires 0.5 m/s of total impulse. The configuration can therefore be changed or undone several times during the mission.

Many alternate orbital configurations have been proposed to create adequate stable baselines. Some propose a slight change of inclination (creating a "flat" ellipse), sometimes combined with eccentricity (producing a slanted ellipse). It seems that, unlike eccentricity, any change of inclination requires some fuel for

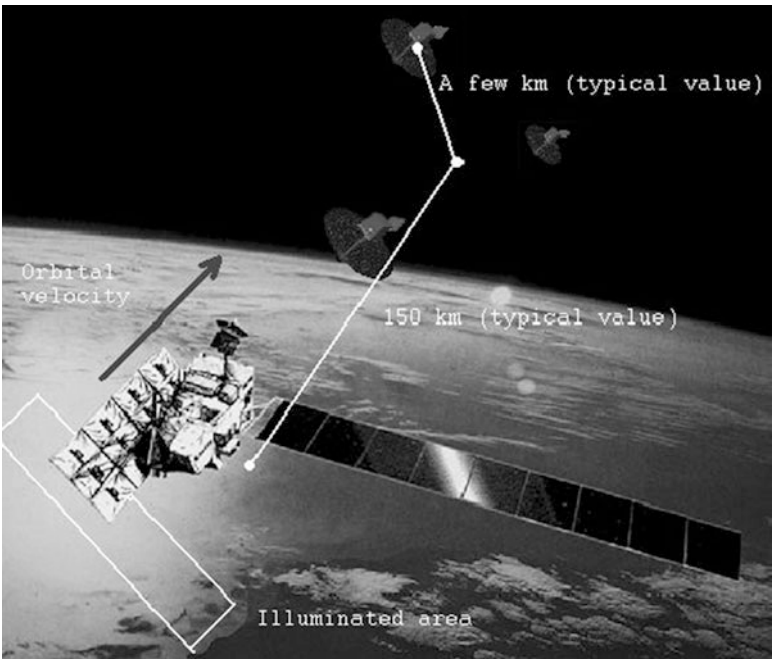


Fig. 14.1 Artist's impression of the Cartwheel configuration

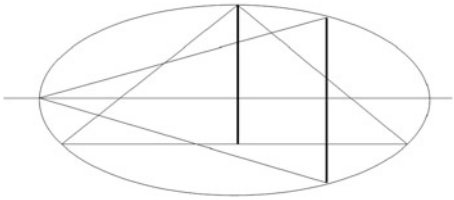


Fig. 14.2 Cartwheel basic geometry

correction, in addition to the fuel needed for compensating atmospheric drag. The "vertical" ellipse of the Cartwheel seems to be the only fully stable one without systematic drift.

Detailed studies show that the perigee of any given satellite will not remain at the same place as time goes. There is a slow secular motion that rotates the Cartwheel and creates typically one additional rotation every 4 months.

Finally, if we consider that the set of three satellites "rolls" on the orbital path, it seems to roll counter clockwise as the lower satellite runs faster than the centre of the constellation. This feature reminded me the cartwheels used by the pioneers in traditional western movies, where stroboscopic effects often created this counter clockwise appearance, hence the name of the design.

14.3 Specificity of Processing

The products of the Cartwheel have specific features. First the time of pulse emission by the Transmitter is not known, nor, in theory, its range sampling rate. Second the level of azimuth ambiguity is larger than in the Transmitter's image, due to the generally smaller azimuth size of the antenna. Third the power of the signal is smaller due to a generally smaller antenna size. Finally, if we consider that the Cartwheel cruises at safe distance from the Transmitter (typically 100 km), the mean Döppler of the receiver's images will be large.

The two last points do not pose serious problems as linear range migrations can be compensated for in the initial step of radar processing at no computing nor memory cost. The signal power might be a problem especially if super resolution is sought. However the global antenna surface of the constellation may remain higher than the one of the Transmitter. Preliminary industrial studies of the concepts featured circular reflector antennae with 2.5 m diameter, which totals almost 15 m^2, equal or larger than the antennae of ERS1 or RADARSAT.

The synchronicity with the Transmitter must be addressed. Each local oscillator within the Transmitter or any receiver is a clock with its own specific features, that is a basic frequency and a drift. The problem exists to some extent with other systems such as ERS1-2 tandem data. However, the difference can be larger here. In addition, the start time of the pulse emission is not known to the receiver, nor the pulse repetition frequency if the Transmitter did not provide it. In the basic Cartwheel design, there is a continuous recording which include the pulse of the Transmitter. Of course the receiver and the Transmitter do not aim at each other, but the direct pulse is very strong and will be recorded despite the poor diagram budget. If necessary, an additional feeder can be added in the direction of the Transmitter. Since each receiver records the Transmitter's pulse and since their relative positions are very well known, a perfect reconstruction of the signal is possible. Data compression can be implemented to avoid recording blank periods in the continuous recording (i.e. no pulse, no signal). Also keeping a record of all pulses is not required.

The high level expected from azimuth ambiguities is a potential problem for image quality. Let us assume that the Transmitter protects against azimuth ambiguity by its large antenna combined with its illumination law by, say, 24 dB. Then the small antenna of the receiver might offer no protection on the way back, reducing the protection to 12 dB. Nothing can be done to kill the ambiguous "energy" which will be poured in the receiver's image. However, we shall explain why ambiguities are not likely to contribute to any coherent combination. Therefore the interferometric products will not be affected by their higher level. The reason lies with the migration phenomenon, a well known parameter in SAR processing [6]. "Parabolic" migrations and "linear" migrations have to be corrected during the processing in order to align the contributions of the given target on a single azimuth line at a given range for further processing. An azimuth ambiguous contribution behaves in a similar way as a nominal target except in the linear part of the migrations. As a consequence the ambiguous contribution will be spread on several

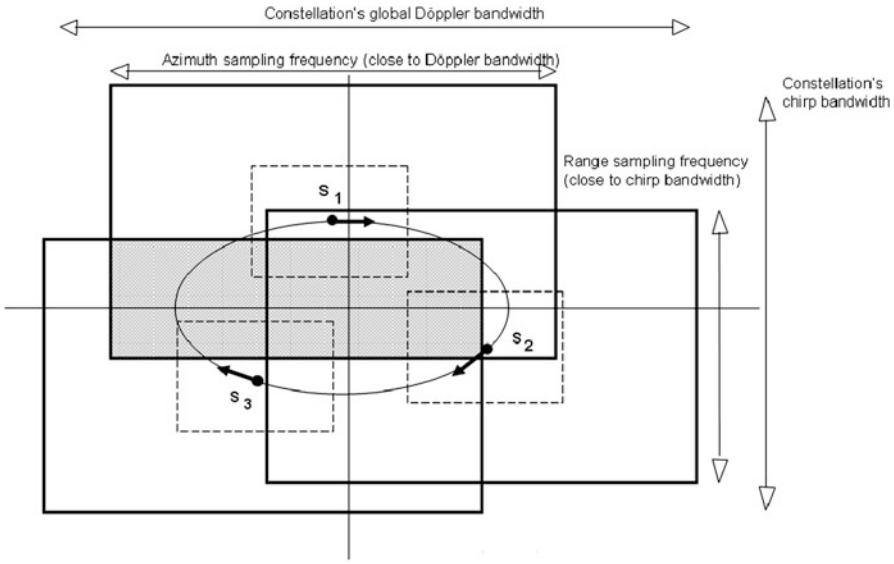


Fig. 14.4 Frequency plane

range gates after the migrations have been corrected for nominal contributions. It is easy to understand that the range resolution of the ambiguous target will be degraded because of this spreading which blurs it in range. However, since the contribution crosses several range gates during azimuth processing, it never sees the full azimuth bandwidth in a single range cell, but only a part of the "azimuth phase history". Much like what is done for multi-looking, the resolution is degraded in azimuth as well. As we will see in the next section, the degradation of the resolution in range and azimuth will reduce the typical area made from the combination of the critical baselines (see the smaller, dotted rectangles in Fig. 14.4 compared to the nominal rectangles surrounding the receivers). The choice of the baseline might be such that the nominal rectangles intersects, allowing for phase connection and coherent processing, while the smaller rectangles attached to ambiguous contributions do not. The ambiguous contributions might be brighter in the receiver's images than in the Transmitter's, but they will not contribute at all in the interferometric products. Similarly, in case of super resolution products they will not see their resolution improved.

14.4 Products

Indeed, the products of an interferometric Cartwheel system do not consist only in across track interferometry but may include along track interferometry [7], super-resolution in azimuth and super-resolution in range. Volume scattering is also

accessible as the loss of coherence in the product is linked solely to the geometric feature of the target rather than any time-dependent change.

We will not detail the features of across track or along track interferometry with the associated metric (orthogonal baseline, altitude of ambiguity, speckle time stability for ocean applications). It is however, worth mentioning that interferograms can be formed not only with the two optimal images: that is the ones that form the selected stable baseline at a given time, but also with the image of the third receiver. The third image is a "free" bonus that can help for instance, unwrap the phase pattern of the nominal interferogram. A raw (i.e. not unwrapped) interferogram shows fringes that are not numbered. In other terms there is an unknown integer number hidden behind each fringe. This number is usually guessed by reconstructing it by continuity across the image, an operation called phase unwrapping [8], with a high probability of errors. If an existing terrain model is accurate to within the difference of height represented by a fringe, it can be used to create and then subtract a fringe model that brings the interferogram to values within a fringe, where no ambiguity exist anymore, thus allowing to refine the terrain model. Noting B the (almost) fixed Cartwheel vertical baseline, the third image can combine to the two other images with baselines the sum of which is B. For topographic applications at a given incidence angle, the altitude of ambiguity h (i.e. the change in altitude that corresponds to a full topographic fringe) is inversely proportional to the baseline, so that, labelling the images by the receiver's number and the interferograms by their two images, we have:

$$h_{13} = \frac{1}{\frac{1}{h_{12}} + \frac{1}{h_{23}}}$$

Two wrapped interferograms can be added or subtracted, but can only be multiplied by an integer factor, in order for the unknown fringe number to remain an integer. For instance if we multiply an interferogram by 2, an additional fringe will be created between each existing fringes, the initial fringe numbers will turn even and the intermediate, new fringes will have odd numbers. Let us assume that any interferogram can be multiplied by $\{2,3,-1,-2,-3\}$ (larger integer factors may increase the noise too much). By adding the results we can create any altitude of ambiguity h such that, if j and k belong to the above integer ensemble:

$$h = \frac{1}{\frac{j}{h_{12}} + \frac{k}{h_{23}}}$$

It can be shown [9] that the choice of h is so rich that a match can be found with the accuracy of any available terrain model, practically removing the need for unwrapping and the associated errors. This trick is especially appealing with

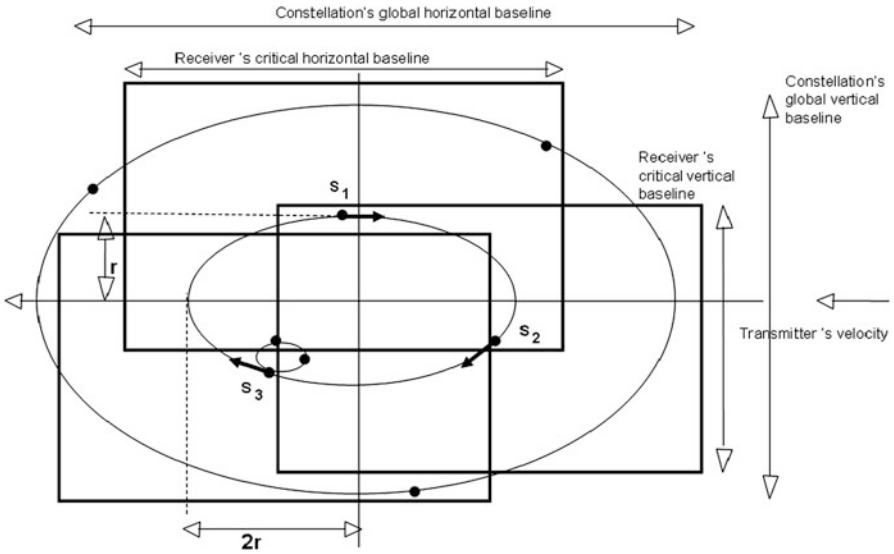


Fig. 14.3 Orbital plane

simultaneous interferometry, where fringes are purely topographic and do not include signals such as from the atmosphere, that are unrelated to the baselines.

The independence of two images acquired simultaneously from two-point with an along-track separation is characterized by the critical horizontal baseline. It is such that the range difference between the receivers and a given ground target changes by one wavelength from one pulse to the next. For coherent combination the along track separation of two receivers must stay below this critical horizontal baseline. The same way, we define a critical vertical baseline such that the range difference seen by the second receiver, across the range pixel of the first receiver, differs from the range pixel size by one wavelength. Again, for coherent combination of the passive images made by each receiver, the vertical separation of the receivers must stay below this critical vertical baseline.

We consider the rectangular area in the plane containing the orbit of the satellites, centred on each satellite, and having the critical vertical baseline as a height and the critical horizontal baseline as a width. The resolution of an interferometric product results from the coherent part common to the intersecting areas of the two receivers involved in the combination. It is inversely proportional to the respective sizes of the intersection in azimuth and range. Similarly, the resolution of the combined images of two receivers with intersecting areas is inversely proportional to the respective sizes of the union of their areas in azimuth and range. The situation is illustrated in Fig. 14.3. To summarize, if ρ is the nominal resolution of the Transmitter in range, and if α is the proportion of non-overlapping of the areas of two receivers in range, with $0 < \alpha < 1$, the resulting interferogram resolution

will be $\frac{\rho}{1+\alpha}$. The same formula applies in azimuth. The result of "super-synthesis" in range and azimuth is most efficient if the rectangular area is twice as large in azimuth, in order to match the orbital ellipse ratio. Since any super resolution processing needs to work coherently, i.e. with images connected in phase, the area of any image involved in the reconstruction shall intersect at least with one other image. The achievable resolution is therefore at most twofold better than the initial resolution, except with a richer Cartwheel configuration (next section).

The relationship between the receiver's areas and the resolution of radar products comes from the fact that the geometric Fig. 14.3 can be turned into an equivalent frequency Fig. 14.4, where the critical baselines are replaced by the corresponding bandwidths of the pulse repetition frequency (in azimuth) and chirp content (in range). For the sake of simplicity we do not distinguish here between the actual bandwidths and the corresponding sampling frequencies adapted to them. In Fig. 14.4 the shaded area represents the resolution of the interferogram formed between images from receivers 1 and 3. In this example, its azimuth resolution would be close to the one of the original images, while its range resolution is degraded by about 2.5 due to a bandwidth reduced to 40% of the nominal chirp.

14.5 Advanced Applications

Although more complicated architectures are unlikely to fly before a first, basic Cartwheel system is launched, such architectures are very appealing.

An advantage of using a synthetic band in range is the possibility to go beyond the bandwidth allocation for any given bandwidth. This advantage is specially true for the wavelengths with the smallest bandwidth allocation. Following a system which uses all the allocated bandwidth, for instance 80 MHz in L-band, with a range super-resolution factor of four (which requires at least six micro-satellites for permanent operations, orbiting on two concentric Cartwheels, see Fig. 14.3), would create images that literally should not exist. In particular, very unusual ratio between final range and/or azimuth pixel size and the wavelength would be achieved. Applying a super resolution factor of almost 4 on a 80 MHz is equivalent to 300 MHz, which in turn is about a quarter of the carrier frequency. In this example, the above ratio is about 4, meaning that phase unwrapping could be achieved by exploiting the results of image correlation (a correlation accuracy down to 25% of a pixel size seems easy to achieve). This would also give access to some new textural information and might revert some of the traditional preference of the military toward shorter wavelengths.

Departing from the initial concept calling for using an opportunistic Transmitter, we could imagine a specially designed Transmitter unable to receive its own signal, equipped with a large antenna capable of implementing a very strong protection against azimuth ambiguities. This "transmission only" satellite could be easier to build because it does not have to mix strong and weak signals; it has no data telemetry; it is not discrete but insensitive to jamming.

Similarly, the Cartwheel design could efficiently implement full or partial polarimetry using a Transmitter not necessarily fully polarimetric. This aspect can be very powerful if combined with the volume scattering assessment. In simultaneous interferometry, the blurring of the fringes, measured mathematically by the coherence, can only be due to a weak signal, which can be confirmed or not by the amplitude of the images, or by volume scattering. In the later case and taking a very simple point of view, if the radar signal penetrates the vegetation by 3 m while the baseline creates a 12 m altitude of ambiguity (values taken as examples), the fringe value will wobble within each pixel by up to 25% of a cycle, resulting in partial blurring. The analysis of this blurring (i.e. the coherence), possibly reinforced by the use of the third image which generally gives two additional altitudes of ambiguity, in addition to polarimetric capabilities, is likely to give unprecedented insight of the vegetation structure and properties [10].

Finally, for environment survey, the optimal azimuth separation for azimuth super resolution is generally not optimal for ocean currents monitoring by along track interferometry. We already mentioned that a smaller Cartwheel of appropriate size can be created from the nominal one at moderate fuel cost. However if we want both missions permanently operated, then we could add two more receivers and create a small, additional Cartwheel using one of the receiver of the large Cartwheel (see Fig. 14.3 around s_3).

With the Cartwheel design, a variety of missions can be generated from a relatively simple brick that can itself be generic to a high degree.

References

1. Massonnet D, Rossi M, Carmona C, Adragna F, Peltzer G, Feigl K, Rabaute T (1993) The displacement field of the Landers earthquake mapped by radar interferometry. Nature 364:138–142
2. Goldstein R, Engelhardt H, Kamb B, Frohlich RM (1993) Satellite radar interferometry for monitoring ice sheet motion: application to an Antarctic ice stream. Science 262:1525–1530
3. Massonnet D, Feigl KL (1998) Radar interferometry and its applications to changes in the earth's surface. Rev Geophys 36(4):441–500
4. Farr TG, Kobrick M (2000) Shuttle radar topography mission produces a wealth of data. Am Geophys Union EOS 81:583–585
5. Massonnet D (2001) Capabilities and limitations of the interferometric cartwheel. IEEE Trans Geosci Remote Sens 39:506–520
6. Massonnet D, Souyris JC (2008) Imaging with synthetic aperture radar. EPFL Taylor and Francis, Boca Raton
7. Goldstein R, Zebker H (1987) Interferometric radar measurements of ocean surface currents. Nature 328:707–709
8. Ghiglia D, Pritt MD (1998) Two-dimensional phase unwrapping: theory, algorithms and software. John Wiley, New York
9. Massonnet D, Vadon H, Rossi M (1996) Reduction of the need for phase unwrapping in radar interferometry. IEEE Trans Geosci Remote Sens 34(2):489–497
10. Cloude S, Papathanassiou K (1998) Polarimetric SAR interferometry. IEEE Trans Geosci Remote Sens 36:1551–1565

Chapter 15
Sabrina

Antonio Moccia, Marco D'Errico, Alfredo Renga, and Giancarmine Fasano

Abstract SABRINA mission was conceived as a dual satellite mission based on COSMO/SkyMed constellation to perform and exploit bistatic Synthetic Aperture Radar in both interferometric and large baseline modes. Analysis of identified application and techniques are presented along with the relative trajectory selection, pointing strategies and safety.

15.1 Introduction

SABRINA (System for Advanced Bistatic and Radar INterferometric Application) mission concept was based on BISSAT (Bistatic and Interferometric Sar SATellite) satellite flying in formation with one of the satellites of the COSMO/SkyMed constellation. BISSAT was supposed to be equipped with a X-band Synthetic Aperture Radar (SAR) able to work in bistatic configuration with the COSMO/SkyMed one. The system was conceived to improve and complete measurement capabilities of monostatic SAR in large-scale Earth observation. The orbital planning and the formation flying control were designed to realize different observation geometries characterized by inter satellites distances, baseline, variable from hundreds meters to some hundreds of kilometers. In a so vast scenario, a wide range of bistatic techniques are applicable and testable, ranging from cross and along track interferometry to large baseline bistatic observations.

A. Moccia (✉) • A. Renga, Ph.D. • G. Fasano, Ph.D.
Department of Aerospace Engineering, University of Naples "Federico II",
P.le Tecchio 80, 80125 Naples, Italy
e-mail: antonio.moccia@unina.it; alfredo.renga@unina.it; g.fasano@unina.it

M. D'Errico, Ph.D.
Department of Aerospace and Mechanical Engineering, Second University of Naples,
Via Roma 29, 81031 Aversa (CE), Italy
e-mail: marco.derrico@unina2.it

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_15,

Initially the BISSAT mission had been thought as a small mission with an experimental nature [1]. The first BISSAT concept consisted of a small satellite, derived from the Italian MITA [2] bus (100–300 kg mass range), and from a receiving-only radar payload designed downsizing the sensor of COSMO/SkyMed mission. During the following years, and specifically during SABRINA phase 0/A study, the idea of a wider range of interferometric and bistatic experiments and applications was investigated [3], leading to an alternative definition of the BISSAT satellite relying on PRIMA platform (400–1,500 kg mass range) [4] equipped with a radar system offering additional features with respect to the first generation COSMO/SkyMed architecture. In this context the experimental nature of the mission can be matched to the dual and industrial nature of the COSMO/SkyMed program. SABRINA mission was then shifted to be a potential enhancement of COSMO/SkyMed second generation.

The chapter is organized as follows: first of all potential applications and techniques will be analyzed for both small (allowing for interferometric techniques) and large distances (signal phase difference cannot be exploited) in Sects. 15.3 and 15.2, respectively. Then, design of relative trajectory will be summarized along with an analysis of pointing strategies (Sect. 15.4). Finally, tradeoffs and safety considerations will be presented (Sect. 15.5).

Part of this study was conducted with the financial support of the Italian Space Agency and of Thales Alenia Space Italia.

15.2 Large Baseline Bistatic Applications and Techniques

The Large Baseline Bistatic (LBB) phase of the SABRINA mission envisages the collection of monostatic-bistatic data with separations between the monostatic and the bistatic sensors larger than the critical baseline. In this case the generated couples of monostatic-bistatic SAR images are phase decorrelated and only the amplitude differences between the images can be taken into account (see the Chaps. 1 and 2 for the details about the critical baseline). Such incoherent combinations are used, in the framework of SABRINA mission, to investigate and experiment a wide range of LBB techniques. The following subsections report a description of the techniques and individuate potential products and applications.

Before analyzing more into details each LBB technique it's worth underlining some useful characteristics of LBB observation geometry. For the sake of simplicity Fig. 15.1 is used as a reference, where Θ_{Tx} is the incidence angle of the transmitted signal, Θ_{Rx} is receiver incidence angle, Φ is the out-of-plane angle, and β is the bistatic angle. In addition, Z is the local vertical to the target area and YZ is the range elevation plane of the monostatic transmitting antenna.

Observation geometries close to the in-plane condition (i.e. $\Phi \cong 0°$ or $\Phi \cong 180°$) with both bistatic and monostatic sensors working in side-looking mode are able to significantly minimize the pixel skewing effect generated by bistatic acquisition [5], thus improving the quality of the matched monostatic-bistatic couples. In addition,

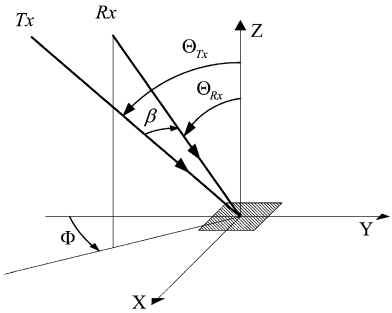


Fig. 15.1 Illustration of the main LBB observation parameters

according to [6], an improvement of in-plane scattering coefficients can be achieved if $\Theta_{Tx} > \Theta_{Rx}$. This combination is met by both same-side ($\Phi \cong 0°$) and opposite side ($\Phi \cong 180°$) geometries. However, bistatic ground range resolution is notably degraded in opposite observation [5] and bistatic range ambiguities can appear [7] depending on local surface relief variations, making very difficult the distinction of scattered signals from different regions of the illuminated area.

As far out-of-plane geometries are concerned, they are considered low-priority in the framework of the SABRINA mission, due to the meaningful expected signal drop-off, the relevant image distortions and the lack of significant literature references (in particular with respect to in-plane LBB observations). On this basis all the selected LBB techniques to be investigated and experimented are thought for observation geometries close to in-plane condition, in which the out-of-plane baseline component is a small fraction and cross-track and radial ones.

Table 15.1 summarizes the results of the performed LBB analysis and lists, for all the selected techniques, the relevant applications, products and impact on system definition. Further details can be found in following sub-sections.

15.2.1 Radargrammetry of LBB Radar Signals

The application of stereo-radargrammetric techniques to monostatic-bistatic SAR data gathered with large ($>10°$) bistatic angles represent a robust approach for Digital Elevation Model (DEM) generation on wide areas. Actually monostatic-bistatic radargrammetry is able to avoid temporal decorrelation, and unlike SAR interferometry (InSAR) does not work with phase domain data so it is not affected by phase unwrapping problems and it is characterized by relaxed relative positioning requirements. For instance, 1 m baseline component error is assumed in the following, that is about two orders of magnitude less than typical requirements for InSAR.

The actual possibility to combine monostatic and bistatic data reflected by common covered areas or targets depends on the possibility to co-register the images, that is to project and resample them onto some common reference systems.

Table 15.1 Summary of LBB SAR techniques and application

Technique	Application	Products	Main system parameter (operative mode, baseline...)	Starting standard product
Radargrammetry of LBB radar signals	Topography	DEM	– Stripmap with baseline longer than 130 km – Baseline uncertainty 1 m	Raw and/or SLC[a]
Velocity measurements by Doppler analysis	Oceanography Velocity field	Velocity map	– Stripmap with baseline longer than 130 km – Satellite antenna velocity uncertainty 0.05 m/s – Doppler centroid frequency uncertainty 1 Hz – Attitude angle uncertainty 10^{-3}	Raw and/or SLC
High-resolution measurements of comp. of sea wave spectra	Oceanography	Wave-motion map	– Stripmap with baseline larger than 200 km	SLC and/or higher level
Investigation of RCS in LBB radar signals	Recognition and Identification	– Surface slope map – Surface roughness map – Discrimination map among rural, urban and forest areas	– Stripmap and spotlight – Minimum useful range of bistatic angles 5°–20°	SLC and/or higher level
Attitude determination by using LBB-SAR raw data	Image quality	Measurement of attitude angles	Stripmap	Raw and/or SLC
Improvement of classification and pattern recognition procedures	Image quality	Improvement of quality of COSMOSky-Med products	Stripmap and spotlight	SLC or higher level data

[a]single look complex

Co-registration accuracies of 1/10 of an image pixel or higher are routinely achieved in interferometric, i.e. short baseline, applications. However LBB geometry represents a completely different, and new, situation with respect to interferometric case. In a LBB scenario well-assessed interferometric co-registration algorithms can fail because of the different LBB scattering properties. Moreover the effect of local and irregular slope can make heavier the risk of a failure in the co-registration processes. This means that novel algorithm shall be developed. The same co-registration accuracy as in InSAR can be therefore obtained only on particular areas characterized by favourable bistatic angles, topography and surface characteristics, but performance degradation, up to 1/4 of an image pixel, must be taken into account for LBB co-registered products.

From a practical point of view, and as reported in the Chap. 1, two different approaches exist for surface relief determination from LBB data:

– *Parallax-based methods*, which principle is borrowed from optical photogrammetry.
– *True stereo radar methods*, which rely on the peculiarities of radar and SAR acquisition.

An error budget has been developed for both methods, considering COSMO/SkyMed operating in nominal stripmap mode [8] with 25°–50° incidence angles and the bistatic sensor collecting bistatic echoes from in-plane same-side geometry with the bistatic angle ranging from 10° to 45° (50–550 km baseline). The results show that the performance of parallax-based methods is strongly dependant on the capability to reduce co-registration errors: only decametre-scale height accuracy can be robustly achieved. On the contrary metric scale (<4 m) height accuracy can be obtained by true stereo methods with the performance very weakly influenced by baseline length, transmitting and receiving incidence angles and image matching errors.

The reported results obviously refer to uncertainty of the measure or relative accuracy. As far as the absolute error of the generated DEM is concerned, considering the bistatic single-pass scenario and the proper utilization of a limited number of ground control points, it is reasonable to assume less than 20 m absolute height errors with improved accuracy, lower or close to 10 m, under favourable conditions.

According to Fig. 15.2, where U.S. National Imagery and Mapping Agency (NIMA) specifications on DEM accuracy [9] are reported, the possibility to comply with high-resolution terrain information (HRTI)-3 specifications on particular areas is clear, whereas NIMA digital terrain elevation data (DTED)-2 can be met under less restrictive condition and, hence, on global scale.

Finally, it is important to point out that the reported results are derived with reference to the nominal COSMOSkyMed stripmap mode and a further performance improvement at the cost of a reduced covered area can be guaranteed by the exploitation of higher resolution spotlight monostatic-bistatic image pairs.

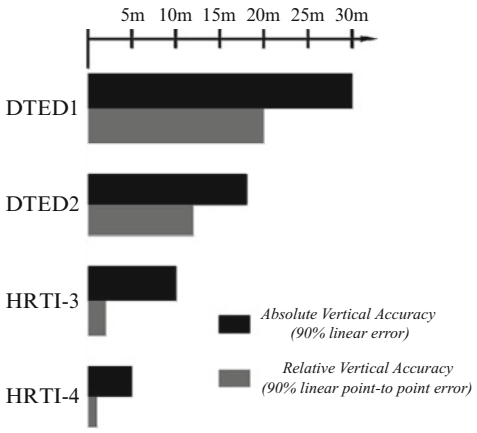


Fig. 15.2 Standard levels for DTED and HRTI provided by NIMA specifications

15.2.2 Velocity Measurements by Doppler Analysis

A procedure to compute the two slant range components of the target velocity is described in Ref. [10]. The procedure, different from along-track interferometry, is based on the computation of monostatic and bistatic Doppler centroid frequencies and in the equations relating the centroids to the parameters of monostatic-bistatic geometry, such as position/velocity vectors of both monostatic and bistatic sensors and monostatic/bistatic slant ranges.

The application of the technique is subsequent or simultaneous with the solution of the stereo intersection problem, therefore similar to the previous section are the constrains about acquisition geometry and baseline determination accuracy. In addition, the adaptation to the SABRINA mission of the error budget analysis carried out in Ref. [10], assuming an uncertainty of 0.05 m/s about monostatic and bistatic satellite antenna velocity, an uncertainty of 1 Hz about monostatic and bistatic Doppler centroid frequency and 0.01° attitude angle uncertainty, leads to target velocity measurement uncertainty lower than 1 m/s.

More attractive applications of this technique are oceanographic, being possible to generate velocity and direction maps of sea wave. The most suitable operative mode is stripmap, even if monostatic bistatic acquisitions can be used to derive higher resolution data.

15.2.3 High-Resolution Measurements of Components of Sea Wave Spectra

Oceanographic applications can take advantages of high-resolution wave-motion maps derived by exploiting some theoretical features of LBB geometry.

Actually LBB observation can potentially offer, with respect to its monostatic counterpart [11]:

- A different Bragg-resonant sea-wave cross-track component (thus providing an additional information on sea state).
- A wider range of sea wave spectrum, in which an approximately linear SAR modulation transfer function is applicable for removing the azimuth shift between a scattering element real position and its imaged position (due to the mean value of the radial component of the long wave orbital velocity in the resolution cell).
- A reduction in azimuth resolution degradation due to instantaneous radial velocity of sea surface facets within a resolution cell during synthetic aperture.

The possibility to effectively attain this performance improvement is closely connected to the implemented formation flying trajectory. Actually, very large baselines and bistatic angles are needed to assure a meaningful improvement of bistatic performance with respect to monostatic one. For the sake of concreteness, according to the models presented in the Chap. 1, a baseline length of 250–300 km, that is a bistatic angle of about 25°, can guarantee a 5% wider range of linearity region and a 10% reduction in azimuth resolution degradation. Similarly, baselines shorter than 150 km lead to performance improvement smaller than 1–2%.

15.2.4 Investigation of RCS in LBB Radar Signals

The difference between monostatic and bistatic radar cross sections (RCS) of land targets represents the starting point of ad-hoc developed techniques for estimating land surface parameters. Specifically, surface roughness and dielectric constant maps can be generated by collecting couples of single-pass monostatic-bistatic SAR data characterized by large bistatic angles ($>10°$). Actually the scattering from smooth surfaces is extremely different when the same smooth area is observed by two different directions, whereas with the increasing surface roughness this variability becomes very weak. In addition, the same principle can be applied to produce discrimination maps between different areas, such are rural, urban and forested, by processing multi-angle LBB acquisitions. Finally LBB reflectivity can be also used for slop determination since it has been demonstrated [12] that LBB scattering coefficient can be related to the root mean square surface slope within a resolution cell.

As noted above, multi-angle acquisitions are needed to investigate RCS in monostatic-bistatic SAR data. More in detail, it is possible to state that useful bistatic angles start from 5°, but from a scientific point of view the need for a better characterization of bistatic radar cross section requires acquisitions at variable bistatic angles from 5° up to 20°, at least.

15.2.5 Attitude Determination by Using LBB SAR Raw Data

Accurate SAR antenna pointing determination can be conducted on the basis of LBB radar echoes. The technique [13] takes advantage of the Doppler characteristics of the received signal to compute pitch and yaw attitude angles of the spacecraft and it relies on echo amplitude analysis for roll. Neither preliminary attitude estimate is needed, nor instantaneous cooperation from the spacecraft carrying the transmitting radar is required. Several applications can be conceived for this technique, but the main expected one is an accurate SAR antenna pointing estimation to improve LBB SAR data focusing. Actually, in order to generate high quality focused SAR images, antenna pointing is very important because it significantly affects critical parameters of synthetic aperture coherent processing (e.g. Doppler band, Doppler centroid, Doppler frequency rate, range migration) and image calibration (e.g. directional antenna gain compensation in the observed swath).

A test of this technique was performed in Ref. [13] based on SIR-C real SAR data with regards to the configuration of coincident receiving and transmitting antennas, therefore in a monostatic repeat-track scenario. The experimentation showed the possibility to achieve root mean square deviations in the order of 0.01° for yaw and pitch angles and of 0.1° for roll angle, but an important performance improvement is expected for single-pass bistatic acquisitions.

The major goal of the experimentation of this technique within the framework of SABRINA mission is to validate a simple technology for SAR antenna pointing estimation. This aims at the validation of an innovative concept for SAR-based space remote sensing, that is equipping the spacecraft with low-cost, medium-performance attitude sensors for mission phases when relaxed requirements exist, to switch to high-accuracy pointing during SAR operation through the exploitation of data produced by the radar instrument. It would be of great benefit from several points of view: cost, number of on-board instruments, power consumption, mass, which are important especially in the case of small platforms. Concerning this, high accuracy attitude sensor embarked on the BISSAT satellite can could be used as reference measurements to definitively validate the technique.

15.2.6 Improvement of Classification and Pattern Recognition Procedures

The quality of standard monostatic COSMOSkyMed products can benefit from the integration with LBB SAR data. First of all, signals from retro-reflectors that are very strong in monostatic images are smoothed in LBB SAR data, so weak signals appear more prominent. Consequently more and different details can be detected in LBB-SAR images thus improving the combined, monostatic-bistatic, mapping capabilities. Moreover the proper exploitation of different polarization channels,

both in transmission and in reception, represent an additional possibility to improve classification and pattern recognition procedures. Finally an increase of the signal-to-noise ratio (SNR) can be obtained by incoherent integration of monostatic-bistatic, phase decorrelated, signals. The improved SNR influences positively some essential image quality parameters, such as geometric/radiometric resolutions and integrated/peak side-to-lobe ratio.

15.3 Interferometric and Polarimetric Applications

The implementation and validation of a wide series of techniques based on the coherent combination of monostatic and bistatic data is planned within SABRINA program. This includes, but is not limited to, along and cross-track interferometry, coherent resolution enhancement, tomography, polarimetric SAR interferometry. Important literature is available about potentials and challenges of such techniques, for these reasons the following sub-sections will focus only on the peculiar aspects of SABRINA, with specific reference to along-track and cross-track interferometry.

15.3.1 Along-Track Bistatic SAR Interferometry

The operation of an along-track interferometer (ATI) is based on the acquisition of two SAR images taken under identical geometry of observation, but separated by a short time interval. When this is accomplished, any difference between the two images results from changes in the scene. In particular, the phase difference between the echoes from the same target allows its radial speed to be measured. A crucial parameter of an along-track InSAR is the time lag between the two images, which is determined by the along-track distance between the two antennas, the platform velocity and the mode of operation of the transmit/receive chains [14]. Specifically, in order to avoid ambiguities and issues related to phase unwrapping procedures, phase differences must not exceed 2π, and this establishes a constraint on the most suitable baseline. In addition phase noise puts a limit on the minimum detectable velocity.

Several applications can take potentially take advantage of this technique [3]:

- *Marine science studies and climate modeling*—in order to obtain clear signatures of surface current variations, the incidence angle of an ATI system should be as high as possible. Ideal incidence angles for current measurements by ATI with reasonable nonlinearities and coherence should be between about 35° and 45°. Sea surface current monitoring is also important to understand the influence on the weather and the global climate.
- *Flood*—a flood map presents the extension of an area which has been flooded during a specific high discharge event. Such information is very useful for civil protection activities because it immediately provides the list of areas and

Table 15.2 Applications, products for ATI techniques

Application	Product	Baseline min (Speed)	Baseline max (Speed)	SAR operative mode
Marine studies	Current monitoring	15 m (28 km/h)	40 m (11 km/h)	ScanSAR
Flood	Flood velocity map	10 m (70 km/h)	15 m (28 km/h)	Stripmap/ScanSAR
Flood	Flood extension map	40 m (11 km/h)	400 m (1 km/h)	Spotlight/Stripmap/ScanSAR
Traffic	Traffic monitoring	12 m (50 km/h)	20 m (20 km/h)	Spotlight/Stripmap
Fisheries	Ship detecting	20 m (20 km/h)	40 m (11 km/h)	Stripmap/ScanSAR
Fisheries	(Slow) Ship monitoring	200 m (2 km/h)	400 m (1 km/h)	Stripmap/ScanSAR

structures struck by similar events. Classification, segmentation and composite color images techniques can be improved by exploiting the interferometric coherence between a couple of SAR images. ATI can also contribute to realize flood velocity map applied to large rivers and streams by the same techniques used for marine application.

– *Fisheries management*—the application of synthetic aperture radar for surveillance of commercial fishing can aid in the detection of illegal fishing activities and provide more efficient use of limited aircraft or patrol craft resources. Many nations have vast economic enterprise zones that cannot be monitored for fishing activities with the available patrol resources. With SAR, large sea areas can be monitored on frequent revisit schedules, thus allowing detected ships to be observed and identified by patrols that are vectored on these targets from cuing derived from the SAR information. Along-track interferometric SAR allows the enhanced detection of moving targets on the ocean surface including ships and ship wakes. In this context, it is important to remark that the velocity measurements provided by along-track In-SAR are independent of the RCS. Consequently, InSAR allows detection of small ships which may not be much brighter than the background ocean clutter. If such ships have a velocity component in the radial direction which is different to that of the surrounding ocean, they will show up as anomalies in the velocity image. An added advantage is that one can estimate true velocities from radial velocities if the ship's heading, which can be determined from either its shape or its wake, is known. In ATI images there will be little correlation of the ocean backscatter in the two images but there will be a large correlation in the backscatter from ships. By using a measure of the correlation between the two acquisitions it is possible to detect ships.
– *Traffic management*—traffic monitoring from space is quite rare, however first experimentations [15–17] have been already carried-out.

Table 15.2 reports the best along-track baseline range for each application along with the most suitable SAR operative modes. The detection of high velocities requires very short time separations. With regard to SABRINA mission scenario

Table 15.3 COSMO/SkyMed interferometric DEM performance [8]

		SAR operative mode			
Vertical Accuracy	Specification	Spotlight (m)	Stripmap (m)	ScanSAR (Wide) (m)	ScanSAR (Huge) (m)
Relative	90% linear point-to-point error	12	28	56	123
Absolute	90% linear error	13	28	46	95

very short along-track components (20–100 m) can be reached, without compromising the overall safety requirements, only together with a meaningful cross-track one. In this case local topography can represent an additional contribution to monostatic-bistatic phase difference other than target velocity. When observing marine targets or ocean areas the effect of topography is expected to be not much significant, but topographic terms must be removed on solid Earth by exploiting a digital elevation model.

15.3.2 Across-Track Bistatic SAR Interferometry

An important possibility to complement and complete COSMO/SkyMed measurement capabilities by SABRINA mission deals with DEM generation. COSMO/SKyMed mission has not been designed to produce high quality DEM. According to Table 15.3, where the expected accuracy of topography DEM generated by COSMO/SkyMed [8] is reported, COSMO/SkyMed could generate, in the best conditions, using high resolution spotlight acquisitions, DTED 2 topographic data. In addition it is worth noting that COSMO/SkyMed was not designed to produce DEM attaining international standards because, in general, the achievable relative accuracy is too close to the absolute one. Finally, wide areas, i.e. global, DTED 2 topographic data can be gathered with much difficulty because of the spotlight operating mode and the need of good coherence values in repeat-pass acquisitions, of correct baselines, incidence angles and terrain slopes.

The COSMO/SkyMed-BISSAT formation, instead, can be exploited also as a single-pass cross-track SAR interferometer for high quality DEM generation. Indeed, DEM generation can greatly benefit from the simultaneous acquisition of the two images in the bistatic configuration as compared with the repeat-pass case. In fact, a considerable drop in the attainable performance results from the decorrelation of interferometric image pairs acquired with time separation, due to radar wavelength-scale changes occurred in the scene between the observations. In addition, atmospheric artifacts equally affect the two acquisitions of the bistatic configuration, thanks to the simultaneity and the slight difference in viewing geometry. Analogous perturbations result in both images, specifically in the echo signal phase, so that the resulting interferometric product is nearly insensitive to them. In addition also digital elevation models and slope maps produced by LBB

radargrammetric techniques can be integrated and can support the phase unwrapping process in critical areas characterized by high terrain slope, irregular topography or vegetation.

The quality of a DEM generated from a single pass-cross track spaceborne interferometer mainly depends on the observation geometry, the formation flying knowledge and control, and the phase error of monostatic-bistatic image pairs. Several error budget models are available to assess DEM accuracy and precision and some references are also reported in this book (see for example the Chap. 2). Those methodologies have been applied to SABRINA case considering two different scenarios:

– Single-pass bistatic interferometric acquisitions, with the baseline representing a small fraction of the critical baseline, for medium-to-high quality DEM generation.
– Multi-passes bistatic interferometric data with proper combinations of small and large baseline acquisitions for the highest quality DEM generation.

In both cases 1 mm baseline component knowledge was assumed along with 25° phase error. In the single-pass bistatic case the selected orthogonal baselines were 400–700 m for spotlight acquisitions and 200–600 m for stripmap ones, whereas for multi-passes DEM generation the possibility to exploit up to 1 km orthogonal baselines for spotlight and 800 m for stripmap was analyzed. The conclusion of this preliminary analysis has been that SABRINA mission is potentially able to generate a global digital elevation model according to NIMA DTED 3 standard, and moreover HRTI 4 topographic data can be gathered on specific area under favorable conditions.

15.4 Mission Analysis

15.4.1 Small Baseline Phase

15.4.1.1 Relative Trajectory

A number of relative orbital trajectories have been proposed to perform SAR interferometry. TOPSAT [18, 19], a joint NASA/ASI program, was the first mission to be studied with this goal and made use of the "parallel" orbit concept, that is to say orbits only separated by ascending node right ascension and time of passage on the ascending node. In more recent years Massonnet [20, 21] introduced the cartwheel concept (see Chap. 14 for details) building on eccentricity and adequate phasing of orbit perigees to achieve prescribed satellite distance. Then, the pendulum concept [22], based on differences in ascending node right ascension and orbit inclination, came up as a cartwheel generalization. DLR also introduced the helix concept [23] building on differences of the ascending node right ascensions and the eccentricity, which has been applied in the Tandem-X mission (see Chap. 13) and also discussed in Chap. 3.

Different relative trajectory options to complement the Italian mission COSMO/SkyMed with interferometric capabilities have been extensively analyzed and compared by Moccia and Fasano [24]. The approach was to define suitable baseline ranges for interferometry and to evaluate the useful orbit fraction guaranteed by the different formations. Additional analysis were later performed by Fasano and D'Errico [25, 26] defining optimal baselines in view of accuracy and ambiguity requirements and trying to achieve these baselines at all the latitudes along the orbit.

On the basis of the latter approach, a helix formation was selected as the operational one, while formation geometries derived in Ref. [24] were considered to experiment autonomous formation flying and to achieve larger baselines useful in multi-passes scenarios.

We herein summarize the relative trajectories which were selected as candidates to test both the different aspects of formation flying and bistatic SAR techniques (along-track interferometry, ATI; cross-track interferometry, XTI):

- Cartwheel XTI—bistatic satellite and a COSMO/SkyMed satellite in a cartwheel-derived configuration, with nominal eccentricity and a perigee separation ($\Delta\omega$) of 20.74°;
- Cartwheel ATI—as before but perigee separation of 0.52°;
- Δe XTI—bistatic satellite on an orbit with different eccentricity ($e = 7.62 \times 10^{-4}$) but both perigees at 90°;
- Δe ATI—as before but $e = 1.169 \times 10^{-3}$;
- $\Delta\omega$ ATI—bistatic satellite with different perigee anomaly ($\omega = 89.9988°$ instead of 90°);
- Pendulum XTI—satellites on pendulum configuration with $\Delta\Omega = 0.0162°$ and $\Delta M = -5 \times 10^{-3\circ}$;
- Pendulum ATI—as before but $\Delta\Omega = 4.18 \times 10^{-5^\circ}$, $\Delta M = -6.18 \times 10^{-4^\circ}$.
- Helix XTI—$\Delta\Omega = 4.9 \times 10^{-3^\circ}$ ($\Delta M = 0°$, $\Delta\omega = 0°$) and $\Delta e = 1.28 \times 10^{-4}$

As illustrated in Fig. 15.3, the selected helix formation allows obtaining an effective baseline in the range 500–700 m at all the achievable latitudes, since positive and negative latitudes are observed in different phases of the orbit (ascending/descending). As a consequence of this, interferometric performance in terms of ambiguity and resolution has a limited dependency on latitude. Height accuracy and height of ambiguity are depicted in Fig. 15.4 as a function of time.

15.4.1.2 Maneuvers and ΔV Budgets

Relative orbital trajectories for SAR interferometry require small distance between satellites, which can be achieved by a single launch strategy. With reference to COSMO/SkyMed, one could also maneuver one satellite in the constellation to bring it closer to the following/preceding one (satellites are placed in the same orbital plane at a nominal anomaly shift of 90°). In this latter case, a maneuver strategy can be established [28]: if a satellite has to gain (lose) an anomaly shift to reach the preceding (succeeding) satellite, it must be speeded up (down), thus its

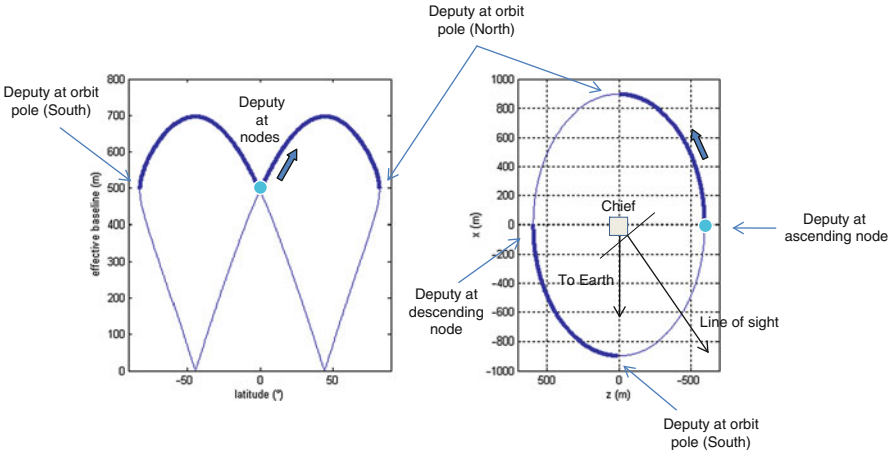


Fig. 15.3 Effective baseline as a function of latitude with $\Delta\Omega = 0.0049°$ and $\Delta e = -1.28 \times 10^{-4}$ (Reproduced with kind permission from Springer Science + Business Media B.V. © 2009 Springer Published in Ref. [27])

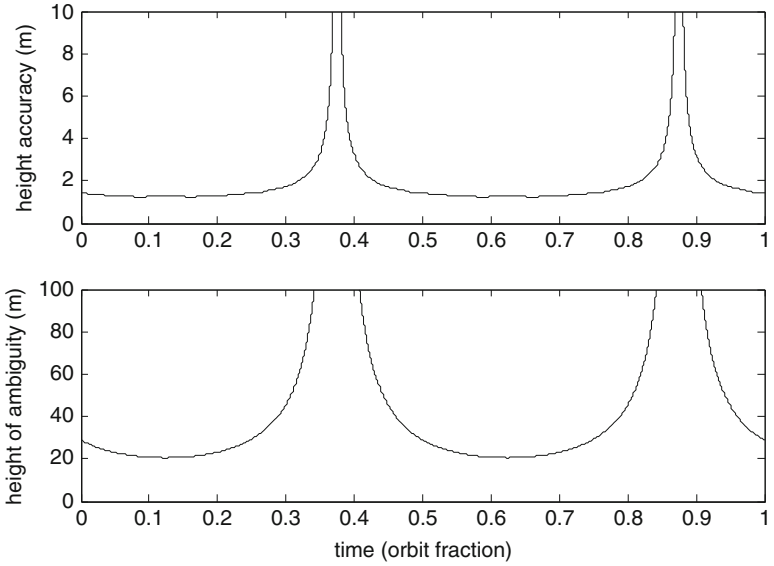


Fig. 15.4 Accuracy and ambiguity along the orbit for the considered helix formation (phase noise taken into account, time computed from the passage at perigee)

orbital period must be reduced (increased), which in turn implies a reduction (increase) of the semi major axis. In particular, the maneuvering satellite has to gain (lose) a given time. Thus, the maneuvering satellite can be transferred over an elliptic orbit with a lower perigee (higher apogee) and apogee (perigee) at the initial

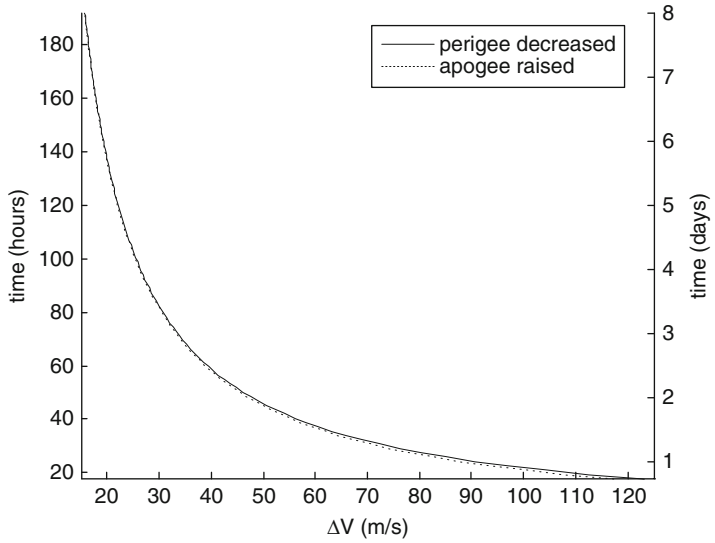


Fig. 15.5 ΔV as a function of the maneuvering time required to reach the following/preceding COSMO/SkyMed satellite

altitude with an orbital period so long that the leader satellite completes one orbit minus (plus) this time shift. This approach allows to perform the maneuver in a short time (approximately one nominal orbit period), but in case of large anomaly shift produces large semimajor axis variation and, consequently, large ΔV. To reduce the required ΔV to acceptable levels, one has to spend multiple orbital periods over the maneuvering ellipse, trading propellant versus time. In the case of COSMO/SkyMed, Fig. 15.5 shows required ΔV versus time spent on maneuvering ellipse. A maneuvering time of 6 days was preliminary selected for a first guess ΔV budget of about 19 m/s for this phase.

In order to estimate the required ΔV to perform different formation flying and payload operation campaigns, different maneuver approaches were studied and modeled [28]. Their application to the selected potential interferometric relative trajectories showed that a ΔV of the order of a few meters per second and maximum for non coplanar transfers should be budget for any trajectory change. All maneuver times are shorter than 200 min.

15.4.2 Large Baseline Phase

15.4.2.1 Large Baseline BSAR Concept

As discussed in the previous sections, the large baseline bistatic SAR relies on two satellites flying at distance quite larger than the ones required for interferometric applications. Selection of the candidate relative trajectory has been discussed in

general in Chap. 3 (Sect. 3.4.3). We herein only recall that satellite separation is to be achieved in "horizontal", i.e. along approximately the elevation direction of the radars. This condition can be typically achieved by a separation between the two ascending nodes ($\Delta\Omega$) and an adequate selection of the in-plane satellite separation (Δu), i.e. of an adequate time shift between the two satellite node crossing time.

It is clear that large baseline bistatic SAR strongly depends on the capability to superimpose the transmitting and the receiving radar swaths, i.e. on the pointing of the two radars. Superimposition can be in general achieved by electronic pointing of the radars, or by attitude maneuvers of the two satellites, or by a combination of both. In designing such pointing, one has to consider that SAR satellites typically perform a yaw steering maneuver to align the antenna with the spacecraft-atmosphere relative velocity to the aim of reducing aerodynamic drag. A general analysis of potential solutions to guarantee adequate overlap between the two radar swaths for large baseline bistatic SAR has been proposed in Ref. [29] and applied in Ref. [30] for the case of a small, parasitic satellite flying in formation with a large SAR satellite. In the framework of SABRINA mission, both radars would share the same pointing capabilities, achieved thanks to their active phased antennas able to steer the beam electronically both in elevation (off nadir angle can be varied between 23.3° and 43.7°, right looking) and in azimuth direction ($\pm 3°$). These capabilities have been integrated on COSMO/SkyMed radar to implement different radar working modes, but they can be also used to program radar pointing adequate to guarantee swath overlap.

When both radars can point the line-of-sight, flexibility in swath selection is greatly improved, as shown in Fig. 15.6 where the graphical concept of the large baseline bistatic SAR is presented in a simplified geometry with both radars sharing the same elevation plane. In particular, the elevation steering capability generates a potential area (access area) where the nominal radar swath can be selected. Therefore, if both radars can be steered in elevation, a bistatic access area (BAA) can be also be defined as the one where both swaths can be positioned.

BAA dimensions decrease with growing horizontal separation between the platforms. If BAA is larger than radar real swath, bistatic target can be selected within BAA by COSMO and BISSAT pointing, otherwise pointing variability is lost. When the latter condition is reached at zero latitude, the maximum horizontal separation along the orbit is achieved, which also maximizes the bistatic angle. In fact if horizontal separation is further increased, bistatic acquisitions over the equator are no longer possible. This orbital condition will be referred to as "final bistatic configuration" (in the case of the "dynamic" large baseline formation, see Sect. 3.4.3). Finally, satellite horizontal separation, bistatic angle, and BAA also vary along the orbit due to changing relative geometry.

15.4.2.2 Relative Trajectory Options (Static and Dynamic)

As recalled above, the "static" or "final" bistatic configuration is referred to as a large baseline formation that allows bistatic observations at the equator with maximum bistatic angle.

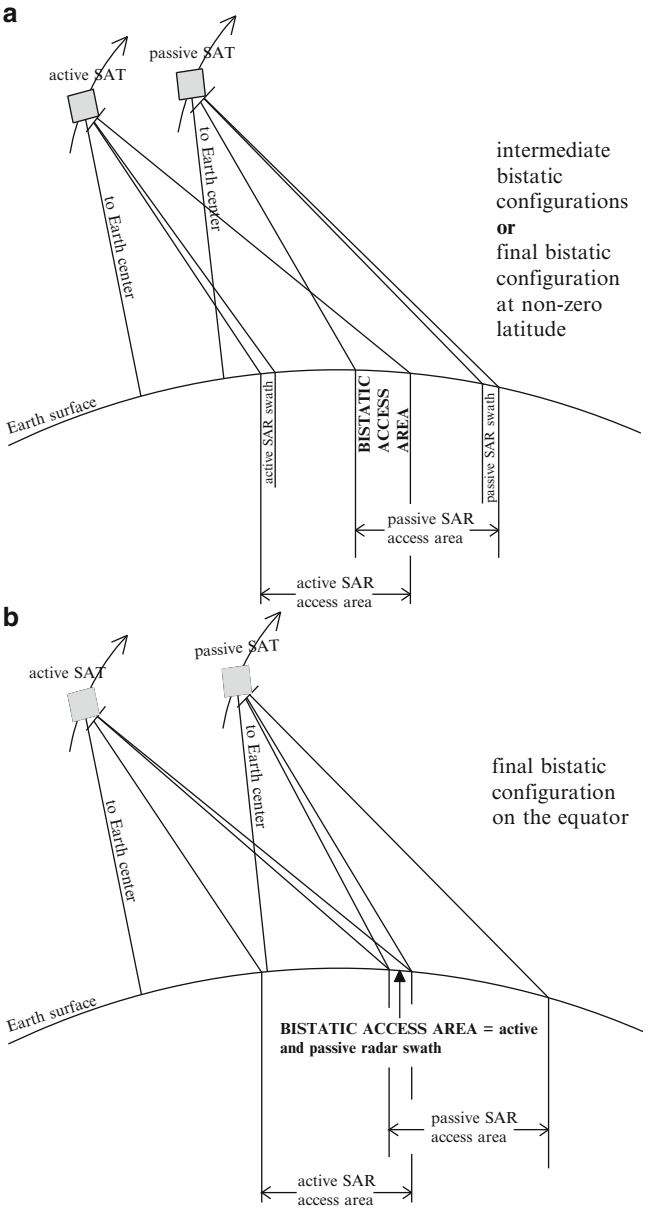


Fig. 15.6 Qualitative view of bistatic observation geometry during ascending phase

Differences in orbital parameters required for achieving this geometry can be calculated from mathematical models presented in Ref. [29] by selecting the maximum off nadir angle for the active satellite (43.7°) and the minimum one for the passive satellite (23.3°). Of course, satellites separation can be further increased

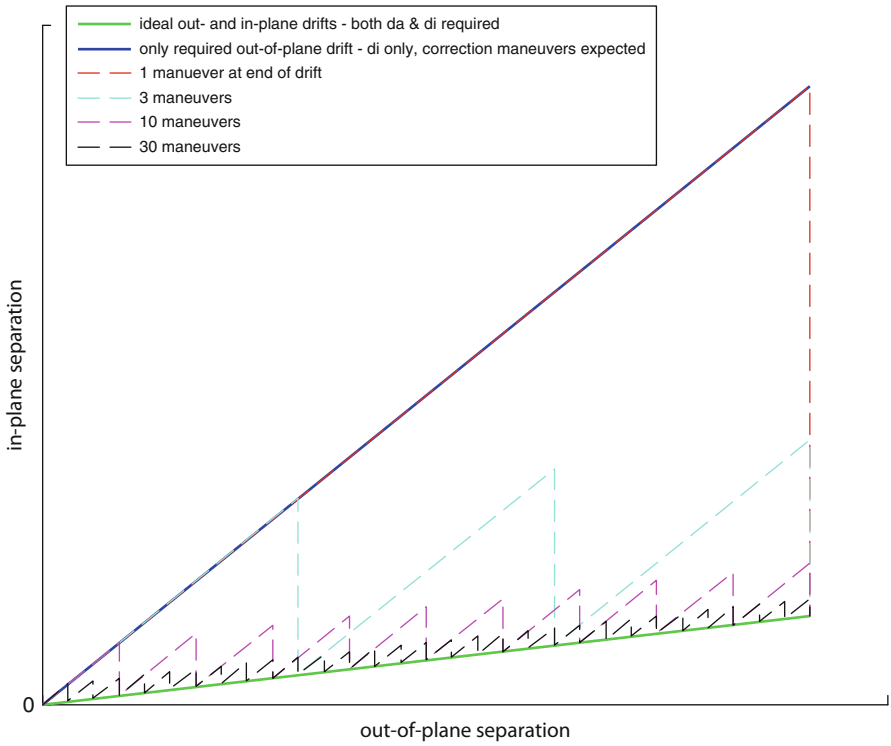


Fig. 15.7 Qualitative view of different strategies for achieving drifting bistatic formations

without losing bistatic coverage at the equator if roll maneuvers are foreseen for the two satellites. If elevation steering is only considered, one gets that $\delta\Omega = 3.19°$ and $\delta M = 0.65°$ are needed.

This final bistatic configuration can be reached starting from a close formation dedicated to SAR interferometry, if differential J_2 effects due to a small change in inclination are properly exploited. Intermediate geometries are referred to as "intermediate", "drifting", or "dynamic" bistatic formations.

As discussed in Sect. 3.4.3, the excessive along-track drift that derives from the change in inclination can be compensated either by applying periodical anomaly corrections, or by using fine semi-major axis control to tune the along-track drift in the desired way.

Of course, anomaly corrections have a cost in term of additional ΔV, which can be reduced if a smaller number of (longer) correction maneuvers is applied. The cost to be paid is then a larger maximum along track separation, which implies larger attitude/point angles required for swaths superimposition.

Figure 15.7 depicts the different options from a qualitative point of view, while Fig. 15.8 reports the total ΔV needed to pass from a close formation to the final

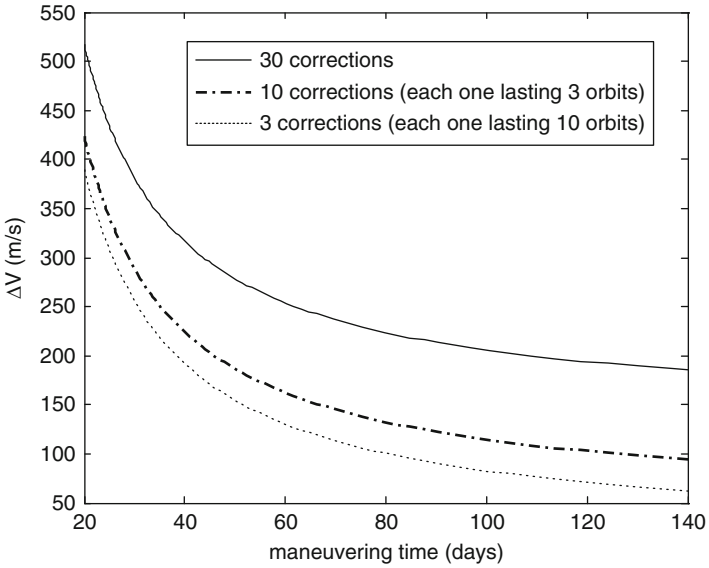


Fig. 15.8 ΔV required to pass to the final bistatic configuration as a function of manoeuvring time and number of foreseen in-plane corrections of anomaly relative drift

bistatic configuration as a function of maneuvering time, considering different anomaly correction strategies.

Indeed, the along-track drift can be tuned by means of a δa of the order of a few meters, and this orbit correction has a negligible cost in terms of ΔV and time: thus, the combined δi–δa strategy is the best candidate for generating dynamic bistatic configurations, which allow acquisitions with increasing bistatic angle over selected test sites. For example, the ΔV expense for activating a drifting BSAR formation with a δi of 0.1° is of about 13 m/s. As noted in Sect. 3.4.3, a small difference in eccentricity (order 1×10^{-4}) can also be used in order to separate the orbits at the poles (argument of perigee 90°). An additional ΔV of 13 m/s is required if nodal drift has to be stopped once the final bistatic configuration has been reached.

15.4.2.3 Safe Establishment of Relative Trajectory

A critical aspect when considering dynamic large baseline BSAR formations is the safe establishment of relative trajectory at the beginning of the maneuver.

In fact, depending on the initial formation geometry, establishment of δi (order 0.1°) with a single impulse commanded at a node of the orbit can dramatically reduce the separation in the radial/cross-track plane. This creates a collision risk since the two orbits can intersect and satellite separation is then only based on along-track baseline control.

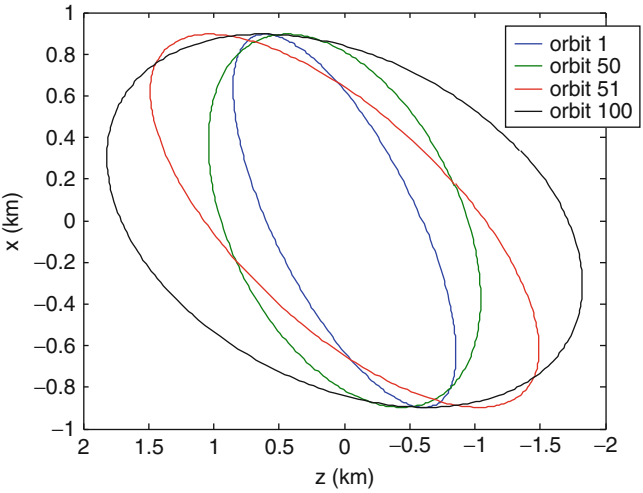


Fig. 15.9 Relative trajectory projection on the radial/cross-track plane during phases 1 and 2 (Reproduced with kind permission from Springer Science + Business Media B.V. © 2010 Springer Published in Ref. [31])

Let us consider a XTI helix as the initial formation geometry. Collision risk can be mitigated by applying a very small δi to let $\delta\Omega$ increase for some time before going to the operational δi. For example we can consider three phases:

1. δi of 0.005° ($\delta e = 1.28 \times 10^{-4}$ still stands): minimum distance in cross-track plane is slightly less than 500 m (Fig. 15.9). In 50 orbits $\delta\Omega$ increases from 0.0049° to 0.007°.
2. δi increased to 0.01°: minimum distance in cross-track plane is about 600 m (Fig. 15.9). This phase also has a duration of 50 orbits.
3. δi to 0.1°, $\delta a = 9.7$ m: drifting formation. After 1,500 orbits $\delta\Omega \approx 1.27^\circ$ (Fig. 15.10).

Drifting bistatic formation stability in along-track is evident in Fig. 15.11, which is a very good results because at each nodal passing the deputy remains in the vicinity of the chief radar elevation angle allowing a feasible pointing strategy and SNR. It is worth noting that a certain along-track drift is needed to ensure that for increasing nodal separation the passive satellite lies near the active radar elevation plane at equator passing.

15.5 Relative Distance and Safety

In order to analyze safety conditions, time history of the satellite formation in the different configurations discussed in Sects. 15.4.1.1 and 15.4.2 have been simulated and verified with respect to expected shapes derived from analytical models (Chap. 3).

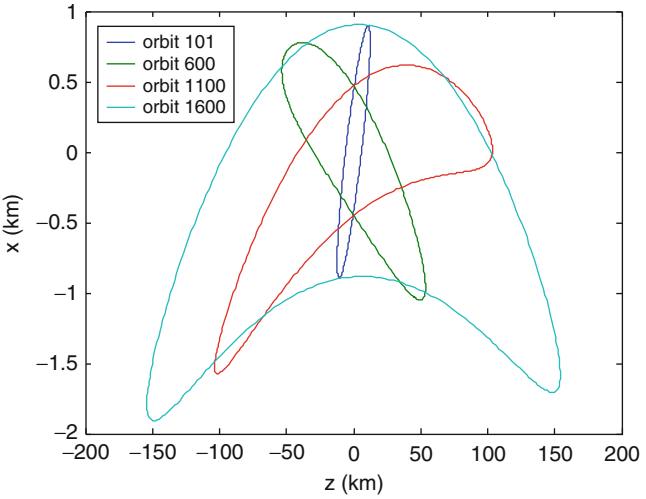


Fig. 15.10 Relative trajectory projection on the radial/cross-track plane during phase 3 (Reproduced with kind permission from Springer Science + Business Media B.V. © 2010 Springer Published in Ref. [31])

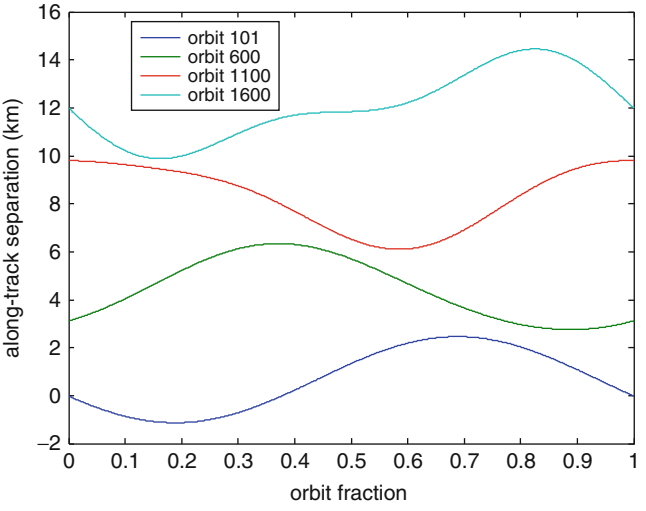


Fig. 15.11 Along track separation during phase 3 (Reproduced with kind permission from Springer Science + Business Media B.V. © 2010 Springer Published in Ref. [31])

It is worth noting that for the interferometric relative trajectories, the minimum satellite distance is less than 100 m in some cases, of the order of few hundred meters in other cases, and of few kilometers in some other cases. On the other hand, for large baselines, the minimum distance is less than 1 km and of the order of several hundred

Table 15.4 Summary of expected distance behaviors

Technique	Configuration	Distance behavior
Cross track INSAR	Cartwheel XTI	Oscillating between 3 and 6 km
	Δe XTI	Oscillating between 3 and 6 km
	Pendulum XTI	Oscillating between 300 m and 2 km
	Helix XTI	Oscillating between 1 and 2.5 km
Along track INSAR	Cartwheel ATI	Oscillating between 75 and 150 m
	Δe ATI	Oscillating between 75 and 150 m
	$\Delta\omega$ ATI	Almost constant at about 150 m
	Pendulum ATI	Almost constant at about 75 m
Large baseline BSAR	Drifting to BSAR (Phases 1–2)	Oscillating between 500 m and 3.5 km
	Drifting to BSAR (Phases 3)	Oscillating between 500 m and kilometers
	Drifting BSAR	Oscillating between 10 and 250 km
	Static BSAR	Oscillating between 25 and 400 km

Table 15.5 Preliminary definition of formation flying domains and requirements

Minimum distance	Formation flying domain	Control requirement	Measurement requirement
<200 m	Tight	1 m	1 cm
<500 m	Close	3 m	3 cm
>500 m	Fair	10 m	10 cm
>10 km	Loose	100 m	1 m
>100 km	Far	1 km	10 m

meters in first orbit cycle when drifting is established, larger than 10 km when large BSAR acquisitions become feasible and larger than 100 km for standard BSAR distances. On the grounds of this outcomes (Table 15.4), different candidate formation flying domains and relative requirements can be preliminarily identified (Table 15.5). In particular, the following rules of thumb have been used: (a) control accuracy of two orders of magnitude less than minimum distance foreseen for the considered domain; (b) relative position measurement (on-board, real time) at least at one order of magnitude less than control accuracy (two orders of magnitude assumed). With these definitions, it is expected to have always tight formation flying for along track interferometry and close or fair formation flying for cross-track interferometry. Whereas, in the case of large baseline applications, formation flying gradually evolves from fair to loose and far.

Relative distance is clearly connected to collision risk, although the latter one strongly depends on relative orbit. In fact, if the minimum distance is kept only for short times per orbit cycles (as an example near points of orbit intersections) the system has enough time to measure and identify possible risks and make decisions well in time to avoid collision. If, on the other hand, satellite are very close, system reaction time could be longer than the time left before collision. But even in these cases, it should be considered that, for configurations with stable relative orbits, relative velocities are smaller at shorter distances. Therefore, a-priori very conservative assumptions could be well beyond real risks.

Table 15.6 Potential qualitative strategies to reduce collision risk

Configuration	Technique
Cartwheel XTI	BISSAT perigee anomaly $> 90°$ if BISSAT ballistic coefficient is larger than COSMO
Cartwheel ATI	
$\Delta\omega$ ATI	
Helix XTI	*See below*
Δe XTI	BISSAT eccentricity less than COSMO if BISSAT ballistic coefficient is larger than COSMO
Δe ATI	
Pendulum XTI	BISSAT with larger ballistic coefficient
Pendulum ATI	
Drifting to BSAR (Phases 1–2)	
Drifting to BSAR (Phases 3)	
Drifting BSAR	
Static BSAR	

Rather, orbits should undergone a finer selection process, after a first level including J_2-only perturbations, to analyze effects of additional perturbations which could act as safety enhancer (or the other way round). As an example, aerodynamic drag, which can be the main relative perturbation in case of satellites with different ballistic coefficient, can be used as a design resource in the orbit design process. As an example, in the case of all large baseline orbit configurations, BISSAT satellite flies through orbit interceptions before COSMO. Thus, choosing BISSAT ballistic coefficient larger than COSMO one is safer since drag causes faster decay and faster velocity: distance naturally tends to increase. The same considerations apply for pendulum orbital configurations.

For the orbit configurations which imply a perigee anomaly difference, the combined effect of J_2, J_3 and drag perturbations should be analyzed. As an example, it must be considered that COSMO flies on a frozen orbit with (at a first approximation) constant eccentricity vector (constant ω and e). Thus, if BISSAT perigee anomaly is not 90°, its orbit will not be frozen. Since eccentricity decreases faster for the satellite with worse ballistic coefficient due to the aerodynamics, J_3 effect should be selected in order to, even though partially, counteract drag effect. Thus, as an example if BISSAT has larger ballistic coefficient, drag produces a faster (with reference to COSMO) decreases of eccentricity which can be partially compensated by having BISSAT perigee anomaly larger than 90° so that J_3 produces an increase of eccentricity.

Then, for the orbital configurations foreseeing a Δe, it would be desirable to have a naturally increasing Δe (which leads to a naturally increasing separation). To this end, the satellite with worse ballistic coefficient must have the lower value of eccentricity. Finally Helix XTI configuration bears similarities to both Δe configuration and large baseline BSAR configuration. Table 15.6 summarizes these ideas.

15.6 Conclusion

An overview of the SABRINA mission study performed at the Universities of Naples has been presented. It has been in particular focused on the potential applications and radar techniques which such mission could enable. Then, the orbit design process has been reported and major collision challenges have been discussed, highlighting potential strategies to minimize the risk with the help of the orbital environment. SABRINA mission to be flown in parallel with COSMO/SkyMed constellation has been later cancelled for a possible partial implementation in the COSMO/SkyMed follow-on mission.

References

1. Moccia A, Rufino G, D'Errico M et al (2001) BISSAT: a bistatic SAR for earth observation. Phase A study – final Report, ASI research contract I/R/213/00
2. Sabatini P, Aceti R, Lupi T, Annoni G, Della Vedova F, De Cosmo V, Viola F (2001) MITA: in-orbit results of the Italian small platform and the first earth observation mission, Hyseo. In: Proceedings of the 3rd international symposium of the IAA on small satellites for earth observation, Berlin, Germany, pp 71–74
3. Renga A, Moccia A, D'Errico M et al (2008) From the expected scientific applications to the functional specifications, products and performance of the SABRINA mission. In: IEEE radar conference, Rome, Italy, pp 1117–1122
4. [Online] available: http://rsdo.gsfc.nasa.gov/images/catalog2010/prima.pdf
5. Moccia A, Renga A (2011) Spatial resolution of bistatic synthetic aperture radar: impact of acquisition geometry on imaging performance. IEEE Trans Geosci Remote Sens 49 (10):3487–3503
6. Willis NJ (1991) Bistatic radar. Artech House, Boston
7. Renga A, Moccia A (2009) Effects of orbit and pointing geometry of a spaceborne formation for monostatic-bistatic radargrammetry on terrain elevation measurement accuracy. Sensors 9:175–195
8. Italian Space Agency (2007) COSMO-SkyMed system description & user guide. ASI-CSM-ENG-RS-093-A, 47pp
9. [Online] Available: http://earth-info.nga.mil/publications/specs
10. Moccia A, Chiacchio N, Capone A (2000) Spaceborne bistatic synthetic aperture radar for remote sensing applications. Int J Remote Sens 21(18):3395–3414
11. Moccia A (2008) Fundamentals of bistatic synthetic aperture radar. In: Cherniakov M (ed) Bistatic radar: emerging technology. Wiley, Chichester
12. Khenchaf A (2001) Bistatic scattering and depolarization by randomly rough surfaces: application to the natural rough surfaces in X-band. Waves Random Media 11(2):61–89
13. Rufino G, Moccia A (1997) A procedure for attitude determination of a bistatic SAR by using raw data. 48th international astronautical federation conference. Paper IAF-97-B.2.04, pp 1–8
14. Moccia A, Rufino G (2001) Spaceborne along-track SAR interferometry performance analysis and mission scenarios. IEEE Trans Aerosp Electron Syst 37(1):199–213
15. Breit H, Eineder M, Holzner J, Runge H, Bamler R (2003) Traffic monitoring using SRTM along-track interferometry. In: Proceedings of the IGARSS03, vol 2, Toulouse, France, 21–25 July 2003, pp 1187–1189

16. Meyer F, Hinz S, Laika A, Bamler R (2005) A-priori information driven detection of moving objects for traffic monitoring by spaceborne SAR. Int Arch Photogram Remote Sens Spat Inf Sci 36:89–94
17. Palubinskas G, Runge H (2008) Detection of traffic congestion in SAR imagery. In: Proceedings of Eusar 2008, Friedrichshafen, Germany, 4pp
18. D'Errico M, Moccia A, Vetrella S (1994) Attitude requirements of a twin satellite system for the global topography mission. In: 45th Congress of the international astronautical federation, Jerusalem, Israel, 9–14 Oct 1994, p 10
19. Zebker HA, Farr TG, Salazar RP, Dixon TH (1994) Mapping the world's topography using radar interferometry: the TOPSAT mission. Proceedings of the IEEE 82:1774–1786
20. Massonnet D (2001) Capabilities and limitations of the interferometric cartwheel. IEEE Trans Geosci Remote Sens 39:506–520
21. Massonnet D (2001) The interferometric cartwheel, a constellation of low cost receiving satellites to produce radar images that can be coherently combines. Int J Remote Sens 22:2413–2430
22. Krieger G, Fiedler H, Mittermayer J, Papathanassiou K, Moreira A (2003) Analysis of multistatic configurations for spaceborne SAR interferometry. IEE Proc 150:87–96
23. Krieger G, Moreira A, Fiedler H, Hajnsek I, Werner M, Younis M, Zink M (2007) TanDEM-X: a satellite formation for high-resolution SAR interferometry. Trans Geosci Remote Sens 45(11):3317–3341
24. Moccia A, Fasano G (2005) Analysis of spaceborne tandem configurations for complementing COSMO with SAR interferometry. EURASIP J Appl Signal Process 2005(20):3304–3315
25. Fasano G, D'Errico M (2006) Relative motion model accounting for J_2 effects: derivation and application to the design of formation-based INSAR missions. In: Proceedings of the 27th IEEE aerospace conference, Big Sky, Montana, 4–11 Mar 2006, p 12
26. Fasano G, D'Errico M (2006) Design of formation missions for Earth observation: relative motion model, validation, and application. In: Proceedings of the 57th international astronautical congress. Paper IAC-06-C1.P.8.1, Valencia, Spain, 2–6 Oct 2006
27. Fasano G, D'Errico M (2009) Modeling orbital relative motion to enable formation design from application requirements. Celest Mech Dyn Astron 105(1–3):113–139
28. D'Errico M, Fasano G (2008) Design of interferometric and bistatic mission phases of COSMO/SkyMed constellation. Acta Astronaut 62(2–3):97–111, ISSN 0094–5765
29. Moccia A, D'Errico M (2008) Spaceborne bistatic synthetic aperture radar, Chap 2. In: Cherniakov M (ed) Bistatic radars: emerging technology. Wiley, Chichester, pp 27–66. ISBN 978-0-470-02631-1
30. Moccia A, D'Errico M (2008) *Bistatic SAR for earth observation*, Chap 3. In: Cherniakov M (ed) Bistatic radars: emerging technology. Wiley, Chichester, pp 67–94. ISBN 978-0-470-02631-1
31. D'Errico M, Fasano G (2010) Relative trajectory design for bistatic SAR missions. In: Sandau R, Roeser H-P, Valenzuela A (eds) Small satellite missions for Earth observation. Springer, Berlin-Heidelberg, pp 145–154

Chapter 16
TOPOLEV and C-PARAS

Tony Sephton and Alex Wishart

Abstract The aim of the European Space Agency funded study on "Concepts for demonstration of advanced techniques and technologies on an EO small mission" was to assess ideas for EO missions compatible with implementation on a small satellite such as PROBA, and which may benefit from Formation Flying. The output of the study was a definition of various small satellite missions and their required developments. The study was led by Astrium Ltd, with support from Astrium SAS, Astrium GmbH, ENVEO, GMV and Verhaert Space. Following initial selection in the first part of the study, three candidates were analysed in detail in Phase 2. Of these three candidates, two of the missions (the Topographic Levelling mission "TOPOLEV", and C-band PAssive RAdar Satellite(s), "C-PARAS") require Formation Flying for single pass SAR interferometry, and these are presented in this chapter.

16.1 Overview of ESA's EO Small Mission Study

This chapter presents a summary of work that was performed as part of a Definition Study[1] for the European Space Agency (ESA) on "Concepts for Demonstration of Advanced Techniques and Technologies on an EO Small Mission" under contract number 20395/06/NL/JA.

The aim of this Definition Study was to assess ideas for EO missions compatible with implementation on a small satellite such as PROBA, and which may benefit

T. Sephton, Ph.D. (✉)
Department of Integrated Applications, European Space Agency, Keplerlaan 1, 2200 AG Noordwijk, The Netherlands
e-mail: tony.sephton@esa.int

A. Wishart, Ph.D.
Future Programmes Department, Astrium Ltd, Gunnels Wood Rd, Stevenage SG1 2AS, UK
e-mail: alex.wishart@astrium.eads.net

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_16,

from Formation Flying. "Small Mission" here implies small satellite platforms in the 150–200 kg spacecraft mass range. The output of the study was a definition of various small satellite missions and their required developments. The study was led by Astrium Ltd, with support from Astrium SAS, Astrium GmbH, ENVEO, GMV and Verhaert Space.[1]

In Phase 1 of this Definition Study, about 40 ideas for EO missions compatible with implementation on a small satellite and which may benefit from formation flying were identified, assessed and traded off. This led to the selection of three candidates to be studied further in Phase 2: the Topographic Levelling mission (TOPOLEV), a C-band PAssive RAdar Satellite(s) (C-PARAS) mission, and an Aerosol Land Mission (AERL). As the AERL mission could also be performed without formation flying it is not further considered here.

The Topographic Levelling Mission (TOPOLEV) is aimed at mapping surface topography and its temporal change at very high accuracy (less than 1 m in height) and high spatial resolution. The focus of the mission is on low and moderate relief terrain, including water surfaces. The mission concept applies single-pass SAR interferometry with two satellites flying in close formation in a Cartwheel configuration (see more details below). The active SAR (Master) operates on the main platform, and the receive-only SAR (Slave) on the second satellite. Ku-band (13.6 GHz) was proposed as the SAR frequency because surface penetration (e.g. in snow, ice, dry soil, etc.) is reduced, and the backscatter signal of water and other surfaces is higher compared to lower radar frequencies. At this frequency a relatively small antenna size can be used.

C-PARAS is a variant on the TOPOLEV mission, being a free-flying formation of three small SAR spacecraft (with radar receiver only) flying in a Cartwheel configuration (see more details below) with each other and providing a bistatic SAR observation capability at C-band in formation with GMES Sentinel-1 which provides the scene illumination. Potential benefits of C-PARAS include acquisition of multiple interferograms in a single pass (overcoming problems of phase unwrapping in steep terrains), along-track and multi-baseline interferometry for ocean current and forestry applications, and spatial resolution enhancement. Both for this mission and TOPOLEV, key aspects are the need for very precise knowledge of the satellite locations, and accommodation of the radar antenna on small spacecraft.

The two missions are both designed for PROBA-class small spacecraft, with conventional platform equipments and structure as far as possible in order to minimise development costs and risk. The driver is accommodation of the relatively large, non-deployable SAR antenna which is fixed to the body of the spacecraft. Both missions have been designed to use the VEGA launch vehicle. TOPOLEV and C-PARAS are dual and triple spacecraft launches, respectively, and in both cases the large size of their fixed (non-deployable) SAR antenna means that

[1] Now QinetiQ Space NV.

the complete VEGA fairing volume is required to accommodate the spacecraft. The two missions both employ sun-synchronous orbits at altitudes between 600 and 700 km, and inclinations of approximately 98°.

The two missions use Formation Flying because this provides the opportunity for single pass SAR interferometry, thus avoiding the temporal effects of scene de-correlation and changes in the atmospheric phase screen. The baseline in both cases is the interferometric Cartwheel, although other configurations are possible (e.g. the helix). The Cartwheel maximises cross-track separation when the along-track separation is a minimum. These are inertial or Keplerian orbits set up in such a way that the relative motion of the spacecraft with respect to each other forms a Cartwheel once per orbit. The Cartwheel offers periodic across-track and along-track baseline viewing geometries suitable for the science applications. Periodic manoeuvres are required to maintain the Cartwheel formation, but this orbit control is open loop by ground station command.

Precise navigation information is required, in terms of the absolute position of the spacecraft (accuracy of order 1 m) and their relative position (accuracy of order 1 mm, with an inter-satellite separation of order 100–500 m). The accuracy of the baseline measurement between the spacecraft is a major driver for DEM quality. Both cross-track and radial components of the baseline vector must be measured accurately in order to reconstruct the viewing geometry (which means that a pure inter-satellite ranging sensor is not sufficient). It is proposed to use differential GPS techniques for the baseline measurement. In this respect C-PARAS is slightly more favourable than TOPOLEV because the longer radar wavelength gives a proportional relaxation in accuracy for C-band compared to Ku-band.

Absolute GPS measurements will be used for orbit determination, which can be achieved with accuracy of the order of 5–10 cm. A validation of the precise orbit determination done with laser ranging measurements can yield accuracies of the order of 3 cm. One of the spacecraft will be the reference and its orbit considered fixed. Using differential GPS measurements the orbit of the other spacecraft will be determined, enabling a precise reconstruction of the baseline.

GPS carrier phase and pseudo-range measurements collected onboard each of the spacecraft will be downlinked and processed on the ground following a dedicated processing flow for the determination of the relative baseline. Based on previous studies for the GRACE and TanDEM-X missions, it should be possible to obtain accuracies in the mm range as required.

It is to be noted that as yet neither the TOPOLEV nor C-PARAS mission has been selected for implementation, though further analysis of the C-PARAS mission has been performed in the ongoing “Earth Observation Sentinel Convoy” study (Theme 1—Ocean & Ice) that is being led for ESA by Astrium Ltd under contract AO/1-6146/09/NL/FC.

16.2 TOPOLEV Mission

16.2.1 Introduction

The Topographic Levelling Mission (TOPOLEV, Fig. 16.1) was originally proposed by ENVEO, and is aimed at mapping surface topography and its temporal change at very high accuracy (less than 1 m in height) and high spatial resolution. The focus of the mission is on low and moderate relief terrain, including water surfaces.

There is high need of precise elevation data for a wide range of scientific and operational applications. The focus of the mission is on low and moderate relief terrain, including water surfaces. The main mission objectives are given below:

Primary Mission Objective

Studying geophysical and environmental phenomena and processes that reveal temporal changes in surface topography at time scales ranging from weeks to seasonal and annual changes. The main targets of interest include:

- Measuring water level in lakes, rivers, and wetland areas.
- Precise topographic mapping of floodplains and other alluvial terrain, and the temporal changes caused by erosion and deposition processes (for applications in geoscience and hazards assessment).
- Precise topographic mapping of inter-tidal zones and coastal wetlands (tideland), and temporal changes caused by tidal action (erosion and deposition).
- Measuring mass balance of glaciers, ice streams and ice caps.

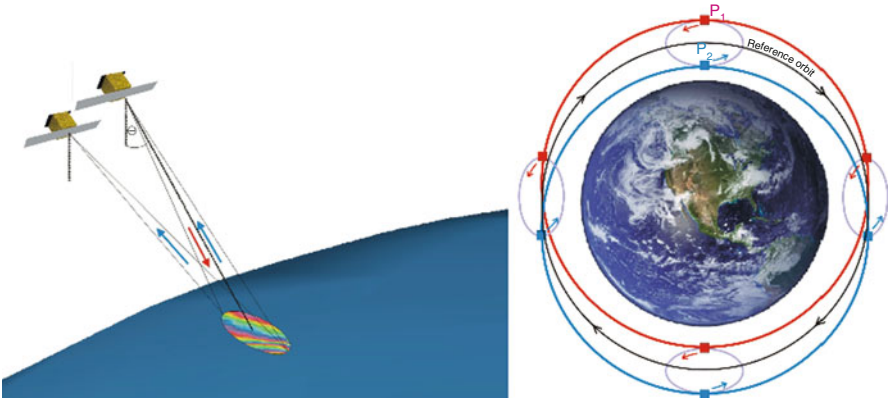


Fig. 16.1 TOPOLEV mission configuration (LHS) and sketch of TOPOLEV Cartwheel orbit configuration with two satellites (RHS)

- Mapping geological structures.
- Observation of surface topographic changes caused by subsidence or inflation phenomena (volcanism, mining, karst erosion, permafrost, etc.).
- Measuring the height of vegetation.

These applications require on the one hand the production of very accurate topographic base maps, and on the other hand the measurement of temporal changes in topography.

Secondary Mission Objective

- The retrieval of very accurate digital elevation data, with emphasis on low or moderate relief terrain, in order to fill gaps or improve digital elevation data from other data sources (e.g. SRTM, TanDEM-X, SPOT stereo, etc.).

In addition, the mission has the objective of advancing and demonstrating coordinated, close formation flying technology.

TOPOLEV will be complementary to the TanDEM-X mission, launched in June 2010. TanDEM-X, operating in formation with TerraSAR-X, will provide high accuracy DEM data aimed at full coverage of global land surfaces within 3 years of operation [3]. In addition, TerraSAR-X and TanDEM-X address many other objectives, based on the capability of the system to operate in many different modes. In view of the need for precise repeat topographic data for a wide range of applications, there are many tasks to be fulfilled by a dedicated topographic mapping mission such as TOPOLEV, complementary and as follow-on to the measurements of TanDEM-X.

The mission concept applies single-pass SAR interferometry with two satellites flying in close formation in a Cartwheel configuration. The active SAR (Master) operates in Stripmap mode, and a receive-only SAR (Slave) on the second satellite (note that for redundancy purposes, and to maintain the same ballistic coefficient, the second satellite will also have a transmit capability). Ku-band (13.6 GHz) is proposed as the SAR frequency because surface penetration (e.g. in snow, ice, dry soil, etc.) is reduced, and the backscatter signal of water and other surfaces is higher compared to lower radar frequencies. At this frequency a relatively small antenna size can be used.

Higher backscattering at water surfaces and swamps is also the driver for the selection of VV polarisation. Atmospheric effects at Ku-band are higher than at X-band but are still sufficiently small. Signal degradation is expected in areas of heavy rain, which will be excluded in the processing of the DEMs.

16.2.2 Requirements Analysis

The mission priority is on measurements of high precision topographic data over low and moderate relief terrain, and its temporal changes, for applications in geoscience and environmental monitoring. It is also planned to measure the water

Table 16.1 Requirements for geoscience and for topographic base maps (DTED-3)

Parameter	Specification	Requirement
Geoscience&elevation change		
Spatial resolution	Independent pixels	25 m (Category 1); 50 m (Category 2)
Relative vertical accuracy		≤1.0 m minimum; ≤0.5 m target (for slope < 20%)
Swath width	Across-track, km	25 km (minimum), 30 km (target)
DTED-3		
Spatial resolution	Independent pixels	12 m (0.3333 arc sec)
Absolute horizontal accuracy	90% circular error	<10 m
Relative horizontal accuracy	90% circular error	<3 m
Absolute vertical accuracy	90% linear error	<10 m
Relative vertical accuracy	90% linear point-to-point error over 1 × 1°	<2 m (for slope < 20%)
Spatial resolution	Independent pixels	25 m
Relative vertical accuracy		≤1.0 m minimum; ≤0.5 m target (for slope < 20%)

level in rivers, lakes, reservoirs and wetlands. These applications require very precise measurement of surface topography. The spatial raster in (X, Y) depends on the target, with a typical range of 25 m for alluvial planes, erosion zones, volcanic zones, small glaciers, etc. (Category 1 in Table 16.1) versus about 50 m for wetlands, lakes, large ice streams, etc. (Category 2).

As a general geodetic product, the aim is to match the Digital Terrain Elevation Data, Level 3 (DTED-3) specifications. These require rather high resolution in (X, Y) (12 m), but have relaxed requirements for vertical resolution compared to geoscience. For geophysical and hydrological applications the vertical accuracy is more important, whereas the horizontal spacing can be scaled down. DTED-3 refers mainly to military and commercial applications (including man-made targets), where high spatial resolution is very important.

The swath width requirements are driven by the dimensions of the phenomena to be studied and also by the data with which the DEM products are combined. A typical minimum swath width for this type of products is 20 km (the target is ≥30 km).

Because for most of the targets under consideration the surface elevation changes very slowly in time, coverage of large areas can be achieved through sequential orbits. For survey of water bodies a sampling strategy needs to be defined.

Two orbit phases (sun-synchronous, dawn/dusk orbits) are proposed:

1. Global Levelling phase (GL-Phase): 91 day repeat orbit (1,394 tracks) enabling near global access to land areas at seasonal intervals. This orbit corresponds to the ICESat orbit, thus enabling validation and improvement of the absolute

Table 16.2 Proposed orbit specifications for the TOPOLEV mission

Orbit parameters	Orbit phase GL	Orbit phase HP
Altitude	600 km	628 km
Inclination (deg.)	97.8 sun-synchronous	97.9 sun-synchronous
Repeat cycle	91 days	15 days
Coverage	Near global access	Subset of land surfaces

Table 16.3 Selected orbit for studying the technical concept of TOPOLEV

Orbit parameters	Orbit phase HP
Altitude	628 km
Inclination (deg)	ca. 98°, sun-synchronous
Repeat cycle	15 days
Coverage	Subset of land surfaces

Table 16.4 TOPOLEV sensor specifications

Parameter	Selected
Centre frequency	Ku-band (13.6 GHz)
Polarisation	VV
NESZ	$\leq$–20 dB (Ku-band)
Operation mode	Stripmap
Nominal swath width	$\geq$25 km (30 km goal)
Incidence angle at swath centre	25°–35° (within this range)
Nominal pixel size	3 × 2.5 m (azimuth × slant range, single look)

height accuracy of the TOPOLEV DEM product over stable surfaces by means of ICESat data. The primary aim of this mode is to map seasonal or annual changes of surface height, e.g. for glaciers, river plains, wetlands, inter-tidal zones, etc. In addition, it can contribute to filling gaps in the global DEM.

2. Hydrology and geophysical process studies (HP-Phase): 15 day repeat orbit to study specific geophysical and hydrological phenomena with higher temporal accuracy (these are mainly processes related to water and ice). This orbit provides only partial coverage of the global land surfaces. A sampling strategy needs to be developed to obtain relevant statistics for various climate zones and regions.

Orbit and sensor specifications for the TOPOLEV mission are given in Tables 16.2, 16.3 and 16.4.

The optimum perpendicular baseline requirements for bistatic SAR geometry depend on the topography. For Level and Moderate Topography (priority objective of the mission) an inter-satellite distance matching the following baseline is proposed:

- Perpendicular Baseline (Across-track) 400 m ± 100 m.

The main latitude band for achieving this baseline is:

- Land areas between 45° and 65°N.

16.2.3 Mission and System Analysis

TOPOLEV is a constellation of 2 SAR spacecraft in a Cartwheel orbit configuration. The TOPOLEV formation design parameters are given in Table 16.5.

The viewing and baseline geometry between the "hub" spacecraft and "rim" spacecraft is shown in Fig. 16.2.

The Cartwheel plane intersects the (X,Y) plane along X, and makes an angle α to the (X,Z) plane. B_p and B_n are the parallel and perpendicular components of the baseline B, and θ is the look direction. The wheel radius is chosen as a consequence of the orbital parameters so that the perpendicular baseline B_n lies within the desired range during the period when the angle ψ sweeps through the range of values about 90° when the constellation is over the particular latitude of interest.

Nominally the look angle is chosen to be 30°, which then allows a pure Cartwheel type of formation. The out-of-orbit plane separation (B_Z in Fig. 16.2) can be calculated, depending on the chosen Cartwheel size. The separation varies with the angle from the intersection of the reference plane and the satellite's orbit plane. Where the planes intersect, the maximum cross-track separation is zero. It is then at a maximum when the satellite is at 90° true anomaly from that point. In this context, the reference plane is the plane of the "hub" of the formation.

Table 16.5 TOPOLEV requirements and baseline

Parameter	Value	Comment
Number of spacecraft	2	Spacecraft are similar with the same ballistic coefficients
Formation type	Cartwheel Radius of cartwheel: 500 m	Formation shape is maintained by periodic manoeuvres (typically one per several days)
Angle between formation plane and local horizontal (α)	30°	Standard cartwheel, consistent with Look angle requirement
SC2 eccentricity difference w.r.t. SC1 to generate formation	3.43E-05 Argument of perigee of both spacecraft = 90°	
SC2 inclination vector difference w.r.t. SC1 to generate formation	RAAN difference = 0 Inclination difference = 0.0034°	
Orbit altitude	628 km	
Inclination	98°	Sun-synchronous
Ground track repeat cycle	15 days	
Perpendicular baseline over which useful measurements are made: B_n (see diagram)	400 ± 100 m	
Latitude over which baseline is in required range	40°−82°N	SSO is at 98° inclination, so max latitude of ground track is 82°N
Look angle θ	30°	

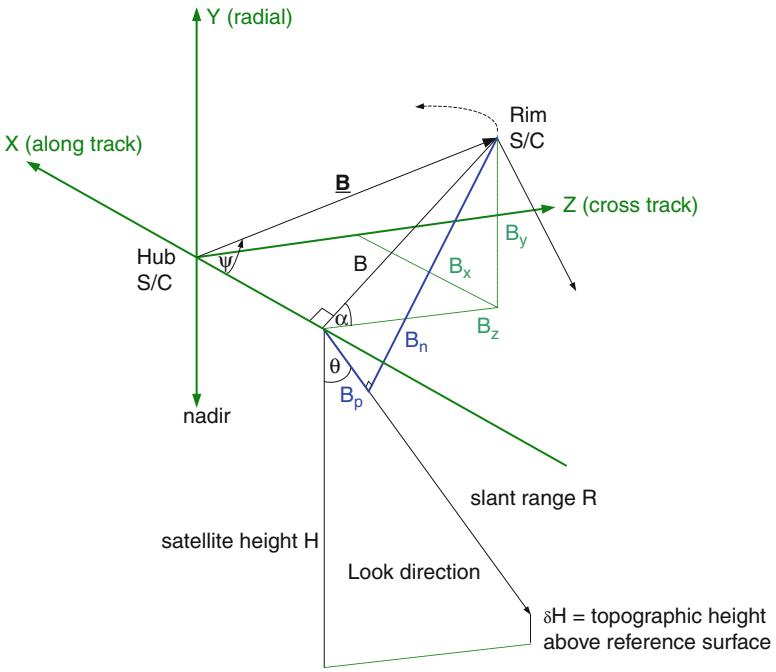


Fig. 16.2 Viewing geometry showing concept of cross-track baseline requirement

The location of the planes intersection can be chosen freely (i.e. at any orbit longitude). For a sun-synchronous orbit, this means the intersection can be at any latitude between approximately −82° and +82°.

The baseline perpendicular to the reference plane is between 300 and 500 m. Figure 16.3 shows that this baseline can be achieved with a Cartwheel radius of 500 m, over a range of angles between approximately 40° and 140°. If a larger radius Cartwheel is selected then the angular range covered is reduced. Choosing the best case allows the reference plane and satellite orbit plane intersections to be placed at the equator. The required out-of-plane separations are then achieved from latitudes 40° up to a maximum of 82°, and then descending to 40° again.

Note that the Cartwheel rather than helix orbit has been chosen because of the requirement for the primary DEM application to have minimum along-track separation when the cross-track baseline is maximised (unlike the helix where the maximum cross-track separation occurs when the along-track separation is also at a maximum). Figure 16.4 shows the case for two spacecraft and the relative motion of the "rim" spacecraft w.r.t. the "hub" spacecraft. The motion in the (X,Y) plane is always an ellipse, with amplitude in X (along-track) twice the amplitude in Y (radial). The Cartwheel and helix formations are then distinguished by the phasing of the cross-track and along-track motions.

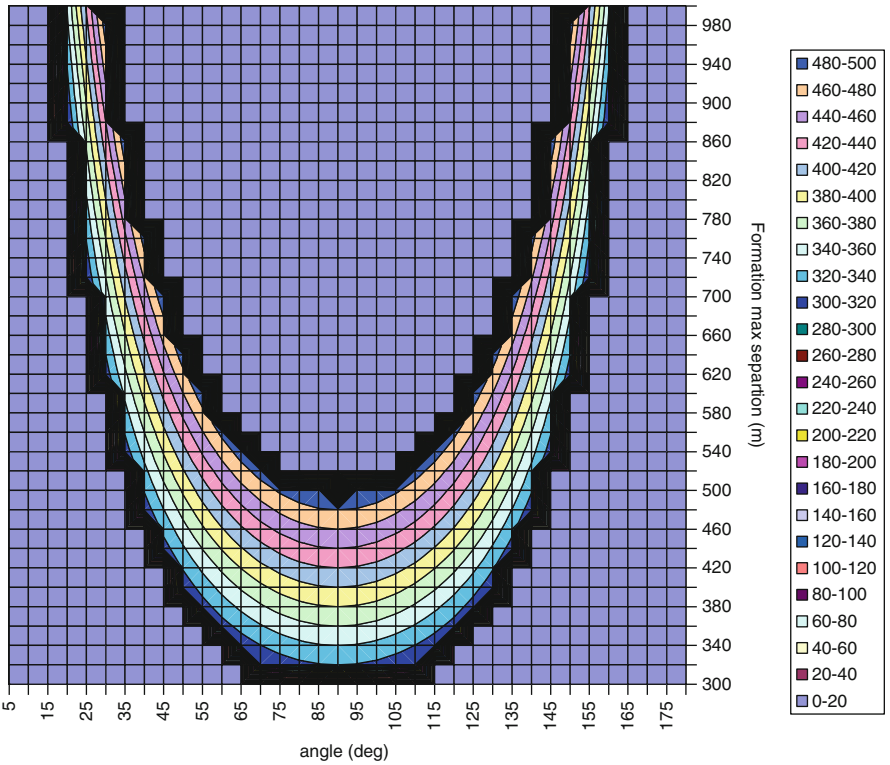


Fig. 16.3 Displacement perpendicular to the reference plane (contours) as a function of angle from the reference plane (hub orbit plane) and satellite orbit intersection and the radius of the selected Cartwheel formation (vertical axis)

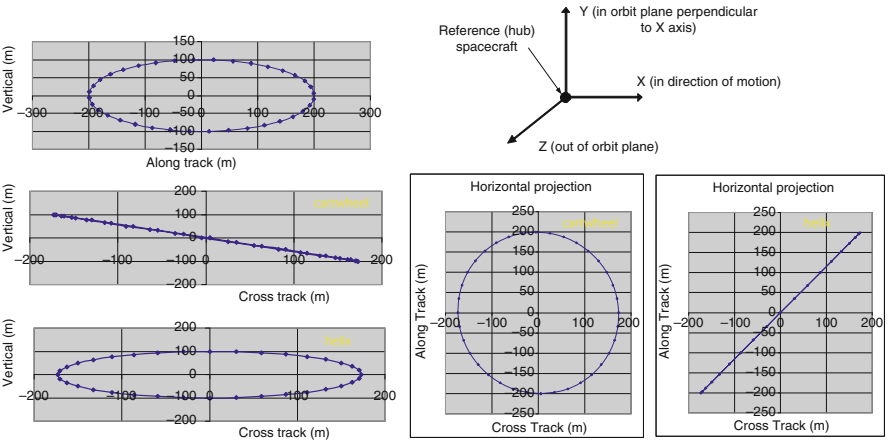


Fig. 16.4 Cartwheel and helix motions

16.2.4 Analysis of Payload and Payload Performance

16.2.4.1 SAR Payload

The SAR instrument for TOPOLEV is a single-polarisation Synthetic Aperture Radar (SAR). The instrument is divided into the following subsystems:

- Antenna, with slotted waveguide radiators (vertical polarisation), cross-feeder and distribution network.
- RF Front-End with Rx/Tx circulator, calibration network, synchronisation horns and fixation, Transmit Receive Module (TRM) for synchronisation horns, and 1:6 switch for the synchronisation horns.
- HPA with TWT and Electronic Power Conditioner (EPC).
- RF Electronics and Central Electronics with up- and down-converter, frequency and chirp generator, LNA and limiter, ADCs, local oscillator, and instrument and timing control.
- GPS receiver and antenna with GPS receiver box, GPS antenna, and external LNA.
- Solid state mass memory.

The synchronisation horns required for the phase referencing among the satellites are part of the RF front-end subsystem. All of the RF equipment must be kept at a stable temperature, especially the local oscillator.

16.2.4.2 Preliminary SAR Antenna Layout and Budgets

The instruments are based on long and narrow slotted waveguide antennas (5.4 0.54 m). To avoid highly complex folding mechanisms a fixed antenna configuration is proposed. This allows for dual launch of the two satellites in a single VEGA launcher.

The TOPOLEV SAR antenna will be a passive slotted waveguide antenna (comparable with the ERS-1 SAR antenna). It will consist of eight identical panels assembled in one row. Each panel will be built by 35 slotted waveguides arranged as 1 column. Each slotted waveguide will have 44 slots ("displaced shunt slots"). It will be centre-fed and of the resonant type. To achieve the necessary surface accuracy (<±0.5 mm) a CFRP design will be necessary. The antenna has to be mounted fixed on the spacecraft (Fig. 16.5).

A summary of the TOPOLEV payload characteristics is given in Table 16.6.

16.2.4.3 Redundancy Philosophy

For the TOPOLEV payload no internal redundancies are foreseen. In case of a failure in the transmit path of a satellite, however, inherent redundancy exists due to

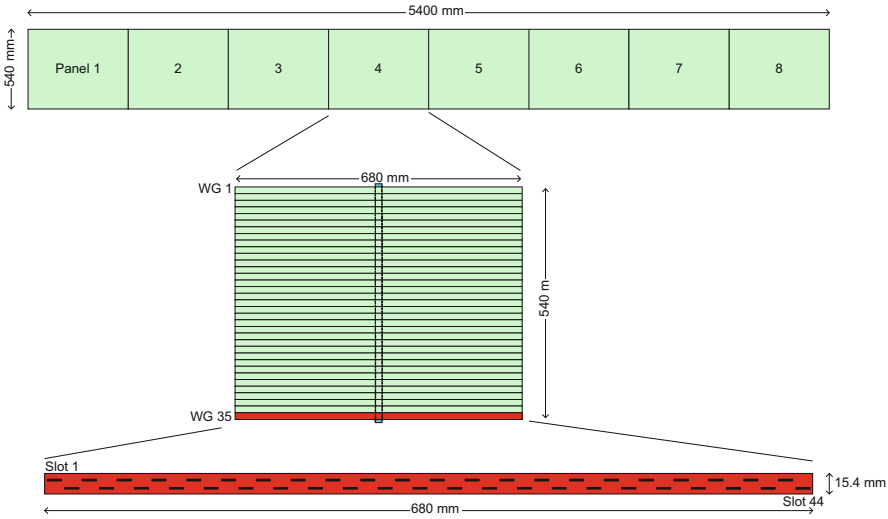


Fig. 16.5 TOPOLEV antenna configuration

Table 16.6 TOPOLEV payload characteristics

Frequency, polarization	Ku-band, VV polarisation
Measurement geometry	31° incidence, stripmap, 25 km swath width
Key performance parameters	NESZ ≤ -20 dB
Mass	81 kg (incl. 20% margin)
Power (payload)	922 W peak/128 W mean over orbit (incl. 20% margin)
Antenna dimension	5,400 mm $\times$ 540 $\times$ 50 mm (fixed, non-deployable)
Data rate	254 Mbps (compressed)
Data/orbit	55 Gbits (compressed, incl. 20% margin)
Telemetry	X-band

the presence of the other (identical) satellite. The transmit part can be shifted to the other satellite without impairing the mission goals.

16.2.4.4 Phase Referencing

For interferometric SAR, the relative phase drift between the two free-running oscillators of the instruments has to be compensated in order to keep the coherence within reasonable limits (Fig. 16.6). As a consequence, the pulse repetition intervals (PRIs) are not simultaneous and some measurement data is not available for interferometric processing (Fig. 16.7).

For TOPOLEV, a phase referencing is proposed between the two satellites (as done for the TanDEM-X mission). The two satellites will be equipped with six horn antennas each, allowing the transmission and reception of reference signals. The

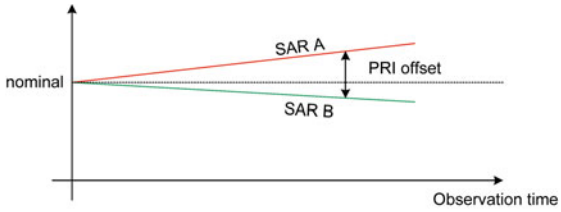


Fig. 16.6 Drift of Local Oscillator (LO) frequencies lead to an offset in the timing

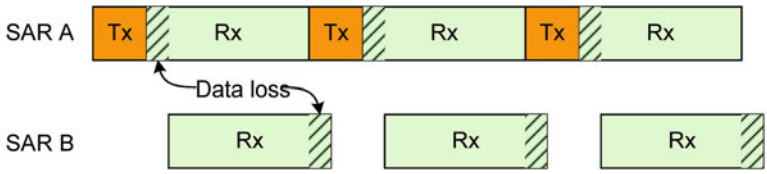


Fig. 16.7 By non-synchronous timing measurement data is lost

horn antennas will provide near omni-directional coverage, but only the horn that is looking towards the tandem satellite is activated through a switch matrix.

For a bistatic interferometric SAR configuration, one satellite transmits and both receive. In regular intervals pulses (5–10 pulses per second) dedicated to phase referencing are exchanged and it is possible to synchronise both phase and PRI timing for offline on-ground processing.

The reference pulses are the same chirp signals that are used for SAR operation. As the power of the HPA is too much for the short distance, the chirp signal is taken from before the HPA and amplified in a dedicated Transmit and Receive Module (TRM). The output power will be about 5 W. For TerraSAR-X/TanDEM-X a phase accuracy of 10° has been achieved. Because of the higher frequency for TOPOLEV a reduced accuracy of 15° is expected.

16.2.4.5 Inter-Satellite Distance Measurements

For the required high-accuracy baseline determination, a dual-frequency GPS receiver will be used. The GPS receiver consists of a receiver box, a GPS antenna and a small external LNA. TerraSAR-X/TanDEM-X experience has indicated that the required baseline accuracy of 1 mm can be achieved on-ground by fitting the GPS measurements to a highly complex orbit model.

Two qualities of baseline determination are available:

- Rapid baseline determination (for quick assessments): Available a few hours after the data-take.

- Precise orbit determination (for full accuracy): Only available from up to a month after the data-take.

For the high-accuracy orbit it is necessary to collect orbital data over a long time range and to average over a range of perturbation effects.

16.2.4.6 TOPOLEV Instrument Performance

The SAR imaging requirements are fulfilled by the TOPOLEV instruments. The accuracy of the relative height measurements meets the requirements in most cases. Height measurement over soil and rock ground is a typical example for a surface with no (or very low) volume de-correlation. This results in the coherence budget shown in Table 16.7, with minimum coherence above 0.8 even in the worst case.

Further model parameters are listed in Table 16.8 together with the resulting relative height error.

Table 16.7 Coherence assessment over soil and rock

	Beam 1	Beam 2	Remark
SNR	0.856–0.905	0.849–0.904	(across the swath)
Quantisation	0.995	0.995	
Range and azimuth ambiguity	0.995	0.993	
Coregistration azimuth	0.984	0.984	(assumption, 10% co-registration error)
Coregistration range	0.984	0.984	(assumption, 10% co-registration error)
Volume de-correlation	1.000	1.000	(at 31° angle of incidence; 3.0 dB/m assumed)
Temporal de-correlation	1.000	1.000	
Processing loss	0.985	0.985	(Tandem-X)
Total coherence	**0.816–0.856**	**0.808–0.856**	

Table 16.8 Assessment of relative height error over soil and rock, no slope

	Beam 1	Beam 2	Remark
Height of ambiguity (m)	12	12	
Post spacing (m × m)	25 × 25	25 × 25	
Area post spacing	625	625	
Multi-looks	41.4	41.4	
Instrument contribution to phase error (deg)	15	15	Assumption, 90% confidence
Relative height error (m)	**0.25–0.33**	**0.25–0.40**	Without contribution of baseline error
RMS baseline error (mm)	1.0	1.0	
Contribution to relative height error (m)	0.89	0.89	90% confidence
Total relative height error (m)	**0.89–0.95**	**0.89–0.98**	With contribution of baseline error

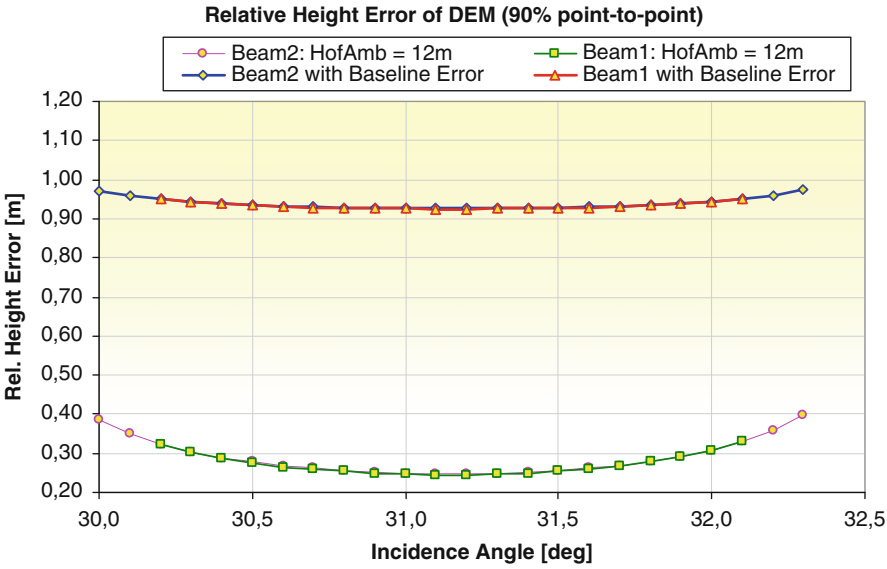


Fig. 16.8 Relative height error over soil and rock, no slope

The relative height error that results from instrument internal phase errors (especially the non perfect synchronisation of the two radars), the signal-to-noise ratio and the de-correlation effect stays below 0.33 m inside the nominal 25 km swath, and grows in the worst case up to 0.4 m at the edges of the extended (goal) 30 km swath. But this error is dominated by the impact of the 1 mm baseline knowledge uncertainty which finally gives a total relative height error below but close to 1 m (cf. Fig. 16.8).

For terrain with 20% slope the performance is slightly degraded (Fig. 16.9) but still below 1.01 m inside a 25 km swath and below 1.06 m inside a 30 km swath. Both results are very close to the 1 m requirement.

For some cases, however, such as water surfaces in combination with low wind (2 m/s), the relative height error reaches 1.25 m in the nominal swath width (25 km) and even 1.5 m in the extended swath (30 km), against the requirement of 1 m relative height accuracy.

16.2.5 Mission Performance

Investigations performed on some typical scenarios indicate the feasibility of achieving the main TOPOLEV mission goal in generating digital elevation models with high accuracy. Based on the SAR design and its sensitivity, the interferometric performance has been studied assuming backscatter properties of bare soil and rock, and taking various error sources (including baseline uncertainty) into account. The 90% point-to-point error was estimated for two proposed DEM products,

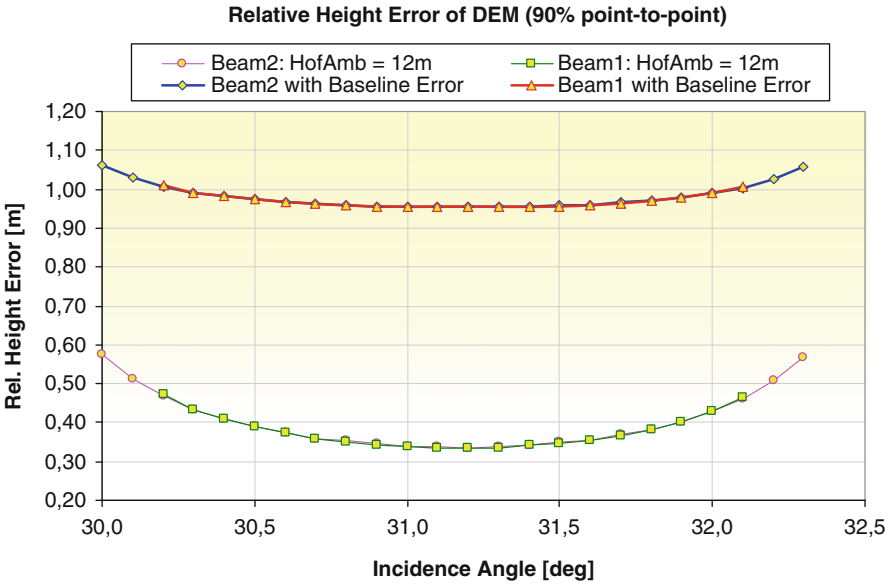
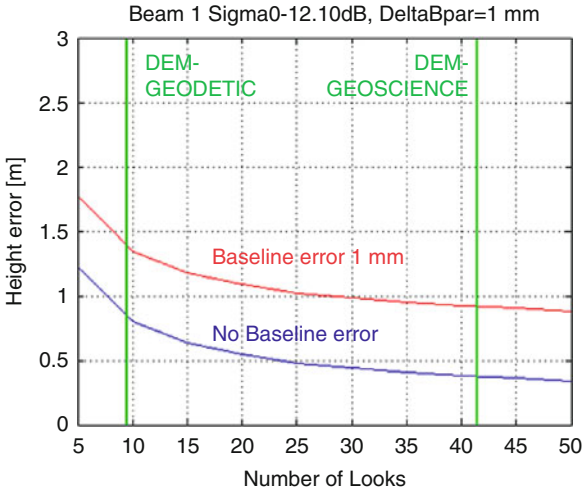


Fig. 16.9 Relative height error over soil and rock, 20%

Fig. 16.10 Height error in dependence on number of looks for bare soil and rock (−95% confidence backscatter signature, at 30° incidence angle, 25 km swath width, flat terrain)



the DEM-GEODETIC and the DEM-GEOSCIENCE product. It was found that baseline uncertainty is a major contributor to the DEM relative height error, a baseline error of 1 mm increasing the height error by about 0.5 m.

The proposed DEM GEODETIC products have a horizontal resolution of 12 m at 9.6 independent looks. The 90% point-to-point height error is $\leq$1.55 m, assuming a baseline uncertainty of 1 mm (Fig. 16.10). This height error is well within the

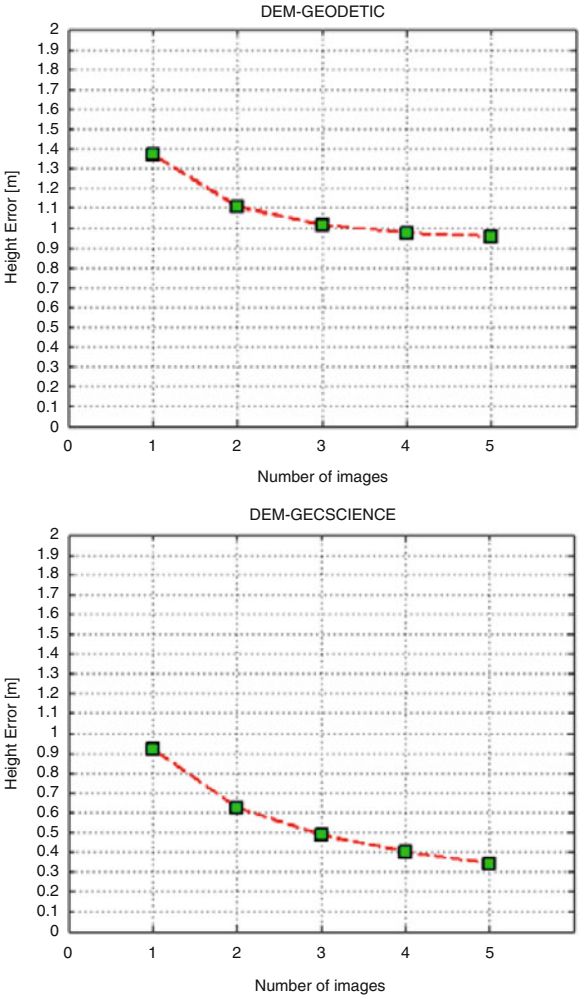


Fig. 16.11 Height error for combining multiple data sets of DEM-GEODETIC and DEM-GEOSCIENCE (each acquired with the same imaging parameters) and 25 km swath width

DTED-3 specifications. In stable areas, the vertical height error can be further reduced to below 1 m by combining two or more interferograms from different over-flights (cf. Fig. 16.11).

The DEM GEOSCIENCE product is dedicated to various applications in geosciences including ecology, hydrology, erosion mapping, glaciology, etc. The product has a resolution of 25 m, with 41.4 independent looks. The 90% point-to-point height error is <1 m (in case of a baseline uncertainty of 1 mm). By combining two interferograms from different over-flights the vertical relative height error can be reduced to 0.6 m, and with five interferograms to 0.4 m.

16.2.6 Critical Technologies

Almost all instrument equipment can be derived from systems with flight heritage (e.g. ERS, TerraSAR-X/TanDEM-X), or Sentinel-1. The only part which needs a new development is the slotted waveguide radar antenna. Here it should be investigated if manufacturing in aluminium would be sufficient (from the thermal stability point of view) or if CFRP technology would be needed.

16.2.7 Summary and Exploitation

Topographic data represent basic information of fundamental importance for a variety of geo-information systems, human security, infrastructure planning, as well as for modelling and understanding Earth processes. For most parts of the world, high accuracy DEMs as would be provided by TOPOLEV are not available. In this respect, TOPOLEV would be able to play an important role in filling the data gaps. However, the main driver for TOPOLEV is the delivery of precise topographic information and its change in time to improve the knowledge on Earth System processes.

For many topics in Earth System Science, reliable and up-to-date data on Earth surface topography is needed. Various international initiatives and programmes (e.g. GEOSS, IGOS, Earth Observation Summit, GMES) have stressed the need for up-to-date, consistent, accurate, high resolution DEM data that are available to the science community without restrictions. TOPOLEV, being dedicated to topographic mapping with high precision, would be able to fulfil a significant part of these needs.

Table 16.9 summarizes the main available global DEM data sets. GTOPO30, a low resolution data set with ca. 1,000 m pixel size based on various sources of elevation data, is a global data set.

These available data sets are well below the spatial and accuracy requirements of many applications.

GTOPO30, SRTM and ASTER-G are single data sets, and do not allow to map elevation changes over time. Combining 2 of these data sets (e.g. SRTM and ASTER-G) to study elevation changes results in even larger errors than for a single data set, rendering it useless for most applications. The TOPOLEV DEM products have several advantages over these data sets, including higher vertical accuracy, improved spatial resolution, and the repeat observations capability, enabling the precise mapping of elevation changes in time. In the proposed two orbit configurations, TOPOLEV monitors a subset of land surfaces with a repeat interval of *15 days in the Geoscience Phase* and has complete access to the land surfaces (except the polar gap) within *91 days in the Global Levelling Phase*. Both phases offer the option of repeat observation for measuring topographic changes.

Table 16.9 Overview of available DEMs (global coverage—from http://www.ersdac.or.jp/GDEM/E/2.html)

	ASTER G-DEM	SRTM (V3)	GTOPO30
Data source	ASTER	Space shuttle radar	From organizations around the world that have DEM data
Generation and distribution	METI/NASA (planned)	NASA/USGS	USGS
Release year	ca. 2009 (planned)	2003 ~	1996
Data acquisition period	From 2000 to ongoing	11 days (in 2000) DEM grid 90 m (30 m) (3 arc-sec)	1,000 m (30 arc-sec)
DEM accuracy 90% point-to-point	±10 to ±30 m Depending on GCPs	ca. ±16 m	ca. ±48 m
DEM coverage	83°N 83°S	60°N–56°S	Global
Area of missing data	Areas with no ASTER data due to persistent cloud cover	Topographically steep areas (due to radar characteristics)	None

16.2.7.1 DEM-GEODETIC Product

The DEM-GEODETIC product is dedicated to operational (e.g. construction activities) and scientific applications where high resolution topographic data are needed. The product is characterized by a pixel size of 12 m and a vertical relative 90% point-to-point height accuracy of 1.55 m in level and sloping terrain (up to slope angles <20°). The accuracy decreases gradually towards steep terrain. The product matches the requirements of the DTED-3 specifications. With the same spatial resolution the vertical height error can be further reduced by combining repeat-pass products, e.g. using two interferometric data sets reduces the height error to less than 1 m. The DEM-GEODETIC product will be of significant interest for geotechnical work, planning of infrastructure such as traffic routes and hydropower reservoirs in remote areas, hazard zonation, etc. The product is also of interest for geoscience. The Global Levelling Phase will enable access to any land area (except the interior of Antarctica).

16.2.7.2 DEM-Geoscience Product

The DEM Geoscience Product is dedicated to providing accurate and up-to-date elevation information for applications in many geoscience areas, including hydrology, glaciology, ecology, forestry, limnology and oceanography. The resolution of this product is 25 m, and it has a vertical relative 90% point-to-point height accuracy of better than 1 m in level and moderate terrain (up to slope angles <20°), with the accuracy gradually decreasing towards steep terrain. Compared to

DEM products available for science (e.g. the ASTER-G and SRTM DEM data sets) the DEM-GEOSCIENCE product has the following advantages:

- Significantly improved relative and absolute vertical accuracy (better than 1 m for a single data set, and better than 0.5 m for a combination of multiple DEM data sets).
- Improved horizontal resolution (25 m).
- Repeat observation capability enables accurate monitoring of temporal elevation changes.
- Very few data gaps in mountainous areas caused by SAR imaging geometry (as are common in mountainous regions in SRTM).

For this product, both proposed orbit configurations are relevant, depending on the type of phenomena to be studied. The Geoscience Phase will provide 15 days repeat observations of a subset of the land surfaces. This frequent repeat capability is of interest to studying geophysical and hydrological phenomena that are subject to high temporal change. These are mainly processes related to water and ice (lake levels, wetlands). The 91 day repeat orbit addresses the observation of seasonal or annual changes of surface height, e.g. for glaciers and ice streams, river plains, inter-tidal zones, volcanic areas, erosional features, etc.

16.3 C-PARAS Mission

16.3.1 Introduction

C-PARAS is another mission for precise topographic mapping, being a free-flying formation of three small SAR spacecraft (with radar receiver only) flying in a Cartwheel configuration with each other and providing a bistatic SAR observation capability at C-band in formation with Sentinel-1 which provides the scene illumination. Potential benefits of C-PARAS include acquisition of multiple interferograms in a single pass (overcoming problems of phase unwrapping in steep terrains), along-track and multi-baseline interferometry for ocean current and forestry applications, and spatial resolution enhancement.

The principal objectives of the C-PARAS mission are:

- Production of basic topographic maps (DEMs), in order to fill gaps and improve available global DEMs (by applying single-pass, multi-baseline interferometry).
- Precise mapping of elevation changes by means of precision DEMs produced at repeat intervals of weeks to years for geoscience applications, e.g. volcanic and geological structures, floodplains, wetlands, tidal zones, glaciers (by applying single-pass, multi-baseline interferometry).
- Estimation of forest parameters by means of multi-baseline interferometry (tomography and/or PolInSAR techniques).
- Observations of motion of ocean currents and fast moving sea-ice drift using along-track interferometry.

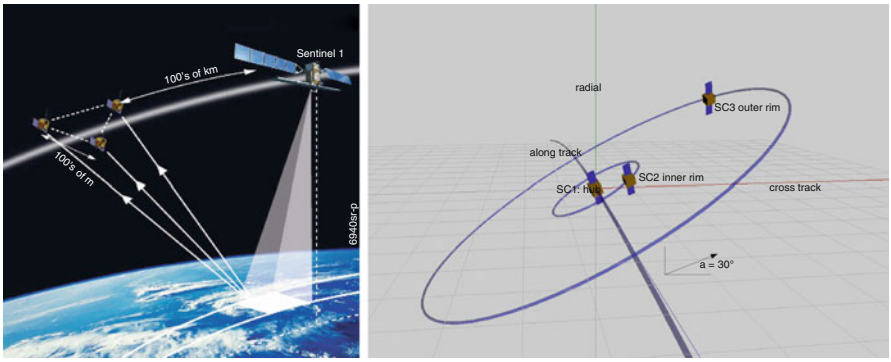


Fig. 16.12 C-PARAS concept of operation and Cartwheel orbit configuration

- Generation of very high resolution SAR image products for mapping and geoscience applications, applying across-track and/or along-track resolution enhancement.
- Improvement of SAR image products of man-made structures (that are rich in corner reflectors and specular reflectors) through exploitation of bi-static reflectance.

In the C-PARAS formation, the hub SC1 follows the orbit of Sentinel-1 at an along-track separation of ~150 km and the two "wheel" spacecraft SC2 and SC3 orbit the hub at radii of 120 and 480 m, respectively (cf. Fig. 16.12). The distance of 150 km from Sentinel-1 ensures good separation at the C-PARAS receivers of the nadir and wanted echoes from the Sentinel-1 illuminations, but exceeds the critical baseline for interferometric SAR imagery (the images are completely uncorrelated) so only C-PARAS SAR image pairs can be used for interferometric products. The three spacecraft perform periodic propulsive manoeuvres to maintain both the formation and its along-track position with respect to Sentinel-1.

Since the Sentinel-1 space segment is non-cooperative (by design), the option of using a two-way synchronisation link is precluded, and other means have to be used to establish system coherence. The separate USOs, although nominally at the same frequency, will inevitably suffer from frequency drifts and will have different random phase noise. It is assumed, however, that the Sentinel-1 datation and radar database information will be available for C-PARAS.

16.3.2 Requirements Analysis

In Table 16.10, the observation requirements are listed for the main applications. The DTED-2 mapping requirements are relevant for producing DEMs over large areas for which data gaps exist in the global DEM (e.g. regions north of 60°N which

Table 16.10 Observation requirements for geoscience and geodetic applications

Parameter	Specification	Requirement/Comment
Topographic base map		
Large-scale product DTED-2		
Spatial resolution	Independent pixels	30 m (1 arc sec at Earth centre)
Absolute horizontal accuracy	90% circular error	<23 m
Absolute vertical accuracy	90% linear error	<18 m (<10 m goal)
Relative vertical accuracy	90% linear point-to-point error over 1° × 1°	<12 m (for slope < 20%) (<5 m goal) <15 m (for slope > 20%) (<10 m goal)
DTED-3		
Spatial resolution	Independent pixels	12 m (0.33 arc sec at Earth centre)
Absolute horizontal accuracy	90% circular error	<10 m
Relative horizontal accuracy	90% circular error	<3 m
Absolute vertical accuracy	90% linear error	<10 m
Relative vertical accuracy	90% linear point-to-point error over 1° × 1°	<2 m (for slope < 20%)
Elevation change—Geoscience		
Spatial resolution	Independent pixels	25–50 m
Relative vertical accuracy		≤1.0 m (for slope < 20%) ≤3.0 m (for slope < 20%)
Ocean currents		
Spatial resolution	Independent pixels	100 m
Velocity	Standard deviation	±0.2 m/s
Sea-ice drift (fast moving)		
Spatial resolution	Independent pixels	50 m
Velocity	Standard deviation	±0.1 m/s
Forest parameters		
Spatial resolution	Independent pixels	25 m
Tree height	Standard deviation	±2 m or ≤10% for h > 20 m

were not mapped by SRTM). DTED-3 is the mapping standard for high spatial resolution products. The requirements are demanding but should be achievable with C-PARAS by applying Sentinel-1 Stripmap mode as master and applying multi-baseline techniques. The geoscience applications require very high vertical accuracy (relative), whereas the spatial requirement is relaxed compared to DTED-3.

The accuracy requirements for measurements of ocean currents and of sea-ice are rather demanding. Typical drift velocities for sea-ice are ≤1 m/s. Ocean currents may reach higher velocities, in particular in regions with strong tidal currents. This means that the requirements should be adjusted to the phenomena to be observed.

The general orbit parameters of C-PARAS are determined by the orbit of Sentinel-1, which is specified in Table 16.11.

As the C-PARAS satellites would not be allowed to operate in close distance to Sentinel-1 due to the danger of collision, a minimum distance of 50 km along-track from Sentinel-1 is imposed.

Table 16.11 Sentinel-1 orbit specification

Orbit parameters	Sentinel-1 orbit specification
Altitude	ca. 700 km
Type	Near-polar sun-synchronous
Inclination (deg.)	98.183402°
Repeat cycle	12 days
Mean local solar time at ascending node	18:00 h (nominal)
Coverage	Global access

Table 16.12 C-PARAS sensor specifications

Parameter	Value
Centre frequency	5.405 GHz
SAR mode	Receiving only
Bandwidth	Synchronized with Sentinel-1
Polarisation	Co-pol (VV or HH)
Noise Equivalent Sigma Zero NESZ	−23 dB
Swath width	≥50 km (receiving signal of Sentinel-1 Stripmap or Interferometric Wide swath Mode)
Use of Sentinel-1 Modes[a]	Interferometric wide swath mode
	Minimum requirement: swath IW1
	Target: selectable IW modes (IW1, IW2, IW3)

[a]With IW1 global access is not feasible, which is a major restriction for science and applications. This minimum requirement is suitable for technological in-orbit demonstration

The use of two instead of three receiving-only satellites (at far along-track distance to Sentinel-1) would enable the formation of one bistatic single-track interferogram with the baseline determined by the position of the two passive satellites. This causes some reduction of the quality of DEMs as no multi-baseline processing is possible. However, it is an option that would nevertheless still be useful.

The proposed characteristics of the C-PARAS receivers are given in Table 16.12.

The baseline requirements for bistatic SAR geometry for the primary (across-track) applications are:

- Perpendicular baseline (across-track) of three image pairs: 150, 450, 600 m (±30%).
- Along-track distance between receiving-only satellites should be as close as possible to increase coherence and minimize Doppler shift of azimuth spectrum.
- The main latitude band for achieving this baseline is land areas between 45° and 65°N.

16.3.3 Mission and System Analysis

C-PARAS is a constellation of three passive small S/C, flying about 150 km behind Sentinel-1 and forming SAR images from the echoes of the Sentinel-1 transmissions (Table 16.13).

Table 16.13 C-PARAS requirements and baseline

Parameter	Value	Comment
Number of spacecraft	3	Spacecraft are identical (but very different from Sentinel-1)
Sentinel-1 orbit altitude	693 km	3-SC constellation follows along the track of Sentinel-1, trailing by 150 km
Sentinel-1 inclination	98°	Sun-Synchronous, mean local solar time at ascending node of 18:00 h
Sentinel-1 repeat cycle	12 days	
Formation type	Concentric cartwheels: SC1 at hub SC2 on inner rim: radius = 130 m SC3 on outer rim: radius = 520 m	Formation shape is maintained by periodic manoeuvres (typically one per several days)
Angle between formation plane and local horizontal (α)	30°	i.e. Standard cartwheel, Is consistent with Look angle requirement
Eccentricity difference w.r.t. SC1 to generate formation	SC2: 9.28E-06 SC3: 3.71E-05 Argument of perigee of all spacecraft = 90°	
Inclination vector difference to generate formation	RAAN difference = 0 SC2: Inclination difference = 0.00092° SC3: Inclination difference = 0.0037°	
Perpendicular baseline over which useful measurements are made: B_n	150 m, 450 m, 600 m, ±30%	To form three image pairs between the three passive spacecraft, i.e. SC1–SC2, SC2–SC3, SC1–SC3. Note that no images are formed between Sentinel-1 and any of the passive S/C as the Sentinel-1 separation exceeds the system critical baseline
Latitude over which baseline is in required range	45–82°N	
Look angle (θ)	30°	

The out-of-orbit plane separation can be calculated, depending on the chosen Cartwheel sizes for C-PARAS. In fact two, concentric Cartwheels can be chosen, with the satellites located along a "spoke" of the wheel.

As for TOPOLEV, the location of the intersection of the reference and satellite orbit planes can be chosen freely (i.e. at any orbit longitude). For a sun-synchronous orbit, this means the intersection can be at any latitude between approximately −82° and +82°.

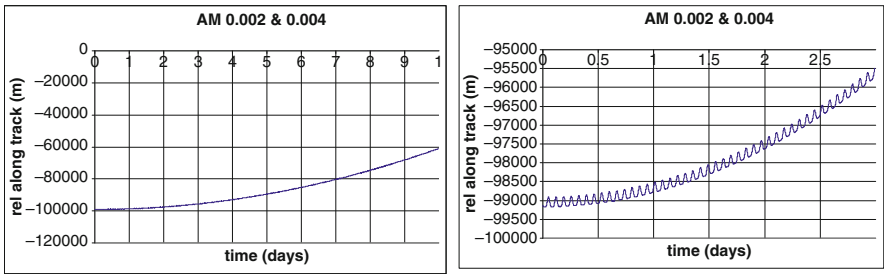


Fig. 16.13 Differential along-track motion with area/mass at 0.002 and 0.004 for Sentinel-1 and C-PARAS, including zoom in on first 3 days

In this context the reference plane is the plane of the "hub" of the formation. The baseline perpendicular to the reference plane is then:

- Between 70 and 130 m for the centre to inner spacecraft, this baseline can be achieved with a Cartwheel radius of 130 m, over a range of angles between approximately 35° and 145°.
- Between 210 and 390 m for the inner to outer spacecraft, this baseline can be achieved with a Cartwheel radius of 390 m (for outer spacecraft relative to inner spacecraft), over a range of angles between approximately 35° and 145°.
- Between 280 and 520 m for the centre to outer spacecraft, this baseline can be achieved with a Cartwheel radius of 520 m, over a range of angles between approximately 35° and 145°.

If a larger radius Cartwheel is selected then the angular range covered is reduced.

Choosing the best case allows the reference and satellite orbit intersections to be placed at the equator. The required out-of-plane separations are then achieved from latitudes 35° up to a maximum of 82°, and then descending to 35° again.

16.3.3.1 Orbit Control Relative to Sentinel-1

Sentinel-1 and the C-PARAS spacecraft will have different Area/Mass values. The C-PARAS formation is designed to follow Sentinel-1 at an along-track distance of typically 150 km. Differential drag effects will cause the whole formation to drift with respect to Sentinel-1. Sentinel-1 will periodically perform station-keeping manoeuvres to maintain its mean semi-major axis. This could be intervals of typically 2 days and is driven by the control band requirements.

No stringent requirement for relative along-track control between Sentinel and C-PARAS exists. However, a generic requirement that the along-track separation should be maintained within 10% of nominal could be appropriate.

The relative motion due to differential drag can be analysed. Figure 16.13 shows this motion at a period of high solar activity where atmospheric drag effects will be relatively high. The Area/Mass ratios chosen are 0.002 and 0.004.

The quadratic nature of the drift with time means that over a period of 2 days (typical Sentinel-1 manoeuvre frequency) the along-track drift is 1.5 km in this example, yet reaches nearly 40 km after 10 days.

Therefore, C-PARAS should aim for a station-keeping frequency in the region of 2–3 days to remain within typically 10% of its designated along-track position.

16.3.4 Analysis of Payload and Payload Performance

It is likely that Sentinel-1 will operate primarily in the Interferometric Wide Swath (IWS) mode. This is a ScanSAR TOPS mode in which Sentinel-1 implements electronic steering of both azimuth and elevation beams. In order to provide simplification, coarse mechanical beam steering in *elevation* is proposed in the C-PARAS satellites, enabling imaging of the selected sub-swath during any given imaging sequence. In *azimuth*, mechanical steering in yaw is proposed, together with electronic scanning to track the Sentinel-1 beam in its ScanSAR mode. Antenna timing control is derived from the platform GPS.

Satellite configurations suitable for triple launch in VEGA have been identified, and lead to the recommendation of an antenna with dimensions 3.6 1.4 m. The performance achievable with such an antenna has been shown to provide a sensitivity ($NE\sigma_0$) of about −23 dB with the addition of Scan-in-Receive (a technique for electronically implementing very fast elevation scanning, in which the receive beam is scanned across the image swath in synchronisation with the transmit pulse). For high accuracy baseline determination (within 2 mm), a dual-frequency GPS receiver will be used.

Figure 16.14 shows the across-swath NESZ pattern for IW1, IW2, and IW3 for an antenna of size 3.6 1.40 m implemented with Scan-in-Receive. The worst case NESZ is:

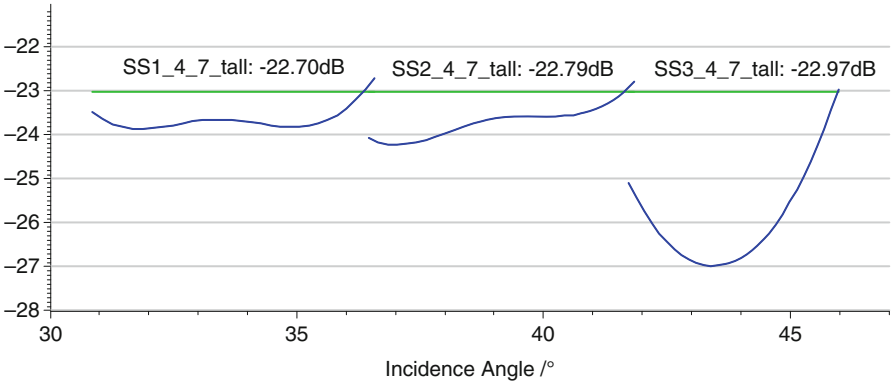


Fig. 16.14 NESZ in dB for C-PARAS IW1, IW2, and IW3 for 3.6 × 1.4 m antenna with scan-in-receive enabled

- IW1: NESZ = −23.64 dB
- IW2: NESZ = −22.79 dB
- IW3: NESZ = −22.97 dB

16.3.4.1 C-PARAS Payload Configuration

The C-PARAS payload is based where possible on re-use of the relevant elements of the Sentinel-1 payload. This includes much of the receive chain and instrument controller with additional sub-systems to enable inter-satellite synchronisation, (optional) direct path timing synchronisation with Sentinel-1, and sub-systems within the antenna structure to enable slow time azimuth scanning (to track the TOPS illumination provided by Sentinel-1) and fast time elevation scanning to enable the Scan-in-Receive functionality.

The main elements of the C-PARAS payload are:

- Principal equipments and subsystems of the Front-End, which include:
  - The principal echo reception antenna with its associated azimuth and elevation steering elements.
  - The *optional* C-PARAS antenna and associated electronics (low noise amplifier, amplitude control antenna and isolating switch) of a direct timing synchronisation link between each C-PARAS satellite and Sentinel-1.
  - The small monopole (or dipole) antenna, transmit signal amplifier, Tx/Rx switch, and low noise amplifier associated with phase synchronisation within the C-PARAS satellite constellation.
- Principal echo signal receive chain.
- Ultra Stable master Oscillator (USO/STALO).
- Frequency generator (which generates the various clocks and Local Oscillator signals needed by the payload).
- Radar instrument controller.

The overall satellite configuration shown in preparation for launch is presented in Fig. 16.15, and the main payload characteristics given in Table 16.14. Because the illumination provided by Sentinel-1 when operating in the IWS mode to any one of the component sub-swaths comes in the pulse-bursts that characterise ScanSAR operations, only modest spatial resolution (20 m) can be achieved in the along-track direction.

16.3.4.2 Timing and Phase Synchronisation Issues

Sentinel-1 has no provision for bistatic operations, so unlike TOPOLEV there is no direct means of establishing synchronisation between C-PARAS and Sentinel-1. C-PARAS receivers record continuously, and ground processing recovers timing from Sentinel-1 datation. Inter-satellite links between the C-PARAS spacecraft are used for the symmetric exchange and recording of their respective oscillator signals

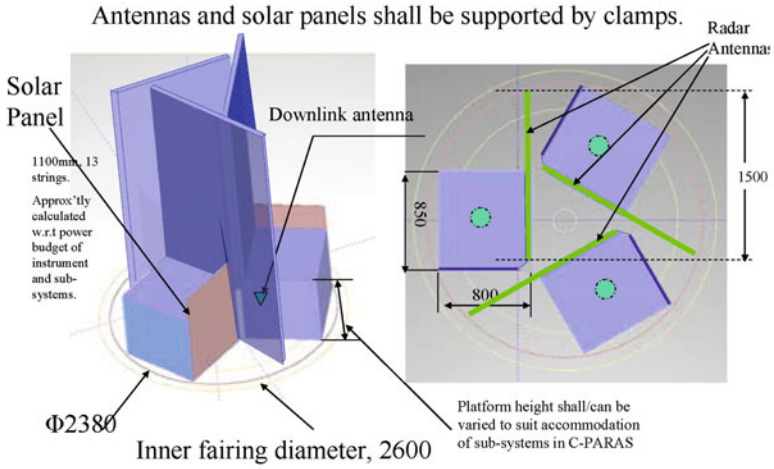


Fig. 16.15 C-PARAS launcher accommodation on VEGA

Table 16.14 C-PARAS payload characteristics

Frequency, polarisation	C-band, VV polarisation
Measurement geometry	IWS mode, incidence angle 30° for IWS1
Key performance parameters	NESZ ≤ -23 dB (including Scan-in-Receive)
Mass	111.8 kg (incl. 20% margin)
Power (payload)	112 W peak/6.2 W mean over orbit (incl. 20% margin)
Antenna dimension	3,600 mm × 1,400 mm (fixed, non-deployable)
Data rate	<200 Mbps
Data/orbit	35 Gbits
Telemetry	X-band

to provide phase compensation information for use in ground interferometric product generation.

Interferometric products will only be generated from the SAR images captured by the C-PARAS satellites themselves. Although the C-PARAS receivers are not phase coherent with the transmitter on Sentinel-1, it is the relative phase differences between the three C-PARAS receivers themselves which matter for the interferometric products (N.B. C-PARAS images cannot be combined with images taken by Sentinel-1 because the separation exceeds the critical baseline). The synchronisation data obtained via the C-PARAS inter-satellite links allows the term due to the Sentinel-1 transmitter to be cancelled out in forming interferograms from pairs of C-PARAS images.

There are two levels of timing synchronisation required between C-PARAS and Sentinel-1:

1. Mode synchronisation to ensure that the C-PARAS data-take occurs during the desired Sentinel-1 operational mode (i.e. IWS mode 1).

2. Alignment of the received radar echoes with the correct transmit pulses from Sentinel-1 and the construction of the range-azimuth matrix as a precursor to formation of the SAR image, and steering the C-PARAS antenna so that it efficiently captures the echo of the Sentinel-1 illumination.

The mode synchronisation can be achieved using GPS timing. A secondary (optional) synchronisation scheme linking event timing between C-PARAS and Sentinel-1 would be the inclusion of an antenna to attempt to receive directly the radar pulse signal emitted from Sentinel-1. This is expected to be very difficult because of the relative orientation of C-PARAS to Sentinel-1 which places the C-PARAS satellites in a null of the Sentinal-1 antenna pattern. A low power transmit and receive system will also be implemented on each of the C-PARAS satellites so that each C-PARAS satellite has knowledge of the relative phase of the master oscillators on the other two satellites.

The antenna steering and data recording are more involved. The C-PARAS antenna will be steered in elevation by rolling the spacecraft, and in azimuth by yawing the spacecraft to counter the effect of the Earth's rotation. In addition the C-PARAS beam must be steered electronically in azimuth to track the azimuth steering of the Sentinel-1 ScanSAR antenna pattern during IWS mode. Without this, the sensitivity is degraded. If Scan-in-Receive (SIR) is implemented, this will also require electronic steering in elevation.

The C-PARAS processor can align its operational mode to Sentinel-1 using GPS data, however the C-PARAS receiver Timing Control Module (TCM) is not synchronised with its counterpart in Sentinel-1 and so precise gating of the C-PARAS receiver window to match the transmitted pulses is not possible because of the inevitable timing drifts between the two USOs. It is proposed therefore that the C-PARAS receiver simply records continuously during the radar data-take. The digitised sample stream comprises alternating segments of noise and radar echoes. The entire dataset is stored and forwarded to the ground for post processing, where the range-line sample windows can be accurately aligned with the transmitted pulses using the detailed timing information in the Sentinel-1 datation.

In Sentinel-1 IWS mode, the echo window duration (swath period) is typically <50% of the PRI. In the Sentinel-1 RX processor which has access to the precise pulse timing from the TCM, and so can gate its sample reception window precisely to the echo, this results in an ~twofold reduction in mean data rate out of the processor to the Solid State Mass Memory (SSMM) unit, compared to the free running system proposed for C-PARAS. However the C-PARAS receiver should be able to define a sample reception window which is accurate enough to envelope the pulse echoes associated with a given IWS mode, based on its GPS timing and its TCM, as illustrated in Fig. 16.16; this will reduce the mean data rate from the processor by ~ a factor of 3 to ~200 Mbps.

It is nonetheless proposed that the processor retain the flexibility to record the complete data take during initial operations of C-PARAS; if necessary, the down-link TM constraints can be solved by either taking a shorter data-take period or taking data only every third orbit. Once the system is proven, then the extended data-take can be performed with on-board windowing of the selected IWS mode.

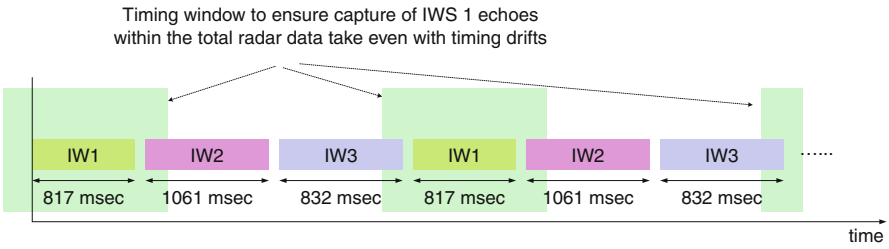


Fig. 16.16 Coarse receiver timing window to capture IWS1 echoes

Table 16.15 Relative height error of DEM–BASIC (for 600 m baseline)

Error in baseline	Sensitivity (NEσ^0)		
	−23 dB	−20 dB	−17 dB
0 mm	2.8 m	3.4 m	4.3 m
1 mm	3.7 m	4.2 m	5.2 m
2 mm	4.5 m	5.1 m	6.1 m

Summary of height errors predicted versus sensitivity baseline error (σ^0-mean)

16.3.5 Mission Performance

Interferometric performance for DEM generation has been studied based on various options of the SAR antenna design and its sensitivity. The 90% point-to-point height error was estimated for two DEM products (DEM-BASIC and DEM-PREC) for mean backscatter and 90% probability, assuming backscatter properties of short vegetation and taking various error sources (e.g. baseline uncertainty) into account.

The DEM-BASIC product (30 m resolution and nine independent looks) has a 90% height error of about 3.5 m for mean backscatter conditions. At 95% probability backscatter the 90% point-to-point height error is below 5 m for most parts of the swath using the 3.6 1.4 m antenna with Scan-in-Receive implemented. By assuming a baseline error of about 1 mm, the height error increases by 0.5 m at near range of IWS1 to about 1.5 m at far range of IWS3.

The DEM-PREC product (resolution of 90 m and 81 independent looks) provides improved height accuracy at reduced spatial resolution. By neglecting the baseline inaccuracy, the 90% point-to-point error is below 2.1 m for the two antennas with Scan-in-Receive implemented. Increasing the baseline error by 1 mm, the height error increases by about 0.5 m at near range of IWS1, and by about 1 m at far range of IWS3 (Table 16.15).

16.3.6 Critical Technologies

The major development activities required for C-PARAS are associated with the radar antenna. About half of the instrument equipment can be derived from systems

with flight heritage (e.g. ERS, TerraSAR-X/TanDEM-X), or Sentinel-1. The parts which need new development are the lightweight patch array antenna, the Scan-in-Receive module and the intra-C-PARAS satellite sync link antenna.

16.3.7 Summary and Exploitation

The passive exploitation of radar signals transmitted by spaceborne SAR systems is able to provide valuable information for a wide range of applications. The C-PARAS mission represents a Cartwheel formation consisting of three receive-only satellites, making use of scene illumination provided by the Sentinel-1 C-band SAR. The generation of digital elevation models has been identified as the primary application.

The main C-PARAS products are digital elevation models with different specification for spatial resolution and vertical resolution. Accurate up-to-date digital elevation models are the basis for many applications in geoscience and technology, e.g. in hydrology, geology, environmental sciences, construction activities, planning of roads and supply networks, etc. The availability of accurate DEMs is also a pre-condition for high precision geocoding of satellite data. Table 16.16 summarizes the main available global DEM data sets.

GTOPO30, is a global low resolution data set with about 1,000 m pixel size, based on various sources of elevation data. For many applications, the resolution and accuracy is not sufficient.

Table 16.16 Overview of available data sets DEMs (global coverage) (from http://www.ersdac.or.jp/GDEM/E/2.html)

	ASTER G-DEM	SRTM (V3)	GTOPO30
Data source	ASTER	Space shuttle radar	From organizations around the world that have DEM data
Generation and distribution	METI/NASA (planned)	NASA/USGS	USGS
Release year	ca. 2009 (planned)	2003 ~	1996
Data acquisition period	From 2000 to ongoing	11 days (in 2000) DEM grid 90 m (30 m) (3 arc-sec)	1,000 m (30 arc-sec)
DEM accuracy (stdev)	±7 m	±10 m	±30 m
DEM accuracy 90% point-to-point	±10 to ±30 m Depending on GCPs	Ca ±16 m	Ca ±48 m
DEM coverage	83°N 83°S	60°N–56°S	Global
Area of missing data	Areas with no ASTER data due to persistent cloud cover	Topographically steep areas (due to radar characteristics)	None

SRTM is an interferometric bistatic SAR-based DEM data set acquired in 2000. Although in the science community SRTM is widely used, it has several disadvantages related to data gaps due to radar layover and shadow, which are found especially in mountainous areas. Another limitation of SRTM is that this data set does not cover high latitudes. These regions are becoming increasingly important as they are strongly affected by climate change.

The ASTER-G DEM is a global DEM data set derived from ASTER stereo image pairs. Despite global coverage up to 83° of northern and southern latitude, there are data gaps due to cloud cover. The accuracy varies significantly from region to region. Toutin [2] estimated that the expected geometric and stereoscopic performances of the standard DEM products are 10–30 m for the geo-positioning and elevation accuracy depending upon the number, quality of GCPs used, and type of terrain.

For the C-PARAS mission two types of elevation data sets are proposed: the DEM-BASIC and the DEM-PRECISION.

The **DEM-BASIC** product has a horizontal grid size of 30 30 m, and a 90% point-to-point vertical height error below 4.5 and 6 m for areas with high and low backscatter respectively, for most parts of the swath assuming an antenna with NESZ <−23 dB (i.e. for a Scan-in-Receive implemented antenna of size 3.6 1.4 m) and a baseline uncertainty below 1 mm. This DEM addresses the sequential filling of gaps and improvement of the global DEM. The horizontal resolution is slightly lower than the DTED-3 requirements, whereas the vertical accuracy fulfils the requirements. The DTED-3 mapping requests 12 m horizontal grid, which is particularly relevant for built-up areas. Outside of built-up areas, the 30 m grid size should be sufficient for most applications. This product would be attractive to a wide range of scientific and operational applications, as mentioned previously.

The **DEM-PRECISION** product will be primarily generated for level and moderate topography terrain, where spatial resolution is less important than the vertical accuracy. This DEM has a horizontal grid size of 90 90 m, and a vertical 90% point-to-point error of 2 and 3 m for areas with high and lower backscattering properties respectively (baseline uncertainty below 1 mm, Scan-in-Receive implemented antenna of size 3.6 1.4 m). It is expected that this product would replace the DEM-BASIC product in moderate terrain areas for applications where vertical accuracy is of prime importance. In addition, a particular focus will be on repeat mapping for applications in geoscience, e.g. to study changes in the volume of glaciers and ice streams, to map accumulations of alluvial deposits, to map inflation and deflation in volcanic areas, etc.

The C-PARAS DEM products will provide significant improvements compared to the SRTM and ASTER-G DEMs:

- Improved spatial accuracy and vertical 90% point-to-point height error when compared to SRTM and ASTER-G.
- Filling of data gaps in mountain areas caused by the SAR imaging geometry.
- C-PARAS has the potential for near global access.

- C-PARAS has the potential for repeat coverage in order to monitor elevation changes.

As C-PARAS is operating in formation with Sentinel-1, the DEM products will also support directly the geocoding process of Sentinel-1 SAR data and the retrieval of calibrated backscatter coefficients, where an accurate measurement of the local incidence angle is needed.

Secondary applications for the C-PARAS mission include the estimation of motion fields using along-track interferometric pairs between the C-PARAS members. The use of three satellites with different along-track distances would enable the observation of a wide range of velocity magnitudes. With this option the along-track SAR interferometric technique can be exploited, including the application for mapping the velocity of ocean streams and fast moving sea-ice.

16.4 Acknowledgments

The work here described was performed under ESA Contract 20395/06/NL/JA as part of the European Space Agency's GSP Programme. The study was led by Astrium Ltd, with the support of Astrium SAS, Astrium GmbH, ENVEO, GMV and Verhaert Space.[2] Grateful thanks is given to the ESA Study Management of Karsten Strauch[3] and Kristof Gantois,[4] and to the rest of the technical team comprising Helmut Rott,[5] Thomas Nagler,[6] Bernhard Grafmueller,[7] Georg Egger,[8] David Hall,[9] Steve Kemble,[10] Alice Robert,[11] Marline Claessens[12] and Christina de Negueruela Alemán.[13]

[2] Now QinetiQ Space NV.

[3] ESTEC, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands.

[4] ESTEC, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands.

[5] ENVEO, ICT Technologiepark, Technikerstrasse 21a, A-6020 Innsbruck, Austria.

[6] ENVEO, ICT Technologiepark, Technikerstrasse 21a, A-6020 Innsbruck, Austria.

[7] Astrium GmbH, 88039 Friedrichshafen, Germany.

[8] Astrium GmbH, 88039 Friedrichshafen, Germany.

[9] Astrium Ltd, Anchorage Road, Portsmouth, PO3 5PU, England.

[10] Astrium Ltd, Gunnels Wood Road, Stevenage, SG1 2AS, England.

[11] Astrium SAS, 31 rue des Cosmonautes, Z.I. du Palays, 31402 Toulouse Cedex 4, France.

[12] Verhaert Space, Hogenakkerhoekstraat 99150, Kruibeke, Belgium.

[13] GMV AEROSPACE AND DEFENCE S.A., Isaac Newton 11, P.T.M. Tres Cantos, 28760 Madrid, Spain.

Appendix

Abbreviations and Acronyms

ACT	Across-Track
ADC	Analogue to Digital Converter
ADPMS	Advanced Data and Power Management System
AERL	Aerosol Land Mission
AIT	Assembly, Integration and Test
AIV	Assembly, Integration and Verification
ALT	Along-Track
AOCS	Attitude and Orbit Control System
APE	Absolute Pointing Error
ASAR	ENVISAT Advanced Synthetic Aperture Radar
AVP	Anti-Velocity Panel
BUP	Bus Panel
CFRP	Carbon Fibre Reinforced Plastics
COMS	Communication, Ocean, and Meteorological Satellite
C-PARAS	C-band PAssive RAdar Satellite(s)
DEM	Digital Elevation Map
DSP	Digital Signal Processor
DTAR	Distributed Target Ambiguity Ratio
DTED	Digital Terrain Elevation Data
EPC	Electronic Power Conditioner
ERS	European Remote Sensing Satellite
FDIR	Fault Detection, Isolation and Recovery
FDS	Flight Dynamics System
FEE	Front End Electronics
FF	Formation Flying
FFCC	Formation Flight Command and Control
FM	Flight Model
FOAMO	Foam Insulated Master Oscillator
FOS	Flight Operations Segment
FOV	Field of View
GCP	Ground Control Point
GOCI	Geostationary Ocean Colour Instrument
HPA	High Power Amplifier
HPGP	High Performance Green Propellant
IF	Intermediate Frequency
InSAR	Interferometric SAR

IOD	In-Orbit Demonstration
ISL	Inter-Satellite Link
IWS	Interferometric Wide Swath
LEO	Low Earth Orbit
LEOP	Launch and Early Operational Phase
LHS	Left-Hand Side
LNA	Low Noise Amplifier
LO	Local Oscillator
LST	Local Solar Time
LTAN	Local Time at Ascending Node
MERIS	Medium Resolution Imaging Spectrometer
MSI	MultiSpectral Instrument
NESZ	Noise Equivalent Sigma Naught
NRT	Near Real-Time
OLCI	Ocean and Land Colour Instrument
PCDU	Power Conditioning and Distribution Unit
PDGS	Payload Data Ground Segment
PDHS	Payload Data Handling Subsystem
PLP	Payload Panel
PM	Particulate Matter
PolinSAR	Polarisation InSAR
PRI	Pulse Repetition Interval
PROBA	Project for On Board Autonomy
PSU	Power Supply Unit
RHS	Right-Hand Side
RF	Radio Frequency
ROIC	Readout Integrated Circuit
RPE	Relative Pointing Error
S/C	Spacecraft
SAR	Synthetic Aperture Radar
SIR	Scan In Receive
SNR	Signal to Noise Ratio
SRTM	Shuttle Radar Topography Mission
SSMM	Solid State Mass Memory
SSO	Sun Synchronous Orbit
STALO	Stable Local Oscillator
STR	Star Tracker
TCM	Timing Control Module
THEOS	THailand Earth Observation System
TOA	Top of Atmosphere

TOPOLEV	Topographic Levelling mission
TOPS	Terrain Observation with Progressive Scans (ScanSAR with azimuth sweep-scanning)
TRL	Technology Readiness Level
TRM	Transmit Receive Module
TWT	Travelling Wave Tube
USO	Ultra Stable Oscillator
UTC	Universal Time, Coordinated
VP	Velocity Panel
WV	Water Vapour
ZP	Zenith Panel

References

1. EO Small Missions Study (Phase 2) (2009) Study Final Report, ESTEC Contract No.: 20395/06/NL/JA, Iss 1 Rev 1, 18 Aug 2009, Astrium Ref. EOSM-ASU-RP22
2. Toutin T (2008) ASTER DEMs for geomatic and geoscientific applications: a review. Int J Remote Sens 29(7):1855–1875
3. Zink M, Krieger G, Fiedler H, Moreira A (2008) The TanDEM-X mission concept. In: Proceedings of the EUSAR 2008 conference, vol 4, Friedrichshafen, 2–5 June 2008, pp 31–34

Chapter 17
The SAR Train

Jean Paul Aguttes

Abstract The concept implements the coherent combination of N separate SAR flying along a same orbital arc as seen from ground. A "Signal Cleaning" mode of SAR train keeps unchanged the antenna area requirement of each individual SAR and brings a factor N advantage that applies on SNR and ambiguity protection. The main formation flying constraint is the width of the tube containing the satellite trajectories. The multiplication by N of the total antenna area is the other counterpart to these advantages. A "Antenna Dilution" mode of SAR train enables the distribution of an unchanged total antenna area into N smaller elementary antennas, together with the multiplication by N of the SAR Merit Factor (Swath over Resolution ratio). With respect to the first mode, the tube width constraint is increased and the space-time separation along the track has to be very accurate. Use of appropriate spread spectrum waveforms instead of conventional pulse waveforms removes the major part of the extra orbit constraints introduced by the "antenna dilution" class. A train of N SAR in visibility with a single transmit makes the concept more robust against lost of coherence and eases the metrology of the formation (DGPS). Moreover the global energy efficiency is increased by N since with only a single transmit SAR the same performance is achieved. However, the along track separation constraints for antenna dilution are made more stringent because restrained to the space domain, which reinforces the spread spectrum interest. As part of its applications, the concept can circumvent the matter of huge antenna size for SAR mission in very low frequency (P band) or at high altitude (surveillance).

J.P. Aguttes (✉)
CNES, Toulouse Space Center, 18 Av. E. Belin, 31401 Toulouse, France
e-mail: jean-paul.aguttes@cnes.fr

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_17,

17.1 Introduction

It is well known that antenna area and power are the major design constraints of a spaceborne SAR. Both limit the Swath over Resolution ratio that is the key performance criteria and that we call here merit factor.

There is a particular SAR sizing called critical sizing [1] which maximizes the achievable merit factor at a given antenna area. Moving to over critical sizing causes great extra antenna area for small extra merit factor (+100% for +50%) whereas the merit factor must be sacrificed well under the critical one to get a substantial antenna area saving (−75% for −30%). Merit factor at critical sizing is about 10,000 for a maximum incidence in the range 50°–60°. Compared to antenna area, the power sensitivity to the merit factor is more linear.

There are modes and design [2] for circumventing these constraints although they cannot simultaneously relax antenna area and power constraints.

SAR train [3, 4] is a concept of collaboration of several SAR along the same track with two cumulative purposes:

- To dramatically push the practical limitation for power or antenna area by diluting them on several bodies.
- To remove the SAR merit factor barrier and to multiply the merit factor achievable under a given total antenna area.

The N SAR fly on the same orbital arc in an earth frame, which means that the SAR are generally on different orbit planes because of the across orbit earth rotation between successive passes. This time delay is supposed to be under the limits which cause the temporal decorrelation of the backscatter, the SAR are not necessarily in visibility although visibility provide a particular implementation case.

17.2 A Train of N Monostatic SAR

17.2.1 Basic Advantage: Energy Addition

The N SAR provide N identical Synthetic Antennas (SA) focused on the processed point at different times. The addition of the N passes of the same (synthetic) antenna brings an improvement by N of the energy budget and of the SNR.

We will now look in more detail to improve the understanding and discover the other properties and advantages of such SAR Trains.

17.2.2 Modelling as a N Element Antenna Array

We use the representation of a synthetic antenna; that is to say the addition along the orbital arc of S identical signal samples received by the real antenna at S position along the orbital arc and previously corrected (SAR processing) from the

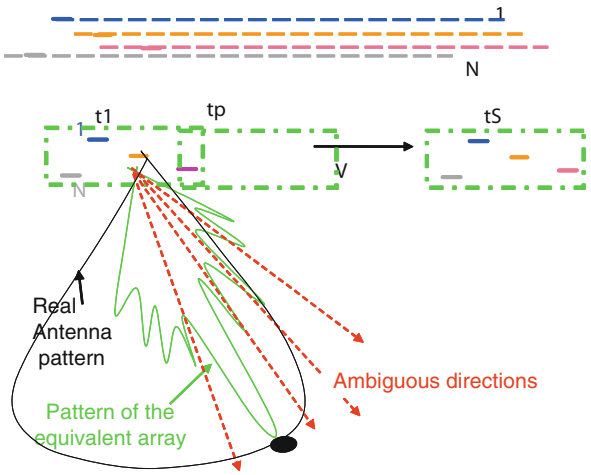


Fig. 17.1 Array model of a SAR train

range variation to the processed pixel. This correction makes the orbital arc as circularly enrolled around the pixel. One can therefore consider the addition of the N synthetic antennas as a first addition of N samples coming from different SAR and previously corrected and then a second addition of the successive positions along the orbital arc of that composite N sample. The composite sample is equivalent to the sample issued from a single array antenna whose the N elements are the antennas of the N SAR located at the N sampling points. Thanks to the SAR processing correction, the array automatically matches a pattern steered toward the processed pixel. For this equivalence, we consider that each element receives only the ground echo of its own transmit signal and nothing from the others which is already the modelling used for the long array representation of the synthetic antenna. The way the N samples are selected to form an array is indifferent providing that, by the end of the addition, we forget none of the NS samples and we do not count one several times.

We can consider a fixed array geometry flying all along the SA in spite of edge effects (where some elements get out of SA) provided that the Wave Form (WF) is periodic (case of pulse waveform) and SA length a multiple of the recurrence (V·period). Indeed, each sorting element can be replaced by the same element when it was entering without changing the array response neither on the processed pixel nor on any of the ambiguous pixels.

As shown by Fig. 17.1, the array pattern provides a coherent addition of the N useful energies from target, which corresponds also to an improvement by N of SNR if the antenna area is kept unchanged; whiles it reduces the ambiguous energy.

As seen later, with efficient array geometries or/and Waves Forms (WF), the ambiguities are so reduced than they can be tolerated within the real antenna pattern for merit factor improvement and elementary antenna area reduction.

17.2.3 Random SAR Separation and Ambiguity Reduction: Signal Cleaning (SC) SAR Train Mode

We call section the sampling interval of a given SAR p0 and PRF the sampling frequency. The section length V/PRF is normally a bit smaller than the half antenna length. Because the N pulse trains are not synchronised, within a section we find a distribution of N − 1 other samples made later or earlier by a delay ($t_{p0} - t_p$). We consider for a while an array model with the N elements of a same section.

For a synthetic antenna normally sampled at V/PRF, the ambiguous directions are off steered by kλ PRF/2V. Here the ambiguity k is received by a N elements array according to a range difference between consecutive elements that is a portion of kλ since the elements spacing is a portion of V/PRF. The distribution of the element separations is random and the N contributions of each ambiguity are added non-coherently while the N contributions from the processed pixel are added coherently. Together with the improvement by N of the SNR, there is an improvement by N of the ambiguity protection.

17.2.4 Accurate Time Space Separation of Satellites (Modulo V/PRF): Antenna Area Dilution (AD) and Merit Factor Multiplication SAR Train Mode

17.2.4.1 Synthetic Antenna Sampling Understanding

We suppose now that the spacing within the array is fixed and equal to V/NPRF, that is to say that the real spacing between a given satellite and each of the N − 1 others, at the (not necessary same) times they transmit a same sample is K V/PRF + p V/NPRF with K undetermined and p = 1 to (N − 1). The superposition of N sampling of SA makes a single synthetic antenna whose sampling interval is reduced by N. We can therefore reduce by N the real antenna length, increase by N the synthetic antenna length and improve by N the azimuth resolution without any impact on ambiguity with respect single SAR.

17.2.4.2 Antenna Array Understanding

A second way to understand that is to see that the ambiguity k is received N times according to a range deviation distribution as

$$2\ p\ V/NPRF{\cdot}k\ \lambda\ PRF/2V = pk\lambda/N$$

with p = 0 to N − 1.

The result of the complex addition is null except when k is multiple of N where the addition is coherent. Therefore with a reduction by N of the antenna length and

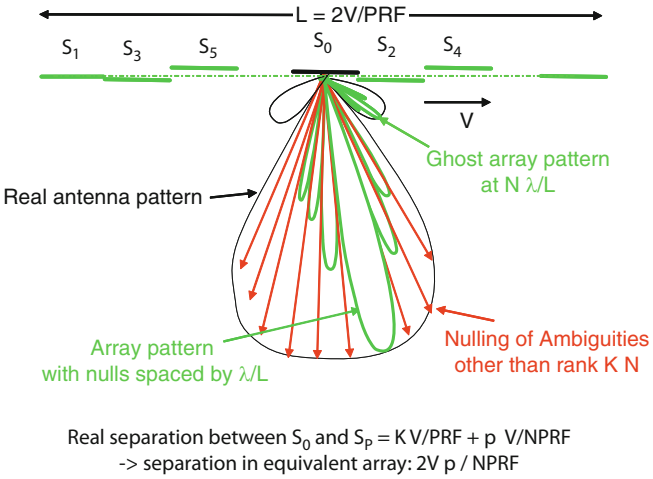


Fig. 17.2 SAR train with accurate spacing (pulse WF)

a stretching by N of both the antenna pattern and the synthetic antenna length we get the same ambiguity configuration as in single SAR, but with a resolution improved by N at unchanged swath since PRF is unchanged.

Without antenna length reduction, there is ambiguity reduction since the number of ambiguities is divided by N. This reduction is greater than in the case of random separation. Indeed, rather than having a statistical reduction by N of each ambiguity, we have a complete suppression of the dominating ambiguities close to the antenna beam together with a reduction by N of the rest of the ambiguous energy on the side lobe plateau. In this SAR train concept, we have to choose between improvement of merit factor and those of ambiguity protection.

We can reach the same conclusion by considering the array pattern. We consider an array model whose samples are taken in a given sampling interval V/PRF for half of the N SAR and in the next one for the others. We see again (Fig. 17.2) that the array pattern nulls all the ambiguities except those of rank multiple of N that fall on the grating lobes of the array.

Any mismatch of the Time Space SAR separation affects the quality of the cancellation of ambiguities other than kN and create extra ambiguity with respect to usual single SAR protection. This brings the SAR formation flying constraints analysed later.

17.2.4.3 Concept Merits

The basic point of the concept for merit factor improvement is that it enables the N first ambiguities to be within the real antenna pattern without changing the total ambiguity budget. This situation can either result from an antenna length reduction

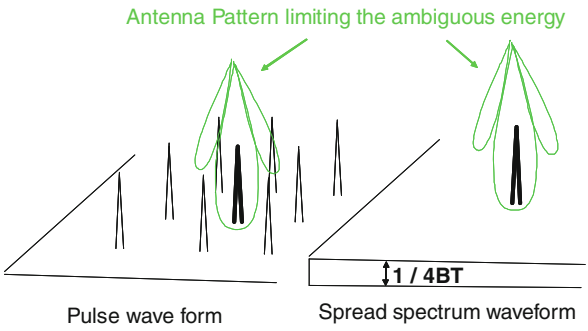


Fig. 17.3 Ambiguous energy seen by antenna pattern

for a resolution improvement, or a PRF reduction by N, a swath improvement by N and an antenna elevation size reduction by N at equal range ambiguity. In both case, this corresponds to a Merit improvement by N and an elementary antenna area reduction by N.

The reduction by N of the antenna area reduces by N^2 the level of the elementary sample collected by each antenna. In spite of the coherent signal addition made by the SAR train, SNR is N times lower than the one available on each SAR with the original antenna area.

We can then summarize the concept and say that a regular interleaving of the sampling of the N SAR enables simultaneously:

- An improvement by N of the merit factor.
- A reduction by N of the antenna area or, in other words, a dilution of an unchanged total antenna area into N smaller parts.
- A degradation by N of the SNR for a total power increased by N or, in other words, a total transmit power N^2 times greater for a same SNR.
- An increase by N of the SAR data volume in line with the increase of the merit factor.

Note that the only other way to increase the merit factor should be to have N independent T/R antennas imaging N juxtaposed swaths and requiring a total antenna area N times greater.

17.2.5 Spread Spectrum Wave Form

As shown by Fig. 17.3, with standard pulse WF the ambiguity is concentrated in pikes and dominated by the first ones which need to be precisely maintained on the fall off of the main lobe in standard single SAR. This is the origin of the criticality of SAR design.

We saw here before that the SAR Train enables the pikes within the real pattern only if they are quite nulled (at price of accurate space-time separation). A continuous spread spectrum waveform can be used in SAR, the ambiguity is

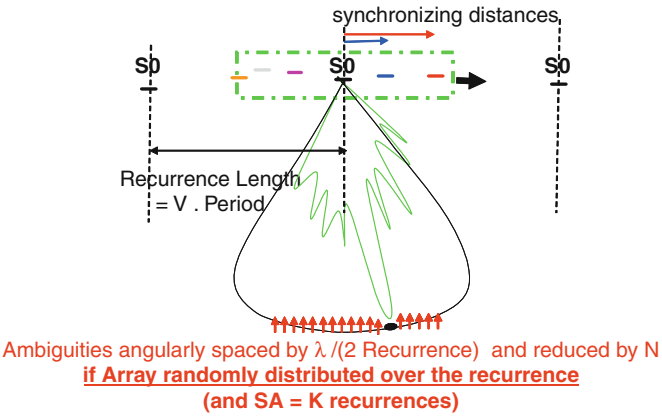


Fig. 17.4 SAR train with a single spread spectrum WF

then distributed at low level on all the pixels and of course dominated by the part present within the real antenna footprint. It can be shown [see annex] that for a single SAR this spread ambiguity brings the same constraints on merit factor and minimum antenna area. However, the relationship is now linear: any reduction of ambiguous energy that can be achieved on any of the pixel (by the array pattern in case of SAR train) enables at unchanged ambiguity performance a proportional increase of the real antenna footprint and of the merit factor and decrease of the antenna area.

If the N SAR use non-correlated waveforms, the signal addition is coherent for the N contributions issued from the processed pixel while it is non-coherent for the N contributions coming from any of the other pixels, which corresponds to a relative reduction by N of the ambiguous energy present on each pixel. With SS WF, we have the same behavior as for pulse WF with the same choice between Signal Cleaning (ambiguity reduction and SNR improvement) and Antenna Dilution (+Merit multiplying), except that here antenna dilution does not bring constraints on the satellite space time separation.

When the N SAR use a same WF, the array model helps in understanding (Fig. 17.4). As said, it works with elements transmitting the same signal (not necessary at same time), i.e. spaced by synchronizing distances. An array length much greater that SA (Synthetic Antenna) corresponds to a case of several uncorrelated WF.

For an array length of SA/q the array pattern is less sharp and does not ensure reduction on the ambiguity issued from the q pixels around the processed pixel and therefore the Train benefit is affected: we cannot improve the real foot print by N without degrading the ambiguity.

However if the WF is periodic with a recurrence length SA/q, the ambiguity is then concentrated on pixels of rank $\pm K \cdot q$ (with respect to the processed one) that are all statically reduced by N provided that the array is randomly distributed over the recurrence (or a multiple of the recurrence), which is not actually a high

formation constraint. SA must also be a multiple of the recurrence for keeping the Train properties even on SA edges (see Sect. 17.2.2).

17.3 Formation Geometry Knowledge and Control

The N SAR do not fly exactly on the same orbital arc in earth frame and do not respect exactly the relative position assignment, if any, along the orbital arc.

We suppose that the N SAR trajectories are randomly distributed within a tube of width ω (at 3σ).

17.3.1 Example of SAR Train Orbital Implementation

The N satellites use true polar orbits with identical parameters except the ascending node and the anomaly whose respective deviations $\Delta\Omega$ and ΔM verify a particular relation.

Ts and Te are the satellite and earth rotation periods and R is earth radius. At a given latitude L, the tracks on an earth supposed without rotation are separated by R Cos(L) $\Delta\Omega$. During the time delay Ts $\Delta M/(2\pi)$ between the two passes at the latitude L, the earth rotation at this latitude is R Cos(L) Ts ΔM/Te. Therefore the two satellites pass over the same point if Ts/ Te $= \Delta\Omega/\Delta M$.

17.3.2 Basic Conditions for Coherent Signal Addition

The vector formed by two elements of the equivalent array considered at the middle of the synthetic antenna has three components. One is along the speed vector and is corrected in term of delay deviation to the processed pixel by the SAR processing correction. One is along the line of sight. The third one is orthogonal to both the speed vector and the line of sight and is called the interferometric base B_{orth}. The second component taken alone give a phase difference between synthetic antennas outputs common across a significant part of the swath that can be easily identified by averaging and removed, but the effect of the interferometric base must be corrected first.

If the interferometric base is too large the samples from a same pixel are not at all coherent, this first working condition is the one of the InSAR (cross track interferometry) and is shown to be directly linked to the range resolution [5]. The two SAR must be within the 3 dB aperture of the equivalent antenna made by the range cell. Since the present concepts are aimed at improving the SNR or the merit factor, we can assume that they are applied in a context of rather good resolution. Therefore, this boundary does not really limit the application domain. For 1 m

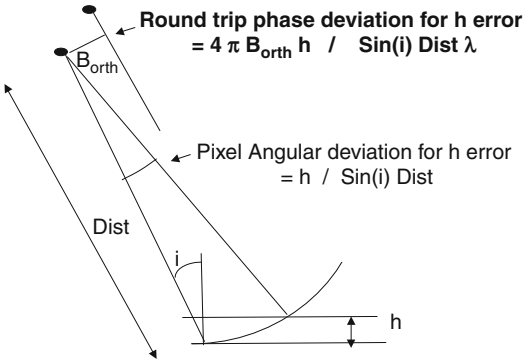


Fig. 17.5 Phase deviation versus relief knowledge

resolution and 600 km range at 20° incidence, the base must be smaller than 10 km in X band and 120 km in L band.

However, under that coarse boundary the phase difference between two SAR contributions from a processed pixel varies across the swath and can be corrected only if the relief is known. Figure 17.5 gives the phase difference sensitivity to the relief knowledge accuracy and to the interferometric base. The corrections are made for each of the N SAR with respect to a reference and virtual trajectory in the center of the tube. The interferometric bases with respect to this reference trajectory are random, with a null mean value and a standard deviation ($\omega/6\sqrt{2}$). After the correction, the N contributions from a given pixel whose altitude knowledge is h (1σ) has a residual phase deviation with a RMS value ε given by:

$$\varepsilon = 4\pi \text{ h } \omega/ \text{ 6 Sin(i) Dist } \lambda\sqrt{2}$$

The effect of this phase deviation on the module of the complex summing of N contributions is:

$$\sum \text{Cos}(\varepsilon_p)/N = \text{e}^{-\varepsilon^2/2}$$

Assuming the relief knowledge better than ±50 m (3σ) and worse case conditions with small distance to target (600 km), small incidence (20°), small wavelength (X band = 3 cm), the tube width (3σ) must be smaller than 90 m if we went to limit to 0.5 dB the statistical effect on power. This is within the state of art of formation flying.

The knowledge error δB_{orth} on interferometric base length gives a phase error after the interferometric correction:

$$4\pi\, \delta B_{orth}\ h_{max}/\text{Sin(i) Dist } \lambda,$$

where h_{max} is the terrain height variation in the image. Assuming that h_{max} is smaller than 1,000 m, this error can be brought negligible if δB_{orth} is smaller than 10 cm. Such accuracy can be easily achieved with differential GPS or Doris.

Note that the flight constraints are here computed in a very worse case. The concern for energy budget or merit factor rather applies at high incidence and high distance. For Dist = 1,500 km and i = 50°, the tube width limit becomes 500 m. If now the system works in L band, the width limit turns to be 4,000 m.

17.3.3 Conditions for Ambiguity Reduction (Std Pulse Waveform)

The ambiguity reduction is based on the fact that the N contributions of the ambiguity k are not coherent when the contributions from the processed pixel are made coherent by the processing and the interferometric corrections. A random spacing along the track introduces randomness between the ambiguous contribution paths and the tube width does not matter. There is no tube width constraint else than those related to the good interferometric correction.

17.3.4 Conditions for Merit Factor Improvement and Antenna Dilution (Std Waveform)

17.3.4.1 Impact of Along Track Assignment Error

The samples of the N SAR are not regularly interleaved as described in Sect. 17.2.4. The standard deviation of the phase of the N contributions is α radians when the standard deviation of the sampling error is $(\alpha/2\pi)$ V/PRF. The ambiguity cancellation is disturbed, the complex summing of the N contributions of the ambiguity k lets a sum of N vectors regularly separated in phase and whose module (positive or negative) is random with standard deviation $k\alpha$ (for small value of $k\alpha$) if 1 is the module of the contribution. The residual part of the received power of the ambiguity k is $N\ k^2\alpha^2$. The power on processed pixel is increased by N2, therefore the residual protection on the ambiguity k is $k^2\alpha^2/N$.

In a standard unambiguous design, the two first ambiguities (forward and backward) are close to the first nulls of the beam when the antenna is at the middle of the synthetic antenna. We consider now that the merit factor is increased by N. At the middle of the synthetic antenna, there are $N - 1$ forward ambiguities and $N - 1$ backward within the beam aperture (defined between the two first nulls). At any one position along the synthetic antenna only half of these $2(N - 1)$ ambiguities are present within the 3 dB antenna aperture. The real antenna pattern halves the total residual ambiguous energy integrated along the arc. The total residual protection against these $2(N - 1)$ ambiguities is therefore:

$$(\alpha^2/\ N)\ \Sigma\ k^2{}_{k=1\ \text{to}\ N-1}$$

The residue of the cancellation by the array pattern of the ambiguities brought within the real antenna pattern (when the merit factor is increased by N) is an extra ambiguity with respect to usual single SAR. In order not to affect too much the usual required Doppler ambiguity protection typically around −25 dB, we should have the residual protection within the array pattern better than −30 dB.

The constraint increases rapidly with N and is already very stringent with N = 2. Indeed α must be smaller than 1/22 radians and the synchronization of the sampling must be better than a part out of 138 of the sampling interval or of the half of N antenna lengths, that is to say in the order of the centimeters.

However, we deal with a space time requirement that can be rather managed in Time domain by locking the transmit time (i.e. the PRF) of the different SAR to the precise measurement of their relative location along the track.

17.3.4.2 Impact of the Trajectories Tube Width

If the altitude of the ambiguity k is equal to the estimated altitude considered for the interferometric correction on the processed pixel, the destructive combination of the N contributions of the ambiguity k is perfect, except for the error coming from base line knowledge but this contribution is well above altitude error. Therefore, the trajectory tube width does not affect the ambiguity cancellation described in Sect. 17.2.3. An altitude deviation H_k with respect to this estimate introduces a random phase deviation of standard value ε_k on each of the N contributions, with:

$$\varepsilon_k = 4\pi H_k \omega/6\sqrt{2}\ \mathrm{Sin(i)}\ \mathrm{Dist}\lambda$$

The complex nulling lets a residual part of the received power of the ambiguity k of value N $\varepsilon_k{}^2$ and a residual protection on the ambiguity $\varepsilon_k{}^2$/N. The total residual protection against the 2(N − 1) ambiguities is

$$(2/9)(\pi\omega/\ \mathrm{Sin(i)}\ \mathrm{Dist}\lambda)^2\sigma^2{}_H(N - 1/N)$$

with σ^2H = the mean square of the altitude deviations of the 2(N − 1) first ambiguities (located forward and backward) with respect to the altitude considered for the interferometric correction.

In X band at 50° incidence and 1,500 km distance, with σ_H = 50 m, the tube width must be smaller than 32 m when N is great and must be smaller than 23 m for N = 2 to get the required residual protection of −30 dB. These width values are proportional to λ and to the distance.

The antenna dilution + merit factor improvement application highly tightens the constraint on trajectory width with respect to what is only required for the basic coherent signal addition. It brings also constraints on terrain smoothness σ_H. However the practical implementation of merit factor improvement is primarily

limited by the along track constraint, except in the case of N = 2 or in the case of use of appropriate waveforms.

17.3.4.3 Case with N = 2

The spacing SD between the two satellites must be (2K + 1)V/2PRF.

This can be achieved by locking the PRF on distance measurement. Assuming a rather long antenna of 10 m, the maximum sampling error is ±3.5 cm (see Sect. 17.3.4.1) and is removed by a relative PRF adjustment of $\pm 3.5 \times 10^{-5}$ if the distance between the satellites is 1 km. Such fine PRF adjustment does not conflict with the much coarser one required for running the instrument. The sampling constraint relies only on the accuracy of the distance measurement in the order of 1 cm that can be achieved, when the two satellites are in visibility, with differential GPS or any laser or RF ranging system. The particular aspects of trains with SAR in visibility will be considered later.

The tube width constraint and the terrain smoothness constraints exposed previously can be relaxed since there is only one of the two ambiguities at any one time within the 3 dB aperture. It is possible to make a interferometric correction not any more based on the altitude of processed pixel but based on the forward ambiguity at the beginning of the synthetic antenna and on the altitude of the backward ambiguity at the end (linearly variable correction). The ambiguity cancellation is now driven by the ambiguity altitude knowledge (instead of the overall terrain smoothness σ_H) whereas the actual altitude difference with those of the processed pixel (which is not anymore corrected from interferometric error) affects the coherent signal summing, which is a less sensitive matter. We assume that the altitude RMS knowledge can be reduced to 20 m ($\sigma_H = 20$ m in Sect. 17.3.4.2) then the tube width requirement is 58 m (3σ) and the altitude difference H (3 h in equation of Sect. 17.3.2) between processed and ambiguous pixel must be smaller than 430 m, all this for X band at 50° incidence and 1,500 km distance. The concept is not any more restrained to quite flat areas.

17.3.5 Use of Spread Spectrum Waveforms

For the constraint of coherent signal addition there is no difference with respect to a pulse wave form. The interest of spread waveforms is related to the ambiguities since they are anywhere and benefit from non coherent addition as long as the waveforms are non correlated or, for identical wave form, as long as the synchronizing distances (see Sect. 17.2.5) are random at the scale of the recurrence length and the synthetic antenna is a multiple of the recurrence. This makes rather easy condition in comparison with pulse waveform case. These conditions are not affected by the tube width and the later is only constrained by the coherent signal

addition. All these drastic relaxations of the formation control represent the major advantage of this waveform particularly, as we will see now, in the case of a train with N SAR in visibility.

17.4 Train with N SAR in Visibility

17.4.1 A Practical Case of SAR Train Implementation

All what is said before suppose that neither the backscatter from the processed pixel nor the atmospheric delay change between the different passes. Both actually can vary and this is known as the major limitation of the multi-pass SAR interferometry, unless the time delay between passes is small. A SAR train with several SAR in visibility is a robust configuration against such constraints. This also eases the ranging of the formation.

On the other hand, if all the SAR transmit, each antenna receives N signals that cannot be separated except by using appropriate waveforms. A radical solution is to have a single transmit SAR.

17.4.2 SAR Train with a Single Transmit: Equivalent Train of T/R SAR

As shown by Fig. 17.6, each couple formed by a transmitter and a receiver corresponds on phase (or range) round trip viewpoint to an equivalent T/R antenna located at midpoint between the Transmitter and the Receiver. If the Transmit is also Receive, the equivalent T/R is the real one.

We can then consider an equivalent SAR train formed by the equivalent T/R. Because of this midway location of the equivalent T/R satellite, all the formation constraints previously analyzed are relaxed by a factor 2. However, since there is a single transmit signal, the along track Space Time separation requirement can be

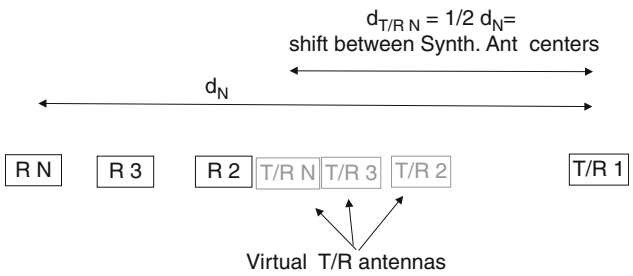


Fig. 17.6 Train of N satellites (N − 1 receive only)

managed only in space domain. The single transmit case makes the concept very uneasy (Antenna dilution + merit factor improvement). For the particular case of two satellites, we have described here before (Sect. 17.3.4.3) a practical time domain solution based on the locking of the PRF from the measurement of the inter-distance. For other cases, this constraint of regular interleaving makes the concept very uneasy. We can imagine a weighting of the elements of the array to provide an array pattern with a very low side lobes level and another type of ambiguity reduction less sensitive to SAR spacing than the very nulling approach. However, to get the same level of reduction requests a great number of extra array elements and satellites without any extra benefit on merit factor or on SNR.

A radical solution for keeping all the concept merits for $N > 2$ is to use a spread spectrum wave form.

All the N antenna need to look toward the same ground area illuminated by the Transmit antenna. All the equivalent synthetic antennas start and finish in the same time but are shifted with respect to those of the T/R by the half of the distance between the T and each R. The SAR combination can be done only within the overlapping part of the equivalent synthetic antennas and the global length is therefore reduced by the virtual formation length or half of the real formation length. This means that the formation length has to be small with respect to the synthetic antenna length.

The transmit signal is added coherently N times. Although there is a single Transmit, the SNR of the N SAR combination is the same as for a train of N T/R without visibility. All the concept merits of a standard train without visibility apply here too but the total transmit power is reduced by N.

With spread spectrum waveforms, the synchronizing distances (Sect. 17.2.5) is much smaller than the synthetic antenna length. Therefore, the Train advantages require a periodic waveform and equivalent T/R randomly distributed along the recurrence length, that is to say real receive satellites randomly distributed on twice the recurrence length.

17.4.3 SAR Train in Visibility with Several Transmits

We consider now that all the SAR of a "visibility Train" transmit and receive and use uncorrelated spread spectrum waveforms. Each transmits satellite drives a SAR Train. There is a supplementary coherent addition between the N SAR Trains with non-coherent addition of ambiguities since the waveforms are uncorrelated. The merit factor is again increased by N with a total antenna area now N times smaller than single SAR at price of a new degradation by N^2 of the SNR. With respect to single SAR, the total antenna area is divided by N, the merit factor increased by N^2, but the total power (now distributed on N bodies) must be increased by N^3 for a same SNR.

17.4.4 A Particular Application Case: GMTI (Ground Moving Target Identification)

It has been shown [6] that an along track formation of four SAR (in visibility) is a good approach for providing GMTI products while saving antenna area. Indeed, the GMTI performance is driven by the SAR antenna length and an along track formation of N antennas allows the construction of a very long antenna, although the gaps between antennas introduce some performance degradation with respect to a true long antenna. Since a GMTI SAR has also to provide standard SAR products, the along track formation can also be used for improving the SAR performance by using the present SAR train technique (N SAR in visibility). In Ref. [6] we use the two modes of the SAR train. The Antenna Dilution (AD) mode which multiplies by N the equivalent SAR antenna Area extends the maximum incidence up to the point that should require an antenna height N time greater in monostatic SAR to keep the same ambiguity ratio. With $N = 2$ (because a AD SAR Train involves a very tight spacing control except if $N = 2$, as seen before) the maximum incidence is increased from 62° to 71°, which gives a dramatic improvement in term of coverage and revisit with exactly the same hardware. For incidence lower than mono-static maximum incidence (62°), we can use the Signal Cleaning SAR Train mode. With $N = 4$ (no tight spacing control in SC SAR Train) either the SNR or the range resolution is improved by 4 together with an improvement of 6 dB of the ambiguity protection.

17.5 Conclusion and Further Works

We have shown the capability of a SAR Train to dilute the power or antenna area constraint on several bodies, and to multiply the merit factor achievable under a given antenna area. We have identified the associated constraints on formation control together with the merit of spread spectrum waveforms to relax them and to extend the application of the concept. More work is needed in several areas, for example:

- Practical implementation of spread spectrum with real ambiguity functions and research of randomness criteria for the synchronizing distances.
- Application in connection with GMTI (Ground Moving Target Identification) and/or in case of predominance of point target concern.
- Application for SAR missions requiring large antenna area (P band and/or high altitude and incidence) and high coverage capability (monitoring).

Appendix: Spread Spectrum SAR

As shown by Fig. 17.3, the Doppler/range ambiguity function with a spread spectrum waveform looks like a thumbtack, with a narrow spike surrounded by a low pedestal of thickness 1/4BT (B = bandwidth, T = integration time) while the same ambiguous energy is concentrated on a grid of well-separated spikes without pedestal in the pulse waveform case. The ambiguous energy AE affecting a given pixel is the sum of the energy on each pixel multiplied by 1/4BT.

We can consider that the polluting pixels are only those present in the main beam footprint. This number is:

$$(\mathrm{T\ V}/r_{az})\left(\mathrm{S}/r_{rg}\right) = (\mathrm{T\ V}/r_{az})(2\ \mathrm{S\ B\ sin\ (i)}/\mathrm{C})$$

with:

V = ground speed
r_{az} = azimuth resolution
r_{rg} = range resolution = C/(2 sin(i) B)
S = swath
Then AE = (S/r_{az}) (Vsin(i)/2C)

We see that, like the case of pulse waveforms, the merit factor S/r_{az} and therefore the minimum antenna area are directly constrained by the ambiguity performance.

One difference with respect to pulse waveforms is that the relationship is linear. The sizing criticality in case of pulse waveforms (see Sect. 17.1) results from the fact that the first lateral (rank 1 ambiguity) spikes dominate the ambiguous energy. Indeed, unlike the other spikes, they cannot be rejected on the plateau of antenna side lobes unless a dramatic increase of the antenna area and reduction of merit factor. They are located in much steeper part of the pattern on the fall off of the main lobe.

The other difference is that for a same ambiguity protection AE about −20 dB, the constraints looks much more severe. Indeed, S/r_{az} should be limited by C/(50 V Sin(i), that is to say by about 1,200 instead of 10,000 with pulse wave form. In fact, the actual ambiguity ratio is the product of the ambiguity protection by the level difference between the ambiguous pixel and the useful one. Because of the large averaging effect, we can consider that this level difference is null in case of spread waveforms. With pulse waveform we need to be much more prudent since the ambiguity is concentrated in only four areas (the rank 1 ambiguities). Assuming a 20 dB total radiometric it makes sense to consider situations where the four areas are globally 10 dB higher that the low level pixels. By taking a 20 dB protection in pulse waveform case and only 10 dB in spread case, we can expect a quite similar statistical result on the actual ambiguity ratio with a value around 10 dB. We can conclude that the two waveforms lead to equivalent constraint for merit factor and minimum antenna area.

Note that in case of a periodic waveform, the ambiguity function is duplicated in time domain with a repetition equal to the period Te and the pedestal is a comb of frequency lines spaced by 1/Te. The spreading effect and the properties with respect to pulse waveform are still valid since the antenna pattern sees only one recurrence in time domain (Te $>>$ range aperture/C) and plenty of lines in frequency domain.

Note also that the ambiguity criteria considered here corresponds to the case where the imaging interest is on scattered targets. A point target context reinforces the spread spectrum merit since all the ambiguities are made scattered.

References

1. Aguttes JP (2001) Radically new design of SAR satellite: short vertical antenna approach. IEEE Trans Aerosp Electron Syst 37(1):50–64
2. Aguttes JP (2002) New designs or modes for flexible space borne SAR. In: Proceedings of the international geoscience and remote sensing symposium (IGARSS 02), Toronto, Canada, June 2002
3. Aguttes JP (2002) The SAR train: along track oriented formation of SAR satellites. In: International symposium formation flying missions and technologies, Toulouse, France, October 2002
4. Aguttes JP (2003) SAR train concept: antenna distributed over N satellites smaller, performance increased by N. In: Proceedings of the international geoscience and remote sensing symposium (IGARSS 03), Toulouse, France, 2003
5. Zebker H et al (1994) Mapping the world's topography using radar interferometry: the TOPSAT mission. Proc IEEE 82(12):1774–1786
6. Aguttes JP et al (2006) Romulus: along track formation of radar satellites for MTI (Moving Target Identification) and high SAR performance. In: Proceedings of the international geoscience and remote sensing symposium (IGARSS 06), Denver, CO, 2006

Chapter 18
P-Band Distributed SAR

Giancarmine Fasano, Marco D'Errico, Giovanni Alberti, Stefano Cesare, and Gianfranco Sechi

Abstract This chapter discusses a spaceborne P-band synthetic aperture radar concept based on a distributed architecture and formation flying technologies. This approach can in principle allow overcoming physical constraints that limit the performance of monolithic SARs, leading in the P-band case to huge antennas and hard swath/resolution trade-offs. The proposed SAR is based on a larger transmitting satellite and a set of lightweight receiving-only platforms. This architecture also enables multi-mission capabilities. In particular, forests observation and biomass estimation based on side-looking SAR data can be in theory combined with near nadir interferometric ice sounding. Payload concept is clarified, and a preliminary performance analysis in terms of ambiguity and coverage is proposed. Then, mission analysis, preliminary spacecraft design, and formation control architecture are briefly described.

G. Fasano, Ph.D. (✉)
Department of Aerospace Engineering, University of Naples "Federico II",
P.le Tecchio 80, 80125 Napoli, Italy
e-mail: g.fasano@unina.it

M. D'Errico, Ph.D.
Department of Aerospace and Mechanical Engineering, Second University of Naples,
Via Roma 29, 81031 Aversa (CE), Italy
e-mail: marco.derrico@unina2.it

G. Alberti, Ph.D.
Consortium for the Research on Advanced Remote Sensing Systems (CORISTA),
V.le Kennedy 5, 80125 Napoli, Italy
e-mail: giovanni.alberti@corista.eu

S. Cesare • G. Sechi
Thales Alenia Space Italia, Strada Antica di Collegno 253, 10146 Torino, Italy
e-mail: stefano.cesare@thalesaleniaspace.com; gianfranco.sechi@thalesaleniaspace.com

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_18,

18.1 Introduction

Low microwave frequencies in general and P-band in particular are widely considered of high interest for a range of scientific applications: biosphere and bioclimatology studies, glaciology, geophysics. Nonetheless, technical challenges have prevented the implementation in space of low frequency imaging radars because they require very large antennas (order of 100 m^2 in P-band) to achieve adequate power and reduce ambiguities. Moreover, a traditional sensor implementation has strong limitations for scientific applications. As an example, full polarization features, necessary to correct Faraday rotation, lead to double the radar Pulse Repetition Frequency (PRF) for a given level of azimuth ambiguities. Therefore, in order to keep an acceptable level of range ambiguities, the swath has to be reduced, which in turn impacts global coverage and revisit time capabilities with consequent limitation of system effectiveness.

"Biomass" is one of the three ESA Earth Explorer mission candidates and it is currently undergoing phase A [1]. It is based on a P-band side-looking SAR, and different concepts are being explored to fulfill scientific requirements [2, 3], such as very large planar or reflector antennas.

Compared with a monolithic sensor implementation as the one foreseen in Biomass, a distributed SAR can in principle overcome some constraints. If the signal transmitted by a radar antenna is received by a multiple antenna system, the global system PRF depends on actual PRF and on the number of satellites, provided that accurate positioning and synchronization among all the satellites is guaranteed.

In this framework, potential scientific applications of a P-band distributed SAR are briefly summarized and the distributed payload concept is presented. After a performance analysis, mission design, formation flying dynamics, and platform-related aspects are then briefly discussed. Finally, a short critical analysis is presented to focus the main issues for spaceborne low frequency SARs based on formation flying. Part of these results has been derived under contract with the Italian Space Agency in 2007.

18.2 Scientific Applications

P-band is of interest to Earth science thanks to the fact that its wavelength (longer than current spaceborne X, C, and L band SARs) is able to penetrate forest canopy and ice, depending on a number of parameters (salinity, humidity, etc.). Two main application fields can benefit from such capabilities:

- Forest areas classification and biomass estimation;
- Ice sheets sounding and subsurface analysis.

It is also worth noting that low frequency SAR capability to map the subsurface down to several meters in arid areas enables a great potential for terrestrial applications, such as hydrology, geology, water and oil resources, and archaeology.

18.2.1 Forest Areas Classification and Biomass Estimation

Low frequency radars can help to retrieve forest biomass, also monitoring disturbances and flooded forests, because canopy penetration is significant and the radar scattering is mainly caused by large scattering elements, such as trunks and large branches where most of the above-ground biomass is stored. As a consequence, P-band backscattering is connected to the so called "woody" biomass. From this point of view, a P-band spaceborne radar can be considered as a unique instrument to provide global coverage of both boreal and tropical forests, with the possibility of biomass estimation up to limits which, due to signal saturation [4–10], are unachievable using higher frequencies, which on the other hand can be used to estimate leaf biomass. Biomass information is of great importance with regard to the requirements of the terrestrial carbon cycle scientific community, filling a crucial gap in the data requirements for coupled models of the Earth System. Inversion techniques to retrieve biomass information from P-band backscatter are well documented in literature and have been applied both in boreal and in tropical forests [4, 8, 9] and several airborne experiments have been carried out over selected test sites.

The first airborne SAR system operating below 1 GHz was AIRSAR, which has been run by NASA's Jet Propulsion Laboratory (JPL) since 1987, with a P-band (440 MHz) mode [11]. Over the past few years other systems have been developed operating in the same range of frequencies, such as RAMSES and E-SAR [11]. The several airborne campaigns that have taken place in a wide variety of forest biomes have shown that the cross-polarized Horizontal-Vertical (HV) backscatter has the largest dynamic range and the highest correlation with biomass [1]. Recent results from the BioSAR flight campaign have also shown that HH backscatter has a similar behavior in hemiboreal forests, while L-band data are of less utility both in terms of correlation coefficients and dynamic range [9]. Other recent techniques to estimate biomass in the saturation region exploit polarimetric SAR interferometry and tomography [9].

Besides biomass estimation, the strong contrast in P-band between forests and unforested areas also allows realization of accurate forest and deforestation maps, contributing to estimate the rate of deforestation and re-growth in tropical areas. Moreover, P-band measurements can give improved information on forest inundation cycles thanks to better penetration into canopy.

18.2.2 Ice Sheets Sounding and Subsurface Analysis

Since P-band radiation is also capable of penetrating fresh ice up to depths of a few kilometers, a spaceborne P-band sensor can provide a global 3D mapping of the whole Antarctica, with subsurface information on ice thickness, glacial topography and internal layering. It is worth noting that knowledge of these parameters is at the moment limited to a few areas because such measurements have been obtained by means of airborne or ground-based low frequency radars. On the other hand,

improvement of Antarctica subsurface knowledge would be very important in the framework of climate and sea level studies, due to the key role played by the dynamics of large ice sheets.

First studies about spaceborne implementation of low frequency ice sounding focused on a nadir pointing synthetic aperture radar, which allows suppression of the received surface clutter and isolation of the internal layers echoes [12, 13]. In more detail, the surface clutter can be reduced by the antenna pattern in the across-track direction and by Doppler filtering in the along-track direction.

Recent studies carried out within the GISMO project [14, 15] focused on the possibility to use a near nadir cross-track interferometer to remove surface clutter preserving basal echoes, so as to generate a 3d model of the bedrock and a global mapping of ice thickness. The latter measurement technique has the potential to be applied in a distributed sensor framework.

Both the nadir sounding and the cross-track interferometric observation modes have recently been demonstrated in airborne experiments in Greenland [16, 17] where it has resulted that the number of receiving channels is an important variable for surface clutter suppression.

18.3 Concept Development

18.3.1 Distributed Payload Concept

In terms of observation technique, forests monitoring and biomass estimation require a conventional side-looking SAR geometry. In this case, the distributed payload is based on a (quasi) linear formation comprised of a number of satellites moving on the same trajectory with respect to an Earth-fixed reference frame. The basic concept of distributed SAR is to enhance azimuth sampling capability without impacting range ambiguities [18–21]. This is why performance limits of monolithic SAR systems can be in principle overcome.

The multi-platform payload is comprised of a number of cooperating antennas receiving the radar signal transmitted by one or more of them and reflected by the Earth surface. As a consequence, it is possible to define a "global" system PRF that depends on the actual transmitted PRF and on the number of receiving satellites (N_{sat}). Figure 18.1 explains the difference between the monolithic and the distributed SAR implementation. In the traditional SAR concept (Fig. 18.1a), the synthetic antenna is built on the basis of the echoes gathered by the radar antenna along its trajectory. If V is the satellite velocity, the spacing between the acquired signal samples is then V/PRF. In the distributed concept (Fig. 18.1b, c), different satellites contribute to the signal sampling along the trajectory, so that the same azimuth sampling capability can be guaranteed by using a lower transmission PRF on board single satellites. In particular, under ideal conditions, the global PRF is directly proportional to N_{sat}.

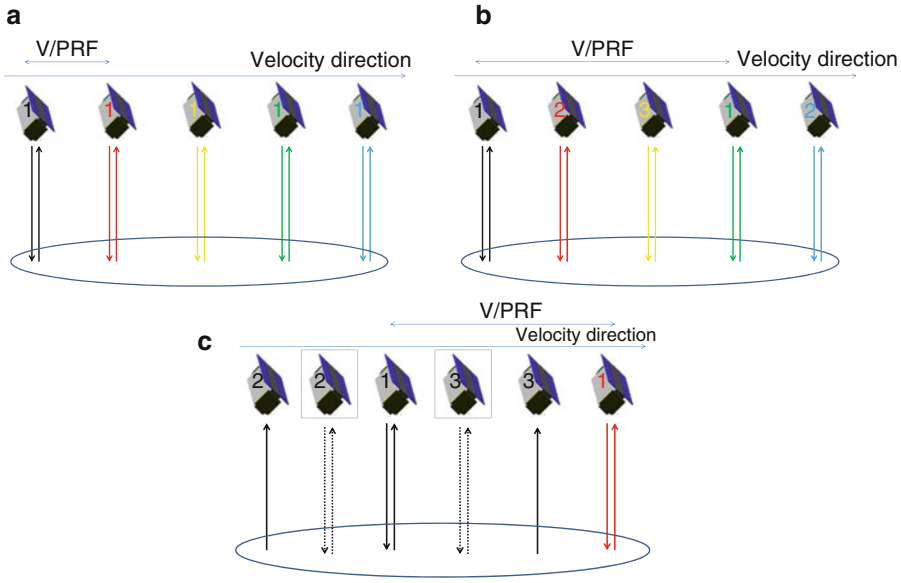


Fig. 18.1 Classical monostatic SAR concept (**a**); distributed SAR with multi-monostatic geometry (**b**); distributed SAR with multistatic geometry (**c**). Different colours are used for the different signal transmission instants. *Dotted squares* and *arrows* identify equivalent monostatic SAR satellites and signals

The main advantage of the distributed concept lies in the fact that azimuth ambiguities depend on global PRF, while range ambiguities depend on actual transmitted PRF [22]. This paves the way to increasing sensor swath and/or improving resolution, while keeping acceptable values for the ambiguity to signal ratios. Of course, the system relies on accurate positioning and synchronization among all the satellites which fly in formation. In fact, along-track positioning is related to uniformity of azimuth sampling. Moreover, each antenna is relatively small and by itself relatively useless, whereas the combined processing of all the received signals leads to high observation performance. The distributed SAR concept can be theoretically achieved by two different configurations, which can be defined as multi-monostatic and multistatic.

In the multi-monostatic system, each antenna is transmitting/receiving its own signal in proper positions along the orbit (see Fig. 18.1b). The main advantage of the multi-monostatic configuration is that tight real time control of the formation is not necessary. In fact, in this case azimuth sampling accuracy is connected to the choice of transmission instants, so accurate real time knowledge (not control) of relative position is required. However, since each antenna has to receive only its transmitted signal 3 dB beams in azimuth must not intersect (neglecting side-lobes). Thus, given the large 3 dB angles corresponding to relatively small antenna dimensions at P-band, required along-track distances should be of the order of 200 km. Of course, this poses major problems in determination and control

of relative position and attitude (besides introducing some temporal decorrelation effects that depend on wavelength). A solution consists in using different orthogonal codes for the transmitters. Indeed, if the single satellites are able to receive not only their own signal, but also the signals transmitted by the other satellites, this leads to a significant increase of the number of sampling points.

Another possible solution is that the different apertures receive the same transmitted signal (multistatic configuration, see Fig. 18.1c where satellites depicted in dotted squares correspond to the location of the effective bistatic phase centers). The main drawback of this architecture is that location of bistatic phase centers depends on satellites along track distance, so that accuracy in real time relative position control is strictly connected to uniformity in azimuth sampling. Within this framework, there exists a number of techniques for dealing with non uniform azimuth sampling [21, 23].

In a multistatic architecture, along track baseline between receiving satellites is flexible and can be chosen according to the following equation [23]:

$$\Delta x_i = \frac{2\,V}{PRF}\left(\frac{i-1}{N_{sat}} + k_i\right), \quad i \in \{2, 3, ..., N_{sat}\}, k_i \in Z \tag{18.1}$$

The same equation holds in the multi-monostatic case without the factor two and considering signal transmission instants.

In the performed study, it was supposed to have a multistatic architecture with a (larger) "mother" spacecraft flying a transmitting-only antenna. The linear formation is then comprised of "children" microsatellites, with low weight receiving-only antennas. This leads to the possibility of optimizing transmitter and receivers separately, avoiding transmit interference problems, and thus enabling larger PRF flexibility. Moreover, a larger antenna allows transmission into a smaller 3 dB lobe of a higher RF peak power, which is important to improve Noise Equivalent Sigma Zero. Finally, as it will be discussed later, from the application point of view this scenario is more suitable for a multi-objective mission.

18.3.2 Preliminary Performance Analysis

Given the principle of the system, different choices are possible in terms of number of receivers, antenna dimensions both for the transmitter and the receivers, processed Doppler bandwidth. Basically, by increasing the number of satellites the following results are obtained: for given ambiguity requirements, actual PRF decreases, ground swath increases, data rate for the single satellite can increase or decrease depending on the actual PRF but also on the desired swath. On the other hand, a larger number of satellites can be used to improve range and ambiguity performance, while keeping adequate data rates and swaths.

Table 18.1 Example: P-band SAR performance for 4.8 m × 1.6 m Tx antenna, 1.6 m × 1.6 m Rx antennas and variable number of satellites

Sat number (receivers)	4	6	8
Patches (range × azimuth)	3 × 3	3 × 3	3 × 3
Physical dimensions Rx antenna (m, range × azimuth)	1.6 × 1.6	1.6 × 1.6	1.6 × 1.6
Example of "optimal" transmission PRF (for each polarization) (Hz)	≈700	≈566	≈425
RAR (dB)	−17, −10	−25, −14	−32, −26
AAR (dB)	−10	−15	−15
Data rate for the single sat (Mbit/s)	150	170	135
Swath achievable without impacting RAR (km)	130	180	200
Off-nadir angle (average) (°)	25	25	25

On the basis of P-band antenna realization technologies (see for example [3]), it comes out that a reasonable choice consists in using, as the basic transmitting/receiving units, patches of a diameter of 44 cm separated by a distance of the order of 0.8λ (wavelength is about 69 cm), which have been assumed as baseline.

Using small antennas with a reduced number of patches is a favorable choice from the satellite engineering point of view, but, due to the large 3 dB angles, in this case a large number of satellites is needed to gain satisfying performance. Combining small receivers with a larger transmitter antenna guarantees a good compromise between system complexity and performance.

For example, let us assume a transmitter antenna of 9 × 3 patches (azimuth × range) and square receiving antennas of 3 × 3 patches. In the considered case, transmitting antenna physical dimensions are of the order of 4.8 m × 1.6 m, while receiving antennas are of the order of 1.6 m × 1.6 m. For the sake of concreteness, achievable performance for this configuration (ideal conditions) is summarized in Table 18.1 as a function of the number of receiving satellites (4, 6, 8). It is important to underline that full quad polarization was considered in all the calculations. Satisfying performance is achieved also with 4 or 6 satellites due to the azimuth radiation pattern of the large transmitting antenna. For example, acceptable resolution and ambiguities are achieved in a 130 km swath using four receiving antennas. Basically, Table 18.1 shows that by increasing the number of receiving satellites it is possible to lower the transmission PRF with positive impact on swath and range ambiguities, while keeping a satisfying azimuth sampling capability and thus an acceptable azimuth ambiguity ratio (AAR).

Figures 18.2 and 18.3 clarify the functional dependences lying behind the numerical budgets. In fact, Fig. 18.2 (valid for any N_{sat}) shows that AAR only depends on the global system PRF and that at least 2,700 Hz are needed to obtain −10 dB. Such a high sampling frequency can be achieved without impacting range ambiguities if actual PRF is much smaller, as shown in Fig. 18.3 (six receivers) which clarifies how range ambiguities affect the achievable sensor swath and also points out the effect of nadir echoes (upper right region of the diagram). It is worth noting that these calculations were carried out considering as a requirement an azimuth resolution

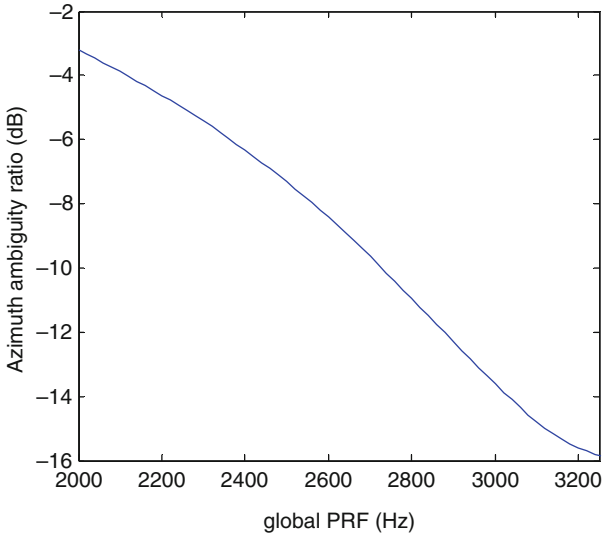


Fig. 18.2 Azimuth ambiguity ratio as a function of global PRF (Reproduced by permission of IEEE © 2008 IEEE. Published in Ref. [25])

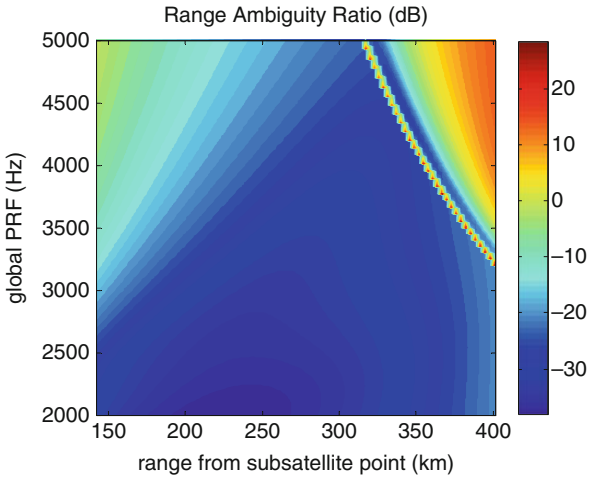


Fig. 18.3 Range ambiguity ratio as a function of global PRF and range from sub-satellite point (six receivers)

of about 35 m with eight looks (e.g. ESA biomass mission foresee a spatial resolution of 50 m $\times$ 50 m with at least four looks [24]). As a consequence, processed Doppler bandwidth is a fraction of the maximum bandwidth achievable by the small antennas. If maximum Doppler bandwidth were processed, azimuth ambiguities would be unacceptable. Of course, for the case of 8, or 12 satellites, because of the

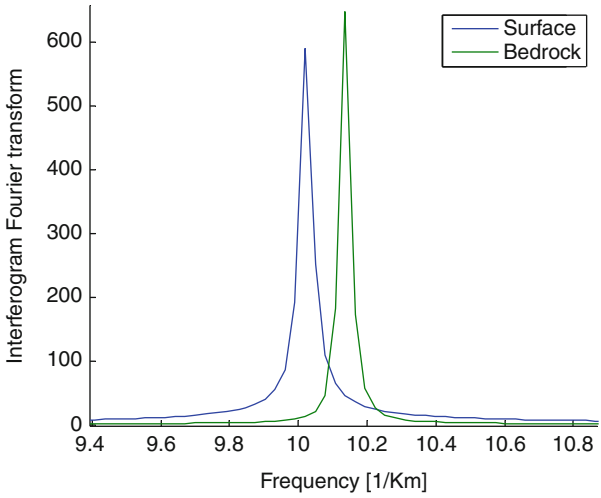


Fig. 18.4 Interferogram Fourier transform (by = 50 m, Bz = 3,500 m, swath of 30 km, bedrock depth of 3,700 m) (Reproduced by permission of IEEE © 2008 IEEE. Published in Ref. [25])

increase in azimuth sampling and thus in global system PRF, processed Doppler bandwidth can be increased to improve resolution and/or number of looks without significant consequences on ambiguity performance.

A distributed implementation of P-band SAR has also the potential to combine biomass estimation with ice sheets sounding, applying an interferometric technique [14–17]. In fact, over ice it is sufficient to add a classical monostatic SAR image to make an interferometric pair with the other SAR image made by the multistatic SAR linear formation. The application requires a quasi-zero off-nadir angle: in these conditions bedrock echo can be separated from ice surface return on the basis of the separation of spatial frequencies in the interferogram. For example, Fig. 18.4 ([25]) reports interferogram Fourier transform for ice surface and bedrock with 50 m of horizontal baseline and 3,500 m of vertical baseline. The separation of the two frequencies allows for surface clutter cancellation. As it will be better explained in the following, the possibility to combine the biomass and the ice sounding applications relies on the fact that the mother satellite can be maneuvered towards a cartwheel-like formation (considering relative motion with respect to the children satellites) [26] in order to gather interferometric data.

18.4 Mission Analysis, Spacecraft Design, and Formation Control

18.4.1 Mission Analysis

Previous analyses showed that assuming a transmitter antenna of 9 × 3 patches (azimuth × range), acceptable resolution and ambiguities are achieved in a 130 km swath using four relatively small receiving antennas. On the basis of these data, a

formation of four receivers ("children" satellites) with a single transmitter ("mother" spacecraft) was selected as the baseline configuration for a technology demonstration mission which is described in the following. The designed formation flying-based P-band SAR mission is comprised of two phases: the first one is devoted to forests observation and biomass estimation, the second one is dedicated to demonstration of interferometric ice sounding. The two applications require different formation architectures, thus a reconfiguration maneuver is foreseen to pass from phase one to phase two. The different architectures are detailed in the following.

The reference orbit design is performed on the basis of observation requirements for the biomass application and on payload preliminary performance. Assuming a swath of the order of 130 km, the requirement of global coverage, and the requested altitude (of the order of 550 km), a suitable choice is a 6 a.m.–6 p.m. sun-synchronous circular orbit at an average altitude of 556 km. The nominal repetition factor is 331/22, thus a maximum revisit time of 22 days is obtained at the equator, which is considered acceptable for the application of repeat-pass interferometric techniques [24]. The latitude belt covered by the ground track is $\pm 82.383^{\circ}$. It is worth noting that the mission was designed with the primary objective of forests observation, thus this orbit can be used to acquire ice sounding data but does not allow interferometric coverage of the whole Antarctica.

As mentioned earlier, the biomass estimation application requires a linear formation made up of a number of receiving satellites moving on the same trajectory with respect to an Earth-fixed reference frame. Thus, all the satellites have the same ground track. This can be obtained by separating the satellites in true anomaly and in right ascension of the ascending node, as shown in Ref. [27]. Along track baseline between satellites is of about 200 m and it was chosen on the basis of a compromise between accuracy of relative navigation and mitigation of collision risk. The distance between the children and the mother satellite is of the order of 1 km. Given these values, in substance all the satellites move on the same orbit. The formation architecture designed primarily for biomass estimation and presented previously can be adapted to demonstrate the ice sounding application if a cartwheel-like motion is generated between the mother and the children satellites. An artistic view of the two phases of the designed mission is shown in Fig. 18.5.

Regarding the first mission phase, a coverage analysis was carried out on the basis of selected formation orbit (sun-synchronous dusk-dawn at an altitude of about 556 km) to evaluate the radar mean orbit duty cycle requested by the application. Starting from a global biomass map based on TURC model [28], a binary longitude-latitude map of regions of interest was derived (Fig. 18.6) and it was used with an orbit propagator to estimate orbit fractions useful for the biomass application. The result is outlined in Fig. 18.7, with a mean orbit duty cycle of 10.6 %. Periodicity is due to the considered orbit repetition factor (331/22).

Considering now the ice sounding application, it comes out that given horizontal (50 m) and vertical baselines (3,500 m) are needed at orbit poles [25]. They can be obtained by introducing in the mother satellite orbit a slight eccentricity (order 5×10^{-4}) with the argument of perigee at 90°, and a very small difference in

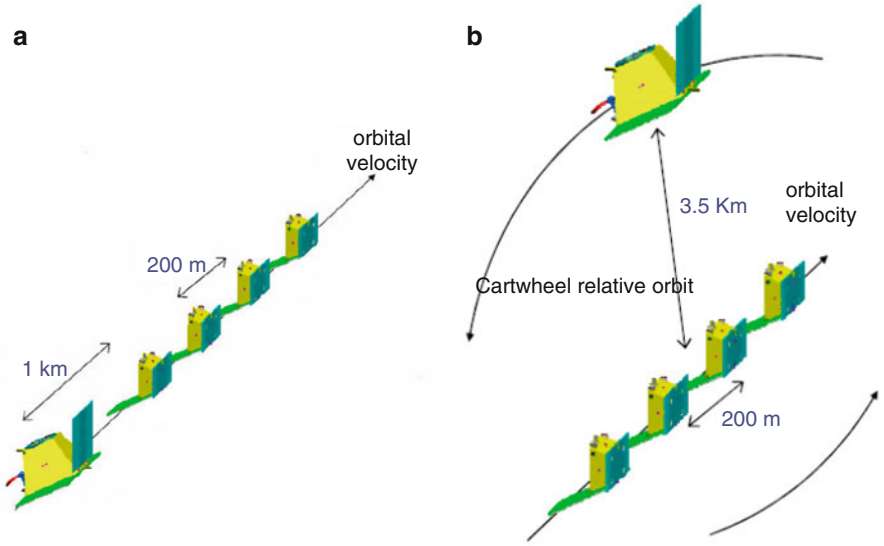


Fig. 18.5 Artistic views (not to scale) of the linear formation for distributed SAR (**a**) and the cartwheel formation for interferometric ice sounding (**b**)

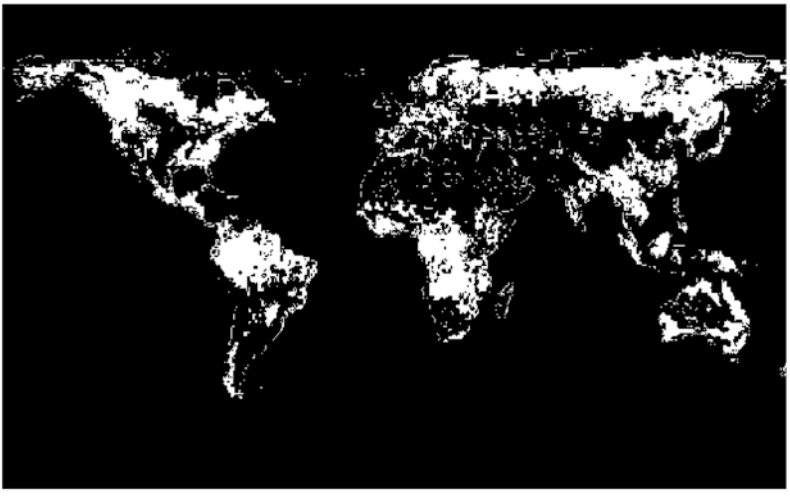


Fig. 18.6 Binary longitude-latitude map with regions of interest in white (Reproduced by permission of IEEE © 2008 IEEE. Published in Ref. [25])

inclination (order $4 \times 10^{-4^\circ}$). These corrections require a very limited propellant expense. Relative motion consists in an out of plane oscillation of about 50 m, whereas the in-plane motion is elliptical with semi-axes 3.5 and 7 km (Fig. 18.8). It is worth noting that the cartwheel relative motion is in theory compatible with the

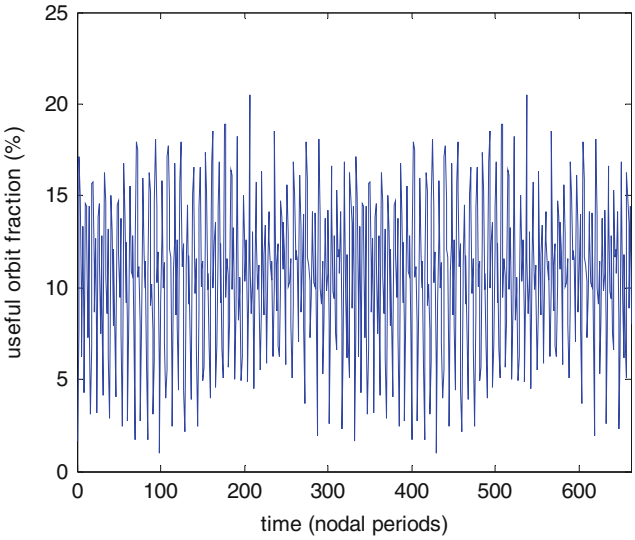


Fig. 18.7 Useful orbit fraction for biomass estimation during two repetition periods (Reproduced by permission of IEEE © 2008 IEEE. Published in Ref. [25])

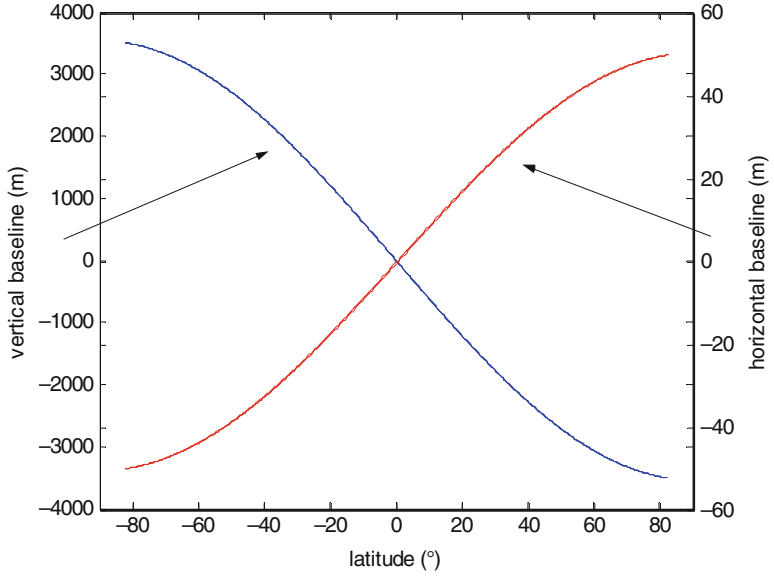


Fig. 18.8 Horizontal and vertical baseline as a function of latitude (Reproduced by permission of IEEE © 2008 IEEE. Published in Ref. [25])

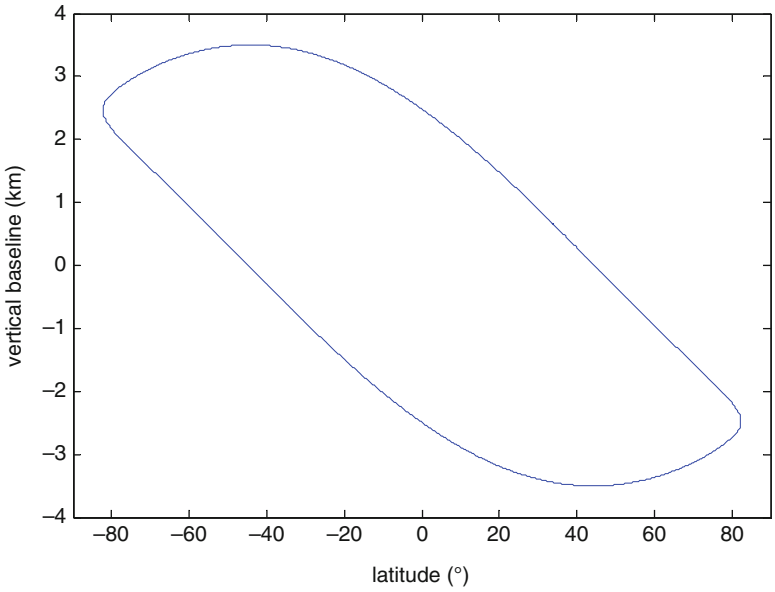


Fig. 18.9 Vertical separation as a function of latitude when $\omega = 45°$

biomass application. In fact, the formation is quite stable from the point of view of differential Earth gravity effects (secular J_2 effects consist in substance in a growing cross-track oscillation caused by the slight difference in inclination, which does not change observation geometry at orbit poles). Instead, perigees' precession has to be counteracted in order to keep required geometry at orbit poles, even if it does not represent a differential effect. For example, supposing that minimum acceptable vertical separation at orbit poles is of about 2,500 m, one gets that at most perigee can change in the range [45°, 135°] as shown in Fig. 18.9 which depicts the vertical baseline as a function of latitude when $\omega = 45°$. Cartwheel was not considered for the entire mission in order to avoid over sizing the inter-satellite link.

18.4.2 Preliminary Spacecraft Design

Preliminary spacecraft design was performed both for the mother and for the children satellites. Mother and child on-orbit configurations are depicted in Fig. 18.10. The mother satellite carries the Tx/Rx radar in P-band with 9×3 printed rectangular patches. The antenna is split in three panels of 3×3 patches each: the central panel is body-mounted; the lateral panels are hinged to the central panel and are folded at launch.

The satellite is equipped with fixed solar arrays which lie on the orbit plane while the radar is looking 25° off nadir (as needed for biomass monitoring).

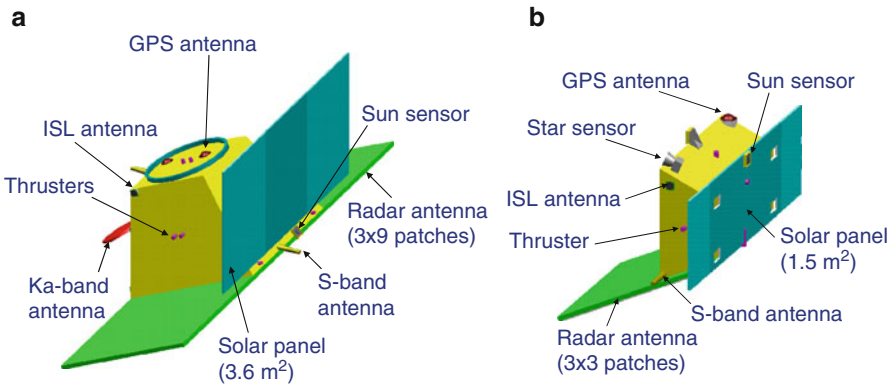


Fig. 18.10 Images of mother (**a**) and child (**b**) spacecraft

The Ka-Band transmitting system enables the high-rate payload data transmission to the ground station: it is comprised of two Ka-Band transmitters, two travelling wave tube amplifiers, and a mechanically steerable dish high-gain antenna. The S-Band system ensures the link for standard command-and-control functions (TM/TC) transmission in any attitude. It is comprised of two S-Band transponders (for redundancy since the Mother spacecraft shall be one-failure tolerant) with their internal RF amplifiers, two low-gain antennas which guarantee omni-directional coverage, and the interconnecting diplexers and switch.

Two wide-coverage Inter-Satellite Link (ISL) antennas allow the Mother satellite to receive the P/L data from the Children satellites and to exchange navigation data (from the on Global Navigation Satellite System (GNSS) receiver) for the relative position determination. The two systems can be operated on the same RF band (i.e. 2.4 GHz) using the same Wireless Fidelity (WiFi) protocol (802.11 g).

The data are stored on flash based mass memories. The on-board data handling is a fully integrated/modular network with system on chip processing modules, layered software architecture and networking protocol. The attitude information is provided by two star sensors, two sun sensors and two 3-axis magnetometers. Twelve cold-gas micro-thrusters are used for the formation control and the orbit maintenance. The configuration of the Mother satellite, for its unique role in the formation, is one-failure tolerant. Its total mass is about 450 kg.

Each Child satellite carries a radar in P-band that only receives the echo of the signal transmitted by the Mother satellite. The radar antenna ($1.5\ m^2 \times 1.5\ m^2$) has 3×3 printed rectangular patches and is split in two panels of 1×3 and 2×3 patches, one of which is folded at launch. The satellite is equipped with fixed solar array.

The collected P/L data are transferred to the Mother satellite through two wide-coverage ISL antennas also used for the exchange of navigation data. Children communicate to Earth with a simplified S-Band system consisting of a single transponder connected to two antennas. This system ensures command-and-control

(TM/TC) link in any attitude. Scientific data are transmitted to the Mother satellite through the satellite-to-satellite link using the WiFi protocol.

The attitude information is provided by one star sensor, two sun sensors and two 3-axis magnetometers. Six cold-gas micro-thrusters are used for the formation control and the orbit maintenance.

For maximum mass saving, the Child satellite is realized with minimum redundancies (no single failure tolerant). In fact the mission can be accomplished (with acceptable degradations) even after the loss of one child satellite. The child satellite total mass is of about 180 kg.

Mother and Children avionics architecture is a Fully Integrated Modular System, based on a Standardized Communication Network. Unlike the traditional partitioning of functions into separate boxes, common resources like Processing Modules (PMs) and Power Supply are shared among functions. The result is a limited number of PMs, on which different tasks shall be allocated. To realize the avionics hardware, extensive use of miniaturization technologies like ASICs and System on Chip, is foreseen in order to achieve significant mass and power consumption reduction.

18.4.3 Formation Control

A preliminary study was also carried out about GNC architecture and formation control. For both the applications, the assumed requirements for the spacecraft relative position control were 2 m for along-track positioning and 10 m for radial and horizontal coordinate. The assumed requirement for (off-line) knowledge of relative position was 0.05 m for all the three coordinates. The above requirements derive from the fact that the error in along-track positioning is connected to the phase centers location and thus to the uniformity of azimuth sampling. As for relative attitude error, 0.1° was assumed as the requirement. These requirements can be satisfied considering the following control architecture:

- Relative position control between spacecrafts according to the leader-follower architecture.
- The relative attitude requirement is satisfied maintaining each spacecraft absolute attitude with an accuracy still compatible with the relative attitude requirement (e.g. absolute attitude accuracy shall be a fraction of the required relative attitude accuracy).

The designed relative position control is based on:

- Global Navigation Satellite System (GNSS) receiver
- Inter Satellite Link (ISL)
- Micro-propulsion sub-system (MPS) based on cold gas micro thrusters

The GNSS receiver provides at 1 Hz the absolute position and velocity (used to define the attitude and angular rate trajectory for the attitude controller), and

spacecraft relative position and velocity. Each spacecraft exchanges GNSS measurements (pseudo-range, carrier phase) with the other satellites in the formation. This allows the implementation of a de-centralized GNSS processing scheme, in which each spacecraft can process autonomously GNSS data for real-time relative and absolute position.

The real-time data-processing of GNSS data has to be compatible with relative position control requirements in term of accuracy and fuel-consumption (relative position accuracy better than 0.2 m). MPS provides the necessary forces in all directions and versus.

Preliminary studies were conducted on formation control and operative mode. Several FF modes were designed for the different phases of the mission (Free Flying Mode, Safe Flying Mode, Formation Acquisition Mode, Formation Keeping Mode).

Free Flying Mode (FFM) is the FF mode after launcher separation. The spacecraft are sufficiently separated (more than km), no collision risk is present. Each spacecraft is under ground control. ISL and relative position determination are activated by ground request. FF control is not active.

In Safe Flying Mode (SFM) all spacecraft are well three axis stabilized, ISL and coarse relative position determination are achieved. In this mode, eventual Collision Avoidance Maneuvers (CAM) may be executed. Altitude control activation is possible. A FF decentralized control scheme may be considered.

In Formation Acquisition Mode (FAM) all spacecrafts maneuver in coordinate way to achieve the required formation geometry.

Finally, the Formation Keeping Mode (FKM) is the normal observation mode. In this condition both the relative position control and the formation altitude control loop are active.

The relative position control was designed according to the state-variable approach, starting from the Clohessy-Wiltshire equations [29], and off line simulations demonstrated that requirements could be satisfied with selected propulsion systems and control laws [30]. In particular, the radial and along-track position errors were found to be of the order of 0.1 m, while the cross-track position error was found to be negligible in the linear formation.

18.5 Outlook and Conclusion

Though the study described in this chapter remained at a pure theoretical level, some lessons learned can be discussed whose interest goes beyond the considered scientific applications and frequency bands.

The performed study showed the potential of the distributed concept but also the difficulty in combining near term implementation and use of assessed technologies, miniaturization and use of small satellites, and significant improvement of observation performance compared with a traditional (monolithic) implementation of the concept.

In fact, low frequencies on the one hand offer the advantage of very coarse requirements on relative position knowledge and control. This allows standard control algorithms and technologies to be effectively used for formation maintenance.

On the other hand, P-band has the drawback, from the formation point of view, that large antennas are required to keep a relatively small lobe width: using existing antenna technologies implies relatively large mass budgets, so that it is hard to consider a micro-satellite implementation. In the discussed study, the necessity to design a near term demonstration mission able to provide data of some scientific interest led to introduce the mother spacecraft concept, which is of course a single point of failure for the formation and thus an undesirable choice from the system point of view.

In other words, full exploitation of the distributed concept probably requires the adoption of innovative technologies such as inflatable antennas.

Another important concept to be underlined is that at low frequencies a relatively high number of receiving antennas is required to achieve satisfying performance with small satellites: launch strategies (single or multiple) represent critical issues for system implementation.

Another limiting factor for system realization is that performance analyses for multi-static configurations in non ideal conditions are not well assessed yet, which hinders mission design because it is hard to clearly identify the trade-offs between trajectory and GNC system design, and application requirements. Furthermore, the necessity to keep an acceptable signal-to-noise ratio (SNR) hinders the reduction of antenna dimensions.

L-band frequency could represent a valuable compromise for a near term implementation of the distributed SAR concept. In fact, if L-band is considered, smaller receive apertures can be used and full polarization is not required, with advantages from the point of view of range ambiguities, while relative position knowledge and control requirements are still coarse. Of course, the potential application fields are in part different from P-band (for example, L-band radiation does not penetrate ice significantly).

References

1. Le Toan T, Quegan S, Davidson MWJ, Balzter H, Paillou P, Papathanassiou K, Plummer S, Rocca F, Saatchi S, Shugart H, Ulander L (2011) The BIOMASS mission: mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sens Environ 115(11):2850–2860, ISSN 0034-4257, 10.1016/j.rse.2011.03.020
2. Hélière F, Lin CC, Fois F, Davidson M, Thompson A, Bensi P (2009) BIOMASS: a P-band SAR earth explorer core mission candidate. In: Proceedings of the IEEE radar conference, Pasadena, CA, USA, pp 1–6, May 2009
3. Ramongassie S, Castiglioni SK, Lorenzo J, Labiole E, Baudasse Y, Svara C, Luigi C, Heliere F, Mangenot C, Klooster KV, Fonseca N, Diez H, Belot D (2010) Spaceborne P-band SAR for BIOMASS mission. In: Proceedings of the IEEE international geoscience and remote sensing symposium (IGARSS), pp 2880–2883, 25–30 Jul 2010, doi: 10.1109/IGARSS.2010.5653156

4. Dobson C et al (1992) Dependence of radar backscatter on coniferous forest biomass. IEEE Trans Geosci Remote Sens 30(2):412–415
5. LeToan T et al (1992) Relating forest biomass to SAR data. IEEE Trans Geosci Remote Sens 30(2):403–411
6. Beaudoin A, Le Toan T, Goze S, Nezry E, Lopes A, Mougin E, Hsu CC, Han HC, Kong JA, Shin RT (1994) Retrieval of forest biomass from SAR data. Int J Remote Sens 15:2777–2796
7. Rignot E, Zimmermann R, van Zyl JJ (1995) Spaceborne applications of P band imaging radars for measuring forest biomass. IEEE Trans Geosci Remote Sens 33(5):1162–1169
8. Santos JR, Freitas CC, Araujo LS, Dutra LV, Mura JC, Gama FF, Soler LS, Sant'Anna SJS (2003) Airborne P-band SAR applied to the aboveground biomass studies in the Brazilian tropical rainforest. Remote Sens Environ 87(4):482–493, ISSN 0034–4257, 10.1016/j.rse.2002.12.001
9. Sandberg G, Ulander LMH, Fransson JES, Holmgren J, Le Toan T (2011) L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest. Remote Sens Environ 115(11):2874–2886, ISSN 0034–4257, 10.1016/j.rse.2010.03.018
10. Saatchi S, Halligan K, Despain DG, Crabtree RL (2007) Estimation of forest fuel load from radar remote sensing. IEEE Trans Geosci Remote Sens 45(6):1726–1740
11. Kramer HJ (2001) Observation of the earth and its environment: survey of missions and sensors, 4th edn. Springer-Verlag, Berlin
12. Herique A, Kofman W, Bauer P, Remy F, Phalippou L (1999) A spaceborne ground penetrating radar: MIMOSA. In: Proceedings of the geoscience and remote sensing symposium (IGARSS '99), Hamburg, Germany, vol 1, 28 June–2 July 1999, pp 473–475
13. Bruniquel J, Houpeit A, Richard J, Phalippou L, Dechambre M, Guijarro J (2004) Spaceborne P-band radar for ice-sheet sounding: design and performances. In: Proceedings of the IEEE international geoscience and remote sensing symposium (IGARSS '04), Anchorage, Alaska, USA, vol 5, 20–24 Sep 2004, pp 2834–2837
14. Rodriguez E, Freeman A, Jezek K, Wu X (2006) A new technique for interferometric sounding of ice sheets. In: Proceedings of the European conference on synthetic aperture radar (EUSAR), Dresden, Germany
15. Jezek K, Rodríguez E, Gogineni P, Freeman A, Curlander J, Wu X, Paden J, Allen C (2006) Glaciers and ice sheets mapping orbiter concept. J Geophys Res 111(E6):E06S20. doi:10.1029/2005JE002572
16. Jezek K, Gogineni P, Wu X, Rodriguez E, Rodriguez F, Sonntag J, Freeman A, Hoch A, Forster R (2008) Global ice sheet mapping orbiter concept: airborne experiments. In: Proceedings of the 7th European conference on synthetic aperture radar Friedrichshafen, Germany, 2008, vol 2, pp 99–102
17. Jezek KC, Gogineni S, Wu X, Rodriguez E, Rodriguez-Morales F, Hoch A, Freeman A, Sonntag JG (2011) Two-frequency radar experiments for sounding glacier ice and mapping the topography of the glacier bed. IEEE Trans Geosci Remote Sens 49(3):920–929
18. Goodman NA, Sih Chung L, Rajakrishna D, Stiles JM (2002) Processing of multiple receiver spaceborne arrays for wide-area SAR. IEEE Trans Geosci Remote Sens 40(4):841–852
19. Goodman NA, Stiles JM (2003) Resolution and synthetic aperture characterization of sparse radar arrays. IEEE Trans Aerosp Electron Syst 39(3):921–935
20. Krieger G, Moreira A (2006) Spaceborne bi- and multistatic SAR: potential and challenges. IEE Proc Radar Sonar Navig 153(3):184–198
21. Li Z, Bao Z, Wang H, Liao G (2006) Performance improvement for constellation SAR using signal processing techniques. IEEE Trans Aerosp Electron Syst 42(2):436–452
22. Curlander JC, McDonough RN (1991) Synthetic aperture radar: systems and signal processing. Wiley, New York. ISBN 047185770X, 9780471857709
23. Krieger G, Gebert N, Moreira A (2004) Unambiguous SAR signal reconstruction from nonuniform displaced phase center sampling. IEEE Geosci Remote Sens Lett 1(4):260–264
24. ESA (2006) SOW for the phase 0 study of the six candidate earth explorer core missions, issue 1 revision 0—EOP-SFP/2006-09-1240, 2006

25. Alberti G, Fasano G, D'Errico M, Cesare S, Sechi G, Cosmo M, Formaro R, Rioli Q (2008) Preliminary performance analysis and design of a distributed P-band synthetic aperture radar. In: Proceedings of IEEE 2008 radar conference, 26–30 May 2008, Rome. ISSN: 1097–5659, Print ISBN: 978-1-4244-1538-0. doi: 10.1109/RADAR.2008.4721015
26. Massonnet D (2001) Capabilities and limitations of the interferometric cartwheel. IEEE Trans Geosci Remote Sens 39(3):506–520
27. Moccia A, Rufino G (2001) Spaceborne along-track SAR interferometry: performance analysis and mission scenarios. IEEE Trans Aerosp Electron Syst 37(1):199–213
28. Ruimy A, Saugier B, Dedieu G (1995) TURC—a diagnostic model of terrestrial gross and net primary productivity based on remote sensing data. In: Guyot G (ed.) Proceedings of the international colloquium photosynthesis and remote sensing, Montpellier, France, pp 261–267, 1995
29. Clohessy WH, Wiltshire RS (1960) Terminal guidance for satellite rendezvous. J Aerosp Sci 27(9):653–658
30. Fasano G, Alberti G, D'Errico M, Cesare S, Sechi G, Mazzini L, Pavia P, Torre A, Esposti M. L., Zin A, Matticari G, Bavaro M, Dionisio C, Cosmo M, Formaro R, Rioli Q (2008) An innovative spaceborne P-band mission based on small satellites and formation flying technologies. In: Proceedings of the 3rd international symposium on formation flying, missions and technologies, 23–25 Apr 2008, ESA/ESTEC Noordwijk. ISBN 978-92-9221-218-6, ISSN 1609-042X

Chapter 19
GRACE

Michael Kirschner, Franz-Heinrich Massmann, and Michael Steinhoff

Abstract The GRACE (**G**ravity **R**ecovery **A**nd **C**limate **E**xperiment) satellite mission is in general to a large extent based on hardware and experience of the German CHAMP (**CHA**llenging **M**inisatellite **P**ayload) satellite mission with the addition of a second satellite plus an ultra-precise intersatellite K-band link. Since March 2002 the two identical spacecrafts are flying on the same orbit 220 ($\pm$50) km apart. The two satellites weigh 485 kg each and were launched to an orbit with an initial altitude of 500 km. After a brief mission overview the main scientific results are presented as well as operational aspects dedicated to the GRACE mission. Here two aspects are addressed more in detail from the flight dynamics point of view, the formation keeping and the swap of the satellite position.

19.1 Mission Overview

GRACE is a joint project between the National Aeronautics and Space Administration (NASA) and the German Aerospace Center (DLR). The mission has been proposed in 1996 jointly by the University of Texas at Austin, Center for Space Research (UTCSR), the German Research Centre for Geosciences (GFZ) in Potsdam and the Jet Propulsion Laboratories (JPL) in Pasadena. In 1997, the Gravity Recovery and Climate Experiment (GRACE) mission was selected by NASA for development under a new Office of Mission to Planet Earth program

M. Kirschner, Ph.D. (✉) • M. Steinhoff
German Aerospace Center (DLR), German Space Operations Center,
Münchner Straße 20, 82234 Wessling, Germany
e-mail: michael.kirschner@dlr.de; michael.steinhoff@dlr.de

F.-H. Massmann
Department of Geodesy and Remote Sensing, German Research Centre for Geosciences,
Telegrafenberg, 14473 Potsdam, Germany
e-mail: fhm@gfz-potsdam.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_19,

called Earth System Science Pathfinder (ESSP). As an innovation, the principal investigator and the mission team were ultimately responsible for developing the flight mission hardware from selection to a launch-ready condition, with minimal direct NASA oversight, for accomplishing the scientific objectives and delivering the proposed measurements to the broader Earth science community and general public as expediently as possible [1, 2].

The gravity field of the Earth is variable in both space and time. The primary objective of the twin satellite GRACE mission is to obtain accurate global models for the mean and time variable components of the Earth's gravity field [21]. The primary product of the GRACE mission is a new model of the Earth's gravity field every 15–30 days. This is achieved by making accurate measurements of the inter-satellite range change between two co-planar, low altitude, near-polar orbiting satellites, using a K-Band microwave tracking system. In addition, each satellite carries a geodetic quality Global Positioning System (GPS) receiver and a high accuracy accelerometer to enable accurate orbit determination, spatial registration of gravity data, and the estimation of gravity field models. The Earth gravity field estimates obtained from data gathered by the GRACE mission provides, with unprecedented accuracy, integral constraints on the global mass distribution and its temporal variations within the Earth system.

The two GRACE satellites have identical hardware and were designed for a nominal lifetime of 5 years. They were launched on-board an Eurockot launch vehicle from Plesetsk, Russia, at March 17, 2002, meaning they have reached nearly the double size of their nominal lifetime. Both satellites were placed in the same nominal circular orbit of about 490 km altitude at an inclination of 89°. Following Launch and Early Orbit Phase (LEOP) operations, the orbits of the two satellites evolve naturally for the remainder of the mission. During the science data collection, the two GRACE satellites point their K-Band feed horns towards each other to a high precision. Over the mission lifetime the two satellites remain in coplanar orbits. Due to differential drag force, the along-track separation varies, and station-keeping maneuvers are required to keep the two satellites within 170–270 km of each other. To minimize the interruption of science data collection, the time between two orbit maintenance maneuvers should be at least 30 days.

In the following the scientific results gained so far as well as operational aspects dedicated to the GRACE mission are presented.

19.2 Science

The GRACE science goals are focusing on Earth gravity mapping and climate monitoring as the mission name expresses.

The gravity field activities comprise the static (mean) gravity field as well as time variable parameters which are an excellent basis for climate monitoring.

An additional climate related goal of the GRACE mission is to enable advances in the atmospheric sciences by the recovery of refractivity (and the derived

quantities of temperature and water vapor profiles) and fine ionospheric structure from the use of GPS radio occultation data measured on-board of the satellites [1].

The GRACE mission gave many breakthroughs in the understanding of changes in the terrestrial water cycle, melting and growing of glaciers and ice sheets, sea level rise and its causes (ice melt, thermal expansion), and solid Earth (past glaciation, large Earthquakes).

The downloaded GRACE measurements (K-Band phase, accelerometer, GPS, attitude) are routinely processed by the GRACE Science Data System (SDS), which is in charge of retrieving, processing and distributing the GRACE data (Fig. 19.1). SDS is composed of three centres: NASA/JPL in Pasadena/USA, UTCSR at Austin/USA and GFZ at Potsdam/Germany. More than 93% of the retrieved data has been processed into Level-2 data products (gravity field and related products), and delivered to the user community via the data centres PODAAC at JPL/USA and ISDC at GFZ/Germany (see also the GRACE web sites at UTCSR and GFZ: http://www.csr.utexas.edu/grace and http://www.gfz-potsdam.de/portal/gfz/Struktur/Departments/Department+1/sec12/projects/grace).

19.2.1 Gravity Mapping

The mass within the Earth and on its surface is not evenly distributed. Molten rocks flow in the Earth core, water masses move in the oceans and on the continents and atmospheric masses are also in continuous movement. Since the gravity force of a body depends on its mass, the irregular mass distribution on our planet causes an inhomogeneous and time variable gravity field.

The GRACE satellite mission concept is using the effect that regions of slightly stronger gravity affect the leading GRACE satellite first, pulling it slightly apart from the trailing satellite, i.e. causing a small change in the distance between the two satellites. For this, GRACE uses a uniquely precise microwave ranging system measuring the distance between both satellites with an accuracy of some microns—about one-tenth the width of a human hair—over a distance of 220 km!

The unprecedented accuracy of the distance measurements allows mapping the Earth gravity field approximately once a month over a period of meanwhile more than 9 years.

The coefficients of the Earth's gravity field model are derived from global (satellite-based, ground-based) observations based on complex mathematical physics. Various global static Earth gravity models, which are the long-term mean of the monthly models, have been computed by a number of institutes (UTCSR, CNES/GRGS, GFZ, AIUB or IGG). The Potsdam gravity-field model, named "EIGEN-5C" (European Improved Gravity field of the Earth by New Techniques—Combined with terrestrial gravity data), has become established worldwide as one of the standards for various satellite missions like Envisat, ERS-1/2 and GOCE [3]. Meanwhile EIGEN-6C is available, which shows an even better accuracy and resolution due to improved and new methods in satellite

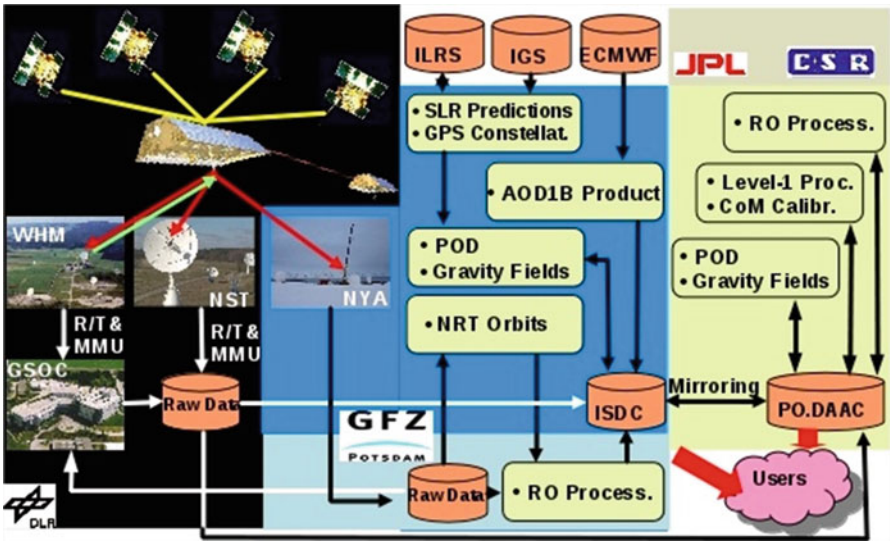


Fig. 19.1 Sketch of the data flow within the GRACE Science Data System (SDS)

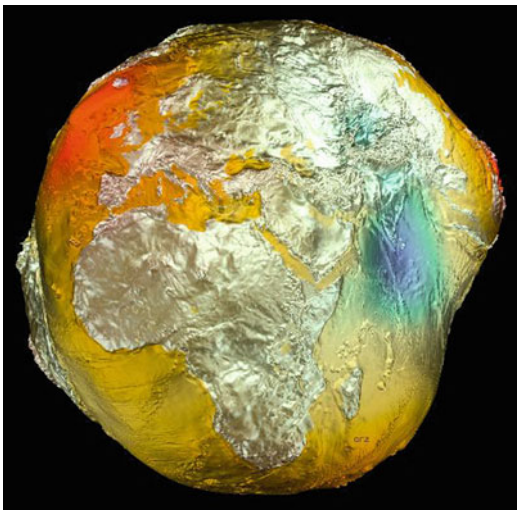


Fig. 19.2 The 'Potsdam Gravity potato' 2011: Geoid based on Lageos, GRACE and GOCE satellite data and surface data (airborne gravimetry and satellite altimetry) [4]

data handling, the inclusion of GOCE data and increased accuracy of Earth surface gravity data (see Fig. 19.2) [4].

The high-precision measurements of GRACE have provided us a much sharper image of our planet. We can now take a look at the Earth's gravity field and its

variations in order to see what conclusions we can deduce regarding the various processes which define the "System Earth".

Integration of GRACE data into the gravity models led to a previously unattained accuracy on a global scale. Variations of the gravity field can be expressed as deviations in the geoid surface from the rotational ellipsoid or as deviations from normal gravity. Together, these observations provide e.g. clues about the topographical and geophysical structures of the tectonic plates. Large blocks, such as the Andes, Himalayas and the North Atlantic Ridge, generate strong positive-gravity anomalies, whereas deep-sea trenches at the edge of the North-West Pacific and off the west coast of South America produce large negative gravity anomalies. In today's gravity anomaly maps, we are able to identify Hawaii as the youngest link in a whole chain of partly undersea volcano cones.

Extensive mountains and valleys in the geoid and the distribution of gravity anomalies are related to structures and processes deep inside the Earth. For example, the geoid domes in the Western Pacific and on the west coast of South America are a consequence of the older, and therefore denser, oceanic lithosphere descending into the mantle at these locations. Other high spots on the geoid and in the gravity anomaly maps are found in areas where hot material, which probably rising within the mantle, is pushing up the lithosphere above it. This occurs e.g. in the North Atlantic, around Iceland and South-East of Africa. The striking deep valley in the geoid to the south of India could be related to the northward movement of the Indian lithosphere plate. This pushes up the Himalayan Massif, thus thinning the mantle mass to the rear of this movement.

It becomes also evident that GRACE can detect the gravity changes associated with subduction zone earthquakes with moment magnitudes larger 8.5 like the Sumatra-Andaman earthquake in December 2004. Unique to the GRACE observational data are the recovery of the long-wavelength components of the deformation field. GRACE observations provide unique opportunities to study the material properties of the subduction zone environment up to several hundred kilometers depth.

A further depression in the geoid is a relict of glaciation around 20,000 years ago, when a thick ice sheet pushed down the lithosphere and upper mantle above Canada and Fennoscandia. This ice melted 6,000 years ago, relieving the enormous weight on the crust, which has continuously risen since then.

Considerable progress has been made recently by the use of GRACE data for modeling this Glacial Isostatic Adjustment (GIA), the slow viscous response of the mantle to the last deglaciation.

19.2.2 Climate Related Science

Highly precise measurement of the temporal variations of the gravity field can also provide information about the climate, the "C" in the name of the mission. Many processes in the climate behavior of our planet are water-driven: Ocean currents transport heat towards the poles and cold towards the equator and thus regulate

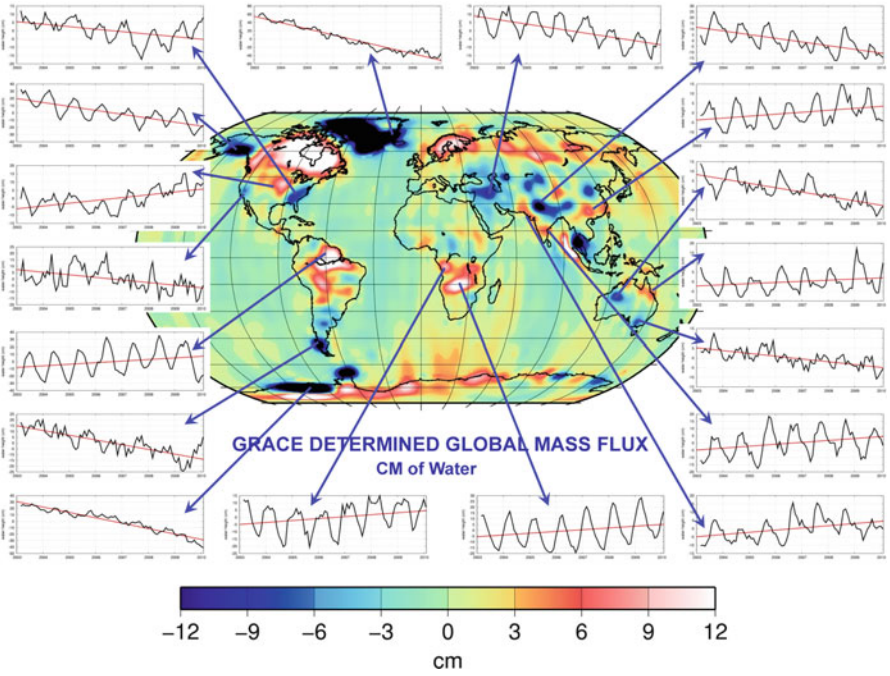


Fig. 19.3 GRACE based monitoring of the continental ice/water balance (GFZ)

ocean circulation. The decrease and increase in ice masses of the large Polar ice sheets are important factors that affect the climate and strongly contribute to sea level rise (see also Ref. [5]). The global hydrological cycle depends decisively on the water balance of large river basins and ground water storage. These changes in the global water distribution correspond to redistribution of masses within the Earth's gravity field. Thus the time series of the GRACE gravity field models provides an excellent base to monitor and study these phenomena. The mass variability is measured with a spatial resolution of about 300 km at the equator with up to sub-monthly resolution. This allows deriving long-term trends, episodic variations and seasonal changes.

From these temporal variations geo-scientists have derived new insight into dynamic processes in the Earth interior, into water mass transfer processes and into the development of ice sheets and glaciers on Greenland and Antartica. With the GRACE mission, for the first time a systematic and thorough monitoring of the amounts of water, ice and matter moving around could be performed and thus a completely new picture of the processes within and on our Earth emerges. On the other hand, these parameters serve to map water transfer processes between land, ocean and atmosphere and are key input for climate modeling.

GRACE provided the first measurements of changes in the global continental water storage (see Fig. 19.3). The data offer more comprehensive information on

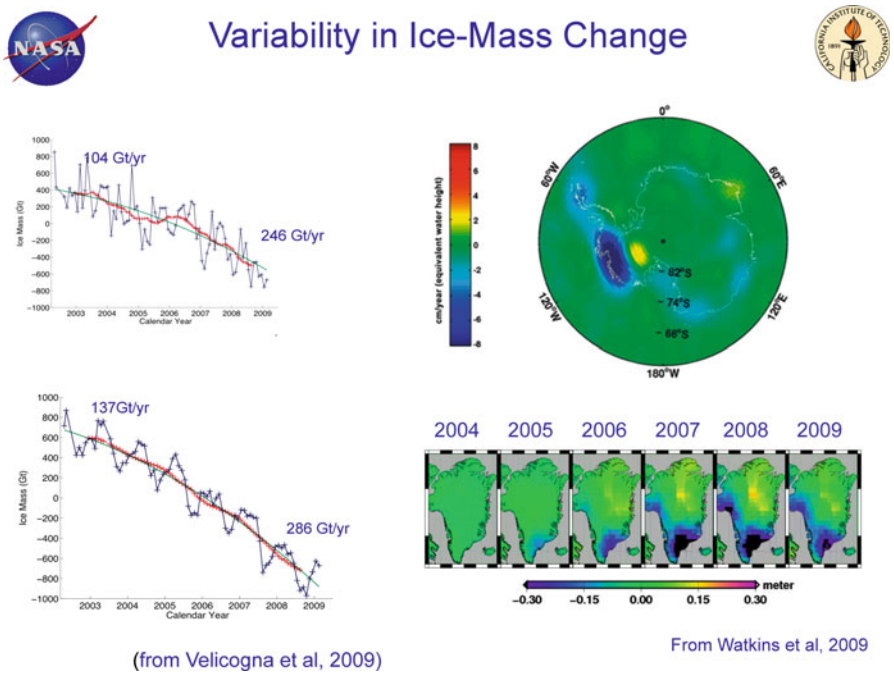


Fig. 19.4 Variability in ice-mass change for Antartica and Greenland

storage changes in groundwater, in snow cover, rivers, lakes and flood plains, than any other ground or satellite-assisted observation system. Analysis of GRACE data show how variability in climate conditions, for example precipitation and air temperature, affect worldwide seasonal and annual variations in water storage within large river catchment areas. It also allows comparison with and adjustment of hydrological models, which will be used to determine climate-induced changes in global and regional water cycles.

As the GRACE time series has lengthened, it has become apparent that the Greenland and Antartica ice mass loss rates are highly variable and show an increasing velocity (see Fig. 19.4) [5]. By combining GRACE measurements with those from other techniques (i.e. IceSat, GPS), investigators are beginning to explore the possibilities of improving snow/ice density models or to better understand GIA effects.

In addition GRACE is also providing mass-change estimates for large continental glacier systems (Patagonia, Alaska, Central Asia). Meanwhile GRACE observations are also used for studies of the permafrost regions. Extended GRACE data records will improve this understanding in order to improve predictions of the ice-sheets in a warming climate.

The GPS receiver on board the trailing satellite (currently GRACE-1) is used to obtain GPS radio occultation measurements. Globally distributed vertical profiles

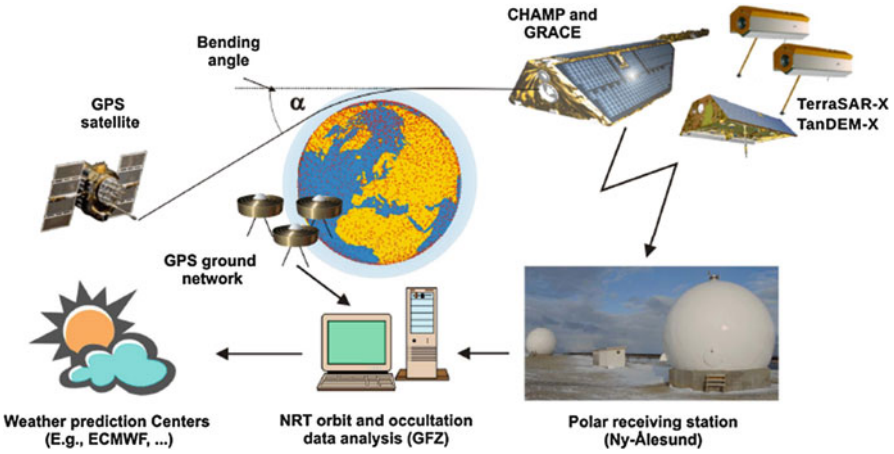


Fig. 19.5 GPS radio occultation data used for operational weather prediction (GFZ)

of atmospheric parameters (bending angles, refractivity, dry temperature) are derived and provided to the national weather services for assimilation into their prediction models (see Fig. 19.5). Since 2006 the data is operationally used by weather services.

19.3 Mission Operations

In this chapter the operational aspects of the GRACE mission within the German Space Operations Center (GSOC) in Oberpfaffenhofen as well as the used tools and applied algorithms are presented. Requirements derived from hardware degradations on-board are also addressed.

19.3.1 Flight Operations

The GRACE system consists of two satellites in orbit. Each of them has two mass memory units one active and the other as backup. Operational use of the back-up unit is not foreseen, but it may serve as temporary storage for data to avoid memory overflow in contingency cases. The spacecrafts are controlled from GSOC. As baseline the ground stations in Weilheim (WHM) and Neustrelitz (NSG) will be used for data dumps, an additional downlink ground station in Ny-Alesund (Norway) is added for NRT (Near-Real-Time) analysis of the radio occultation data plus frequent status monitoring and data is available twice.

Commanding of the satellites is nominally planned only via Weilheim. As the separation of the two satellites is so small that the station visibilities overlap,

the assignment of station to satellite can not be changed during a pass. This means any reaction on dump failures can occur at the next scheduled Weilheim pass.

The Weilheim S-Band 15-m parabolic antenna provides GSOC with Tracking (Angle data) and TM data in addition to the primary function of TC uplink transmission (4 Kbps). WHM extracts real-time data, out of Virtual Channel 0 (VC0) and dump data HK of Virtual Channel 7 (VC7) from the acquired downlink telemetry (32 Kbps or 1 Mbps, respectively).

The Neustrelitz (NSG) 7.5-m antenna is the designated, prime TM reception station. NST acquires the downlink TM, extracts the Housekeeping (HK) data after each pass and provides near real-time, HK files on the NSG FTP Server for GSOC collection. Raw Science data acquired at NSG is extracted and archived onsite and made available to the Science Data System for further processing. Additional ground stations are used in special situations. The NASA Polar Ground Network (PGN). McMurdo (MGS), Antarctica, Svalbard (SGS), Spitzbergen, and the Wallops facility (WGS), provide support during Spacecraft Contingencies and Emergencies. Command (TC) and Telemetry (TM) sessions will be supported, and tracking data will be collected during each contact, however MMU dumps are only exceptional.

19.3.1.1 Power Saving Operations

Due to the long mission duration (almost twice as designed) the batteries experienced some degradation (cell loss). This leads to extra support of the engineers and the operations team. New tools have been compiled always dependant on the changing battery-loading strategies.

Lower heater settings near the cold points of the instruments to serve power of the batteries have been established.

The Instrument Control Unit (ICU) including the accelerometer has been switched off for 2 weeks since end of 2010 during ß = 0 (long shadow period).

New on-board command sequences (Macros) have been stored to make life easier and to cut down the number of time tagged commands for new requirements.

The on-call service became more intensive to observe changes in the behaviour of the satellites and to have a chance to galvanize the on-call person into action.

19.3.2 Telemetry

GRACE (Gravity Recovery And Climate Experiment) is using a pre-PUS (Packet Utilisation standard) TM (TeleMetry) technology. This means the on-board (o/b) collected Housekeeping (HK) and Science (SCI) data are transferred via CCSDS (Consultative Committee for Space Data Systems) TM Frames that contain CCSDS TM Source Packets that again contain CHAMP Application Packets (CAPs) (refer also to the CHAMP website at GFZ: http://op.gfz-potsdam.de/champ).

TM Transfer Frames containing R/T data are transferred via the Virtual Channel 0 (VC0), whereas dump data and idle Frames use VC7. The size of the Frames is fixed to 1,115 byte.

R/T TM source packets are implemented with variable length and make use of packet segmentation. Dump source packets have a fixed length to enable a fast dump from on-board memory.

CAPs have a fixed maximum size for each packet.

The Source Packets contained in the Frames are using a CCSDS standard Source Packet Header, but do not carry any time information (timestamp). The timestamp calculation of the Source Packets, CAPs and TM parameters respectively is performed by evaluation of dedicated GPS synchronous timestamp CAPs that are generated, stored and sent with a frequency of 1 Hz. So, the timestamp precision of CAPs and parameters is 1 s.

GRACE mission distinguishes data into two main categories, HK and SCI. This means in the sense of the PUS, no Diagnostic service is supported. The main categories are further broken down into R/T and dump data. These main and sub categories are accordingly reflected within the source packet and CAP headers to allow appropriate on ground processing.

Upon the categories and sub-categories the data flow of the ground data processing is controlled in the sense of filtering special information and packet routing to different processing stages.

Besides the on-board FDIR (Failure Detection Isolation and Recovery) all parameter processing takes places on ground—this includes also parameter calibration. The on-ground CAP and parameter processing is done by FRAMTEC (FRamework for advanced Monitoring, Telemetry and Control), which extracts the parameter values from the CAPs, performs necessary calibrations, limit checks and provides all necessary information to display these values on the display system with their correct raw, engineering value, status, colour and special formatting option. FRAMTEC offers also a convenient way to derive parameters from actual telemetry or monitoring parameters (direct parameters) via an Operations Language (OL) which is very similar to the C programming language.

One way to monitor the actual status of the GRACE satellites is real-time with the GUI called SATMON (SATellite MONitor), which is fed directly from the FRAMTEC processor. This can be done directly in the control room or via a dedicated web interface called OpsWeb (Operations Web) even from home.

Another way to check against the satellite status and configuration is implemented with the Consistency Checker (ConChk) that allows to pre-define lists of parameters together with their expected values or value ranges and to check against these limits with a single keystroke, e.g. at the beginning of a pass to minimise the time needed for problem detection.

A third—more detailed—way to assess the satellites' status is to analyse offline data provided by downlinked HK store dumps. HK dumps contain data of coherent periods of time and are processed by a dedicated process called Offline Processor, which is triggered shortly after reception of the offline (dump) file and is able to process large amount of data in short time—again to minimise the reaction times to

satellite problems. The offline processing on the one hand allows assessing the satellite's status over long periods of time and on the other hand allows the detection and analysis of problems which can very well be transient, i.e. not visible during the short visibility periods of the GRACE satellites (about 30 min/day aggregate).

19.3.3 Telecommand

The GRACE Command System is one of the major components of the GRACE Mission Operations Segment (MOS) running within the GSOC control centre, in Oberpfaffenhofen. Along with the Telemetry System described in the previous section it forms the main software interface to the operational GRACE 1 and GRACE 2 spacecraft.

The system has a heritage based on many previous missions supported at GSOC not least the precursor mission CHAMP. The system itself is based on a Core Command System concept with a design based on a multi-mission capable architecture. All satellite specific execution control, interface and database requirements were handled by satellite specific software "plug-ins".

A satellite in low earth orbit is visible from a very small fraction of the Earth's surface at any one time and has a high angular rate of motion against the celestial sphere. Ground contact with the satellite is limited in duration and coverage. This means that the command system and supporting network and uplink capabilities must support the following:

- High commanding rate during short contacts
- Uplinking of deferred execution, i.e. time-tagged telecommand sequences
- Verification and correct reception and storage of transmitted telecommands

As GRACE is comprised of two spacecraft, whose functions with respect to telecommanding are identical, this allows the two satellites to share a common source telecommand database. The onboard functionality is not identical due to failures that have occurred to one or other of the spacecraft during the mission life time. The telecommand database is compiled, validated and subsequently installed on the two command servers running GRACE1 and GRACE2 respectively. From a flight operations personnel perspective the two command systems are identical except for a GRACE 1 and 2 specific colour scheme used for each command interface. This colour scheme is consistent with the one used by the telemetry display system.

The majority of commands to be uplinked to the GRACE satellites are derived from products from two of the other components of the MOS, namely:

- Mission Planning; which is responsible for the generation of timelines and the associated time-tagged commands and
- Flight Dynamics; which handles orbit determination and prediction plus attitude determination and manoeuvre planning/execution.

The GRACE Command system then utilises the Weilheim Ground station for the uplink of the following GRACE specific command activities:

- Time-tagged commands for execution of memory dumps and prime star camera switching (to avoid Sun and Moon intrusions).
- Command containing updated orbit information of both satellites to the on-board orbit propagator (daily basis although the limit is $<$120 h).
- Commands from other sources (e.g. special requests, recommendations, etc.).

In line with the operational requirements, the GRACE command system was tailored to support the PUS standard and is fully CCSDS compliant with the CCSDS standards for Telecommanding, such as:

- Part 1 CCSDS 201 channel service—CLTUs
- Part 2 CCSDS 202 data routing—TC-Frames
- Part 2.1 CCSDS 202 command ops proc—COP-1
- Part 3 CCSDS 203 data management—TC-Packets

The GRACE command system is based on the Core Command System architecture and although it is no longer state-of-the-art it has shown it's pedigree by support numerous missions of varying scope over the last 20 years. As we approach the tenth anniversary (2012) of the GRACE satellite's launch, the command system will continue to support the mission until such time as the satellite is no longer in service. To date, after more than 3,500 days in orbit, the GRACE command systems have sent more than 1,000,000 telecommands to the twin spacecraft.

19.3.4 Attitude and Orbit Control System

The GRACE satellites are themselves part of the instrumentation used to probe the Earth's gravity field. This directly affects the attitude and orbit control system (AOCS) in several ways. The inter-satellite distance must be maintained within a window of $\pm$50 around 220 km and the attitude within dead bands of $\leq$3–5 mrad. Disturbances and interruptions of payload operations should be minimized. The continuous fine-tuning and optimization of the control system for the benefit of the scientists implied that GRACE never really entered into a routine operational phase like other satellites. After more than 9 years in orbit—almost twice the projected life time already—several components start to deteriorate, or can not be used any more. This also requires more and more special AOCS operations.

19.3.4.1 Orbit Control

Two thrusters with 40 mN each are located at the back of the satellites. Each thruster is connected to one of two tanks (the cross-connection—a solenoid valve has been kept close since the beginning of the mission). If an orbit maintenance

maneuver has to be performed, the type of the maneuver (altitude increase or decrease) depends on the satellite (leader/follower) and on the relative altitude and distance (refer also to Sect. 19.3.5.1). In addition, a 180° yaw slew is required in the case, if the leader has to perform an orbit increase maneuver or the follower an orbit decrease maneuver. Inclination corrections require a ±90° yaw turn near one of the nodes.

None of these choices or implementations is automated and AOCS tasks comprise:

- Commanding the correct attitude at the correct time.
- Conversion of Δv, delivered by the flight dynamics group, into thruster on-times depending upon predicted pressure and temperature; conversion of thruster on-times into integral on-board counts, which can then be commanded to the OC-thrusters.
- Choice of one or two thrusters, depending upon desired cross-couplings (e.g. for inclination control) and tank contents.
- A posteriori Δv reconstruction with the actually measured pressure and temperature during the burn; administration of all maneuvers and calibrations.

19.3.4.2 Attitude Control

Two sets of six 10 mN thrusters are used for attitude control together with magnetic torque rods on each axis which have a maximum moment of 110 Am^2. Each set of thrusters is connected to its own tank and the torque rods have double coils for redundancy.

The exact pointing of the two spacecraft is derived from daily uploaded two-line elements; orbit information of both satellites is fed to both satellites. It is also possible to set a fixed attitude bias with respect to the orbital frame. AOCS tasks here include:

- Tuning of the ratio of the torque delivered by the rods and by the thrusters. Numerous settings of this ratio and also of several other control parameters were tried at the beginning of the mission until the scientists were satisfied.
- Implementation of so-called magnetic compensation designed to minimize the interaction of the magnetic torque rods and the magnetometer. Before this implementation disturbances were minimized by mixing the measurements with the on-board magnetic field model, whereby several relative weights were tried.
- Optimization of the attitude dead bands in order to minimize the number of thruster firings and to balance the number of thruster cycles per axis.
- Optimization of star camera control parameters in order to minimize noise and drop outs.

19.3.4.3 Other AOCS Activities

The precession of the orbit necessitates a change of the prime star camera every 161 days.

Also twice in the 322d period images are made in two eclipses without Moon with both star cameras in order to determine the number of "hot pixels". The number of pixels showing spurious light detection is slowly increasing with time. The degradation of one of the cameras on GRACE 2 made detection of enough bright stars around the galactic poles to a problem and prompted an even more elaborate switching scheme as mentioned above.

The configuration of the AOCS safe mode must be adapted thrice in the same 161d period (also when the mode is not actually used!).

Centre-of-mass calibrations are performed once or twice per year. Both satellites are being wiggled in one axis after the other by sending alternating currents through the torque rods while at the same time disabling the thrusters. Specific orbit locations are chosen in order to minimize disturbances. Results may lead to mass trim maneuvers; GRACE has three small masses that can be moved with a stepper motor along a rod aligned with one of the three axes thus adjusting the centre of gravity.

The alignment of the microwave assembly was calibrated on both satellites only once at the beginning of the mission. Also the star camera alignments were measured and adapted several times early in the mission.

A number of components show in orbit performance degradation or has become defunct during the mission. In all instances there is a direct impact on AOCS operations. For example the degradation of the batteries impedes the use of the IMU (inertial measurement unit; an optical gyro) and regularly calls for yaw maneuvers in order to turn the panels into specific orientation with respect to the Sun. Malfunctioning of some CESS thermistors prompted the design of an additional safe mode using the magnetic field to align the satellite in flight direction. However, no further thermistors developed problems so far and hence this mode has not yet been used in praxis.

Finally, an administration of the use of resources is made, the most obvious being fuel expenditure. The projected life time can then be compared with the expectancy based upon other factors, such as front-end or battery degradation, or orbit decay (see Fig. 19.6).

19.3.5 Flight Dynamics

Besides the typical flight dynamics tasks performed routinely between several times per day and weekly like e.g. orbit determination based on GPS navigation solution data, generation of ephemeredes files like Keplerian elements, event files like AOS/LOS over ground stations, etc., two dedicated requirements to the GRACE mission have to be fulfilled: (1) formation keeping within a distance of 170 and 270 km and (2) swap of the satellite position after about 3 years. These two aspects are addressed more detailed in the following two chapters.

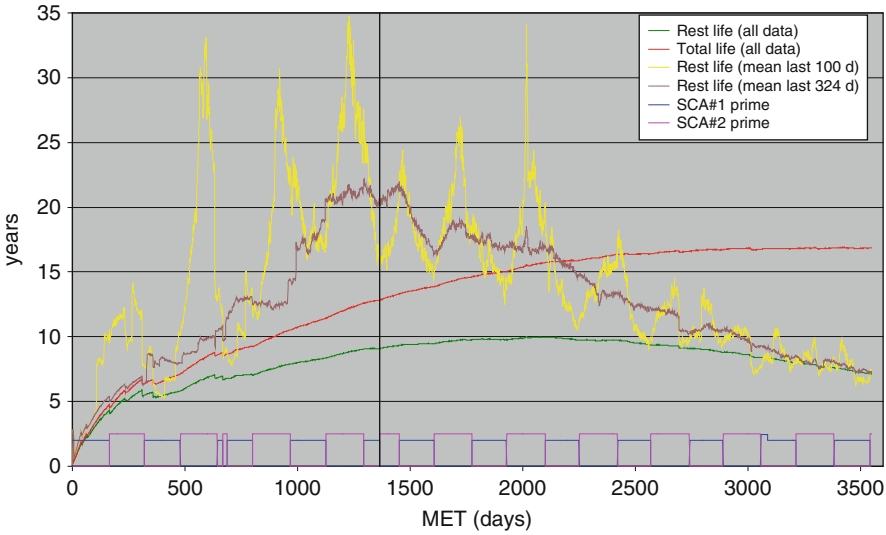


Fig. 19.6 Life expectancy in years based upon fuel expenditure on GRACE1 as a function of mission elapsed time. The *red* and *green* lines show the total and remaining life time. The *dashed yellow* and *orange* lines are predictions based upon the last 3 months and 322 days of data, respectively. The difference in star camera performance (prime is shown at the bottom; *blue* is camera 1, *lilac* camera 2) is apparent. The *vertical black* line shows the moment leader and follower swapped roles (at first GRACE1 was leader, now it is GRACE2)

19.3.5.1 Formation Keeping

Once the initial formation has been achieved in a nominal distance of 220 ± 50 km, the orbits of both spacecraft evolve naturally under the action of the gravitational and non-gravitational forces. During short time scales ($\approx$1 rev), the relative motion is dominated by two effects: (1) the differences in eccentricity, which are limited to less than 1.4×10^{-4} as a consequence of the separation and drift stop strategy, because the two satellites were released from the upper stage by an angle of $+60°/-120°$ w.r.t. the flight direction and an separation velocity of about 28 cm/s for each spacecraft related to the upper stage [6], and (2) the 1.9° difference in eccentric anomaly (scaled by an eccentricity of less than 2.5×10^{-3}).

If a difference Δa of the mean semi-major axes exists, the along-track separation L changes for longer time scales according

$$\frac{dL}{dt} = a \cdot \Delta n = -\frac{3}{2} n \cdot \Delta a. \tag{19.1}$$

by roughly $150 \times \Delta a$ per day.

Under the influence of the atmospheric the altitude of both spacecraft decreases from an initial value of 490 km throughout the mission life-time. In parallel to the secular decrease, the non-equality of the aerodynamic forces acting on the two

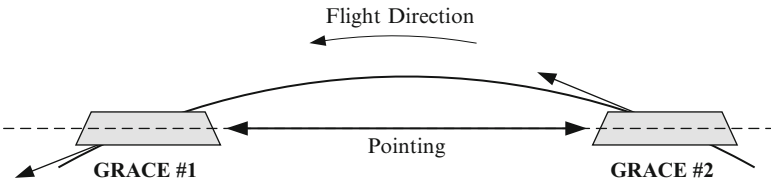


Fig. 19.7 On-orbit constellation of the GRACE satellites

Table 19.1 Pitch angle depending on separation and altitude (bold values apply for the lower and upper boundary of nominal separation)

Separation (km)	100	170	270	500
H = 300	0.43	0.73	1.16	2.15
H = 500	0.42	0.71	1.12	2.08

satellites is responsible for a difference in the relative semi-major axes. Despite the symmetrical design of the satellites, their ballistic coefficients ($B = c_D \times A/m$) differ as a consequence of a slightly different attitude (pitch angle and orientation; cf. Fig. 19.7). To ensure a line-of-sight orientation of the K-band link, pitch angles between 0.4° and 2.2° are required depending on the altitude and relative distance (cf. Table 19.1). Within the nominal operations regime, the associated ballistic coefficient difference ranges from 0.14% to 0.32%, if the mass of both spacecraft is identical. In this case the leading satellite always has a higher drag and, as a consequence, a higher decay rate [7–9].

Different ballistic coefficients in general result in a rate of change

$$\frac{d\Delta a}{dt} = -\Delta B \cdot \rho \cdot v^2 \frac{1}{n} = -\Delta B \cdot \rho \cdot a^2 n \tag{19.2}$$

of the semi-major axis difference (cf. [10]), where $v \sim 7.5$ km/s is the orbital velocity of the satellites and ρ denotes the atmospheric density. Combining (19.1) and (19.2) and assuming a constant atmospheric density, the solution for L is given by the relation

$$L(t) = L_0 + \frac{3}{4} \frac{1}{\Delta B \cdot \rho \cdot a^2} (\Delta a(t))^2. \tag{19.3}$$

In this case, the evolution of the separation shows a quadratic behavior. For a maximization of the time between subsequent formation keeping maneuvers, the relative altitude versus relative separation should be a parabola mirrored to the altitude of the reference satellite. This requires that the two spacecraft should initially be placed at the maximum desirable separation with a semi-major axis offset

$$\Delta a_{\max} = \sqrt{4/3 \cdot \Delta B \cdot \rho \cdot a^2 \cdot \Delta L_{\max}} \tag{19.4}$$

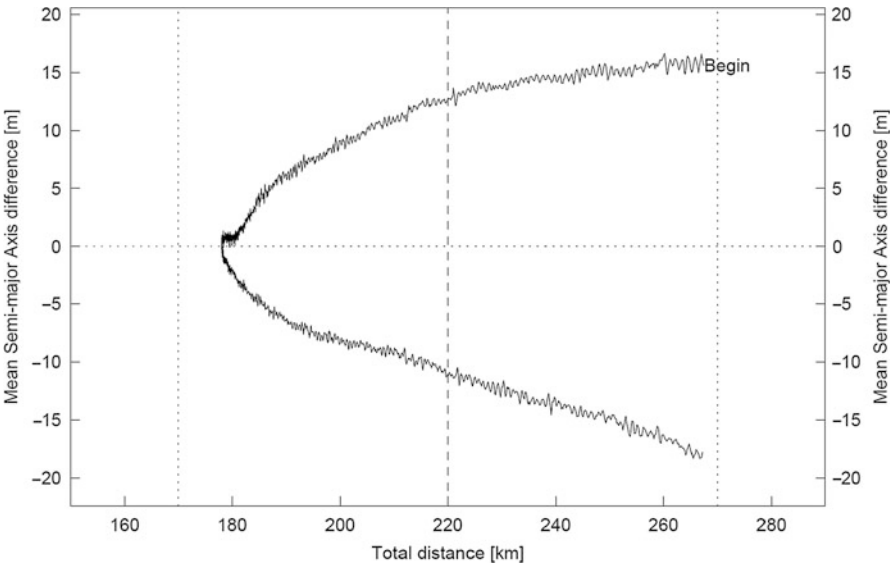


Fig. 19.8 Numerical simulation of the relative motion of the GRACE spacecraft of the first cycle after separation from the upper stage under the influence of differential drag using realistic solar flux profiles and atmospheric density values as well as ballistic coefficients depending on the varying pitch angle

Table 19.2 Comparison of predicted and simulated drift cycle

	ΔB (m^2/kg)	ρ (kg/m^3)	a (km)	n (1/s)	ΔL_{max} (km)	Δa_{max} (m)	t_{cyc} (d)
Prediction	1.56×10^{-5}	2.35×10^{-12}	6,878	0.0011	100	15.3	185
Simulation	Variable	Variable	Variable	Variable	90	15.5	174

of the leading spacecraft with $\Delta L_{\max}$ as the width of the station keeping box. Due to the symmetry of the parabolic motion, the minimum distance is achieved, when both semi-major axes are equal. At the end of the station keeping cycle a small maneuver is performed to raise the semi-major axis of the leading s/c or, alternatively, to lower the semi-major axis of the trailing satellite. Using (19.3) and the maximum allowed variation in the separation $\Delta L_{\max}$, the length of the station keeping cycle, i.e. the time between subsequent correction maneuvers, can be derived as (refer also to Ref. [6])

$$t_{\text{cyc}} = \sqrt{\frac{16}{3} \frac{\Delta L_{\max}}{\Delta B \cdot \rho \cdot a^2 n^2}} \tag{19.5}$$

A simulation of the first cycle performed before launch and based on numerical orbit propagation (cf. Fig. 19.8) illustrates similar results compared to the analytical prediction based on (19.4) and (19.5) (refer also to Table 19.2). Even though the relative motion in general is well explained by the analytical model derived above,

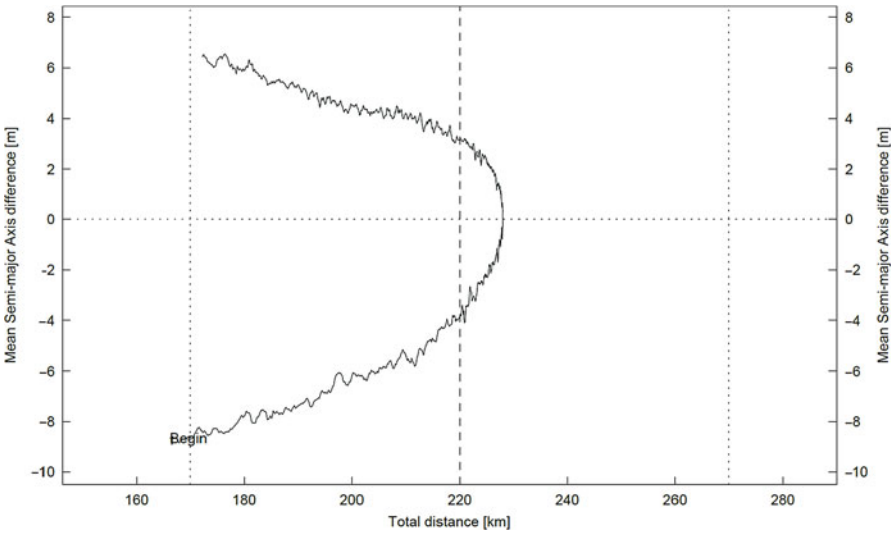


Fig. 19.9 Monitoring of the cycle after the 17th formation maintenance maneuver covering the period from 2011/02/08 to 2011/08/19. GRACE2 is the leading spacecraft after the swap maneuver in 2005 (cf. next section)

the ideal parabolic motion will be perturbed due to realistic density variations for cycle lengths of several months (cf. Fig. 19.8). In case of different attitude activities, the ballistic coefficient difference ΔB will also be influenced due to changes in the s/c mass. Owing to expected cycle lengths of up to 1 year, forecasts of the solar flux values and their uncertainties provided by the European Space Operations Center (ESOC) on a daily and monthly basis have to be used in the planning of an upcoming orbit maintenance maneuver. As a consequence, the available station keeping box will not be fully exploited to leave some margin for unpredictable density variations. Throughout the mission, the relative distance and semi-major axis difference are monitored on a routine basis (cf. Fig. 19.9) as well as the evolution of the mass of both s/c. In parallel, predictions of the expected motion in the L-Δa-plane have been carried out to decide on a suitable maneuver time and size.

For proper maneuver planning, the individual mean semi-major axes must be resolved from the osculating states with an accuracy of better than 1 m. While a purely numerical averaging requires lengthy time intervals to give proper results, analytical models do not, in general, provide sufficient accuracy. It has therefore been decided to use a suitable combination of both techniques to derive the mean semi-major axes of the GRACE spacecraft at any time. To this end, the trajectories are numerically integrated over a time interval of six orbits using a rigorous force model and initial conditions from the latest orbit determination. Using a sampling interval of 10 min, the resulting state vectors y are fit to the SGP4 model by the least square method [11, 12]. The SGP4 orbit model, which forms the basis of the

NORAD twoline element sets [13], is based on the analytical theory of Brouwer and accounts for the Earth gravitational field through zonal terms *J2*, *J3*, *J4*, and the atmospheric drag through a power density function assuming a non-rotating spherical atmosphere. Using this technique, smooth Δa values with an uncertainty of less than 1 m are obtained (cf. Figs. 19.8 and 19.9).

19.3.5.2 Safe Switching of the GRACE Formation Using an Eccentricity-Inclination Vector Separation

Mainly to balance the surface erosion of the on-board K-band radar, both satellites have to exchange their position (leader/trailer) at least once during the mission lifetime. After GRACE has entered the third year of the original nominal 5 years mission, such a position exchange maneuver was planned for the second half of 2004. In an effort to maximize the operational safety of the satellite swap maneuver with no increase in fuel expenditure, an eccentricity and inclination vector separation is applied. The concept of e/i-vector separation has originally been developed for the co-location of geostationary satellites, but can likewise be applied to avoid a collision risk during proximity operations of satellites in low Earth orbits. Using a specific phasing and size of the relative orbital inclination and the relative eccentricity, a minimum radial or cross-track position offset between two spacecraft can always be ensured irrespective of their along-track separation. This is of particular concern if uncertainties in the maneuver calibration or differential drag modeling do not allow an accurate forecast of the along-track motions. Likewise the ever-present risk of maneuver failures can easily be covered by application of an e/i-vector separation.

As a result of natural orbital perturbations, the relative e/i-vectors of the two GRACE satellites vary in time and attain an optimal, parallel configuration once every 47 days. A fuel optimal and safe separation of both spacecraft during the fly-by can thus be achieved by proper selection of the maneuver date. In adopting the e/i-vector separation, single maneuvers are sufficient to initiate and stop the drift required for the swap of GRACE 1 and 2, which notably reduces the operational complexity and workload.

Relative Motion of LEO Satellites

For low Earth orbits (LEO) the relative motion of a spacecraft formation can be approximated by the well-known Clohessy-Wiltshire (or Hill's) equations [10, 14], which describe the location of two spacecraft in a co-moving frame in the radial (R), along-track (T) and cross-track (N) direction. In general, the relative motion is a harmonic oscillation perpendicular to the orbital plane superimposed by an elliptic in-plane motion. If the semi-major axis of the two spacecraft is different, a linear drift in along-track direction is superimposed, too.

As the Cartesian formulation of the Clohessy-Wiltshire equations does not provide immediate insight into some aspects of the relative motion, a description

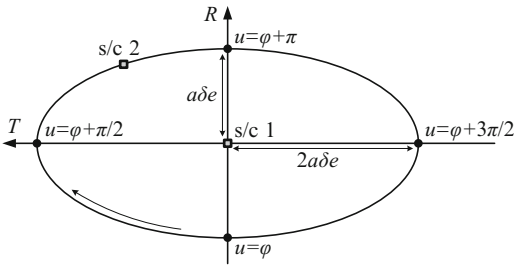


Fig. 19.10 Relative motion of two spacecraft with eccentricity vector separation [17]

in terms of orbital element differences is therefore preferred for proximity analyses. Thus the concept of eccentricity and inclination vectors is introduced followed by a description of the relative motion in terms of the relative e/i-vectors.

Eccentricity Vector

For near-circular satellite orbits, the Keplerian elements e (eccentricity) and ω (argument of perigee) are commonly replaced by the eccentricity vector

$$\vec{\mathbf{e}} = \begin{pmatrix} e_X \\ e_Y \end{pmatrix} = e \cdot \begin{pmatrix} \cos\omega \\ \sin\omega \end{pmatrix} \tag{19.6}$$

It is free from singularities and well suited for the study of orbital perturbations of remote sensing satellites [15]. The relative eccentricity vector $\Delta\vec{\mathbf{e}}$ can be expressed by the difference for two spacecraft

$$\Delta\vec{\mathbf{e}} = \vec{\mathbf{e}}_2 - \vec{\mathbf{e}}_1 = \delta e \cdot \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}, \tag{19.7}$$

which characterizes the relative motion within the orbital plane. For small differences in the orbital elements a linear expansion can be performed using the well known partial derivatives [16] of the Keplerian orbit model. It can then be shown that the relative orbit of s/c 2 with respect to s/c 1 is an ellipse of dimension $\pm 2a\delta e$ in along-track direction and $\pm a\delta e$ in radial direction (Fig. 19.10), where a denotes the (common) semi-major axis.

While δe is a measure of the size of the relative trajectory, the angle φ defines the relative pericenter. Whenever the mean argument of latitude $u = \omega + M$ equals φ, s/c 2 is located right below s/c 1. Subsequently, s/c 2 takes over and is just ahead of s/c 1 as soon as the argument of latitude attains a value of $u = \varphi + \pi/2$ [17].

Inclination Vector

In the same manner, a relative inclination vector can be used to describe the relative motion of two satellites perpendicular to the orbital plane. Here the unit vectors $\boldsymbol{X}_\mathrm{i}$, $\boldsymbol{Y}_\mathrm{i}$, and $\mathbf{Z}_\mathrm{i}$ (i = 1,2) describe a coordinate system aligned with the ascending node

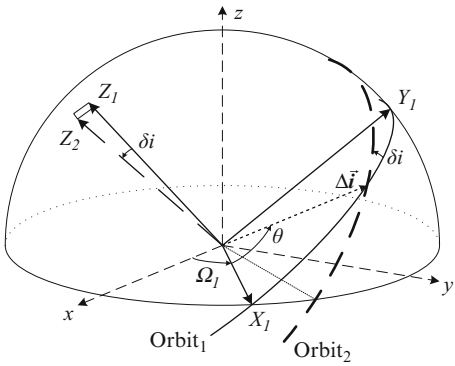


Fig. 19.11 Relative inclination vector [17]

and the orbital plane of the respective spacecraft (Fig. 19.11). These vectors can be expressed with the Keplerian elements i (inclination) and Ω (right ascension of the ascending node) by the equations

$$\mathbf{X} = \begin{pmatrix} +\cos\Omega \\ +\sin\Omega \\ 0 \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} -\cos i \sin\Omega \\ +\cos i \cos\Omega \\ +\sin i \end{pmatrix} \quad \mathbf{Z} = \begin{pmatrix} +\sin i \sin\Omega \\ -\sin i \cos\Omega \\ +\cos i \end{pmatrix}. \tag{19.8}$$

The projection of the orbital plane normal vector $\boldsymbol{Z}_2$ onto $\boldsymbol{X}_1$ and $\boldsymbol{Y}_1$, can now be used to construct the relative inclination vector

$$\Delta\vec{\mathbf{i}} = \begin{pmatrix} -\mathbf{Y}_1^T\mathbf{Z}_2 \\ +\mathbf{X}_1^T\mathbf{Z}_2 \end{pmatrix} = \sin(\delta i) \cdot \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \tag{19.9}$$

which simplifies to

$$\Delta\vec{\mathbf{i}} \approx \begin{pmatrix} \Delta i \\ \sin i \Delta\Omega \end{pmatrix} \tag{19.10}$$

for small differences in the orbital elements. Its modulus equals the sine of the angle δi enclosed by the two orbital planes while θ is the argument of latitude at which s/c_2 crosses the orbital plane of s/c_1 in ascending (i.e. $+\mathbf{Z}$) direction (the relative ascending node).

For small angles δi, the motion of s/c_2 with respect to the orbital plane of s/c_1 may again be linearized. Using appropriate series expansions (see e.g. [18]), the cross-track motion of s/c_2 can then be described by a harmonic oscillation of amplitude $a\delta i$ and phase angle u–θ [17].

Linearized Relative Motion Equations

In addition to the eccentricity and inclination vector difference, differences in the semi-major axisΔa and the mean argument of latitude Δu have an influence on

the relative motion of two spacecraft. These result in a systematic offset of size Δa in radial direction as well as a drift of $-3\pi\Delta a$ per revolution and a constant offset $a\Delta(\omega+M)$ in along-track direction.

Overall, the relative position vector $\Delta\boldsymbol{r}$ of s/c_2 with respect to s/c_1 in a local horizontal frame aligned with the radial (R), along-track (T) and cross-track (N) direction, can thus be described by linearized equations as given in [17]

$$\begin{pmatrix} \Delta r_{\mathrm{R}} \\ \Delta r_{\mathrm{T}} \\ \Delta r_{\mathrm{N}} \end{pmatrix} = a \cdot \begin{pmatrix} \frac{\Delta a}{a} & 0 & -\Delta e_x & -\Delta e_y \\ \Delta l & -\frac{3}{2}\frac{\Delta a}{a} & -2\Delta e_y & +2\Delta e_x \\ 0 & 0 & -\Delta i_y & +\Delta i_x \end{pmatrix} \cdot \begin{pmatrix} 1 \\ u - u_0 \\ \cos u \\ \sin u \end{pmatrix} \tag{19.11}$$

Here u_0 is the argument of latitude at the epoch of the orbital elements and Δl denotes the difference of the mean orbital longitudes of both spacecraft [17].

The use of relative orbital elements is preferable in the present context, since it allows the consideration of short periodic perturbations and can thus be used to plan longitude swap maneuvers with extended drift phases.

E/I-Vector Separation

The concept of e/i-vector separation has originally been developed for the safe collocation of geostationary (GEO) satellites [19] but can also be applied for proximity operations in LEO formations. It is based on the consideration that the uncertainty in predicting the along-track separation of two spacecraft is generally much higher than for the radial and cross-track component. Due to the coupling between semi-major axis and orbital period, small uncertainties in the initial position and velocity result in a corresponding drift error and thus a secularly growing along-track error. Predictions of the relative motion over extended periods of time are therefore particularly sensitive to both orbit determination errors and maneuver execution errors.

To avoid a collision between the two spacecraft due to along-track position uncertainties, the satellites have to be separated properly in radial and cross-track direction. This can be achieved by a parallel (or anti-parallel) alignment of the relative e- and i-vectors.

With the above introduced notation the angles φ and θ are equal for parallel vectors $\Delta\vec{e}$ and $\Delta\vec{i}$. For this case, the arguments of latitude $u = \varphi$ and $u = \varphi + \pi$ mark the orbital positions, at which the two spacecraft show their maximum radial separation. And in addition, these points describe also the line of intersection of both orbital planes, at which the cross-track separation vanishes. On the other side, $u = \varphi + \pi/2$ and $u = \varphi = 3\pi/2$ define the points, where the radial separation vanishes and the cross-track separation has a maximum.

In the case the vectors $\Delta\vec{e}$ and $\Delta\vec{i}$ are orthogonal, the radial and out-of-plane separation disappear together at two points of the orbit (Fig. 19.12; right). If the

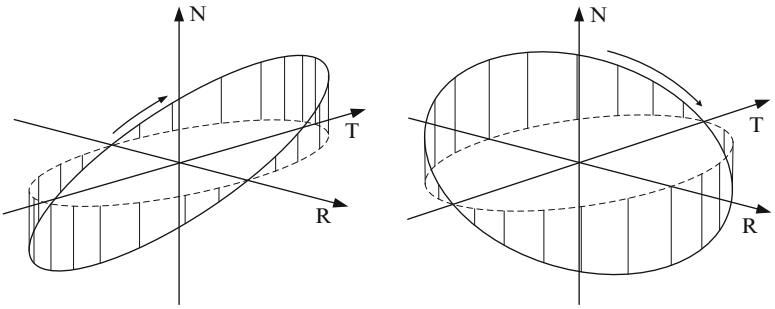


Fig. 19.12 Relative motion for parallel (*left*) and orthogonal (*right*) relative e- and i-vectors [17]

along-track separation at these points is smaller than the along-track position uncertainty of the satellites, the probability of a collision is relatively high.

In the case of no along-track drift (identical semi-major axes), the radial and cross-track separation are given by

$$\begin{pmatrix} \Delta r_{\mathrm{R}} \\ \Delta r_{\mathrm{N}} \end{pmatrix} = \begin{pmatrix} -a\delta e \cos(u-\varphi) \\ a\delta i \sin(u-\varphi) \end{pmatrix} \tag{19.12}$$

for parallel e-/i-vectors. Accordingly the distance between the two s/c is always than $\min(a\delta e, a\delta i)$. This is completely independent of the along-track separation uncertainty. The thresholds will be smaller for drifting satellites due to radial offset. But this can be compensated by a suitably increased eccentricity vector separation.

Orbital Perturbations

Eccentricity Vector

The oblateness of the Earth results in a variety of short periodic, long-periodic and secular perturbations of the orbital elements of a LEO satellite. According to Refs. [15, 19] the short periodic (sp) variations of the eccentricity vector depend primarily on the second order zonal coefficient $J_2 = -C_{20} = 1.082 \times 10^{-3}$ and vary periodically with the argument of latitude u.

An averaged eccentricity vector

$$\bar{\vec{\mathbf{e}}} = \vec{\mathbf{e}} - \delta\vec{\mathbf{e}}_{\mathrm{sp}} \tag{19.13}$$

can be obtained by removing the short-term perturbations. This resulting vector has only a long periodic rotation about the "frozen eccentricity" [10, 15] vector, given by

$$\vec{\mathbf{e}}_{\mathrm{G}} \approx \left(0, -\frac{1}{2}\frac{J_3}{J_2}\frac{R_\oplus^2}{a^2}\sin i\right) \tag{19.14}$$

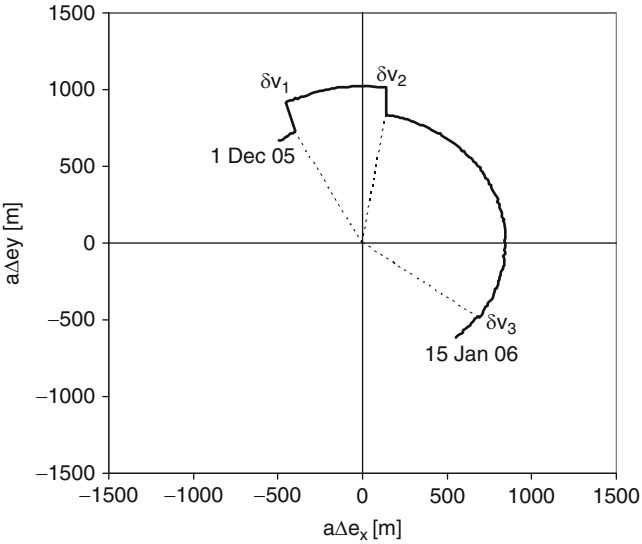


Fig. 19.13 Evolution of the mean relative eccentricity vector of the GRACE formation from 1 Dec 2005 to 15 Jan 2006

The rotation of $\vec{\mathbf{e}}$ about $\vec{\mathbf{e}}_{\mathrm{G}}$ has a period

$$T_{\mathrm{G}} = \frac{4}{3} T \frac{a^2}{R_{\oplus}^2} \frac{1}{J_2 |5\cos^2 i - 1|} \tag{19.15}$$

which is roughly a 1,000 times larger than the orbital period T [15].

In the discussion of proximity operations, it is sufficient to consider only the long-periodic eccentricity vector variations

$$\Delta\vec{\mathbf{e}} = \overline{\delta e} \cdot \begin{pmatrix} \cos\bar{\varphi} \\ \sin\bar{\varphi} \end{pmatrix}. \tag{19.16}$$

These describe a circle of radius $\overline{\delta \mathbf{e}}$ that is centered in the origin of the e-vector plane and traversed at an angular velocity of $\dot{\bar{\varphi}} = 2\pi / T_{\mathrm{G}}$.

Inclination Vector

The inclination and the right ascension of the ascending node experience short periodic perturbations of twice the orbital frequency [15, 19]. Using the simplification of (19.10), the short-term periodic perturbations vanish for polar orbits.

In case of the GRACE formation with an inclination of about 89°, the short periodic perturbations of the inclination vector (Fig. 19.14) are roughly two orders of magnitude smaller ($|\Delta\delta\vec{i}_{\mathrm{sp}}| \approx 10^{-6}$) than the corresponding variations of the

relative eccentricity vector. In the planning of proximity operations using the linearized relative motion equations (19.11), they can therefore be neglected.

GRACE Swap Maneuver Planning

Since the beginning of the GRACE mission, GRACE2 has been following GRACE1 at a distance of typically 220 km. Taking into account the current fuel balance of both satellites, it was decided to perform all required orbital maneuvers with the GRACE2 spacecraft. By lowering the semi-major axis of GRACE2, a positive along-track drift should be initiated and stopped again after passing GRACE1 and achieving the desired along-track separation. For the subsequent mission, GRACE2 would then be leading the formation, followed by GRACE1 in a distance of 220 km.

At the begin of the position exchange, the two GRACE satellites had a separation of about 170 km in early December 2005. As a basis for the swap maneuver operations, it was foreseen to initiate a positive along-track drift of roughly 30 km/day on GRACE 2 by lowering its semi-major axis. Within a week, GRACE2 would then pass by GRACE1 and subsequently lead the formation. The longitude swap was also deemed of interest for collecting additional science data. For small intersatellite distances the K-band radar measurements are particularly sensitive to high-frequency components of the gravity field. It was therefore decided to slow down the drift to roughly 3 km/day after reaching a separation of 70 km and to collect data outside the nominal operational window of 220 ± 50 km.

In a trade-off between fuel expenditure and loss-of science data, a total drift duration of 2 weeks has been adopted by the project. The required drift rate of 2 km/rev corresponds to a semi-major axis offset of 200 m and requires tangential a maneuver of 11 cm/s.

During the GRACE mission several orbit maintenance maneuvers were executed with an accuracy of better than 5%. This performance results in an uncertainty of about 8 h for the expected time of closest approach and makes a precise forecast of the relative along-track position impossible. In such a situation an adequate separation in cross-track and/or radial direction has to be ensured. One option would be to remove eccentricity vector difference to achieve a constant radial separation of both spacecrafts during the flyby. But with the planned Δv for the swap this was not feasible. An additional velocity increment of four times of the planned would have been required. Therefore the e/i vector separation was selected in terms of cost and operational complexity. With this strategy the number of maneuvers had been minimized and a maximum level of passive safety was reached.

During the swap period the relative inclination vector of the GRACE formation was located at a phase angle of 90°, which implied that the two orbital planes intersected close to the north and South Pole of the Earth. The relative inclination at that time resulted in a maximum cross-track separation of 360 m near the equator crossing and remained nearly unchanged for the formation swap. Therefore an appropriate orientation of the relative eccentricity vector had to be used for a

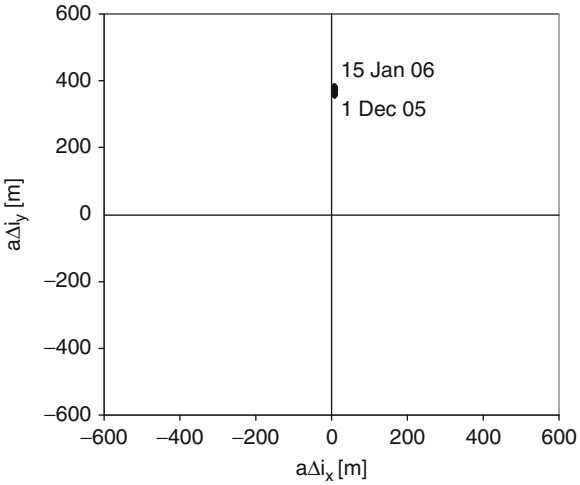


Fig. 19.14 Evolution of the mean relative inclination vector of the GRACE formation from 1 Dec 2005 to 15 Jan 2006

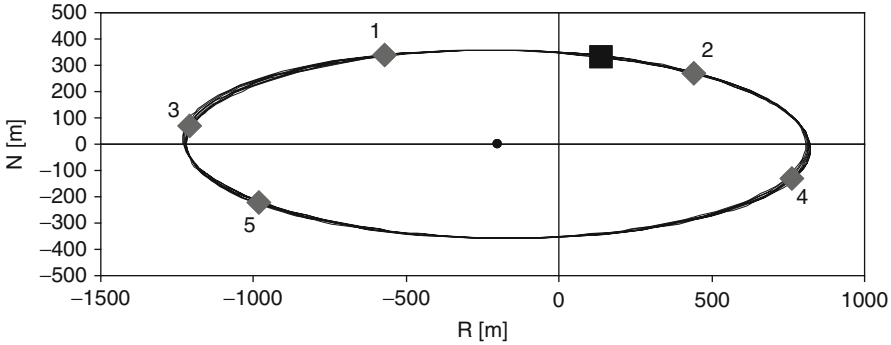


Fig. 19.15 Measured motion of GRACE2 relative to GRACE1 perpendicular to the flight direction near the time of closest approach (Dec 10, 2005, 0–12 h UTC). *Dark grey* diamonds mark the five instances of vanishing along-track separation and the *black square* marks the point of closest approach (431 m)

parallel e/i-vector configuration. Due to the rotation of the relative eccentricity vector with a period of 93 days as shown in (19.15) and Fig. 19.13, the parallel configuration was reached on Dec 10, 2005. This opportunity was selected for the closest approach leading to a drift start maneuver on Dec 3, 2005. As also shown in Fig. 19.13, the maneuver was used to increase the relative eccentricity in order to maximize the radial separation between the two s/c during the proximity phase.

As shown in Fig. 19.15, the relative motion of GRACE2 w.r.t. GRACE1 perpendicular to the flight direction is along an ellipse of the size $\pm$1,020 m $\times$ $\pm$360 m with an offset by $-$200 m in radial direction (marked by a black dot).

The linear drift of about 30 km/day caused by the semi-major axis offset and the superimposed periodic along-track variation of ±2 km caused by the relative eccentricity resulted in total of five epochs at which the along-track separation between the two GRACE satellites disappeared. But due to the e/i vector separation the separation perpendicular to the flight direction was always larger than 500 m, whereas the closest distance of 431 m between the satellites occurred on Dec 10, 2005, at 3:47 UTC.

The successful execution of the GRACE swap maneuver with a safe flyby demonstrated the potential of the eccentricity/inclination vector separation algorithm. As it was foreseen to use the e/i vector separation for upcoming close formation missions at GSOC like TerraSAR-X/TanDEM-X and PRISMA, an outlook on how the method is generalized and used for other applications can be found in Ref. [20].

References

1. Fowler W, Bettadpur S, Tapley B (2000) Mission planning for the twin GRACE satellites. AAS/AIAA space flight mechanics meeting, Paper AAS 00–164
2. Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: mission overview and early results. Geophys Res Lett 31:L09607. doi:10.1029/2004GL019920
3. Foerste Ch et al (2008) EIGEN-5C: a new global combined high-resolution GRACE-based gravity field model of the GFZ-GRGS cooperation. EGU2008-A-03426
4. Foerste Ch et al (2011) EIGEN-6: a new combined global gravity field model including GOCE data from the collaboration of GFZ-Potsdam and GRGS-Toulouse. EGU2011-3242
5. Velicogna I (2009) Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE. Geophys Res Lett 36:19503. doi:10.1029/2009GL040222
6. Kirschner M, Montenbruck O, Bettadpur S (2001) Flight dynamics aspects of the GRACE formation flying. In: Proceedings of second workshop on satellite constellations and formation flying
7. Seywald H, Kumar RR (1998) GRACE station-keeping strategy. Analytical mechanics associates, Inc. (AMA). Report No. 98–3, JPL Contract 961157
8. Mazanek DD, Kumar RR, Seywald H (2000) GRACE mission design: impact of uncertainties in disturbance environment and satellite force models. AAS/AIAA space flight mechanics meeting. Paper AAS 00–163
9. AMA and NASA (1998) Ballistic coefficient and station-keeping study for GRACE D. NASA LaRC, AMA Inc.
10. Vallado DA (2001) Fundamentals of astrodynamics and applications, 2nd edn. Space Technology Library. Kluwer Academic Publishers
11. Herman J (2009) General description of the TLEGEN software for 2-line elements generation. FDS-GEN-1030, Issue 1.5. DLR/GSOC
12. Montenbruck O (2000) An epoch state filter for use with analytical orbit models of low earth satellites. Aerosp Sci Technol 4:277–287
13. Hoots FR, Roehrich RL (1980) Models for propagation of NORAD element sets. Project Spacecraft Report No. 3. Aerospace Defense Command, United States Air Force
14. Clohessy WH, Wiltshire RS (1960) Terminal guidance system for satellite rendezvous. J Aerosp Sci 270:653

15. Micheau P (1995) Orbit control techniques for low earth orbiting (LEO) satellites. Chap. 13. In: Carrou JP (ed) Spaceflight dynamics. Cépaduès-Éditions, Toulouse
16. Montenbruck O, Gill E (2000) Satellite orbits—models, methods, and applications. Springer, Heidelberg
17. Montenbruck O, Kirschner M, D'Amico S, Bettadpur S (2006) E/i-vector separation for safe switching of the GRACE formation. Aerosp Sci Technol 10(7):628–635
18. Montenbruck O, Pfleger T (2000) Astronomy on the personal computer. Springer, Heidelberg
19. Eckstein MC, Hechler F (1970) A reliable derivation of the perturbations due to any zonal and Tesseral harmonics of the geopotential for near-by circular satellite orbits. ESRO SR-13-1970; ESA, Darmstadt
20. D'Amico S, Montenbruck O (2006) Proximity operations of formation flying spacecraft using an eccentricity/inclination vector separation. AIAA J Guid Contr Dyn 29(3):554–563
21. Stanton R, Bettadpur S, Dunn D, Renner K-P, Watkins M (1999) Science and mission requirements (SMRD) document. Revision B, GRACE Document No. 327–200, JPL D-15928

Chapter 20
Next Generation Gravity Mission

Stefano Cesare and Gianfranco Sechi

Abstract After the successful experience of the gravity missions GRACE and GOCE, several activities are on going in preparation of a "Next Generation Gravity Mission" (NGGM) aimed at measuring the temporal variations of the Earth's gravity field over a long time span (up to ~11 years) with high spatial resolution (comparable to that provided by GOCE) and high temporal resolution (weekly or better). Its data will find wide application in geodesy, geophysics, hydrology, ocean circulation and many other disciplines. The most appropriate measurement technique identified for such mission is Low-Low Satellite-Satellite Tracking in which two (or more) satellites flying in "loose" formation in a low Earth orbit act as proof masses immersed in the Earth gravity field. The distance variation between the satellites (measured by a laser interferometer) and the non-gravitational accelerations of each satellite (measured by ultra-sensitive accelerometers) are the fundamental observables from which the gravity field is obtained. Suitable satellite formations for this mission include the "In-line" (the simplest one), the "Cartwheel" and the "Pendulum" (more complex but also scientifically more fruitful), with an inter-satellite distance up to 100 km. Polar, circular orbits with altitudes between ~340 and ~420 km are suitable candidates for the NGGM, providing all-latitude coverage, short repeat cycles/sub-cycles and a still excellent gravity signal compatibly with a long lifetime. Each satellite shall be endowed with a complex control system capable of carrying out several tasks in close coordination: orbit maintenance, formation keeping, provision of a "drag-free" environment to the accelerometers, laser beam pointing and attitude control.

S. Cesare (✉) • G. Sechi
Thales Alenia Space Italia, Strada Antica di Collegno 253, 10146 Torino, Italy
e-mail: stefano.cesare@thalesaleniaspace.com; gianfranco.sechi@thalesaleniaspace.com

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_20,

20.1 Context and Background

The measurement of the Earth gravity field (or of related parameters) has been the objective of several space missions. For example, Lageos 1 (launched in 1976) and Lageos 2 (launched in 1992) allowed to obtain an accurate determination of the lowest coefficients (J_2, J_3, J_{22},..) of the expansion in spherical harmonics of the gravity field and of their time variation, through the measurement of their orbital trajectory by ground-to-space laser ranging.

CHAMP [1], launched on July 2000, mapped the Earth gravity field using the technique of the Satellite-Satellite Tracking (SST) in High-Low mode (HL-SST). The position of the satellite, which flew at an altitude $<$450 km, was accurately determined by means of the GPS network and processed for separating the orbital gravitational perturbations from the non-gravitational ones, the latter estimated from the measurements of an on-board accelerometer.

GRACE [2], launched on March 17, 2002, proved the capability of detecting the temporal variations of the Earth gravity field by exploiting the technique of Satellite-Satellite Tracking (SST) in Low-Low mode (LL-SST). The relative distance between two satellites, flying in a pair along the same low orbit, is accurately measured by means of a K-band Ranging system. Again, the gravitational perturbations on the relative distance are separated from the non-gravitational ones utilising the measurements of an accelerometer embarked on each satellite.

GOCE [3] (Gravity Field and Steady-State Ocean Circulation Explorer), the first Earth Explorer Core Mission in ESA's Living Planet programme, launched on March 17, 2009 has provided the most accurate map of the static Earth gravity field using the technique of the gradiometry. Three pairs of accelerometers arranged along orthogonal axes (gradiometer) measure the differences of the gravitational accelerations between the points in which they are located, so providing the five independent components of the gravity gradient (second spatial derivatives of the gravitational potential).

Since 2003 the European Space Agency (ESA) has promoted studies to establish the scientific requirements, to identify the most appropriate measurement techniques, to start the associated technology developments, and to define the system scenarios for a "Next Generation Gravity Mission" (NGGM). This mission will make use of Low-Low Satellite-Satellite Tracking between satellite pairs flying in loose formation on a low Earth orbit, with more advanced measurement systems. The key requirements and issues related to the control of the satellite orbit, formation and attitude are addressed in the following chapters.

20.2 Mission Objectives and Measurement Technique

In one of the preparatory studies for the NGGM [4] carried out by Thales Alenia Space Italia (TAS-I) for the European Space Agency, with the contribution of some of the most prominent European institutes of geophysics and geodesy, a list of

Table 20.1 Fields of prioritisation for the NGGM with their spatial and temporal resolution and approximate signal magnitudes

	Description	Spatial resolution	Temporal resolution	Signal magnitude in geoid heights
1	Melting of ice sheets (with separation of glacial isostatic adjustment)	100–1,000 km	Seasonal Secular	0.01–1 mm/year (secular)
2	Non-steric component of sea-level variations at seasonal and shorter time scales	Global to basin level	Inter-annual Secular	0.1 mm/year (secular)
3	Ground water (soil moisture and snow) at larger spatial scales	10–200 km	Hourly Seasonal Secular	0.05–1 cm (seasonal)
4	Post-seismic deformation	10–200 km	Sub-seasonal	1 mm (sub-seasonal)

Table 20.2 Requirements in terms of maximum CGE for monthly gravity field retrievals

	Resolution	10000 km	1000 km	200 km	100 km	10 km
	Spherical harmonic degree	$\ell = 2$	$\ell = 20$	$\ell = 100$	$\ell = 200$	$\ell = 2000$
CGE	10 mm			3		
	1 mm	2		4		
	0.1 mm		1			

The numbers in the *grey boxes* correspond to the first column of Table 20.1. The *shaded* areas on the right mark the restriction to more realistic values in case of a monthly satellite gravity field

scientific objectives was identified for this mission. They are provided in Table 20.1 ranked in order of priority. The associated spatial resolution and maximum Cumulative Geoid[1] Error (CGE) requirements are in Table 20.2. To satisfy the requirements of the geophysical applications (solid Earth science, glaciology, hydrology, oceanography, atmosphere circulation, etc.), the NGGM must provide the *temporal variations* of the Earth's gravity field over a long time span (possibly covering a complete solar cycle: ~11 years) with *high spatial resolution* (comparable to that provided by GOCE) and *high temporal resolution* (weekly or better), so as to reduce the level of aliasing of the high frequency phenomena found in the time series of the Earth's gravity field variation provided by GRACE, and to improve the separability of the observed geophysical signals [5]. Moreover, the NGGM shall be able to provide the following CGEs from monthly solutions of the Earth gravity field:

[1] The Geoid is the equipotential surface of the Earth gravity field that best fits (in least squares sense) the surface of the oceans and the seas at rest (i.e. assuming no currents, waves, etc.).

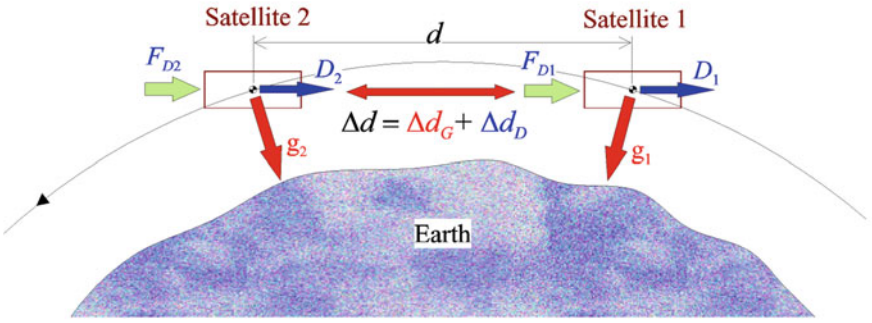


Fig. 20.1 Principle of the LL-SST technique for measuring the Earth gravity field

- CGE $\leq$ 0.1 mm at degree $\ell = 150$ (spatial resolution = 133 km),
- CGE $\leq$ 1 mm at degree $\ell = 200$ (spatial resolution = 100 km),
- CGE $\leq$ 10 mm at degree $\ell = 250$ (spatial resolution = 80 km).

From the wavelengths of the geopotential spherical harmonics to be monitored, the measurement bandwidth (MBW) of the NGGM has been preliminarily established between 1 and 100 mHz (corresponding to a spatial sampling along the orbit between ~77 and ~7,700 km).

The LL-SST is in practice the only observing technique potentially capable of detecting the time variable gravity signal with the required resolution [4, 6]. This technique exploits the satellites themselves as the "proof masses" immersed in the Earth gravity field (Fig. 20.1). The fundamental observable is the distance variation between two satellites centres of mass (COMs) produced by the gravity acceleration, Δd_G, formally obtained as:

$$\Delta d_G = \Delta d - \Delta d_D \tag{20.1}$$

- Δd = total distance variation between the COMs, whatever the source, measured by a distance metrology,
- Δd_D = distance variation produced by non-gravitational (i.e. drag) forces on the satellite COMs along the line joining the COMs themselves:

$$\Delta d_D = \iint \Delta \ddot{d}_D \, dt^2, \quad \Delta \ddot{d}_D = D_1 - D_2 \tag{20.2}$$

- D_1, D_2 = non-gravitational accelerations of the satellite COMs, measured by accelerometer(s).

The measurement instrument formed by the satellite pair can be regarded as a kind of one-dimensional gradiometer with a very long baseline. Thanks to the separation between the satellites it has a higher sensitivity for the phenomena of interest than a gradiometer embarked on a single satellite, especially at altitudes higher than GOCE's one, necessary for a long duration mission lifetime in order to

compensate the drag forces with an affordable amount of propellant. In fact, the gravitational potential V of degree ℓ rapidly decreases with the orbit radius r:

$$V_\ell \propto r^{-(\ell+1)} \tag{20.3}$$

and therefore the instrument baseline must be proportionally increased in order to maintain the same signal-to-noise ratio.

20.3 Mission Outline and Satellite Formations

The mission scenarios that have been devised by TAS-I [4] for the NGGM feature one or two pairs of satellites. In all cases the satellite orbits are circular or nearly circular so that the signal of the Earth gravity field is nearly constant at the satellite altitude.

In the minimal scenario featuring one single satellite pair, the preferred orbit inclination is 90° (polar orbit), in order to avoid coverage gaps around the poles, which are undesirable for ice mass studies. In the scenario featuring two satellite pairs, the inclinations providing the best spatial and temporal sampling of the Earth gravity field are 90° and 63° (Bender architecture [7]).

The mean orbit altitude of the satellite pairs shall be as low as possible (for maximising the gravitational signal) compatibly with a lifetime of ~11 years (too low an orbit would lead to an excessive amount of propellant for the atmospheric drag compensation) and a repeat cycle of about 1 month with a sub-cycle of about 1 week [4] to get a fast and homogeneous coverage of the globe. For instance, polar orbits with mean altitudes of 340 km (32/503 repeat with 7 and 4 day sub-cycles) and 424 km (30/463 repeat with 7 day sub-cycle) are both characterised by such repeat cycles/sub-cycles and are potential candidates for the NGGM.

The formations, each constituted by a pair of satellites, can have typically three configurations:

1. *In-line (or Pearl String) formation* (Fig. 20.2). This is the simplest formation type (adopted by GRACE) in which the satellites follow the same orbital path, keeping a relative distance nearly constant. The main limitation of the in-line formation is its strongly anisotropic sensitivity. Essentially, the inter-satellite ranging data acquired by this mission contains information about gravity differences between the satellite locations. In an in-line formation on a polar orbit (preferred inclination for single satellite pair) the satellites are located for most of the time at nearly the same meridian. Consequently, the observations describe North–south variations of the gravitational field (and mass transport) much better than East–west variations. This anisotropic signal structure leads to the well-known North–south striations in the gravity field solutions obtained from the GRACE measurements. A significant gain in sensitivity and isotropy can be achieved by adding to the polar in-line formation another in-line formation at 63° inclination (Bender architecture).

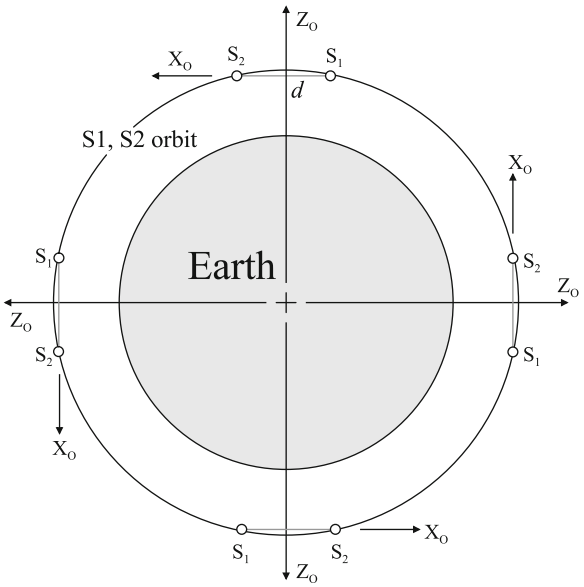


Fig. 20.2 In-line formation geometry (S_1, S_2 = satellite 1, 2; X_O, Y_O, Z_O = LORF axes)

2. *Pendulum formation* (Fig. 20.3). This type of formation is obtained by collocating the two satellites in two circular orbits with either slightly different inclinations or slightly different right ascension of the ascending nodes (RAANs). As a result, the satellite-to-satellite line performs a pendulum-like motion in the X-Y plane of the mean Local Orbital Reference Frame (LORF, X = local horizontal, Z = local vertical, Y = local normal to the orbit plane) synchronised with the orbital period. The angular amplitude of the oscillations is determined by the difference in inclination or RAAN. The pendulum formation mainly captures cross-track and along-track gravity signals and provides higher sensitivity and isotropy in the gravity field solutions. For an aperture angle of 45°, a single satellite pair in polar orbit provides a scientific return comparable to the Bender architecture consisting of two In-line formations.
3. *Cartwheel formation* (Fig. 20.4). This type of formation is obtained by collocating the two satellites on orbits with the same plane, same period and equal (small) eccentricity, with line of apsides coincident and perigee/apogee on opposite sides. As a result, the satellite-to-satellite line oscillates twice per orbit around the direction of the line of apsides (which maintains approximately the same orientation in an Inertial Reference Frame) with amplitude of ~ ±20°. The two satellites appear to revolve around each other on elliptical trajectories (with 2:1 ratio of the semi-major axes and period = orbital period) in a mean LORF of a circular orbit having the same period as those of the satellites. The Cartwheel formation mainly captures radial and along-track gravity signals. It provides a scientific return similar to the Pendulum but leads to the most complex implementation problems.

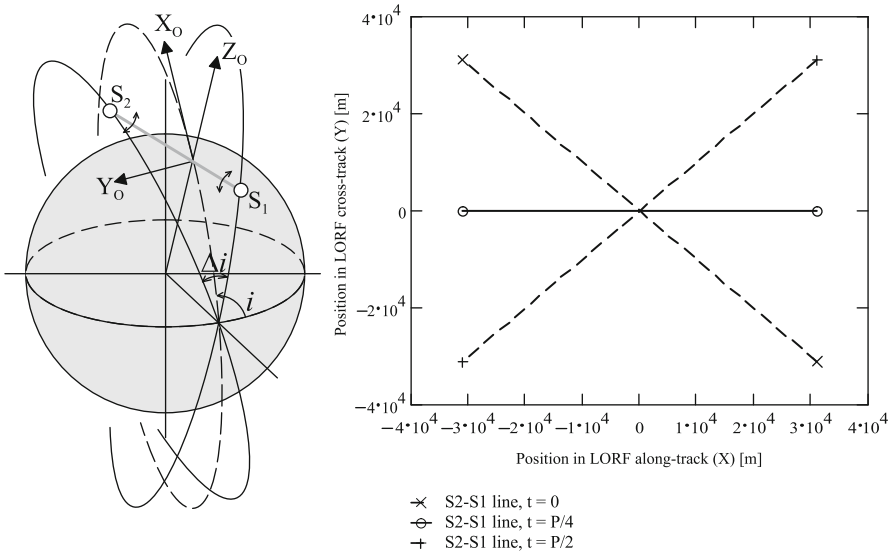


Fig. 20.3 Example of pendulum formation geometry (obtained with inclination separation) and example of the pendulum-like motion of the satellite-to-satellite line with ±45° amplitude in the mean LORF

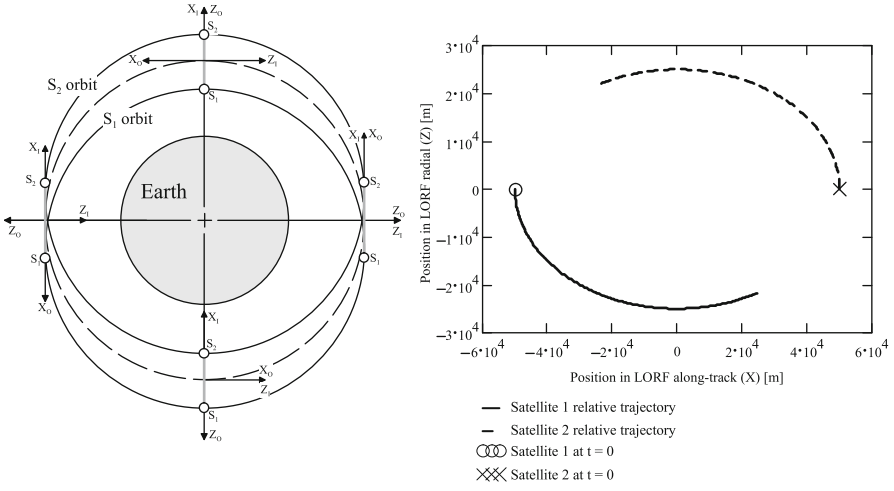


Fig. 20.4 Cartwheel formation geometry and example of satellite relative trajectories in the mean LORF (X_I, Y_I, Z_I = Inertial reference frame axes)

The cumulative geoid error achievable from 1 week and 1 month of measurements of the gravity field using the In-line (alone and in Bender architecture), Cartwheel and Pendulum satellite formations is shown in Fig. 20.5 as function of the spherical harmonic degree. It was obtained from a detailed numerical simulation [4], considering the following formation parameters:

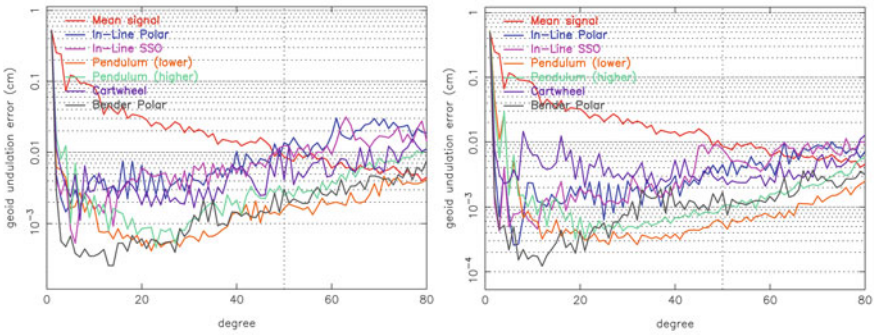


Fig. 20.5 Cumulative geoid error for one week (*left*) and one month (*right*) of measurements respectively

- Mean orbit altitude = 340 km for all the formations with the exception of the Pendulum one, for which a higher altitude option (424 km) was assessed too.
- Orbit inclination = polar for all the formations with addition of a 63° orbit for the Bender architecture (with In-line formations), and the assessment of the sun-synchronous inclination for the In-line formation alone.
- Mean separation between the satellites = 75 km (in the Cartwheel formation the distance varies between 50 and 100 km; in the Pendulum formation the distance varies between 62 and 88 km and the maximum aperture angle is ±45°).

For a given measurement accuracy of the fundamental observables, the In-line Bender architecture displays a good performance at the lowest spherical harmonic degrees even after a single week of measurements (which can be explained by the much better and denser global coverage provide by the two pairs), but is even outperformed by the pendulum formation for degrees larger than about 20. The Cartwheel and Pendulum formations (the latter also in the higher altitude version) provide better performance than the In-line formation over measurements periods of 1 month.

However, although the Cartwheel and Pendulum formations are advantageous for gravity field determination, they impose high demands on satellite/instrument design, orbit/attitude control (including higher fuel consumption). Unlike the In-line formation where the satellites keeps a fixed attitude in the LORF (so it can be oriented with the minimum cross section towards the orbital velocity and the drag force can be minimised), in the Cartwheel and Pendulum the orbital velocity direction turns around the satellite and couples with a variable cross section giving rise to a higher drag force with wider dynamic range.

Moreover, in the In-line formation one satellite keeps a nearly fixed position in the reference frame of the other satellite, and can be easily tracked by the laser metrology. In a Cartwheel formation with 100 km/50 km semi-major axes, the relative distance changes with a variation rate up to ~57 m/s and the line joining the satellites oscillated by ~±20°. In a Pendulum formation with 75 km

mean distance and ±45° oscillation angle of the line joining the satellites, the relative distance changes with a variation rate up to ~30 m/s rate. These variations of the relative distance and angular position impose specific design requirements to the satellite-to-satellite ranging system (which has to cope with a significant Doppler shift) and to the attitude control (which must keep the laser beam pointed towards a moving target). The Cartwheel formation is also affected by a number of problems arising from a non Earth-referenced attitude: the main component of the drag force turns once per orbit around the satellite. Moreover, the inter-satellite line becomes aligned to the satellite-Earth direction twice per orbit, with potential interferences (optical and thermal) between the Earth radiation and the laser metrology.

In conclusion, the In-line formation is the easiest to implement (the scientific limitations of a single In-line pair can be recovered by means of a Bender architecture). The Pendulum is more complex but still viable if the maximum oscillation amplitude is limited to 10°–15° (even with such small angle the scientific return is superior to the In-line case). The Cartwheel appears at the limit of feasibility.

20.4 Payload Requirements and Design Outline

Achieving the scientific objectives of the NGGM described in Sect. 20.2 requires the capability of measuring the satellite-to-satellite distance variation Δd with the error spectral density not exceeding the limit displayed in Fig. 20.6 [4]. Such performance requirement implies replacing the K-band Ranging system of

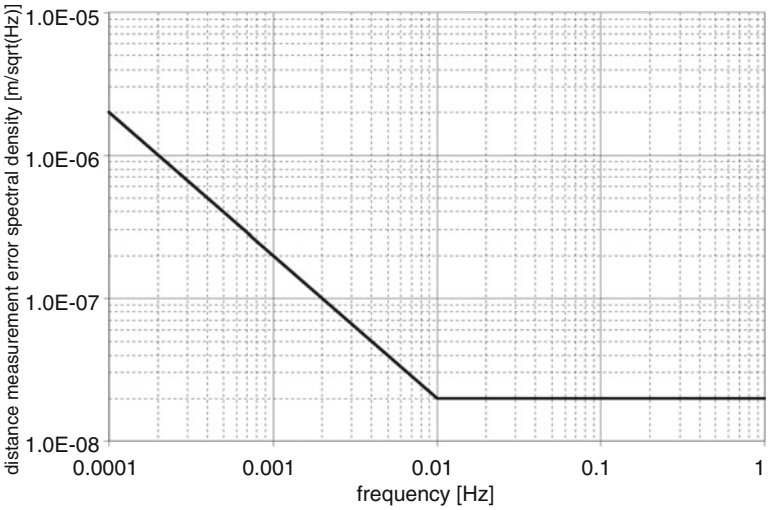


Fig. 20.6 Upper limit to the measurement error spectral density of the satellite-to-satellite distance (Δd)

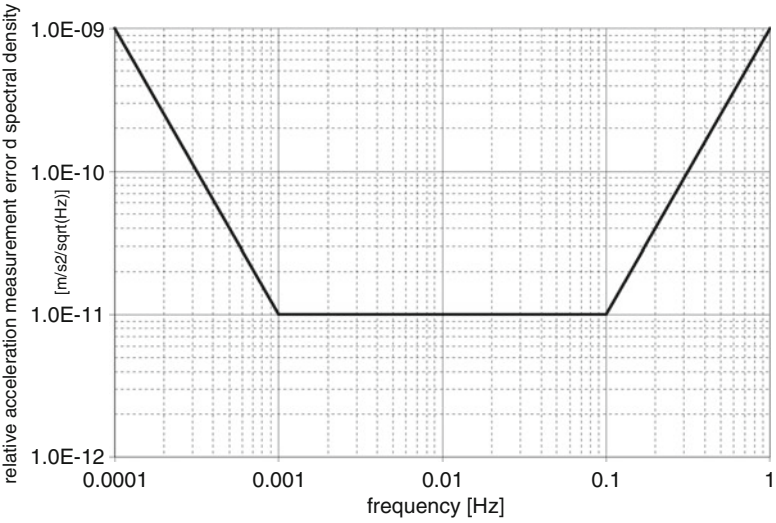


Fig. 20.7 Upper limit to the measurement error spectral density of the non-gravitational relative acceleration ($\Delta\ddot{d}_D$)

GRACE with a laser interferometer. Moreover, the second fundamental observable, i.e. the non-gravitational relative acceleration of the satellite COMs ($\Delta\ddot{d}_D$), must be measured with an error spectral density not exceeding the limit displayed in Fig. 20.7, defined consistently with the requirement on Δd [4]. Such requirement is within the reach of the accelerometers utilised in the GOCE mission [8].

The whole set of instruments concurring to the gravity field measurement in the NGGM is schematised in Fig. 20.8 and includes:

- The laser interferometer, which measures the distance variation between two retro-reflectors installed ideally in the COMs of the two satellites.
- A set of accelerometers, which measures the non-gravitational acceleration of the COM of each satellite.
- An auxiliary angle metrology, which provides the satellite orientation relative to the satellite-to-satellite line (information needed for reporting the interferometer measurements from the retro-reflector position to the COM position, and for projecting the acceleration measurements along the satellite-to-satellite line).
- An auxiliary lateral metrology, which provides the lateral displacements of Satellite 2 (S_2) relative to the laser beam emitted by Satellite 1 (S_1) and feeds the beam pointing controller via a radio-frequency inter-satellite link.
- A Global Navigation Satellite System (GNSS) receiver, which provides the precise orbit determination and the relative position of the satellites (again the GNSS measurements are exchanged between S_1, S_2 via the inter-satellite link).
- Star sensors, which provide the inertial attitude of the satellites.
- A laser retro-reflector, for the precise orbit determination through laser ranging from ground (as complement to that provided by the GNSS receiver).

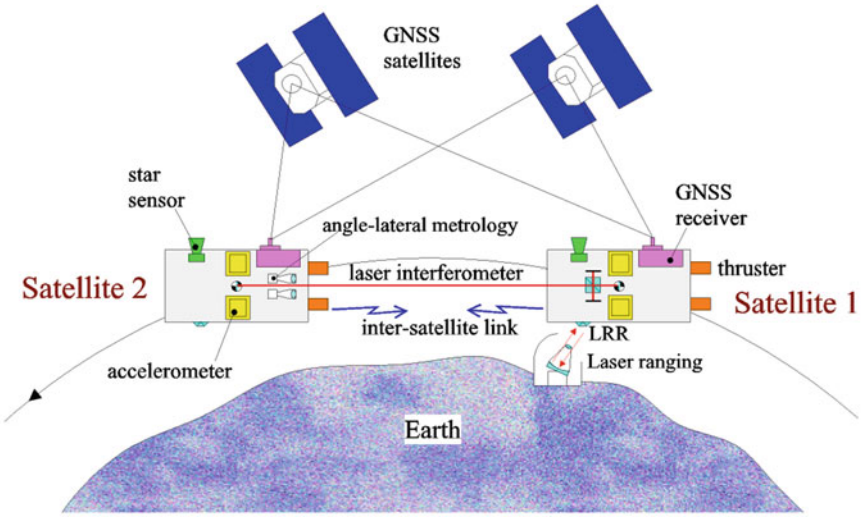


Fig. 20.8 Instruments involved in the measurement of the LL-SST fundamental observables

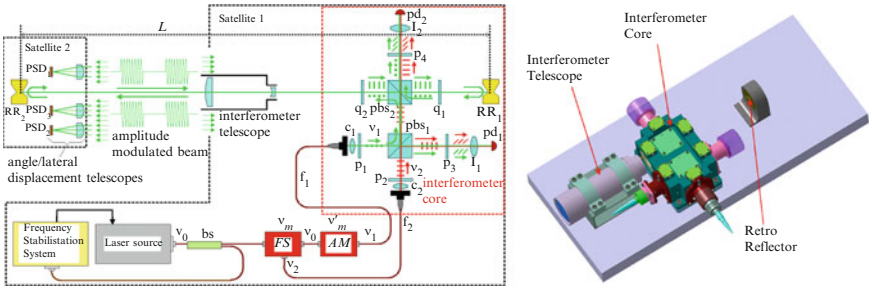


Fig. 20.9 Interferometer functional scheme and configuration

The core of the satellite-satellite distance variation measurement system is a Michelson-type heterodyne laser interferometer based on the retro-reflector scheme (laser transmitted by S_1 and back-reflected by S_2) and adapted for long-distance operation through an on/off amplitude modulation scheme (Fig. 20.9). Alternatively, a transponder scheme can be adopted in which the laser beam transmitted by S_1 and received by S_2 is "regenerated" by a second laser source locked in phase to the incoming beam before being retransmitted towards S_1.

The angle metrology consists of three small telescoped endowed with Position Sensing Detectors and arranged on S_2 in triangular configuration. Each telescope picks up a fraction of the laser beam transmitted by S_1 and measures its incoming direction that depends on the S_2 orientation relative to the satellite-to-satellite line (materialised by the laser beam). The same three-telescope assembly measures the optical power of the intercepted portion of the incoming beam: the lateral

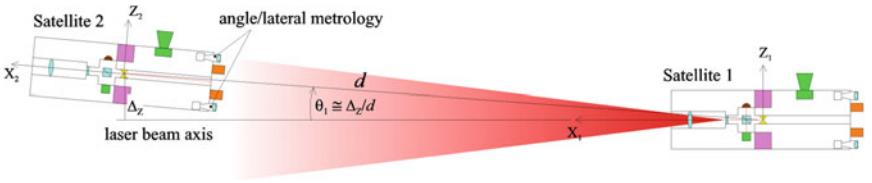


Fig. 20.10 Measurement principle of S_1 orientation (θ_1, ϕ_1) relative to the satellite-to-satellite line from the lateral displacement (Δ_Z, Δ_Y) of S_2 relative to laser beam axis

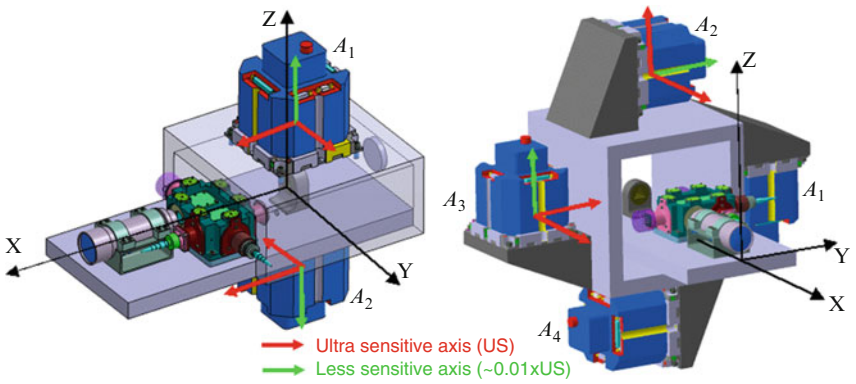


Fig. 20.11 Arrangement of two accelerometers (*left*) and of four accelerometers (*right*). The X axis is nominally aligned to the satellite-to-satellite line

displacement of S_2 from the axis of the laser beam is obtained from the unbalance of the optical powers. As the laser beam transmitted by S_1 is modulated in amplitude, by applying a synchronous demodulation scheme to the output of each detector it is possible to get rid of spurious signals caused by straylight. The S_1 orientation relative to the satellite-to-satellite line is obtained from the lateral displacement of S_2 relative to laser beam axis and from the knowledge of the absolute distance between the satellites (measured by the differential GNSS), as shown in Fig. 20.10.

The laser interferometer and the angle-lateral metrology device have been developed by TAS-I in cooperation with INRIM, breadboarded and subjected to laboratory tests that proved their potential capability to meet the NGGM performance requirements [9].

Accelerometers like those utilised on GOCE can measure the non-gravitational acceleration of the satellite COM with the required precision. A minimum set of two accelerometers arranged like in Fig. 20.11 enables to locate the acceleration measurement point (mid point of the accelerometer centres) in the COM while it is physically occupied by the laser retro-reflector. Moreover, besides the three linear accelerations of the COM, it provides the three angular accelerations of the satellite

with the precision necessary for feeding the control of the angular dynamics (see next chapter), either by combination of the linear accelerations measured by the two accelerometers or from the angular acceleration of the proof mass measured at accelerometer level. By increasing the number of accelerometers from two to four, arranged like in Fig. 20.11, the additional measurement of two components of the gravity gradient (along two axes orthogonal to the satellite-to-satellite line) can be also obtained, besides a more precise measurements of the linear and angular accelerations.

The laser interferometer core, its telescope, its retro-reflectors (two hollow corner-cubes mounted back to back with coincident vertices) and the accelerometers are arranged in a common assembly sharing the same stable mechanical support and the same controlled thermal environment.

The same instrument assembly is replicated on both satellites. By so doing, if a failure occurs on S_1 preventing the interferometer operation, the mission can continue after having exchanged the position of the two satellites (i.e. S_2 becomes the active satellite).

20.5 Preliminary Satellite Configuration

The preliminary configurations of the satellites defined by TAS-I along the NGGN preparatory studies [4] are shown in Fig. 20.12. The configurations were defined so that two satellites can be launched together by a small launcher like Vega, Rockot, Dnepr (two identical satellites back-to-back fill up the fairing).

Each satellite is a slender prism with trapezoidal cross section with area of about 1.3 m. For the In-line formation, this is also the area projected in the direction of motion, so that the atmospheric drag force is minimised. For the other formation geometries also the satellite lateral sides are periodically projected in the direction of motion and contribute to the drag cross section. The internal equipment layout is such as to place the satellite COM approximately in the geometrical centre, where

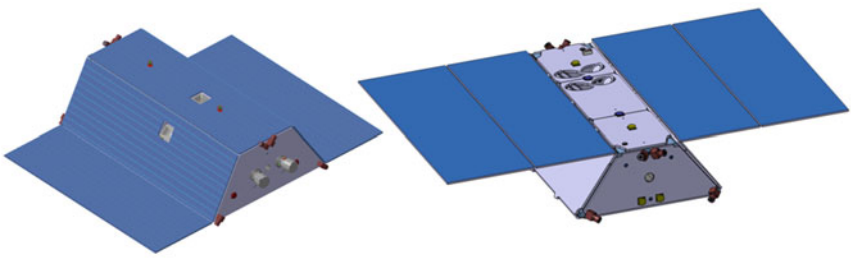


Fig. 20.12 Preliminary configurations of the NGGM satellite: with solar cells partly installed on the satellite body (*left*) and installed on deployable wings only (*right*)

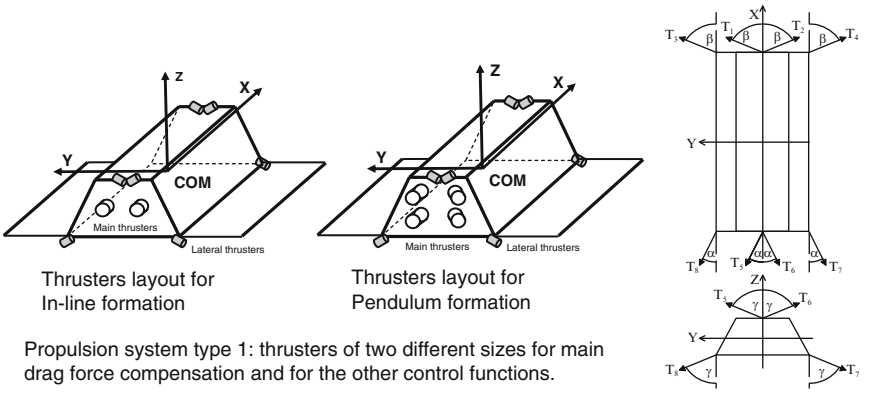


Fig. 20.13 Propulsion system typologies and thrusters layout identified for the NGGM

the payload bay is located. A slightly flared tube crosses the spacecraft from end to end to provide an unobstructed path to the laser beam.

The solar cells can be placed partly on the satellite body and partly on deployable wings, or on deployable wings only. Since the reference orbit of the NGGM is not sun-synchronous, it is necessary to track the seasonal variation of the sun aspect angle either by changing the orientation of the solar panels or that of the whole spacecraft: a rotation about the X axis (roll) is sufficient for maintaining a good illumination of the solar panels along the year. The latter solution is preferred since it avoids the presence of moving parts which displace the COM and disturb the micro-vibration environment of the satellite.

All the satellite control functions are accomplished by means of electrical propulsion (selected for the low propellant consumption, as needed for a long permanence in low orbit), partially complemented by three magnetic torquers which supply torques around the satellite axes not aligned with the local direction of the Earth magnetic field. Actuators based on moving parts, like reaction wheels, have been excluded since their mechanical noise is incompatible with the micro-vibration requirements imposed by the accelerometers. Two typologies of propulsion system have been identified (see Fig. 20.13). The first one, more suitable to the lower orbit altitude (where the atmospheric drag force is larger) is characterised by the use of thrusters of two different sizes: main thrusters (two for the In-line formation, four for the Pendulum formation), placed on the rear side of the satellite and oriented towards the COM for the compensation of the main component of the drag along the X axis, and eight smaller lateral thrusters in the Y-Z plane (four on the front side and four on the rear side) for the compensation of the lateral components of the drag and for the satellite pointing. The second type of propulsion system, more suitable for the higher orbit altitude, makes use of eight thrusters with

the same size of the lateral ones, arranged on the front and the rear sides of the satellite and oriented so to provide forces and torques along and around each axis. The electric thrusters must be capable to supply a force with variable amplitude, similarly to those utilised on GOCE, in order to precisely match the external disturbances to be compensated so as to ensure a very quiet environment to the accelerometers. Potential candidates for the NGGM, for their size and performance, have been identified in the mini radio-frequency ion thrusters (Mini-RIT) developed by the University of Giessen [10].

20.6 Dynamics Control System for the NGGM

20.6.1 Control Requirements

Each satellite shall be endowed with a control system capable to carry out the following tasks:

- Maintenance of the orbit altitude for each satellite, compensating the decay caused by the atmospheric drag force (the dominant non-gravitational force along the satellite orbit).
- Satellite formation acquisition and keeping along the mission, compensating the differential orbital perturbations on the two satellites.
- Provision of a "drag-free" environment to the accelerometers, consisting in attenuating the non-gravitational linear accelerations of the COM in absolute value and in spectral density (in a given spectral domain) below the limits compatible with the accelerometer dynamic range and with the measurement performance when coupled to the gradiometer imperfections (scale factors, sensitive axes misalignments and couplings) and accelerometer non-linearity remaining after the in-flight calibration.
- Acquisition and maintenance of the fine pointing of the laser beam emitted by S_1 towards S_2.
- Control the satellite attitude, and its angular accelerations and rates below the limits established by the measurement performance of the satellite-to-satellite distance variation (Fig. 20.6) and of the non-gravitational accelerations (Fig. 20.7).

The orbit altitude shall be controlled with the tolerance required to maintain the established repeat cycle of the ground tracks as necessary for ensuring a homogeneous coverage patters across the mission. In particular, the maximum separation between the ground tracks of a repeat cycle and those of the successive one shall be maintained within a fraction of the smallest spatial resolution to be achieved (see Sect. 20.2). A mean altitude control within a range of few tens of metres, similar to that applied to GOCE, is compatible with the NGGM needs.

Table 20.3 In-line and Pendulum formation control requirements

Parameter	Requirement	Remark
Maximum distance between the satellites	$\leq$100 km	Set by the maximum working range of the metrology
Admissible variation of the maximum distance	0 –10%	Applicable to both In-line and Pendulum formations
Admissible variation of the maximum aperture angle of the pendulum	$\pm 5^\circ$	Combined with the requirement on the maximum distance this constrains also the minimum distance of the Pendulum
Satellite relative displacement in radial direction	$\leq$1% satellite distance	Applicable to both In-line and Pendulum formations; corresponding to a ~0.6° angular displacement of the satellite-satellite line
Satellite relative displacement in cross-track direction	$\leq$1% satellite distance	Applicable to In-line only. The cross track excursion of the pendulum is constrained by the angle and the maximum distance

No stringent requirements apply to the two-satellite formation control if compared with other formation flying missions. In fact, since in the LL-SST technique the satellites themselves act as proof masses, they should be nominally free to move under the action of the gravity field. In practice, the natural orbital perturbations (of gravitational and non-gravitational nature) if not controlled drive the satellites away from the initial formation geometry with consequences on the gravity field sampling (e.g. due to the modification of the satellite-to-satellite baseline or of the maximum aperture angle of the pendulum) and of the measurements (e.g. the satellite-to-satellite distance can exceed the working range of the laser interferometer). Thus, the satellite formation must be kept bounded around the initial parameters through a "mild" control action that shall not interfere with the scientific measurements ("loose formation" concept). In practice this means that the formation controller shall operate with a bandwidth below the lowest frequency of the NGGM MBW ($<$1 mHz) and shall avoid as far as possible the introduction of non-gravitational linear accelerations occurring systematically at the same frequencies (e.g. the orbit frequency and multiples), which show up in the acceleration measurements as "harmonic" errors superimposed to the random ones. Preliminary requirements established for the In-line and Pendulum formation control are summarised in Table 20.3.

The drag-free control applies to each satellite of the formation; its objective is to reduce the non-gravitational accelerations of the satellite COM below 10^{-6} m/s^2 (each axis) and its spectral density in the MBW below 10^{-8} m/s^2/$\sqrt{\text{Hz}}$ (for the payload composed by four accelerometers) or below 5×10^{-9} m/s^2/$\sqrt{\text{Hz}}$ (for the payload composed by two accelerometers), as shown in Fig. 20.14. Similarly to the non-gravitational linear accelerations, also the angular accelerations and the

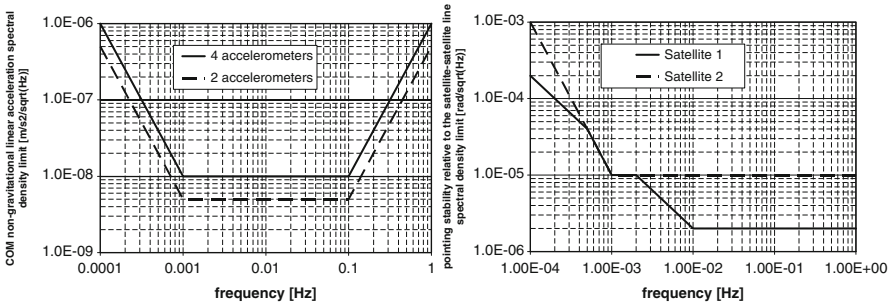


Fig. 20.14 Upper limits on COM non-gravitational linear acceleration spectral density (*left*) and on the satellite pointing stability relative to the satellite-satellite line (*right*)

angular rates of the spacecraft around the COM must be constrained in maximum value ($\leq 10^{-6}$ rad/s^2 each axis, $\leq 10^{-4}$ rad/s roll and yaw, $\leq 1.2 \times 10^{-3}$ rad/s pitch = rate for Earth pointing attitude) and in spectral density in the MBW ($\leq 10^{-8}$ rad/s^2/$\sqrt{\text{Hz}}$ and $\leq 10^{-6}$ rad/s/$\sqrt{\text{Hz}}$ around each axis) in order to not impact the overall performance in the measurement of the non-gravitational acceleration.

The attitude control of S_1 is in charge of pointing the laser beam towards S_2 with a precision $\leq 2 \times 10^{-5}$ rad (defined by the optical link budget of the interferometer) and with a stability $\leq 2 \times 10^{-6}$ rad/$\sqrt{\text{Hz}}$ down to a frequency of 10 mHz (limit established by the error on the distance measured by the laser interferometer produced by the coupling of the laser beam angular jitter with the far-field wavefront distortions). The alignment precision of S_2 with respect to the laser beam can be relaxed up to 1° thanks to the use of the retro-reflector. The pointing stability can be also relaxed to 10^{-5} rad/$\sqrt{\text{Hz}}$ in the MBW (see Fig. 20.14).

20.6.2 Control System Preliminary Design and Performance

The main difficulty in the formation controller design is that the natural control action opposes the relative satellites motion that is the main observable for the Earth gravity field reconstitution. Moreover, the formation controller shall operate in synergy with the drag-free control and shall minimise the additional thrusters force and the propellant consumption. To avoid wasting propellant for compensating the satellite relative displacements produced by the differential bias and drift between the accelerometers which on each satellite feed the drag-free control, these bias/drifts shall be estimated and removed from the command. On the other hand, to prevent that the drag-free control from cancelling out the formation control accelerations, the latter ones must be used as "reference" in the drag-free loops. Of course, formation control accelerations must fulfil the drag-free requirements too.

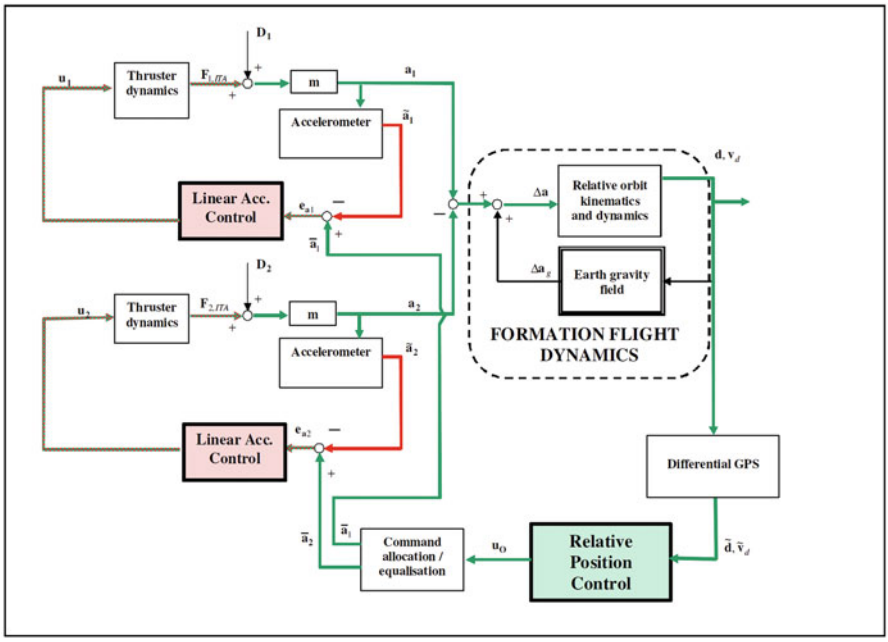


Fig. 20.15 Formation and drag-free control architecture

The architecture of the control system designed to fulfil the performance requirements and the above conditions is shown in Fig. 20.15. Each satellite has three independent Linear Acceleration Controls (one for each axis) and the Relative Position Control. A hierarchical control structure is implemented: the acceleration commanded by the Relative Position Control is the reference acceleration profile which is tracked by the Linear Acceleration Control. This is necessary to avoid that the Linear Acceleration Control cancels out the acceleration commanded by the Relative Position Control. In any case, the reference acceleration for the Linear Acceleration Control and its acceleration error shall meet the requirement on the residual non-gravitational linear acceleration (magnitude and spectral density). This means that the formation control action shall be limited in magnitude and shall be applied as much as possible at very low frequency.

At each control step (0.1 s), any drag-free control receives in input the acceleration commanded by the formation control, and the acceleration measurement; then, it computes the commanded forces to be applied by the thrusters assembly. The linear acceleration control is realised as a linear acceleration observer, based on a discrete-time model of the disturbances to be attenuated and of the delay due to accelerometers and to thruster dynamics. By the observer, one-step-ahead prediction of the disturbance acceleration is computed. Figure 20.16 shows the sensitivity function of the Linear Acceleration Control on X and Y axes (cross-over frequency about 0.3 Hz). This control design adopts the same approach as was followed for the GOCE Drag-Free and Attitude Control System [11, 12].

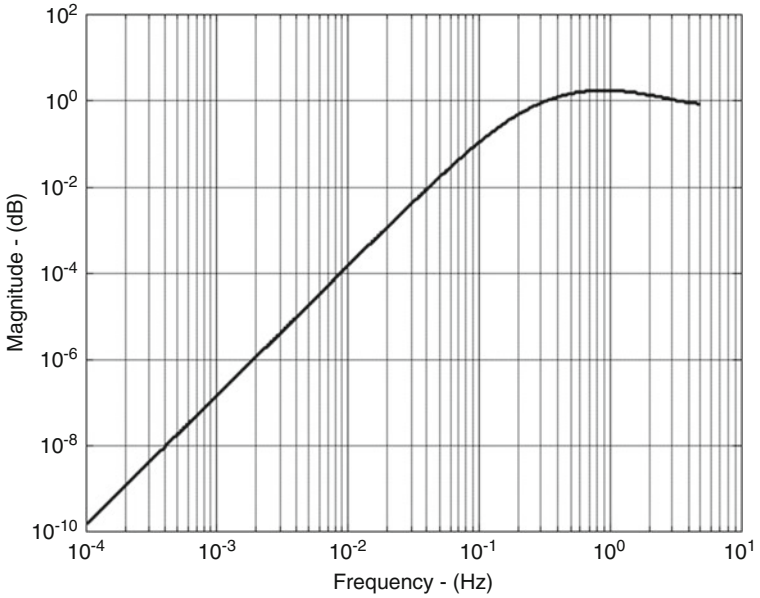


Fig. 20.16 Sensitivity function of the X and Y axes linear acceleration controllers

To justify the formation control scheme it is necessary to recall some points and results of the analysis. Starting from a given satellite relative position and velocity, the formation geometry changes in time because of:

- Differential perturbation due to Earth gravity field (non-spherical components like J2, etc.), atmospheric drag, solar radiation pressure.
- Differential linear acceleration bias of the accelerometers introduced in the drag-free control.
- Non-exact initialisation of the state variables (relative position and/or velocity) in the control system.

The evolution of the formation geometry have been analysed for the In-line and Pendulum configurations, and it turned out that the major destabilising source is the differential acceleration bias that in few days can move the satellites far away from the initial geometry. From that, it is clear that the Relative Position Control shall embed at least the capability to estimate and to compensate the differential biases. The bias estimator shall be fast enough to follow the low frequency drift of the accelerometers differential biases, and, at the same time, slow enough not to provide disturbances at the frequency relevant for the scientific mission.

Since the Relative Position Control shall operate at low frequency (i.e. lower than the orbit frequency), the model used to design the controller may be very simple, without taking into account J2 effect, etc. The Hill-Clohessy-Wiltshire (HCW) model (chief motion almost circular, differential effects related to J2,

drag, etc. not taken into account) was considered sufficient for control synthesis. Using the HCW model, it is possible to show that:

- A residual bias along the X axis (along-track) is responsible for:
  - Relative position along X axis growing in quadratic way;
  - Relative position along Z axis growing in linear way;
- A residual bias along the Z axis (radial) is responsible for:
  - Relative position along X axis growing in linear way;
  - Relative position along the Z axis remaining bounded with magnitude depending on the applied bias.
- A residual bias along the Y axis (across-track) does not destabilise the Y axis relative position that remains bounded with magnitude depending on the applied bias.

As a consequence the proposed formation control for the observation phase becomes very simple. Only the X axis (along-track) relative position is controlled to maintain the formation geometry on a time horizon to be defined and addressed by simulation. The X axis Relative Position Control embeds the capability to estimate the differential bias and provides compensation for it. The radial (Z) and across-track (Y) degrees of freedom do not need any specific control since a differential bias up to 2×10^{-7} m/s^2 (maximum value expected from GOCE-like accelerometers) induces an oscillation with amplitude equal to 0.15 m, which is negligible with the formation initialisation error in the control system.

The design of the Relative Position Control was done following the same approach used for the linear acceleration control (state observer, disturbance model, etc.); the sampling frequency has been fixed to 0.1 Hz, but it may be still reduced. The relative position and velocity are measured with the required accuracy and noise by the differential GNSS technique. The commanded linear accelerations computed by any linear acceleration control, and the commanded angular accelerations required by the attitude control, are used to compute, at each control step, the forces to be applied by the thrusters assembly.

To fulfil the specified requirements (in particular the spectral density limits of the non-gravitational linear accelerations), it is necessary that all the control loops work in linear condition. Saturations due to insufficient thruster force and/or drag force below the minimum level applicable by the thrusters, are responsible for residual accelerations with high frequency components. The performance that the thrusters of the various typologies (ref. Sect. 20.5) must provide in order to ensure the implementation of all the control functions within the specified requirements are summarised in Table 20.4 and Fig. 20.17.

Performance assessment has been done by extending the use of the GOCE End-To-End (E2E) Simulator [13] to the case of two satellites. The E2E Simulator embeds very accurate and representative models for environment (MSIS90 atmosphere model for the computation of the air density, Hickey model for the

Table 20.4 Thrusters preliminary requirements

Parameter	Main thrusters	Lateral thrusters, In-line formation	Lateral thrusters, Pendulum formation	Unit
Minimum thrust	<0.1	0.05	<0.1	mN
Maximum thrust	>6	>1.0	>2	mN
Thrust resolution	<4	0.5	0.5	μN
Thrust noise	see Fig. 20.17	see Fig. 20.17	see Fig. 20.17	
Rise/fall time	<50	<50	<50	ms
Slew rate	>2	>0.25	0.5	mN/s
Update command rate	10	10	10	Hz
Thrust non linearity	<2%	<2%	<2%	
Lifetime	11 years	11 years	11 years	
Specific power	<30	<40	<40	W/mN

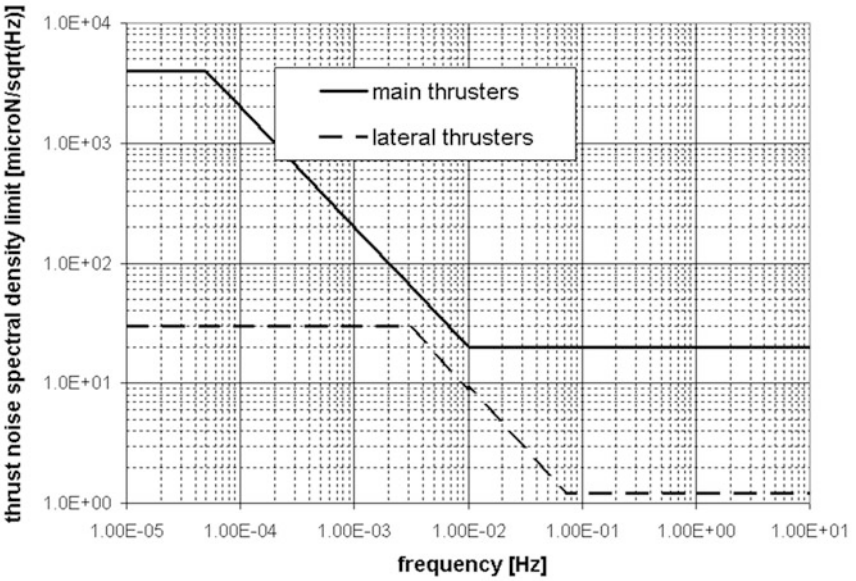
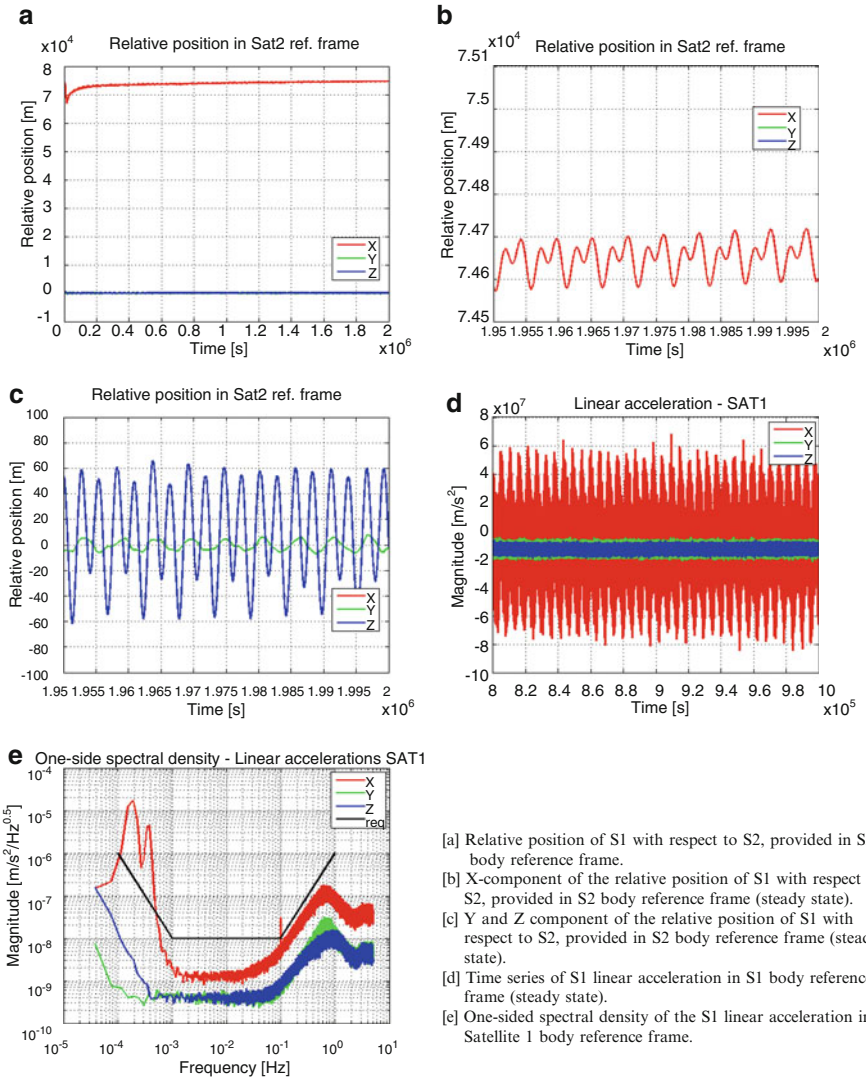


Fig. 20.17 Upper limit to the thrusters force noise spectral density

computation of the small scale air density variations, HWM93 model of the horizontal neutral wind in the upper thermosphere, a model of the Earth's magnetic field derived from the Ørsted mission, EGM96 Earth's gravitational potential model, complete to degree and order 360, Sun and Moon ephemerides), interaction between satellite and space environment, payloads, sensors and actuators. An example of obtained performance in the case of In-line formation with 75 km baseline is shown in Fig. 20.18. Moreover, numerical simulations spanning a time period of more than 60 days have demonstrated the capability of the control to keep stable the geometry of the analysed formations (In-line and Pendulum).



[a] Relative position of S1 with respect to S2, provided in S2 body reference frame.
[b] X-component of the relative position of S1 with respect to S2, provided in S2 body reference frame (steady state).
[c] Y and Z component of the relative position of S1 with respect to S2, provided in S2 body reference frame (steady state).
[d] Time series of S1 linear acceleration in S1 body reference frame (steady state).
[e] One-sided spectral density of the S1 linear acceleration in Satellite 1 body reference frame.

Fig. 20.18 Synthesis of results about the formation and non-gravitational linear acceleration control for the In-line formation

Acknowledgments The material utilised to write this chapter derives from the preparatory studies for the NGGM promoted and funded by the European Space Agency and performed by Thales Alenia Space Italia (team leader) with the contribution of the following sub-contractors/consultants:

- Deimos Space, Spain.
- Delft University of Technology, Delft Institute of Earth Observation and Space Systems (DEOS), The Netherlands.

- Istituto Nazionale di Ricerca Metrologica (National Institute of Metrology Research, INRIM), Italy.
- ONERA, France.
- Polytechnic of Milan, Department of Earth Sciences, Italy.
- Polytechnic of Turin, Department of Automatics and Informatics, Italy.
- Technical University Munich, Institute of Astronomical and Physical Geodesy (IAPG), Germany.
- University of Luxemburg.
- University of Milan, DIIAR, Satellite Geodesy Group, Italy.
- University of Pisa, Department of Mathematics, Italy.
- Universität Stuttgart, Institute of Geodesy (GIS), Germany.

Appendix

List of Abbreviations

CGE	Cumulative Geoid Error
CHAMP	CHAllenging Minisatellite Payload
COM	Centre Of Mass
E2E	End-To-End
ESA	European Space Agency
GNSS	Global Navigation Satellite System
GOCE	Gravity field and Ocean Circulation Explorer
GRACE	Gravity Recovery And Climate Experiment
HCW	Hill-Clohessy-Wiltshire
INRIM	Istituto Nazionale di Ricerca Metrologica
LL-SST	Low-Low Satellite-Satellite Tracking
LORF	Local Orbital Reference Frame
MBW	Measurement BandWidth (from 1 to 100 mHz)
NGGM	Next Generation Gravity Mission
RAAN	Right Ascension of the Ascending Node
RIT	Radio-frequency Ion Thruster
TAS-I	Thales Alenia Space Italia

References

1. http://op.gfz-potsdam.de/champ/index_CHAMP.html
2. http://op.gfz-potsdam.de/grace/index_GRACE.html
3. http://www.esa.int/esaLP/LPgoce.html
4. Assessment of a next generation mission for monitoring the variations of Earth's gravity. ESA Contract 22643/09/NL/AF, Final Report, Issue 2, 22 Dec 2010

5. Koop R, Rummel R (2007) The future of satellite gravimetry. Final report of the future gravity mission workshop, ESA/ESTEC, Noordwijk, The Netherlands, 12–13 Apr 2007
6. Cesare S, Aguirre M, Allasio A, Leone B, Massotti L, Muzi D, Silvestrin P (2010) The measurement of Earth's gravity field after the GOCE mission. Acta Astronaut 67:702–712
7. Bender PL, Wiese DN, Nerem RS (2008) A possible dual-GRACE mission with 90 degree and 63 degree inclination orbits. In: Proceedings of the 3rd international symposium on formation flying, missions and technologies, Noordwijk (NL), April 2008
8. Allasio A, Muzi D, Vinai B, Cesare S, Catastini G, Bard M, Marque JP (2009) GOCE: space technologies for the reference earth gravity field determination. In: Proceedings of the European Conference for Aerospace Sciences (EUCASS), Versailles, France
9. Laser interferometry high precision tracking for LEO – ESA contract 20512/06/NL/IA. Summary report, 10 Sep 2008
10. Feili D, Lotz B, Loeb HW, Leiter H, Boss M, Braeg R, Di Cara DM (2009) Radio frequency mini ion engines for fine attitude control and formation flying applications. In: Proceedings of the 2nd CEAS European air & space conference, Manchester, UK, 20–26 Oct 2009
11. Sechi G, Buonocore M, Cometto F, Saponara M, Tramutola A, Vinai B, Andrè G, Fehringer M (2011) In-flight results from the drag-free and attitude control of GOCE satellite. IFAC 2011 Congress, Milan
12. Canuto E (2007) Embedded model control: outline of the theory. ISA Trans 46(3):363–377
13. Allasio A, Anselmi A, Catastini G, Cesare S, Dumontel M, Saponara M, Sechi G, Tramutola A, Vinai B (2010) GOCE mission: design phases and in-flight experiences, AAS-10-081. In: Proceedings of the 33rd annual AAS guidance and control conference, Breckenridge, CO, 6–10 Feb 2010

Chapter 21
PRISMA

Simone D'Amico, Per Bodin, Michel Delpech, and Ron Noteborn

Abstract PRISMA is a precursor mission for formation-flying and on-orbit-servicing critical technologies. It consists of two spacecraft launched clamped together in low Earth orbit and separated in space after the commissioning phase in August 2010. The mission represents a unique in-orbit test-bed for guidance, navigation, and control (GNC) algorithms, novel relative navigation sensors (GPS, radio-frequency, vision-based), as well as new propulsion systems (high performance green propellant, micro-electro-mechanical). Originating from an initiative of the Swedish National Space Board, PRISMA is run by OHB Sweden with important contributions by the German Aerospace Center, the French Space Agency, and the Technical University of Denmark. After a brief overview of motivations, partners, and objectives, the chapter starts with a comprehensive description of the mission, including spacecraft platform, formation-flying and rendezvous sensors and actuators, as well as GNC key modes and algorithms. The discussion is followed by a summary of the main project phases, including overall schedule, verification process, and mission operations. Actual flight results from the basic PRISMA mission and its numerous GNC experiments are presented along with the achieved relative navigation and control accuracies over a broad range of autonomous operations between 30 km and nearly zero inter-spacecraft separation.

S. D'Amico, Ph.D. (✉)
German Aerospace Center (DLR), German Space Operations Center,
Münchner Straße 20, 82234 Wessling, Germany
e-mail: simone.damico@dlr.de

P. Bodin, Ph.D. • R. Noteborn, M.Sc.
OHB Sweden AB, PO Box 1064, SE-171 22, Solna, Sweden
e-mail: per.bodin@ohb-sweden.se; ron.noteborn@ohb-sweden.se

M. Delpech, Ph.D.
CNES, Toulouse Space Center, 18 Av. E. Belin, 31401 Toulouse, France
e-mail: michel.delpech@cnes.fr

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_21,

21.1 Introduction

21.1.1 Motivation

Several space agencies have identified formation flying and on-orbit servicing as key enabling technologies for advanced scientific and commercial applications [1–3]. The expected benefits of these techniques have resulted in a virtual explosion of envisioned space assets. Although any attempt to list the missions under consideration would be inevitably incomplete, the possible applications can be practically classified in three major domains: space science, planetary science, and technology development. Related space science disciplines include solar system exploration, astronomical search for origins, structure and evolution of the universe. Distributed satellite systems in low planetary orbits can augment and/or enable remote sensing tasks like digital elevation modeling, gravity recovery, and atmosphere characterization. The technological development of space requires the capability to assemble structures in orbit, to re-supply and repair orbital platforms, to capture and de-orbit space debris. These are but a few of the applications which will take advantage of key driver technologies like formation flying and on-orbit servicing.

Common to these classes of missions is the necessity to implement on-board autonomous guidance, navigation, and control (GNC) functionalities which are traditionally performed in a ground-in-the-loop fashion. It is in particular this aspect which motivated the Swedish National Space Board (SNSB) and the Space Systems Division of the Swedish Space Corporation (SSC) (now OHB-SE) to conceive the Prototype Research Instruments and Space Mission technology Advancement (PRISMA) in late 2004 [4, 5]. PRISMA was established with the ambition to test in flight GNC software as well as sensor and actuator hardware technologies for formation flying and rendezvous of two co-orbiting satellites [6, 7]. After the approval of the basic funding by the SNSB in early 2005, the prime contractor OHB-SE established cooperation with a number of national and international partners to realize PRISMA. Potential participants were invited by the prime to share mission time and resources in terms of self defined experiments as a compensation for the costs associated with specific software and/or hardware contributions to the project. Especially the German Aerospace Center (DLR/GSOC) and the French Space Agency (CNES) responded to the call and developed their own GNC software (integrated into the OHB-SE on-board software) in order to execute dedicated closed-loop experiments in orbit during the mission timeline.

21.1.2 Partners

OHB-SE can be considered as prime contractor of the SNSB, and is responsible for the overall design, build, and operations of the mission. The main partners of OHB-SE in the development and scientific exploitation of PRISMA are listed in the following together with their key contributions:

- The German Aerospace Center (DLR/GSOC) contributes the GPS absolute and relative navigation system, comprising both the GPS hardware (Phoenix-S) and on-board navigation software [8, 9]. DLR performs the Spaceborne Autonomous Formation Flying Experiment (SAFE), and the Autonomous Orbit Keeping experiment (AOK) through dedicated software embedded in the overall GNC software [10, 11]. Furthermore DLR is responsible for the precise orbit determination (POD) on-ground verification layer [12].
- The French Space Agency (CNES) contributes the Formation Flying Radio Frequency sensor (FFRF) and the associated software. Also the Spanish agency CDTI is involved in the development of the FFRF instrument. CNES performs autonomous formation flying experiments with dedicated software embedded in the overall GNC software. The complete package is called Formation Flying In-Orbit Ranging Demonstration (FFIORD) [13–15].
- The Technical University of Denmark (DTU) contributes the Vision Based Sensor (VBS) [16].

Other partners in the project are listed below:

- SSC/ECAPS contributes a novel High Performance Green Propellant (HPGP) motor system developed as a replacement system for hydrazine [17].
- SSC/Nanospace contributes a novel cold-gas Micro-propulsion motor system based on MEMS technology [18].
- Techno System Developments (TSD) in Naples, Italy, contributes a versatile high resolution color Digital Video System (DVS) [19].
- The Institute of Space Physics (IRF) in Kiruna, Sweden, contributes a newly developed mass spectrometer with a MEMS velocity filter [20].

The distribution of tasks and responsibilities in the project is depicted in Fig. 21.1.

21.1.3 Objectives

PRISMA is primarily intended to demonstrate autonomous formation flying, homing and rendezvous scenarios, as well as close-range proximity operations. The mission objectives are associated to specific experiment sets which structure the mission timeline and operations [6, 7]. There are four operational and four equipment experiment sets which are described in Table 21.1. The closed-

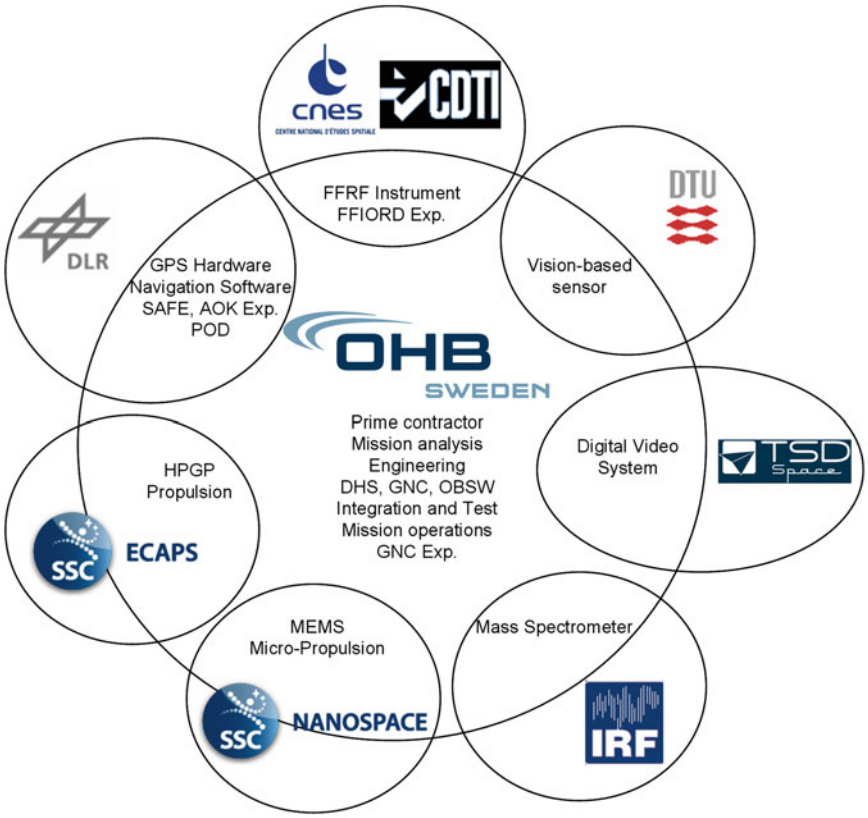


Fig. 21.1 PRISMA partners and contributions

loop orbit control experiments are conducted by OHB-SE, DLR and CNES over a wide range of separations and formation geometries. Each GNC experiment makes use of one of the three relative navigation sensors available on-board (i.e., GPS, FFRF, and VBS). The hardware experiment sets include tests of the HPGP and MEMS experimental propulsion systems (SSC), the vision-based sensor (DTU), and the formation flying radio-frequency metrology sensor (CNES/CDTI).

In addition to the primary objectives of Table 21.1, the following several secondary objectives have been defined:

- Flight test of newly developed Data Handling System and Power Conditioning and Distribution Unit with battery management electronics (OHB-SE).
- Act as model project for new model based development of on-board software (OHB SE).
- Demonstrate Autonomous Orbit Keeping (AOK) of a single spacecraft (DLR).

Table 21.1 Summary of GNC related and hardware related experiment sets (primary objectives)

GNC experiment set	Description	Separation (m)	Key sensors
Autonomous formation flying	Closed-loop autonomous formation acquisition, keeping, and reconfiguration using various guidance and control laws	5,000–20	GPS or FFRF
Homing and rendezvous	Autonomous approaches emulating assembly in space orbits and Mars sample return scenarios	100,000–3	VBS
Precision 3D proximity operations	Technology demonstration using forced motion around virtual structures for on-orbit servicing, inspection and assembly	100–3	GPS or VBS
Final approach and recede maneuvers	Close approaches to demonstrate on-orbit servicing, inspection and assembly capabilities	3–0	VBS

Hardware experiments	Description	Responsible
HPGP (Green propellant)	First flight demonstration and validation in orbit	SSC/ECAPS
Micropropulsion (Cold-gas)		SSC/Nanospace
FFRF (Radio frequency)		CNES/CDTI
VBS (Vision based)		DTU

- Demonstrate a newly developed Ground Support and Operational Support Equipment for multi-vehicle missions, the RAMSES system (OHB-SE).
- Provide a test flight of the Digital Video System developed by Techno System.
- Provide a test flight for a MEMS-based particle mass spectrometer from the Institute of Space Physics in Kiruna, Sweden.

21.2 Satellites

21.2.1 Spacecraft Platform

The PRISMA space-segment consists of an active servicer or deputy minisatellite, named Mango (150 kg), and a passive client or chief microsatellite, named Tango (40 kg), both built by OHB-SE. Mango features six degrees of freedom control through 3-axis attitude stabilization and 3D orbit maneuver capability independent of the spacecraft attitude. Attitude determination and control are based on star trackers and reaction-wheels respectively. Orbit determination and control are

based on three relative navigation sensors (i.e., GPS, FFRF, and VBS) and three propulsion systems (i.e., Hydrazine, HPGP, and Micro-propulsion) respectively. Although the aforementioned sensors and actuators are all alternatively used during experimental phases, the primary relative navigation system is based on GPS (which is always active) and the main propulsion system is based on Hydrazine with six thrusters and approximately 120 m/s delta-v capability. Electrical power for the operation of the Mango spacecraft bus and payload is provided by two deployable solar panels delivering a maximum of 400 W. The central body of Mango has exterior dimensions 750 × 750 × 820 mm. When deployed, the distance between the tips of the solar panels is 2,600 mm.

The Tango satellite has a simplified, yet 3-axis stabilizing, solar magnetic attitude control system and no orbit maneuver capability. The Tango body is 570 × 740 × 295 mm. Tango relies on one body-mounted solar panel providing a maximum of 40 W. Spacecraft operations rely on an S-band ground-to-space link on Mango which supports commanding with a bit rate of 4 kbps and telemetry with up to 1 Mbps. In contrast, communication with Tango is only provided through Mango which acts as a relay and makes use of a bidirectional Inter-Satellite-Link (ISL) in the Ultra-High-Frequency (UHF) band with a data rate of 19.2 kbps.

The core of the data handling system (DHS) on Mango is the spacecraft controller based on a LEON3 microprocessor [21]. LEON3 implements a 32-bit processor compliant with the SPARC V8 architecture, which is particularly suited for embedded applications. In contrast to its predecessor LEON2, LEON3 recognizes bit flips and is fault-tolerant. Its implementation through a field programmable gate array (FPGA) by Atmel Corporation provides a performance of about 20 MIPS and accommodates one floating point unit (FPU). Communication between platform units and the spacecraft controller is implemented via a controller area network (CAN) bus. The PRISMA system is single failure tolerant in all essential functions. A Failure, Detection, Isolation and Recovery system (FDIR) has been implemented which enables autonomous switching to redundant units in most failure situations. Furthermore the FDIR system can detect potential collision or evaporation risk, and go to an inherently collision-free orbit through a minimum number of autonomous maneuvers.

Figure 21.2 shows photos of the PRISMA flight models in split and combined configuration taken before launch in January and June 2010 respectively. For ease of identification the key sensing elements of the three available relative navigation sensors are indicated on the photos. The GPS antennas on Mango and Tango are named GPS-M and GPS-T. The FFRF antennas are named FFRF-M and FFRF-T. The short- and far-range VBS cameras on Mango are named VBS-S and VBS-F respectively.

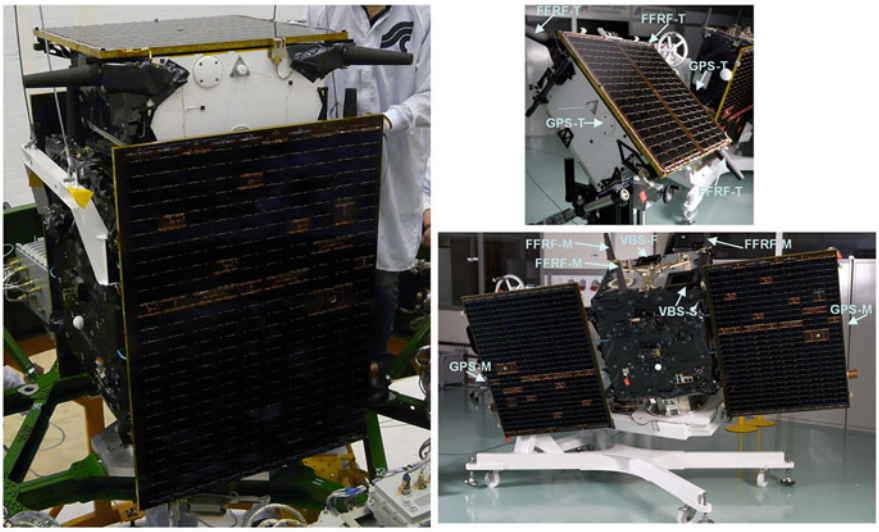


Fig. 21.2 Flight models of the main satellite Mango (*right-bottom*), the target satellite Tango (*right-top*), and the combined launch configuration (*left*). The photos were taken in the SSC clean-room in January 2010. The combined spacecraft was photographed in June 2010 during the mounting of PRISMA on the Russian Dnepr launcher

21.2.2 Formation Flying and Rendezvous Sensors

The primary relative navigation sensor of the PRISMA formation is a GPS-based navigation system developed by DLR/GSOC [9]. Relative GPS serves both as safe mode sensor of the formation to support FDIR tasks like collision avoidance and as navigation source for the on-board feedback controllers to enable autonomous formation flying and rendezvous experiments. Furthermore in absence of instruments characterized by higher technology readiness level, GPS serves also as true reference for sensor-cross validation, actuator characterization, GNC and experiment accuracy evaluation. To this end the GPS raw data are processed post-facto on-ground by the Precise Orbit Determination (POD) facility of DLR/GSOC to routinely generate accurate daily products [12].

The GPS-based navigation system consists of a hardware architecture based on Phoenix-S receivers [8] which is identical on-board the two satellites and a navigation software embedded in the Mango on-board computer for real-time absolute and relative navigation. The GPS hardware is illustrated in Fig. 21.3. For redundancy purposes, each spacecraft carries two independent Phoenix-S GPS receivers that are operated in a cold redundant configuration. Increased flexibility for handling non-zenith attitude pointing is provided by two GPS antennas on each spacecraft, which are selected by an onboard algorithm for optimum GPS coverage or may, alternatively, be set by ground command.

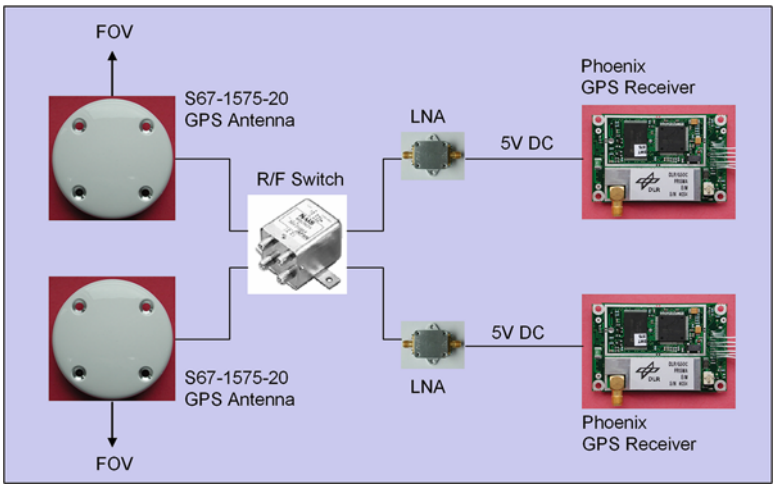


Fig. 21.3 Cold-redundant GPS hardware architecture based on Phoenix-S receivers identical on Mango and Tango

The miniature Phoenix-S is a 12 channel single-frequency GPS receiver based on a commercial-off-the-shelf (COTS) hardware platform and qualified by DLR/GSOC for use in Low Earth Orbit (LEO). The single-frequency (C/A, L1) pseudorange and carrier phase measurements output by the receivers on Mango and Tango for each channel are combined in quasi-ionosphere-free observables and processed by the on-board navigation software on Mango through an Extended Kalman Filter (EKF). A symmetric filter design has been chosen which adjusts the absolute states of both spacecraft through rigorous numerical dynamics modeling [22]. The relative spacecraft state is simply computed by differencing the absolute states (output at 1 Hz) without the need of an explicit relative motion model. The accuracy of the relative navigation is driven by the single-difference carrier-phase measurements, whereas the absolute navigation accuracy is driven by the noise of the GRAPHIC measurements (i.e., average of pseudorange and carrier-phase).

The contribution of CNES to the PRISMA mission is the FFIORD experiment [15]. Its main objective is to perform a first in-orbit demonstration of the FFRF subsystem. In its general configuration the FFRF instrument is intended to perform coarse relative positioning of two to four satellites on formation flying missions. It produces relative position, velocity and line-of-sight (LOS), as inputs to the GNC functions. The subsystem is composed of one FFRF terminal and up to four sets of antennas on each satellite of the formation. A set of antennas can be either a triplet (1 Rx/Tx master and 2 Rx slaves) or a single Rx/Tx antenna. Each S-band terminal transmits and receives dual-frequency signals to and from all the other satellites (1 to 3) of the constellation in a TDMA based pattern. Ranging and angular measurements are extracted from received signals and are used for computing relative position, velocity and line-of-sight. In addition to providing relative navigation measurements, the FFRF subsystem also provides an ISL as auxiliary

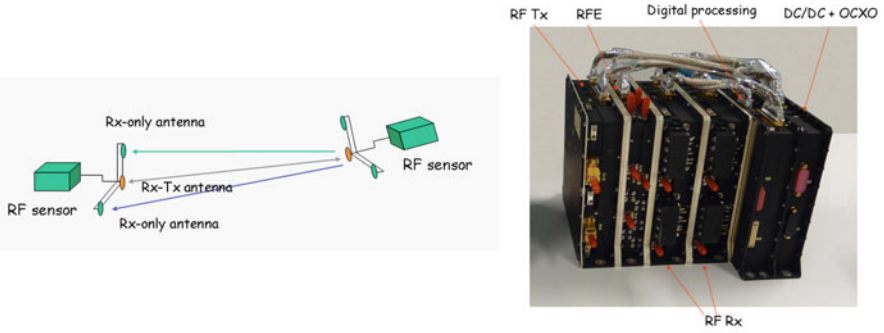


Fig. 21.4 FFRF subsystem configuration on PRISMA (*left*) and FFRF terminal hardware (*right*)

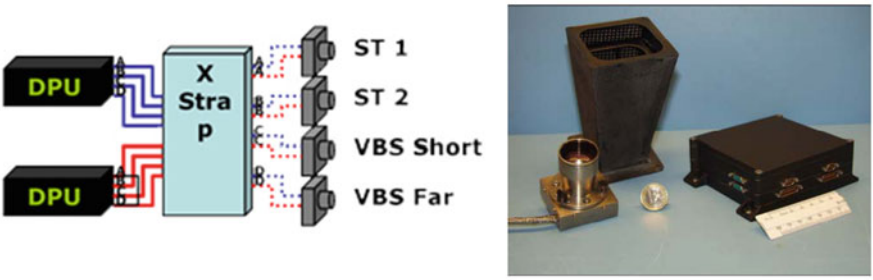


Fig. 21.5 Star tracker and VBS subsystem configuration on Mango (*left*). Camera head and data processing instrument unit (*right*)

functionality (12 kbit/s or 4 kbit/s). As depicted in Fig. 21.4, two configurations for the FFRF terminal are available on PRISMA. On Mango, the FFRF terminal is connected to a unique triplet antenna set, whereas the FFRF terminal is connected to three single Rx/Tx antenna sets on Tango.

Similar to GPS, the FFRF sensor can deliver two types of metrology information. The RF raw data composed of pseudo-code, phase, delta-phase measurements from each terminal along with the clock bias between each terminal, and the relative position, velocity and time for the companion satellite. The FFRF sensor is complemented by a navigation function in the GNC on-board software which filters this latter navigation solution in an EKF that includes direct modeling of the relative dynamics and performs measurement biases estimation [23]. In contrast to the GPS system employed on PRISMA, the FFRF navigation approach avoids the processing of raw data in the GNC software to reduce the complexity and computational load associated with the RF signal filtering and the integer ambiguity resolution functionalities.

The contribution of DTU to PRISMA is based on the microASC, a fully autonomous miniature star sensor (c.f. Fig. 21.5) [24]. The microASC is designed for highly flexible configurations and can host from one to four camera head units

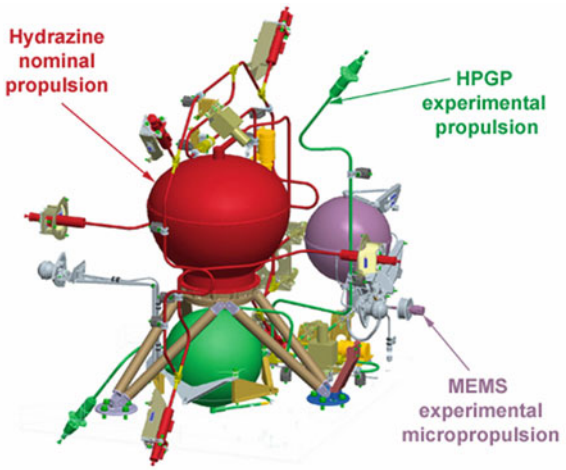


Fig. 21.6 Propulsion systems embarked on the PRISMA mission

(CHU), located at suitable places and directions on a spacecraft, such that a fully redundant blinding free attitude sensor configuration can be achieved.

Onboard PRISMA two CHUs are used as standard attitude reference sensors. Their pointing directions are selected such that simultaneous blinding by Sun, Earth and the Tango spacecraft is avoided during the complex fly-around maneuvers. The third port on the microASC is also equipped with a standard CHU. This CHU is however pointed in the forward direction, such that Tango can be seen in the field of view for most of the mission phases. The fourth port is equipped with a CHU with a modified focal length, iris and electronic shutter, to enable operations at close range with strong light conditions. These latter CHUs are used for vision-based navigation of Mango with respect to Tango and are named VBS-Far and VBS-Short range cameras.

The VBS data processing is performed in the microASC, which is made possible by its huge spare processing power in standard operation. A description of the VBS can be found in Ref. [25]. The basic software has four different modes of operation:

- Far range—Stars and Tango both detectable in field of view.
- Intermediate range—Tango high brightness prevents detection of stars.
- Cooperative short range—Features of Tango discernible in field of view through five LEDs mounted on each surface of the Tango body.
- Non-cooperative short range—Features of Tango obtained using natural illumination from Sun and Earth albedo are matched to a 3D model database stored in the microASC memory.

In short range mode both relative position and relative attitude (pose) information are output by the VBS subsystem for usage within the GNC software.

21.2.3 Propulsion Systems

The Mango satellite is equipped with a nominal hydrazine system, a HPGP system, and a Micro-propulsion system (c.f. colors red, green, and cyan in Fig. 21.6, respectively). The hydrazine propulsion system has been chosen as the nominal system since both HPGP and Micro-propulsion represent new developments which are demonstrated in-orbit for the first time.

The hydrazine propulsion system consists of six 1-N thrusters directed towards the Mango center of gravity, thus providing torque-free translational capability. The propellant tank contains 11 kg of usable fuel which gives approximately 120 m/s delta-v over the mission. Firing times range from 0.1 s to steady state burns of up to 2 min. The hydrazine system has been designed by OHB-SE based on procured components from mainly US suppliers.

The HPGP motor experiment is a new propulsion system introducing environmentally friendly, non-toxic monopropellant ammonium-dinitramide (ADN) fuel which promises up to 10% better impulse and 30% higher density than hydrazine [26]. Indeed the performance in orbit has proved to be better than hydrazine in all modes of operation. The HPGP propulsion system has two 1-N thrusters, also directed towards the center of gravity. The system provides redundancy to the main hydrazine system if any nominal thruster should fail. The propellant tank contains 5 kg of usable fuel and gives approximately 60 m/s delta-v over the mission. The development, including the propellant itself and a compatible 1-N thruster and catalyst bed, is driven by ECAPS, a subsidiary company to SSC, and has been supported by ESA for several years.

The Micro-propulsion system is based on MEMS technology and is developed by NanoSpace, a subsidiary of SSC, on contracts from ESA and SNSB [18]. It delivers accurate thrust ranging from tenths of micro-Newtons up to milli-Newtons. The system is a potential candidate for future missions where extremely low and accurate thrust is requested. The key component is a golf-ball sized thruster module containing a silicon wafer stack with four complete rocket engines with integrated flow control valves, filters, and heaters. Extremely small internal heaters inside the thrust chamber increase the performance of the system in terms of specific impulse. The propellant is Nitrogen. The four thrusters are orthogonally distributed in the equator plane of the thruster module.

21.3 GNC Experiments

21.3.1 Classification and GNC Modes

The GNC experiments conducted by OHB-SE, DLR, CNES, and DTU during the PRISMA mission were intended to demonstrate closed-loop relative orbit control of Mango with respect to Tango using different algorithms and sensors. The

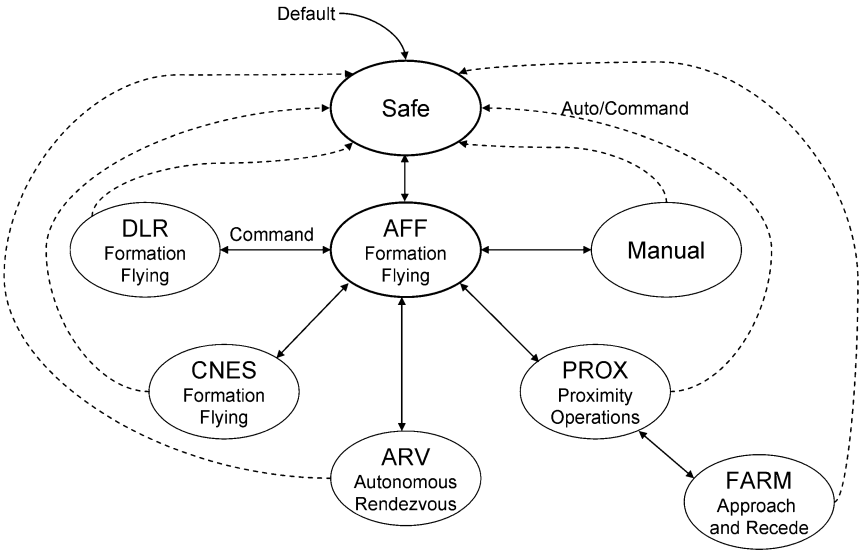


Fig. 21.7 GNC modes diagram on Mango spacecraft

experiments were distributed over the mission length in a sequence with increasing level of complexity, to enable early harvest results for all parties at the beginning of the mission. The overall experiment planning and design was under OHB-SE management. DLR and CNES autonomous formation flying experiments were based on their respective sensor system contributions (i.e., GPS and FFRF respectively). To this end, dedicated GNC software has been developed by these partners and embedded by OHB-SE in the GNC core software. The VBS-based experiments were supported by DTU with the highly sophisticated functionality of the VBS camera system.

As illustrated in Fig. 21.7, the backbone GNC modes are named Safe and AFF. The Safe mode is entered by default upon reboot of Mango on-board computer, or in general upon reset of the platform. Furthermore the Safe mode is automatically set by the FDIR logic in case of contingencies related to the various subsystems (e.g., power, thermal, propulsion, GNC, etc.) and in the presence of high collision risk with Tango. In order to handle these situations, the Safe mode itself is structured in sub-modes like Safe-Sun, Safe-Celestial, and Safe-Orbit. Whereas Safe-Sun and -Celestial may be considered standard functionalities of an ACS subsystem, Safe-Orbit is a unique multi-satellite feature which allows the execution of orbit control maneuvers to minimize the risk of collision with the co-orbiting spacecraft using relative GPS as navigation means.

Under nominal operations AFF is the only GNC mode that can be entered via command from Safe. AFF is considered the central node in the mode architecture (c.f. Fig. 21.7). All other experimental modes (e.g., DLR, CNES) can only be set via command upon transition from AFF which is also used as parking mode during the

Table 21.2 Simplified GNC mode functional configuration on Mango

	Functions used in closed-loop		OHB-SE
GNC modes	Orbit navigation	Orbit control	Responsible
Safe	GPS	Impulsive	OHB-SE
AFF	GPS	Feedback	OHB-SE
Manual	–	Impulsive	OHB-SE
DLR	GPS	Impulsive	DLR
CNES	FFRF	Impulsive or feedback	CNES
ARV	VBS	Impulsive or feedback	OHB-SE
PROX	GPS or VBS	Feedback	OHB-SE
FARM	VBS	Feedback	OHB-SE

preparation of other experiments or during formation standby phases. Similar to Safe, AFF also makes use of GPS for relative navigation. The other GNC modes trigger experimental functionalities and software implemented by DLR (DLR mode), CNES (CNES mode), and OHB-SE (PROX, ARV, FARM modes). In addition, a Manual mode is available to allow the ground to plan and command orbit control maneuvers bypassing the autonomous functionalities. For completeness, Table 21.2 offers an overview of the GNC mode functional configuration. Each GNC mode makes use of specific navigation and control functions. Whereas relative GPS is used by most of the modes, FFRF and VBS are only used in CNES mode and by OHB-SE in ARV and PROX/FARM modes. Depending on the employed algorithm and the desired control tracking accuracy the need for sparse or frequent orbit control maneuvers arises. The consequent thruster activations are triggered by either an impulsive or a feedback control function which are described in the following.

21.3.2 OHB-SE Modes and Key Algorithms

The GNC experiments under OHB-SE responsibility have been grouped in three different modes: Autonomous Formation Flying (AFF), Proximity Operations and Final Approach/Recede Maneuvers (PROX/FARM), and Autonomous Rendezvous (ARV). AFF demonstrates GPS-based passive formation flying. PROX/FARM includes close-range forced-motion operations under geometric flight constraints, where either GPS- or VBS-navigation can be used. The FARM mode makes use of VBS only and is designed to represent the final stage of docking or similar physical delivery. The capability to efficiently switch from passive relative orbits with sparse maneuvering to forced-motion control with quasi-continuous thrust (and vice-versa) is also an objective of the OHB-SE experiments. ARV is designed to demonstrate autonomous rendezvous capabilities from several kilometers distance down to a few meters using VBS as the only navigation sensor.

A detailed explanation of the applied algorithms can be found in Refs. [27, 28]. The AFF orbit guidance function uses passive target orbits up-linked from ground.

The guidance function can also generate its own passive target orbit upon autonomous entry to AFF mode from a higher mode. In this case a "closest" passive orbit is selected and the AFF mode acts as a higher safe or fallback mode. The AFF uses an orbit control function based on a Model Predictive Control (MPC) framework [29]. The MPC function optimizes the next control request in terms of fuel consumption based on orbit propagation over a number of future control points. The orbit propagator is implemented for a general case of elliptic orbits using the Yamanaka-Ankersen equations [30]. Fuel optimal solutions are obtained through on-board solution of the associated simplex problems. The AFF experiment implements relative orbits ranging from several kilometers down to distances around 10 m.

The PROX/FARM experiments consist of forced motion of Mango around Tango at distances around and below 10 m. Mango navigates around a virtual structure centered at Tango's position. The structure is designed to mimic a large space structure such as the International Space Station (ISS) around which a flight map is spanned. This flight map consists of a set of nodes with associated allowed transitions. Based on the commanded transitions, the guidance function finds the optimal allowed path through the flight map by making use of an A* algorithm [31]. The orbit control in PROX/FARM uses the same framework as in AFF but with different settings for propagation horizons and actuation frequency.

The ARV experiment demonstrates a complete autonomous rendezvous starting with Mango approximately 25 km away from Tango. Mango maneuvers autonomously down to a few tens of meters away from Tango ending the experiment with a FARM maneuver that takes Mango to within less than 1 m from Tango. The final steps make use of the PROX/FARM functions described above. Mango autonomously locates Tango at far distance and then performs a series of orbit aligning, closing, and homing maneuvers to finally get within a distance from Tango to allow for the VBS to deliver explicit distance information so that handover to the PROX/FARM functions is possible. All of the maneuvers in the ARV are performed using VBS-navigation only and at far distances, these navigation functions are only based on angular measurements.

21.3.3 DLR Modes and Key Algorithms

The Spaceborne Autonomous Formation Flying Experiment (SAFE) and the Autonomous Orbit Keeping (AOK) experiments of DLR are conducted in the homonymous DLR GNC mode.

SAFE aims at demonstrating fuel-efficient long-term autonomous acquisition, keeping and reconfiguration of passive relative orbits on a routine basis. To this end a convenient parameterization of the relative motion in terms of relative orbital elements is employed. The guidance approach applies the method of relative eccentricity and inclination vector separation to minimize collision risk and to increase the passive stability of the formation under J_2 perturbation [32]. Maneuver

delta-v commands are computed according to an analytical impulsive feedback control function which is characterized by a high level of determinism, and, through its negligible computational burden, is ideally suited for an onboard implementation [10, 33]. A wide variety of formation geometries is exploited by the SAFE experiment, spanning a range from 6 km down to a minimum separation of 30 m.

AOK is designed to demonstrate on-board a single spacecraft the autonomous control of the osculating Longitude of the Ascending Node (LAN) with an accuracy of 10 m (1σ). The target trajectory for orbit control can be either up-linked from ground as a set of desired LANs or generated on-board through numerical propagation. Similar to SAFE, the maneuver delta-vs are computed analytically but are restricted to orbit corrections in velocity and anti-velocity directions [34].

21.3.4 CNES Modes and Key Algorithms

The CNES FFIORD closed-loop experiments are executed in CNES GNC mode and make use of the FFRF metrology subsystem [15]. The main objectives are the demonstration of autonomous rendezvous (RDV, 2MT), stand-by on relative orbits (SBY), and proximity operations capabilities (PROX) like, e.g., station-keeping at different distances and offset positions from the orbit track, low speed translations in-plane and out of plane. Furthermore FFIORD includes collision avoidance (CA) experiments. Each objective translates into a specific GNC function of the CNES flight software whose acronym has been indicated in brackets.

A detailed explanation of the applied algorithms can be found in Refs. [35, 36]. Most of the aforementioned orbit control functions apply the inverse state transition matrix of Yamanaka-Ankersen to compute impulsive maneuvers (SBY, 2MT). Forced-motion control along three axes in the vicinity of Tango is based on linear quadratic control with a typical period of actuation of 200 s (PROX). Safe rendezvous or deployment from a distance of about 10 km to 100 m is based on a MPC-like method with fixed maneuvers dates and L2-norm criteria for the minimization of the propellant consumption. The collision avoidance approach features two options which consist either in a relative drift between Tango and Mango or in a proper phasing of radial and cross-track components to maximize the separation in the plane perpendicular to the flight direction.

21.3.5 Ground Verification Layer

The PRISMA POD facility is based on DLR's in-house GPS High precision Orbit determination Software Tools (GHOST). This software suite has been used routinely to support many missions (e.g., GRACE, CHAMP, TerraSAR-X, TanDEM-X, PROBA-2) and has proved its high-readiness level as well as its ability to

provide reliable and accurate orbit products. The data processing for precise orbit determination follows a four-stage scheme described in the following [37].

First the raw GPS measurements are processed with the Single Point Positioning for Low-Earth Orbit (SPPLEO) program to derive a kinematic navigation solution. The output of the kinematic single point positioning is a discrete set of navigation points affected by errors at the meter level. Second, this kinematic solution is filtered dynamically using a batch least square process, called PosFit, which relies on accurate models of the spacecraft dynamics. Third, a precise orbit determination is done using the previously computed coarse orbit solution for data editing. During the data editing, the coarse reference is used to assess the quality of raw measurements and to select the healthy observations. The program for Reduced Dynamics Orbit Determination (RDOD) implements a batch least-squares filter which processes ionosphere free combinations of pseudorange and carrier phase measurements (so called GRAPHIC measurements) to generate orbit products accurate at the meter level for single-frequency receivers. Finally, a filter for relative navigation of satellites (FRNS) provides accurate relative positioning using the precise orbit products of Mango and Tango as references for data editing. The filter makes use of double difference carrier phase measurements to achieve ultimate relative positioning accuracy. The double difference carrier phase integer ambiguity is solved using the LAMBDA method [38].

The PRISMA mission offers many technical challenges for GPS based navigation when compared with typical space missions [39]: complex and rapidly varying attitude profiles, dense thrusting activities, frequent switching of antennas. This unusual complexity makes the routine generation of precise orbit products difficult. These considerations have triggered the implementation of an intermediate model based on float ambiguity estimation. In this mode the carrier phase ambiguities have float values which are simply estimated as part of the state vector in the FRNS extended Kalman filter. This method is more robust but comes at the cost of a degraded accuracy and is thus used only if the integer ambiguity cannot be applied successfully.

21.4 Project Phases

21.4.1 Overall Schedule

The PRISMA project has been characterized by tight constraints in terms of time and costs throughout the complete life cycle. The original plan was drafted in 2005 and foresaw a target project time of 3 years, aiming at a launch in summer 2008 [5]. As shown in Table 21.3 the actual schedule accumulated an overall delay of only 17 months when the PRISMA spacecraft were finally ready to be shipped to the launch site. The actual launch of the PRISMA satellites from Yasny, Russia was further

Table 21.3 PRISMA project original and actual schedule

Project milestones	Original plan (as of 2005)	Actual schedule (as of 2011)
Mission requirement review	End of April 2005	End of April 2005
System requirement review	Mid of June 2005	Mid of June 2005
Preliminary design review	November 2005	November 2005
Critical design review	November 2006	January 2007
Flight acceptance review	March 2008	May 2009
Ready for shipment to launch site	June 2008	November 2009
Launch from Yasny (Russia)	Mid of 2008	15 June 2010

delayed by approximately 6 months due to political and environmental problems associated with the launch site itself and out of the control of OHB-SE.

According to plan, the PRISMA Phase-B was concluded in November 2005 with a Preliminary Design Review (PDR) after less than 1 year from project start. The PDR was followed by purchasing processes, detailed satellite design and production of engineering models of the satellite subsystems. The bench tests of engineering models in SSC laboratories took place in September 2006. The GNC system simulations started in October 2006, whereas the vibration tests of the structural models were conducted in December 2006.

The Critical Design Review (CDR) milestone was successfully completed in January 2007, less than 2 months behind schedule. This paved the way to the delivery of flight hardware and experiments (also by the partners) in the time frame between June and December 2007 and to the Assembly, Integration, and Test (AIT) activities which were completed in April 2008. Advanced satellite system tests took place between May and October 2008, followed by environmental testing (vibration, thermal, cycling, etc.) until February 2009. The AIT took place in the clean room facilities at SSC engineering center in Solna. Environmental tests at system level were performed in Arboga near OHB-SE's premises and at Intespace in Toulouse, France. Figure 21.8 shows two photos taken during these activities which represent a radio test model of Mango mounted at the antenna test range in Arboga (c.f. Fig. 21.8, left) and flight models of Mango and Tango in the sun beam of the space simulator at Intespace (c.f. Fig. 21.8, right).

After a two-stage Flight Acceptance Review (FAR) in mid 2009, the PRISMA satellites were declared ready for transportation to the launch site in November 2009. Finally, PRISMA has been successfully launched aboard a Dnepr launcher from Yasny, Russia on June 15, 2010. The two PRISMA satellites were released, clamped together in launch configuration, into a nominal dawn-dusk orbit at a mean altitude of 757 km, 0.004 eccentricity and 98.28° inclination. The subsequent 2-day Launch and Early Operations Phase (LEOP) signed the beginning of the first technology demonstration mission for autonomous satellite formation flying and on-orbit servicing in Europe.

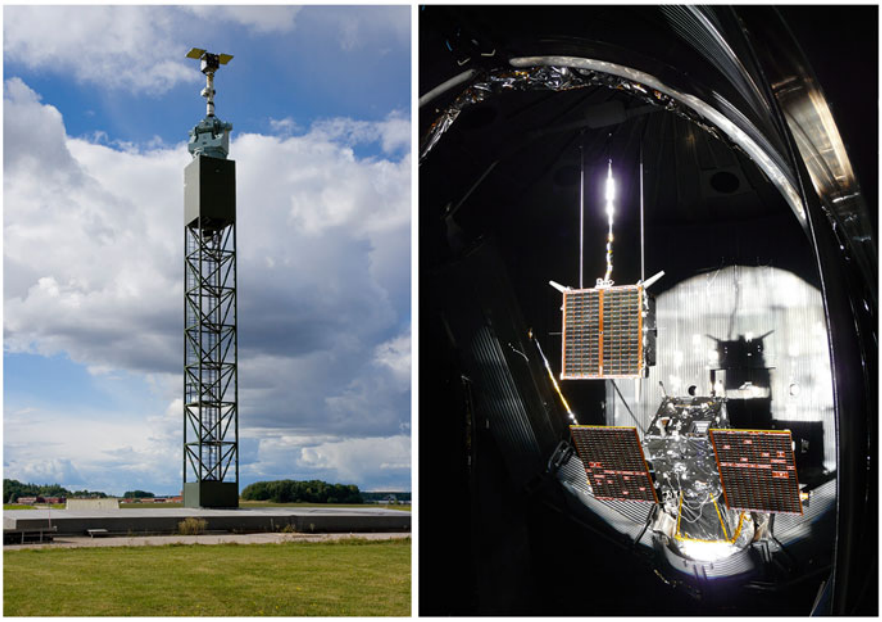


Fig. 21.8 Radio test model of Mango, geometrically correct and with real antennas, at the antenna test range in Arboga, Sweden (*left*, Autumn 2007). Flight models of Mango and Tango in the sun beam of the space simulator at Intespace in Toulouse, France (*right*, Early 2009)

21.4.2 Verification Process

Although the verification and test process has been similar to a single-satellite project, new elements had to be considered which are specific to a multi-satellite mission, especially at system level [40]. The main test verification elements are listed below:

- A Structure Test Model (STM) has been used to qualify the new structure of both Mango and Tango, the separation system between the two satellites, and the propulsion equipment.
- A Bench Test Model (BTM) environment has been used to iteratively debug interface problems and check basic functionality through a mixture of Engineering Model (EM) and Flight Model (FM) units connected to the EM data handling and power systems.
- The SatLab software development and test environment.
- The System Functional and Performance Tests (SFPT) environment.
- The Environmental test campaign.

The software system tests took place in a real-time software validation facility developed at OHB-SE, called SatLab. It consists of one EM of the computer board for each of the spacecraft. The complete onboard software runs on a single

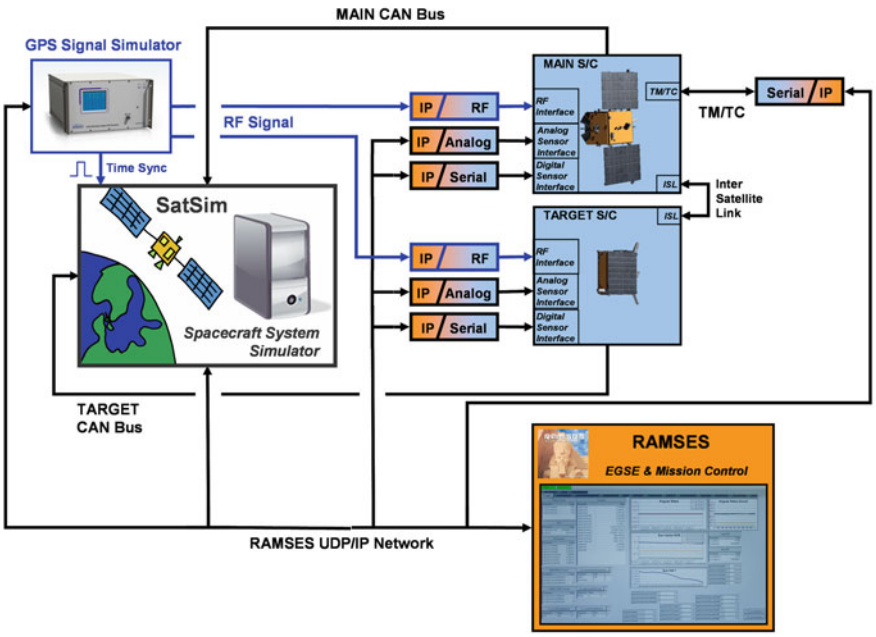


Fig. 21.9 Integrated GNC closed-loop tests using Mango and Tango flight models with GPS hardware in the loop

computer for each of the satellites. In the real case, the boards are connected to the sensor and actuator interface electronics via a CAN bus. The SatLab environment includes a simulator, SatSim [41], which simulates the interface electronics and includes models of all on-board sensors and actuators, spacecraft attitude and orbital dynamics, and space environment for both satellites.

The SFPT included series of open-loop and closed-loop tests performed on the flight model spacecraft. The tests were based on different test scenarios constructed to represent key phases of the mission such as the initial acquisition, commissioning, and the different experimental phases. A number of test cases is executed within the framework of each scenario. The test cases are constructed to verify each on-board function used within the scenario. The whole test setup is commanded and monitored over the UDP/IP network from the OHB-SE developed command and checkout software RAMSES [42]. This system is also used for flight control during the PRISMA mission. The test cases are implemented as PLUTO (Procedure Language for Users in Test and Operations) test scripts. The whole test setup is highly modular and can be configured to include anything from only some of the onboard hardware in the loop up to the complete on-board data handling system with all corresponding interface electronics. In the most complex configuration, the setup includes platform unit simulators (PUSIMs) representing the FFRF-sensors as well as a DLR's RF GPS signal simulator (GSS) that can be used dynamically in closed-loop [43]. For this purpose, SatSim was modified to

take an external time reference interrupt signal from the GSS. Figure 21.9 shows this closed-loop test setup. The scenario based test approach is also constructed to be highly modular so that the same test script can be used in several scenarios. Since RAMSES is also used in flight operations, several of the test scenario scripts have evolved into flight operational procedures.

21.4.3 Mission Operations

The PRISMA mission operations concept is highly modular and flexible. The functional organization consists of a Mission Control Center (MCC) and several experiment control teams (ECT). The MCC was located in Solna, Sweden at the premises of OHB-SE for most of the mission duration, whereas the ECTs were situated at different locations around the world (e.g., DLR/GSOC, CNES, DTU). For approximately 5 months, between March and July 2011, the MCC has been cloned and relocated at DLR/GSOC in an effort to support the mission, and prolong its lifetime. The temporary handover of the PRISMA mission operations to DLR/GSOC was successful thanks to the flexibility of the operations concept and the long experience of DLR/GSOC which, at that time, was already operating the GRACE and TanDEM-X formation flying missions (c.f. Chaps. 19 and 13).

The MCC crew comprised one mission manager and up to three mission experts, while the experimenters were of variable number. As a baseline, mission operations were conducted using the Esrange ground-station in Kiruna, Sweden, which provided an average of ten contacts per day. As mission control center, DLR/GSOC managed to introduce additional ground stations in the loop, namely Weilheim in Germany and Inuvik in Canada. This allowed the planning of multiple and overlapping passages within the same orbit to achieve 30 min long contacts with the spacecraft.

The PRISMA mission operations included the following three phases:

- Launch and early operations (LEOP) and Commissioning phase (66 days).
- Basic mission (388 days).
- Extended mission (to end of mission).

The initial checkout of on-board equipment, the verification of essential on-board functions, and the calibration of navigation algorithms such as attitude, rate estimators, and GPS navigation were conducted during the LEOP and Commissioning. As indicated in Table 21.4, during the first 66 days of the Commissioning Phase, PRISMA was operated as a combined spacecraft where Tango was still mated to Mango. Given the knowledge of the relative position between the centers of mass of the two spacecraft, this phase provided a unique opportunity to verify the quality and accuracy of the precise GPS POD products generated post-facto on-ground. Most of the sensors and actuators embarked on Mango got the opportunity to be powered on and to deliver the first data during this phase, including FFRF, DVS, and the propulsion systems.

Table 21.4 PRISMA mission timeline flown from June 15, 2010 to September 12, 2011

Phase	Activity	Start (dd-mm-20yy)	Finish	Duration (days)	Responsible	Usage
LEOP & commissioning (66d)	LEOP	15-06-10	16-06-10	2	OHB-SE	–
	Combined commissioning	17-06-10	14-07-10	18	OHB-SE	–
	Standby ops (Summer)	05-07-10	02-08-10	29	OHB-SE	–
	TANGO separation	03-08-10	15-08-10	13	OHB-SE	–
	GPS calibration 1	16-08-10	19-08-10	4	DLR	Primary
Basic mission (388d)	HPGP 1	20-08-10	23-08-10	4	ECAPS	Primary
	Microthruster 1	24-08-10	26-08-10	3	NANOSPACE	Primary
	GPS calibration 2	27-08-10	29-08-10	3	DLR	Primary
	AFF early harvest	30-08-10	14-09-10	16	OHB-SE	Primary
	VBS & FFRF validation	30-08-10	14-09-10	16	DTU/CNES	Passenger
	FFRF initialize	15-09-10	17-09-10	3	CNES	Primary
	AFC 1	20-09-10	05-10-10	16	DLR	Primary
	FFRF envelope 1	06-10-10	12-10-10	7	CNES	Primary
	HPGP 2/1	13-10-10	14-10-10	2	ECAPS	Primary
	PROX GPS 1	15-10-10	26-10-10	12	OHB-SE	Primary
	FFRF GNC 1/1	27-10-10	31-10-10	5	CNES	Primary
	ARV coop 1	01-11-10	10-11-10	10	OHB-SE	Primary
	HPGP 2/2	11-11-10	18-11-10	8	ECAPS	Primary
	HPGP 3/1	19-11-10	22-11-10	4	ECAPS	Primary
	PROX GPS 2	23-11-10	06-12-10	14	OHB-SE	Primary
	HPGP 3/2	07-12-10	13-12-10	7	ECAPS	Primary
	PROX/FARM VBS 1	14-12-10	20-12-10	7	OHB-SE	Primary
	Standby ops (Christmas)	21-12-10	10-01-11	20	OHB-SE	–
	AFF completion 1	11-01-11	17-01-11	7	OHB-SE	Primary
	HPGP 3/3	18-01-11	19-01-11	2	ECAPS	Primary
	PROX/FARM VBS 2	20-01-11	07-02-11	19	OHB-SE	Primary

(continued)

Table 21.4 (continued)

Phase	Activity	Start (dd-mm-20yy)	Finish	Duration (days)	Responsible	Usage
	FFRF envelope 2	08-02-11	11-02-11	4	CNES	Primary
	FFRF GNC ½	12-02-11	13-02-11	2	CNES	Primary
	ARV non-coop 1	14-02-11	24-02-11	11	OHB-SE	Primary
	FFRF GNC 2	25-02-11	26-02-11	2	CNES	Primary
	ARV non-coop 2	27-02-11	09-03-11	14	OHB-SE	Primary
	FFRF passenger far	27-02-11	09-03-11	14	CNES	Passenger
	FFRF GNC 3	10-03-11	14-03-11	5	CNES	Primary
	AFC 2	16-03-11	03-04-11	19	DLR	Primary
	ARV coop 2	04-04-11	06-04-11	3	OHB-SE	Primary
	AFF completion 2	07-04-11	13-04-11	7	OHB-SE	Primary
	HPGP 4/1	14-04-11	17-4-11	4	ECAPS	Primary
	PROX/FARM VBS 3	18-04-11	20-04-11	3	OHB-SE	Primary
	Standby ops (Easter)	21-04-11	25-04-11	5	OHB-SE	–
	HPGP 4/2	26-04-11	01-05-11	6	ECAPS	Primary
	PROX/FARM VBS 4	02-05-11	06-05-11	5	OHB-SE	Primary
	HPGP 4/3	07-05-11	25-05-11	19	ECAPS	Primary
	PROX/FARM VBS 5	26-05-11	08-06-11	14	OHB-SE	Primary
	GSOC experiment 1	09-06-11	17-07-11	39	GSOC	Primary
	AOK	18-07-11	16-08-11	30	DLR	Primary
	GSOC experiment 2	17-08-11	23-08-11	7	GSOC	Primary
	HPGP 5	24-08-11	12-09-11	20	ECAPS	Primary

After a 13-day preparation campaign, Tango was successfully separated from Mango on August 15, 2010. This triggered a first GPS navigation calibration campaign based on the reference POD. The careful tuning of the on-board navigation filter took a few days and resulted in an update of the extended Kalman filter settings. These were uploaded to the spacecraft and verified in the second GPS calibration slot between August 27 and 29, 2010. The proper calibration of the GPS relative navigation functionalities was followed by the verification of the central formation flying GNC modes, namely Safe and AFF, and by the official start of the Basic mission phase.

As of September 12, 2011, the Basic mission phase has come to an end (c.f. Table 21.4). All experiments needed to accomplish the primary mission objectives have been successfully executed. The Mango and Tango spacecraft have been reconfigured for the different experiments mostly within the ISL range domain (<30 km). The minimum separation between the spacecraft (after separation) has been reached on January 25, 2011 during the GPS-based PROX/FARM-2 slot. During this experiment Mango approached autonomously Tango in forced motion control from a distance of 20 m to hold a constant position in the orbital frame at 2 m distance for about 5 min and then recede back to a safer orbit.

The maximum separation between Mango and Tango instead has been achieved at the end of the AOK experiment on August 16, 2011. In this phase Mango tried to maintain its semi-major axis to track an external reference trajectory irrespective of the motion of Tango. The resulting drift between the satellites induced an along-track separation of approximately 50 km after 30 days of operations.

All relative navigation sensors available on-board (i.e., GPS, FFRF, VBS) have demonstrated the capability to support closed-loop control of the relative motion at various ranges of operations and with decimeter to centimeter level navigation accuracy throughout the mission.

The project has entered the Extended mission phase in September 2011, including additional operational GNC experiments with old and new partners. The formation will evaporate at the end of this phase which includes also the final propellant depletion activities. For completeness, Table 21.4 shows the full PRISMA mission timeline as flown from June 15, 2010 to the end of the Basic mission on September 12, 2011. Each experiment slot is characterized by the Activity name (second column), a Start and Finish date (third and fourth column), a Duration (fifth column), a Responsible for the experiment (sixth column) and the envisaged Usage in terms of experiment slot precedence (last column). "Primary" means that the Responsible has the responsibility of shaping the experiment in terms of relative orbit trajectory and if necessary also in terms of selecting the attitude profile of the Mango spacecraft. The Responsible organization has also full control over the propellant usage during the experiment. In contrast, an experiment in Passenger mode is executed without any possibility to shape the experiment other than by configuring an instrument to be on, a piece of software to be active, and/or some telemetry packets to be enabled. These configurations must of course be within existing power, data handling, and downlink budgets.

21.5 Representative Flight Results

21.5.1 Relative Navigation

The PRISMA GPS-based navigation system has been commissioned during the first 2 months of the mission in the time frame between the launch of the satellites on June 15, 2010 and the official start of the formation flying experiment timeline in August 2010. In the absence of an independent relative navigation sensor with sufficient technology readiness level and flight heritage, the characterization of the on-board GPS navigation throughout the mission had to rely mainly on comparisons with POD products generated post-facto on the basis of the same GPS data processed on-board. The strategy adopted to commission the GPS navigation system on PRISMA included three fundamental steps. First, the GPS hardware system on Mango and Tango (c.f. Fig. 21.3) was checked out in terms of functionality, GPS signal acquisition and carrier-to-noise density levels. Secondly, the capability and performance of the on-ground POD facility was verified when Mango and Tango were still clamped in a combined configuration. Here, the known relative position between the spacecraft center of mass served as a reference to evaluate the accuracy potential of the precise orbit determined on-ground. Finally, the POD products gave the possibility to evaluate the real-time navigation errors and tune the on-board navigation filter in dedicated calibration efforts to obtain the best trade-off between robustness and accuracy [44, 45].

Up to the end of the Basic Mission, nearly 1.5 years of autonomous operations based on on-board carrier-phase differential GPS has been completed and high navigation performance could be demonstrated both in formation keeping with passive relative orbits as well as proximity operations with forced motion control. Moreover the cross-comparison of relative GPS with other experimental relative navigation sensors embarked on PRISMA (i.e., FFRF and VBS) has led to a centimeter accurate calibration in orbit of novel RF and optical metrology systems which are considered as key technologies for future multi-satellite missions.

Figure 21.10 provides the typical navigation accuracy obtainable through carrier-phase differential GPS on the PRISMA mission. As compared with the POD product, the relative position and velocity errors are about 5.12 cm and 0.21 mm/s (3D, rms) respectively and thus well below the formal requirement of 0.2 m and 1 mm/s defined at the beginning of the mission. It is noted that the tuning of the navigation filter plays a key role in the achievable accuracy, and especially the weights of measurement noise and a-priori standard deviation of the empirical accelerations have to be carefully considered. As an example the filter settings applied during the considered scenarios introduce relative large empirical accelerations in radial direction as compared to the other components. The resulting loose constraint on the relative dynamics causes larger errors in radial direction, but is shown to be beneficial for the absolute orbit determination accuracy which amounts to 2 m and 7.5 mm/s (3D, rms) for position and velocity respectively.

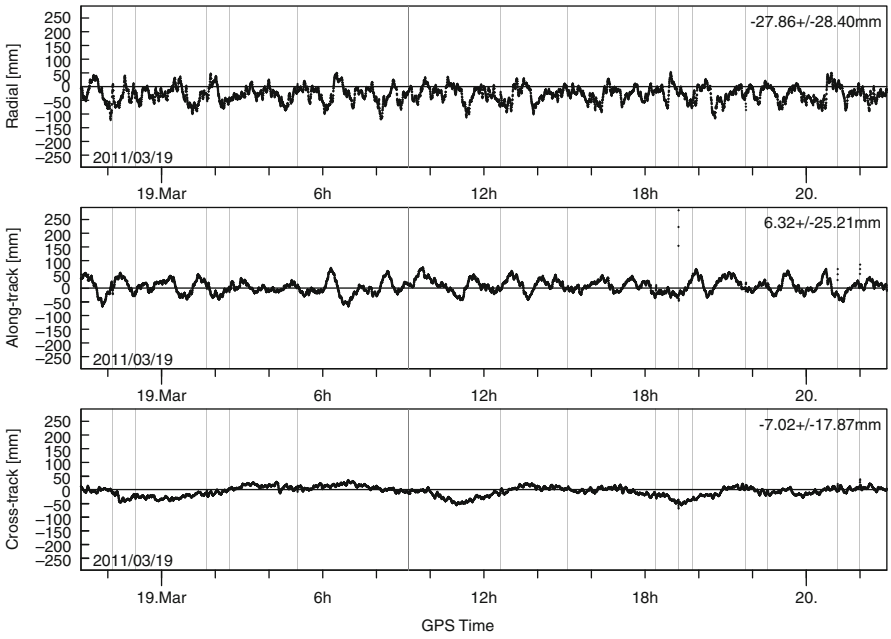


Fig. 21.10 Difference of real-time on-board navigation solution versus post-facto on-ground precise orbit product for PRISMA on March 19, 2011. Orbit control maneuvers for autonomous formation control are indicated by vertical lines. Relative position errors of 5.12 cm (3D, rms), and relative velocity errors of 0.21 mm/s (3D, rms) are obtained

More results on the demonstrated accuracy of the GPS-based relative navigation system are presented in Sect. 5.4.3.

The FFRF sensor and the related navigation functionalities were checked-out, validated, and tuned throughout the mission in a progressive approach through Passenger experiments, and Primary experiments in open- and closed-loop (c.f. Table 21.4). First flight results of the FFIORD experiment have been presented in Ref. [46]. The navigation performance has been evaluated for ranges between 20 m and 30 km through the systematic comparison of on-board relative position and velocity estimates with the post-facto on-ground POD products. Several GNC functions (i.e., RDV, CA, SBY, PROX, c.f. Sect. 21.3.4) and distance ranges (from 7 km in RDV down to 20 m in PROX) have been exercised to complete the characterization of the sensor throughout mission representative operations. The evaluated navigation accuracy spans a range between a few centimeters (3D, rms) during PROX at 20 m separation, up to a few meters cross-track error during RDV activities from 7 km. During formation separation, the cross-track error builds up to 12 m at 30 km range whereas the in-plane error does not exceed 1 m (2D, rms) and this illustrates the strong link between performance and orbital dynamic axes coupling. As expected, the angular errors do not change significantly in the different GNC modes (~0.1°), but it is rather the relative position error that is amplified by the increase of range. The main performance limiting factor has been identified as

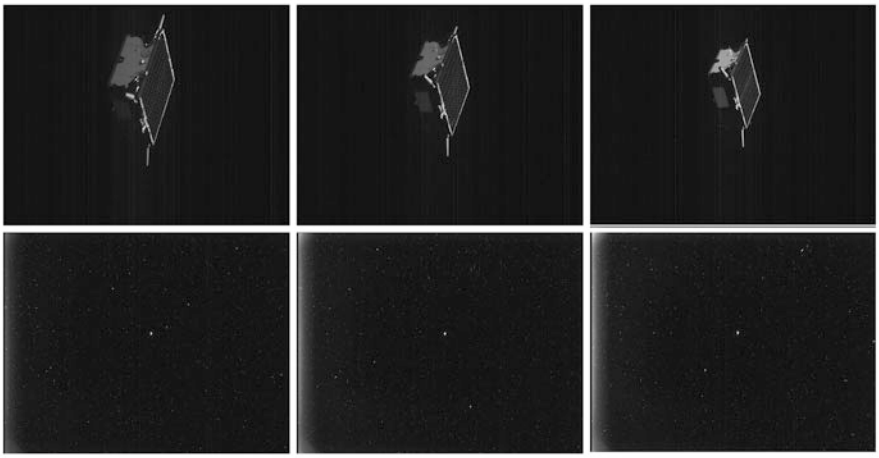


Fig. 21.11 VBS-Short range camera images on October 7, 2011 at 20 m inter-spacecraft separation (*top*). VBS-Far range camera images on October 5, 2011 at 20 km inter-spacecraft separation (*bottom*)

the FFRF line-of-sight bias variation, which is caused by un-modeled multi-path effects during large attitude maneuvers ($>30^\circ$), and temperature variations also due to rotations around the line-of-sight. A specific complexity of FFRF-based navigation is represented by the initialization and initial convergence procedures. As demonstrated in orbit, after completion of the integer ambiguity resolution process and the estimation of a first relative state through least squares filtering, the EKF reaches steady state accuracy in typically half an orbit (3,000 s) in the absence of orbit control maneuvers.

The VBS sensor was commissioned as a Passenger experiment in the time frame between August and November 2010 (c.f. Table 21.4). A brief in-flight characterization of the VBS-Far camera functionality can be found in Ref. [47]. In Far range operational mode (c.f. Sect. 21.2.2), the VBS processing unit is able to isolate a single object from the VBS-Far images, identify it as the target satellite, and provide a line-of-sight vector in either the camera body mounted frame or the inertial reference frame. The latter information is only available if the inertial spacecraft attitude can be determined from visible stars in the same image. A comparison of the sensor output with the POD reference for large separations (~10 km) provided a typical match to within 5–10 arcsec for target intensities below 10,000 Data Numbers (DN), with peaks up to 40–80 arcsec for larger intensities. Assuming a POD accuracy at the decimeter level (2 arcsec), and considering the tight correlation of the error pattern with the target intensity, the primary sources of error in this comparison are believed to be the knowledge of the camera orientation and field of view errors of the VBS attitude measurements. Peaks in the intensity are produced by Sun light flashes under specific orientations of reflective items. Both functionality and performance deteriorate when the size of the target becomes

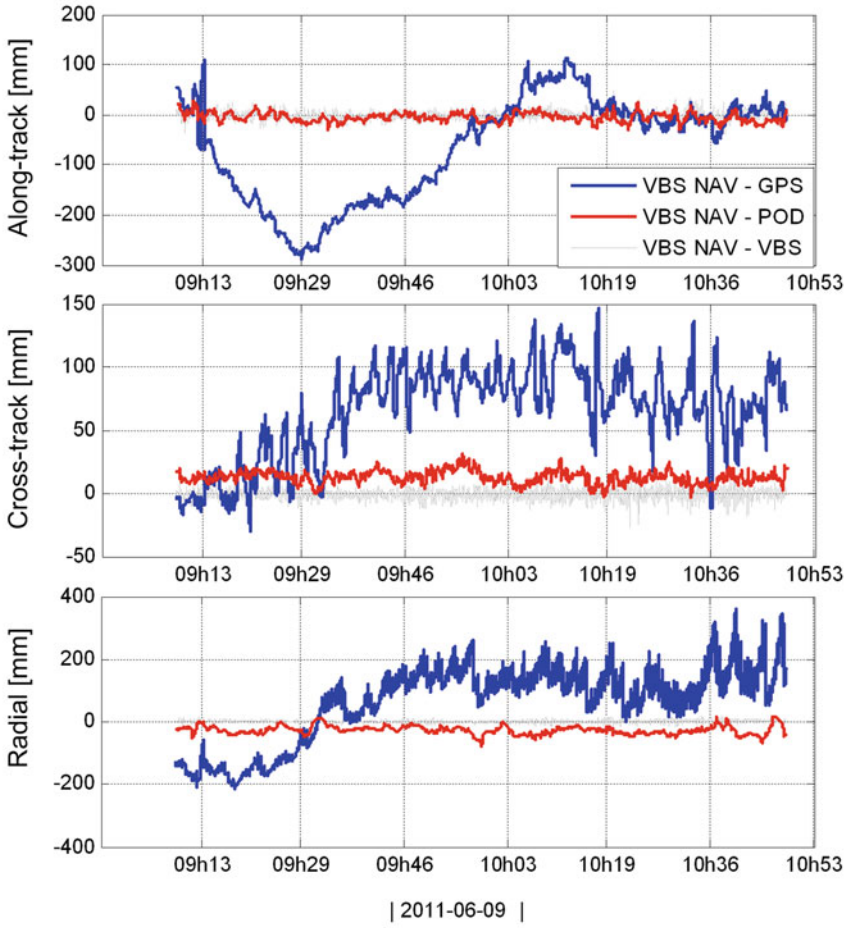


Fig. 21.12 Difference of real-time on-board VBS navigation solution (filtered) versus on-board relative GPS (*blue*), post-facto on-ground POD (*red line*), and VBS sensor measurements output (*gray*) on June 9, 2011. Overall navigation consistency at the centimeter level is obtained during forced motion control at 10 m separation of Mango from Tango

noticeably large on the VBS-Far image. At 50 m separation the delivered line-of-sight switches between several body features (e.g., radiators and antennas), introducing systematic errors of over a degree. Despite the identified limits and reduced robustness, autonomous approaches based on VBS-Far navigation were successfully performed from 5 to 30 km to delivery points at hundred meters separation (c.f. Sect. 21.5.3). Typical images collected in orbit using the VBS-Far and -Short range cameras are shown in Fig. 21.11.

The potential of the VBS-Short camera has been fully exploited in June 2011 during the PROX/FARM VBS five experiment slot (c.f. Table 21.4). Here precise

station-keeping of Mango is performed at 10 m along-track separation from Tango. Figure 21.12 illustrates the comparison between the on-board VBS-based navigation filter output with both the post-facto on-ground POD and the on-board relative GPS navigation. In addition, also the difference between VBS Kalman-filtered and non-filtered on-board solutions is shown. First of all, the plot demonstrates a consistency between VBS and reference POD at the centimeter level. Small offsets in radial and cross-track directions (c.f. Fig. 21.12, red line) are most probably due to uncertainties in the VBS-Short camera orientation and the phase center position of the GPS antennas on Mango and Tango. Secondly, during this scenario, the performance of optical navigation is shown to be better than the on-board relative GPS (c.f. Fig. 21.12, blue line). The degraded GPS relative navigation solution (30 cm maximum error) is due to the heavy maneuvering activity, corresponding to thrusters activations every 50 s. Indeed the on-board GPS orbit determination relies on the knowledge of the systems dynamics and can hardly incorporate delta-v maneuvers at a frequency similar to the sample time of the extended Kalman filter (30 s).

21.5.2 Propulsion Systems

Results from the in-space flight demonstration of the HPGP propulsion system are described in Ref. [48]. The commissioning of the propulsion system took place on June 23–24, 2010, 8 days after launch. The first firing sequence was a pulse train of forty 100 ms pulses at a duty cycle of 1%. The maneuver estimated by the DLR's POD ground facility confirmed a total delta-v variation of 2.1 cm/s and the HPGP was thus declared ready for operations.

Since then, all planned firings of the propulsion system have been successfully completed and calibrated through GPS precise orbit determination. On September 9, 2011 a 75 s continuous firing was performed, the longest yet with a HPGP thruster in space. The accumulated number of pulses exceeded 50,000, 63% of the propellant has been consumed through a total of 363 fire sequences. After 15 months in space, no indications of degradation have been observed and the flight demonstration has evolved into an actual space flight qualification.

The HPGP propulsion system has been operated in several operational firing modes, from Quasi Steady-State (continuous firing) to Pulse Mode (duty factors between 0.15% and 50%), from Off-Modulation (duty factors between 50% and 99%) to Single Pulse (very low duty factors).

The HPGP thruster performance measured in space is coherent with the measurements performed on ground in the near vacuum test stand. Back-to-back in-space comparison between the HPGP and hydrazine propulsion systems shows superior specific and density impulse for the HPGP. The comparison is performed at comparable thrust levels. For Quasi Steady-State the improvement over hydrazine specific impulse is 6% and 12% respectively at begin and end of life. In Single Pulse mode the improvement over hydrazine is 10% and 20% respectively, whereas

in Pulse Mode the HPGP and hydrazine performance are comparable. A theoretical improvement of 6% was expected between HPGP and hydrazine performance, however the back-to-back in space comparison demonstrates higher performance in most cases.

The Micro-propulsion experiment has been conducted in August 24–28, 2010. The experiment consisted in demonstrating cold gas micro-thrusters in the mN range, with highly innovative thruster design etched in silicone wafers and with miniaturized gas management system. This flight demonstration represented the only PRISMA set-back. When commissioning the system, involving opening up of the miniaturized latch-valve between the gas tank and the pressure regulator and thrusters, it appeared as though the pressure had disappeared. The suspicion that this actually could have happened had already been raised during the early days of the mission. During one 24 h period, 2 days into the mission, a strange behavior of the onboard momentum management function indicated an external torque on the system. The temperature of the Micro-propulsion tank dropped slightly in the beginning of this period. Moreover, empirical accelerations estimated by the precise on-ground orbit determination process matched suspiciously the expected effects of a gas leakage. Despite the anomaly, the micro-propulsion team managed to demonstrate that the thruster hardware and control electronics functioned as expected, opening and closing micro-valves and managing thermal control in the thruster pods. Hundreds of cycles have been commanded to the individual thrusters and telemetry data showed nominal signatures of voltages, currents and temperatures. Unfortunately, this potential success could not be manifested in a real thrust demonstration due to the lacking tank pressure. The Micro-propulsion experiments were stopped on August 28, 2010.

21.5.3 Closed-Loop Orbit Control Experiments

Flight results from the GNC experiments conducted by OHB-SE during the PRISMA mission have been presented in Refs. [47, 49, 50]. Here a brief summary of the achievements from the usage of the three dedicated modes (i.e., AFF, PROX/FARM, and ARV, c.f. Sect. 21.3.2) is provided. AFF represents the only GNC mode on-board PRISMA that after the experimental phase has been used on a routine basis for an extended period of time. AFF has accumulated over 4 months of closed-loop cooperative satellite formation-flying time. About 20 days have been dedicated to AFF experiments, while the rest has been operational routine formation-flying, where AFF has been used to support other experiments based on GPS navigation.

Over 110 different formation reconfigurations have been performed so far. The operational range has been from 30 km down to 10 m. Figure 21.13 summarizes the complete AFF usage in PRISMA between August 2010 and April 2011. The illustration is not intended to provide details of the individual trajectories, but rather to provide a fair visualization of the operational range of autonomous

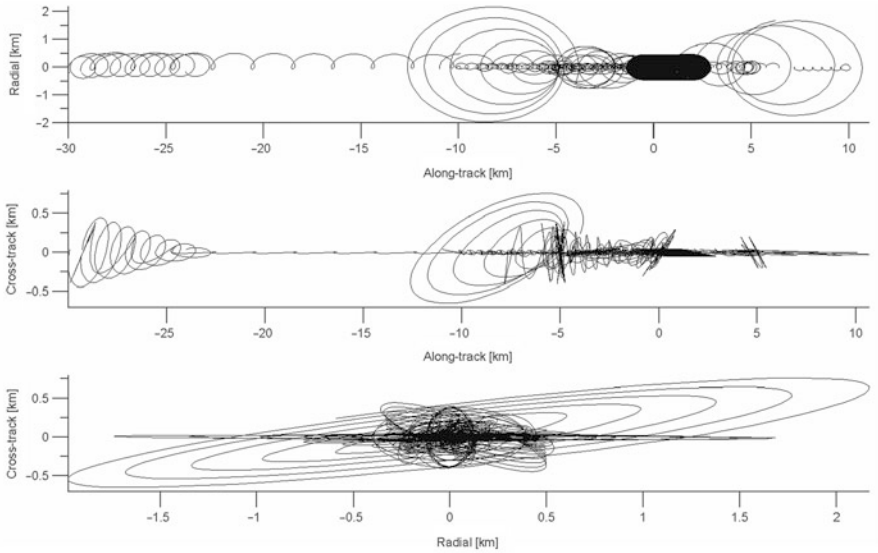


Fig. 21.13 Autonomous Formation Flying (AFF) trajectories flown by Mango w.r.t. Tango during the PRISMA basic mission (Adapted from Ref. [49])

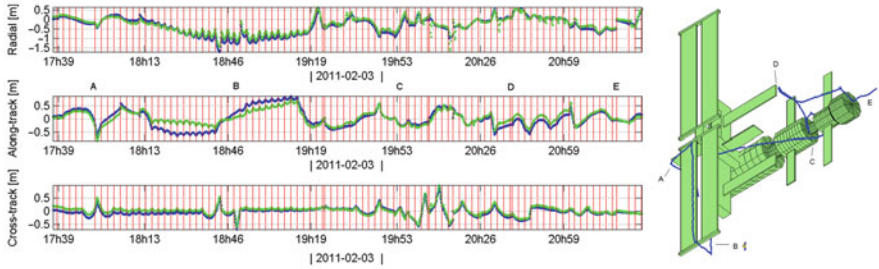


Fig. 21.14 Flight trajectory (*right*) and corresponding control tracking errors (*left*) based on POD (*blue*) and on-board GPS navigation (*green*) on February 3, 2011. *Red vertical lines* represent on-board thruster's activations. An image of Mango has been superimposed in correct scale right next to the trajectory arc B (Adapted from Ref. [49])

GPS-based formation flying activities tested during PRISMA. In contrast to the impulsive and sparse orbit control activities of AFF, PROX/FARM operations consist of forced-motion flight of Mango around Tango in the ranges from about 50 m down to 2 m. The forced motion is either directly around Tango or around a virtual structure defined about the Tango spacecraft. The purpose of the virtual structure is to mimic the circum-flight about a large object with appendages and no-fly zones, such as the ISS or the Hubble space telescope.

Representative flight results from a virtual inspection tour of the ISS are illustrated in Fig. 21.14. Compared to other flown scenarios, here the required delta-v to track a quickly changing and highly non natural trajectory is much larger (i.e., 0.97 m/s in 2.4 orbits with 98 thrusts) and poses challenges in terms of GPS-based navigation, non-convex guidance optimization, and non-linear controller behavior.

The applied MPC box for this scenario is 0.5/1.0/0.5 m respectively in radial/along-track/cross-track, whereas the control period is set to 150 s. As shown in Fig. 21.14, the achieved control tracking errors remain within the prescribed control box most of the time with a performance of approximately 0.5 m (3D, rms) when compared with a reference POD.

Finally two key experiments were performed by OHB-SE in ARV mode to demonstrate autonomous rendezvous with a passive non-cooperative target using optical line-of-sight navigation in April and November 2011 [47]. The ARV mode consists of an initial search for the target, a subsequent centering of this target in the field of view of the VBS-Far camera, several orbital revolutions of angle-only orbit determination, a first alignment of the two orbits to a specified control tolerance, and the final along-track approach to the delivery point from target.

Each of these functionalities has been exercised in orbit leading to errors of ca. 1–2 km in the on-line estimation of the range from target after three orbits of on-board orbit determination at 30 km separation. The subsequent alignment and orbit closing phases led to a reduction of the range estimation error to below 5 m at a representative 50 m separation delivery point. The major challenge faced by the ARV experiment has shown to be the robustness against false or erroneous sensor data at all ranges of operations. Special attention had to be paid to incorrect first detections of target in the field of view at large separations and to systematic line-of-sight errors due to reflecting features of Tango at small separations.

Flight results from the SAFE closed-loop experiment executed by DLR in September 2010 and March 2011 are discussed in Refs. [44, 51]. SAFE has demonstrated the capability of the DLR's GNC subsystem to establish, maintain, and reconfigure arbitrary baselines in space in full autonomy for future remote sensing applications employing formation flying spacecraft. The applied impulsive relative orbit control scheme has been shown to be fuel efficient and passively safe. Moreover the adopted parameterization in terms of relative eccentricity and inclination vectors facilitated all mission phases, from the GNC design, analysis, and implementation, all the way to experiment planning, operations, and post-facto evaluation. A total of 22 different formations have been prescribed via telecommand during the 35-days flight experiment which have all been smoothly acquired and accurately maintained without ground-intervention. The formation geometries span a wide range and are characterized by mean along-track separations between 0 and 5 km, and oscillation amplitudes in radial and cross-track direction between 0 and 400 m. A minimum separation of 20 m between Mango and Tango was acquired and maintained through anti-parallel relative eccentricity/inclination vectors on Match 21, 2011.

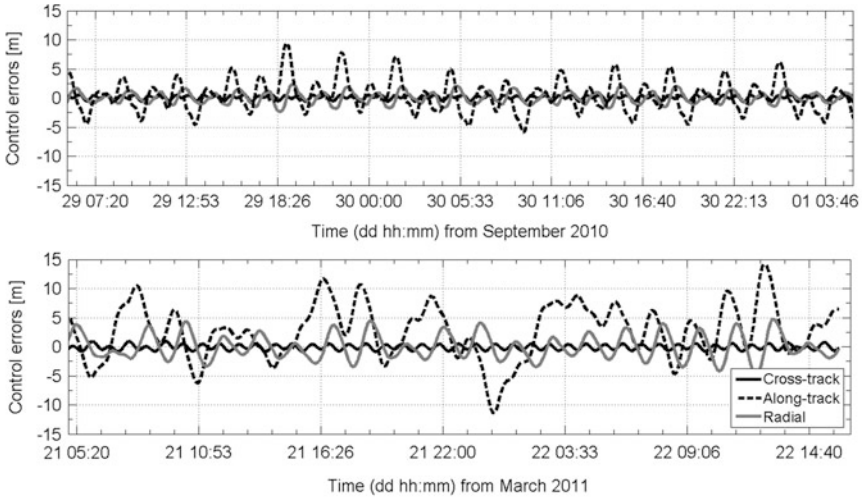


Fig. 21.15 Control tracking errors during maintenance of similar formations through pairs of radial (*top*) and along-track maneuvers (*bottom*) (Adapted from Ref. [51])

The average duration of fine formation keeping for each configuration was 13.4 orbits (22.3 h) for a total of 294.7 orbits (20.4 d) or ca. 58.3% of the experiment duration. The maneuver cycle for in-plane control or the distance between consecutive pairs of maneuvers was two orbital revolutions in average. The demonstrated formation keeping accuracies spanned a range from 1.1/1.7/0.1 m in Radial/Along-track/Cross-track directions at ca. 100 m separations to 30.6/38.0/1.4 m at 5 km along-track distance. Representative control tracking errors during formation maintenance are depicted in Fig. 21.15 and are clearly dominated by second order relative dynamics effects which are neglected by the adopted relative motion model.

Formation reconfigurations were characterized by average convergence times in cross-track direction of 1.2 orbits, and two cross-track maneuvers were in general sufficient to obtain any desired correction of the relative inclination vector. The average convergence time in radial and along-track directions is correlated and amounted to 2.4 orbits for the relative eccentricity vector. Typically 2–3 pairs of in-plane maneuvers were sufficient to enter the fine formation keeping phase. Mean along-track reconfigurations required longer convergence times of 3.7 orbit revolutions in average. The total commanded delta-v during SAFE amounted to 38.58 cm/s for formation keeping and 2.63 m/s for formation reconfiguration. Typical images collected in orbit during the SAFE fly-around and inspection phase using the DVS camera are shown in Fig. 21.16.

The AOK experiment represented the DLR's secondary mission objective (c.f. Sect. 21.3.3). It was executed successfully from July 18 to August 16, 2011, and demonstrated precise autonomous absolute orbit control of a single spacecraft.

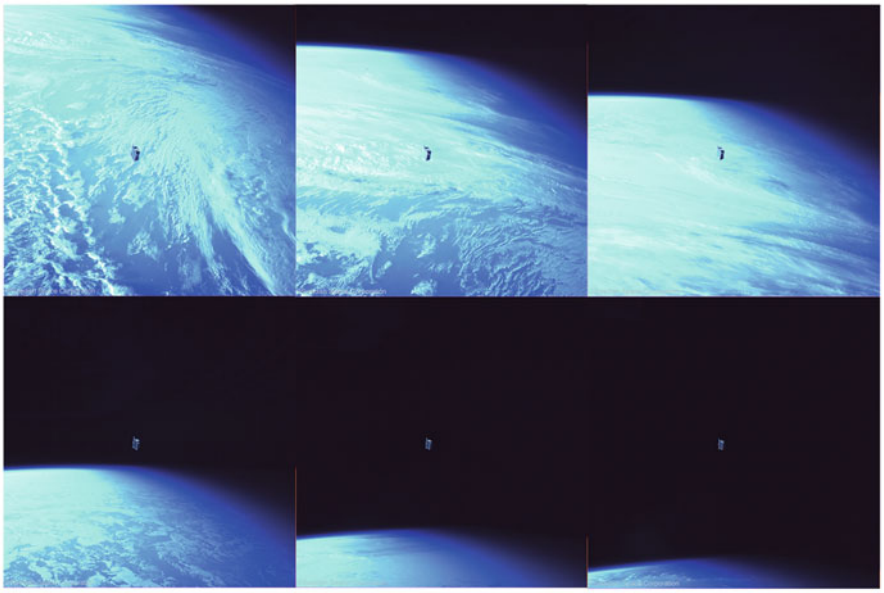


Fig. 21.16 Fly-around/inspection of Tango on March 27, 2011 during SAFE experiment. Inter-spacecraft separations between 30 and 60 m. 75 s image rate. DVS camera

The main performance requirement was a control accuracy of 10 m (1σ) on the LAN with a maneuver budget constraint of 0.5 m/s. The position and velocity vector of Mango at 01:00 UTC on July 18, as estimated by the GPS based POD process, have been taken as initial state for the generation of the reference orbit. In a first phase of the experiment, AOK demonstrated reference orbit acquisition capability reducing gradually the LAN deviation from 300 m to within the prescribed control window. After the convergence phase, at steady-state, the control accuracy requirement was fulfilled with slight margin. The mean value of the LAN error controlled by AOK was -3.6 m with a standard deviation of 9.5 m during the fine control phase. The total delta-v spent during the entire experiment was 0.1347 m/s.

A comprehensive discussion of the FFIORD closed-loop results based on FFRF navigation is available in Refs. [52, 53]. Overall the CNES's objectives in terms of functional behavior and performance have all been achieved through 12 days of experiment operations and a total delta-v allocation of 6 m/s. The rendezvous (RDV) activities cover six approaches from distances of 4–7 km to a few hundreds meters and one deployment from 600 m to 5 km. The single rendezvous durations and delta-v consumptions covered a spectrum from 7 to 14 orbits and from 27 to 9 cm/s respectively. Typical rendezvous accuracies at the delivery point stayed below 10 m and 10 mm/s with propellant consumption in line with expectation values obtained from ground simulations. The standby

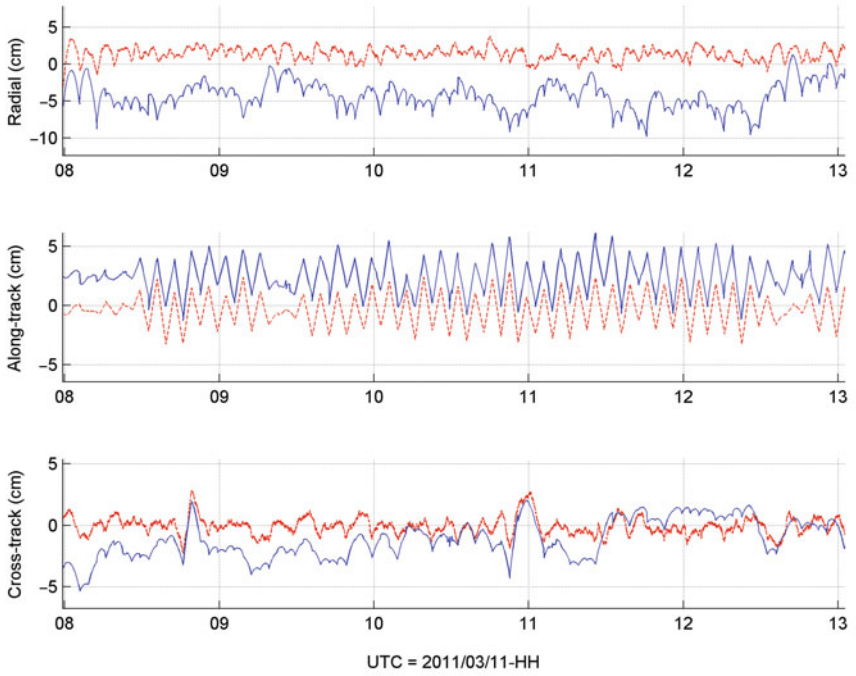


Fig. 21.17 Control tracking errors based on POD (*blue*) and on-board FFRF navigation (*red*) during PROX activities performed at 20 m distance on March 11, 2011 with sub-pulse mode activated

function (SBY) was activated during 40 orbits for an overall cost of approximately 7 cm/s. The demonstrated formation keeping accuracy amounts to a mean of 30 cm in cross-track direction and to a few meters in radial and along-track directions. A total of four successful collision avoidance (CA) maneuvers have been performed with the Mango relative trajectory exiting the prescribed keep-out-zone a few seconds after the maneuver. The ultimate navigation and control accuracy obtained in the FFIORD experiment has been demonstrated by the proximity operations activities (PROX). In order to remove the limitation given by the minimum impulse bit of the propulsion system (i.e., 100 ms thruster burn time), PROX featured a so-called sub-pulse mode where the thrusters were operated in a differential way to realize the requested inputs. Station keeping at 80 and 20 m separations has been exercised with quasi-continuous activations of the thrusters (200 s control cycle). As depicted in Fig. 21.17, navigation and control errors based on the reference POD showed accuracy at the centimeter level in both bias and standard deviation. Although an ideal fixed relative pointing of the FFRF sensor has been imposed to mitigate temperature induced

biases, these GNC accuracies are probably the best that have been achieved during the PRISMA mission so far.

21.6 Conclusion

The growing interest in autonomous formation flying and on-orbit servicing technologies is demonstrated by the deployment in orbit and by the successful flight demonstrations of the last years. The ETS-VII experiment initiated the trend in 1997 with a first validation in low Earth orbit of rendezvous and docking techniques. Orbital Express went a step further in 2007 by demonstrating rendezvous, inspection, and servicing capabilities. TanDEM-X was launched in 2010 and constitutes nowadays the first synthetic aperture radar interferometer involving two formation flying spacecraft at distances down to few hundred meters.

In this context, with its start in June 2010, PRISMA represents the most recent and significant step in the advancement of formation flying and rendezvous techniques. The originality of the mission stems from the integration of a large variety of navigation and guidance algorithms along with new sensors and actuators, and from the implementation of numerous flight tasks within complex operational scenarios in a fully autonomous manner. In addition to the technological breakthrough, PRISMA represents an international success, thanks to its multilateral nature with contributions and independent experiments conducted by OHB-SE, DLR, CNES and DTU.

After more than 1 year in orbit, all primary and secondary objectives of PRISMA have been achieved, and valuable results have been collected in various disciplines which are now under further and deeper investigation. As a matter of fact, PRISMA can be considered as a mission of many firsts. From a relative navigation point of view, PRISMA flies the first GPS differential carrier-phase-based real-time navigation system (SAFE), the first RF based relative metrology subsystem (FFRF), and a novel vision-based relative navigation system (VBS) in closed-loop. Cross-comparisons between relative navigation sensors and post-facto on-ground precise orbit determination have demonstrated real-time navigation accuracies at the sub-decimeter level even during frequent orbit maneuvering activities.

From an actuation perspective, PRISMA embarks the first high performance green propellant propulsion system (HPGP), which has been demonstrated to be superior to hydrazine by 6–10% in specific impulse at similar thrust levels. Novel micro-electro-mechanical propulsion technology has also been shown to function as expected in orbit, even though a leakage of the cold-gas tank prevented an actual thrust demonstration.

PRISMA has shown to be an ideal test bench for formation guidance and control algorithms. It is the first time that arbitrary passive relative orbits are controlled autonomously on-board making use of relative orbital elements and in particular relative eccentricity and inclination vectors. Furthermore novel model predictive control schemes based on the Yamanaka-Ankersen equations have been flight

proven for the first time during sparse orbit maneuvering, forced-motion quasi-continuous control, and collision avoidance activities. Algorithms for autonomous rendezvous of a servicer vehicle with respect to a passive non-cooperative target using only optical line-of-sight measurements have also been demonstrated during the mission.

Overall PRISMA has shown that the major phases of a complete formation flying mission in low Earth orbit can be autonomously performed using state-of-the-art sensors while assuring centimeter-level positioning accuracy. Functional behavior and performances are fully consistent with what is expected from a first-stage metrology system. Potential next customers for these technologies are either cooperative formation flying projects like PROBA-3 (ESA), or non-cooperative on-orbit-servicing missions like DEOS (DLR). In the former case, GPS, FFRF, and VBS metrologies are to be implemented for formation acquisition, maintenance and collision avoidance whereas optical metrology will take over for precision maneuvers. In the latter case, both GPS and VBS relative navigation metrologies are possible candidates to support experiment validation, monitoring, and collision avoidance on one hand (GPS) and autonomous rendezvous to a non-cooperative target on the other hand (VBS).

Flight data and user experience acquired during the PRISMA experiments are a most valuable resource for the design of the navigation approach and for the prediction of related performances. Using the vast collection of data available in various scenarios, representative measurements can be provided for typical trajectories and performances inferred via simulation. Whatever their actual utilization in the coming years, the PRISMA achievements have brightened the future of formation flying by breaking a psychological barrier: maintaining satellites in close vicinity has been demonstrated in a routine manner with limited ground support and at reasonable cost. Lessons learned will therefore provide useful guidance for future multi-satellite projects.

References

1. Seguela D, Brigitte T, Feuillerac L (2011) DEFI TECHNOLOGIQUE, LE VOL EN FORMATION/Pushing the edge of technology, formation flying. Special report, CNES MAG, no 48, Jan 2011, pp 38–53
2. Reintsema D, Thaeter J, Rathke A, Naumann W, Rank P, Sommer J (2010) DEOS – the German robotics approach to secure and de-orbit malfunctioned satellites from low earth orbits. In: Proceedings of the i-SAIRAS 2010, Sapporo, Japan, 29 Aug–1 Sep 2010
3. Bristow J, Folta D, Hartman K (2000) A formation flying technology vision. In: Proceedings of the AIAA space conference, Long Beach, CA. AIAA Paper No. 2000–5194
4. Berge S, Jakobsson B, Bodin P, Edfors A, Persson S (2005) Rendezvous and formation flying experiments within the PRISMA in-orbit testbed. In: Proceedings of the ESA GNCS, Louthraki, Greece
5. Persson S, Jacobsson B, Gill E (2005) Prisma - demonstration mission for advanced rendezvous, formation flying technologies and sensors. In: Proceedings of the 56th international astronautical congress, Fukuoka, Japan. IAC-05-B5.6.B.07, 17–21 Oct 2005

6. Persson S, Veldman S, Bodin P (2009) PRISMA – a formation flying project in implementation phase. Acta Astronaut 65(9–10):1360–1374. doi:10.1016/j.actaastro.2009.03.067
7. Bodin P, Larsson R, Nilsson F, Chasset C, Noteborn R, Nylund M (2009) PRISMA: an in-orbit test bed for guidance, navigation, and control experiments. J Spacecr Rocket 46(3):615–623. doi:10.2514/1.40161
8. Montenbruck O, Markgraf M (2006) User's manual for the phoenix GPS receiver. DLR/GSOC GTN-MAN-0120, version 1.7
9. D'Amico S, Ardaens J-S, Montenbruck O (2009) Navigation of formation flying spacecraft using GPS: the PRISMA technology demonstration. In: Proceedings of the ION-GNSS-2009, Savannah, USA, 22–25 Sep 2009
10. D'Amico S, De Florio S, Larsson R, Nylund M (2009) Autonomous formation keeping and reconfiguration for remote sensing spacecraft. In: Proceedings of the 21st ISSFD, Toulouse, France, 28 Sep–2 Oct 2009
11. De Florio S, D'Amico S (2009) The precise autonomous orbit keeping experiment on the PRISMA mission. J Astronaut Sci 56(4):477–494
12. Ardaens J-S, Montenbruck O, D'Amico S (2010) Functional and performance validation of the PRISMA precise orbit determination facility. In: Proceedings of the ION international technical meeting, San Diego, CA, 25–27 Jan 2010
13. Harr J, Delpech M, Lestarquit L, Seguela D (2006) RF metrology validation and formation flying demonstration by small satellites – the CNES participation on the PRISMA mission. In: Proceedings of the 4S symposium small satellites, systems and services, ESA SP-625
14. Lestarquit L, Harr J, Grelier T, Peragin E, Wilhelm N, Mehlen C, Peyrotte C (2006) Autonomous formation flying RF sensor development for the PRISMA mission. In: Proceedings of the ION-GNSS-2006, Forth Worth, TX, 26–29 Sep 2006
15. Harr J, Delpech M, Grelier T, Seguela D, Persson S (2008) The FFIORD experiment - CNES' RF metrology validation and formation flying demonstration on PRISMA. In: Proceedings of the 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, Noordwijk, The Netherlands, 23–25 Apr 2008
16. Benn M, Jørgensen JL (2008) Short range pose and position determination of spacecraft using a μ-advanced stellar compass. In: Proceedings of the 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, Noordwijk, The Netherlands, 23–25 Apr 2008
17. Anflo K, Möllerberg R (2009) Flight demonstration of new thruster and green propellant technology on the PRISMA satellite. Acta Astronaut 65(9–10):1238–1249. doi:10.1016/j.actaastro.2009.03.056
18. Grönland T-A, Rangsten P, Nese M, Lang M (2007) Miniaturization of components and systems for space using MEMS-technology. Acta Astronaut 61(1–6):228–233. doi:10.1016/j.actaastro.2007.01.029
19. Capuano G, Severi M, Cacace F, Lirato R, Longobardi P, Pollio G, DeNino M, Ippolito M (2008) Video system for prisma formation flying mission. In: Proceedings of the IAA symposium on small satellite systems and services (4S), Rhodes, Greece, 26–30 May 2008
20. Brinkfeldt K, Enoksson P, Wieser M, Barabash S, Emanuelsson M (2011) Microshutters for MEMS-based time-of-flight measurements in space. In: Proceedings of the 24th international conference on micro electro mechanical systems (MEMS), doi: 10.1109/MEMSYS.2011.5734495
21. Gaisler J (2001) The LEON processor user's manual, version 2.3.7
22. D'Amico S, Gill E, Garcia M, Montenbruck O (2006) GPS-based real-time navigation for the PRISMA formation flying mission. In: Proceedings of the 3rd ESA workshop on satellite navigation user equipment technologies, NAVITEC'2006, Noordwijk, The Netherlands, 11–13 Dec 2006
23. Delpech M, Guidotti P-Y, Djalal S, Grelier T, Harr J (2009) RF based navigation for PRISMA and other formation flying missions in earth orbit. In: Proceedings of the 21st ISSFD, Toulouse, France, 28 Sep–2 Oct 2009

24. Denver T, Jørgensen JL, Michelsen R, Jørgensen PS (2006) MicroASC star tracker generic developments. In: Proceedings of the 4S Symposium, Sardinia, Italy, Sep 2006
25. Benn M, Jørgensen JL (2009) Range management of a vision based rendezvous and docking navigation sensor. In: Proceedings of the international astronautical conference (IAC), IAC-09.C1.4.4, 2009
26. Anflo K, Persson S, Thormählen P, Bergman G, Hasanof T, Grönland T-A, Möllerberg R (2006) Flight demonstration of an ADN-based propulsion system. In: Proceedings of the 57th IAC/IAF/IAA (International astronautical congress), Valencia, Spain, IAC-06-C4.1.08, 2–6 Oct 2006
27. Larsson R, Berge S, Bodin P, Jönsson U (2006) Fuel efficient relative orbit control strategies for formation flying rendezvous within PRISMA. Adv Astronaut Sci, 125:25–40; also American Astronautical Society Paper 06–025, Feb 2006
28. Larsson R, Mueller J, Thomas S, Jakobsson B, Bodin P (2008) Orbit constellation safety on the prisma in-orbit formation flying test bed. In: Proceedings of the 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, Noordwijk, The Netherlands, 23–25 Apr 2008
29. Stephen GN, Ariela S (1996) Linear and nonlinear programming. International editions, McGraw-Hill Series in Industrial Engineering and Management Science
30. Yamanaka K, Ankersen F (2002) New state transition matrix for relative motion on an arbitrary elliptical orbit. J Guid Control Dynam 25(1):60–66
31. Tillerson M, Inalhan G, How J (2002) Co-ordination and control of distributed spacecraft systems using convex optimization techniques. Int J Robust Nonlin Contr 12:207–242
32. D'Amico S, Montenbruck O (2006) Proximity operations of formation flying spacecraft using an eccentricity/inclination vector separation. J Guid Control Dynam 29(3):554–563
33. D'Amico S, Gill E, Montenbruck O (2006) Relative orbit control design for the PRISMA formation flying mission. In: Proceedings of the AIAA guidance, navigation, and control conference, Keystone, CO, 21–24 Aug 2006
34. De Florio S, D'Amico S (2009) Optimal autonomous orbit control of remote sensing spacecraft. In: Proceedings of the 19th AAS/AIAA space flight mechanics meeting, Savannah, USA, 8–12 Feb 2009
35. Berges J-C, Cayeux P, Gaudel-Vacaresse A, Meyssignac B (2007) CNES approaching guidance experiment within FFIORD. In: Proceedings of the ISSFD 2007, Annapolis, USA
36. Cayeux P, Raballand F, Borde J, Berges J-C, Meyssignac B (2007) Anti-collision function design and performances of the CNES formation flying experiment on the PRISMA mission. In: Proceedings of the 20th international symposium on space flight dynamics, Annapolis, USA, 24–28 Sep 2007
37. Wermuth M, Montenbruck O, Helleputte van T (2010) GPS high precision orbit determination software tools (GHOST). In: Proceedings of the 4th international conference on astrodynamics tools and techniques, Madrid, 3–6 May 2010
38. Teunissen PJG (1995) The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. J Geod 70(1–2):65–82
39. Ardaens J-S, D'Amico S, Montenbruck O (2010) Flight results from the PRISMA GPS-based navigation. In: Proceedings of the NAVITEC'2010, 5th ESA workshop on satellite navigation technologies, Noordwijk, The Netherlands, 8–10 Dec 2010
40. Bodin P, Noteborn R, Larsson R, Chasset C (2011) System test results from the GNC experiments on the PRISMA in-orbit test bed. Acta Astronaut 68(7–8):862–872. doi:10.1016/j.actaastro.2010.08.021
41. Nylund M, Bodin P, Chasset C, Larsson R, Noteborn R (2011) SATSIM an advanced real-time multi satellite simulator handling GPS in closed-loop tests. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Quebec, 18–20 May 2011

42. Carlsson A (2007) A general control system for both sounding rockets and satellites. In: Proceedings of the 18th ESA symposium on European rockets an balloon programmes and related research, Visby, 3–7 June 2007
43. D'Amico S, De Florio S, Ardaens J-S, Yamamoto T (2008) Offline and hardware-in-the-loop validation of the GPS-based real-time navigation system for the PRISMA formation flying mission. In: Proceedings of the 3rd international symposium on formation flying, missions and technology, ESA/ESTEC, Noordwijk, 23–25 Apr 2008
44. D'Amico S, Ardaens J-S, DeFlorio S (2010) Autonomous formation flying based on GPS – PRISMA flight results. In: Proceedings of the 6th international workshop on satellite constellation and formation flying, Taipei, Taiwan, 1–3 Nov 2010
45. Ardaens J-S, D'Amico S, Montenbruck O (2011) Final commissioning of the PRISMA GPS navigation system. In: Proceedings of the 22nd international symposium on spaceflight dynamics, Sao Jose dos Campos, Brazil, 28 Feb–4 Mar 2011
46. Grelier T, Guidotti P-Y, Delpech M, Harr J, Thevenet J-B, Leyre X (2010) Formation flying radio frequency instrument: first flight results from the PRISMA mission. In: Proceedings of the 5th ESA workshop on satellite navigation technologies and European workshop on GNSS signals and signal processing (Navitec 2010), 8–10 Dec 2010
47. Noteborn R, Bodin P, Larsson R, Chasset C (2011) Flight results from the PRISMA optical line of sight based autonomous rendezvous experiment. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Quebec, 18–20 May 2011
48. Anflo K, Crowe B (2011) In-space demonstration of high performance green propulsion and its impact on small satellites. In: Proceedings of the 25th annual AIAA/USU conference on small satellites, Logan, UT, 8–11 Aug 2011
49. Larsson R, Noteborn R, Bodin P, D'Amico S, Karlsson T, Carlsson A (2011) Autonomous formation flying in LEO seven months of routine formation flying with frequent reconfigurations. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Quebec, 18–20 May 2011
50. Larsson R, D'Amico S, Noteborn R, Bodin P (2011) GPS navigation based proximity operations by the PRISMA satellites - flight results. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Quebec, 18–20 May 2011
51. D'Amico S, Ardaens J-S, Larsson R (2011) Spaceborne autonomous formation flying experiment on the PRISMA mission. In: Proceedings of the AIAA guidance, navigation, and control conference, Portland, USA, 8–11 Aug 2011
52. Guidotti P-Y, Delpech M, Berges J-C, Djalal S, Grelier T, Seguela D, Harr J (2011) Flight results of the FFIORD formation flying experiment. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Quebec, 18–20 May 2011
53. Delpech M, Guidotti P-Y, Grelier T, Seguela D, Berges J-C, Djalal S, Harr J (2011) First formation flying experiment based on a radio frequency sensor: lessons learned and perspectives for future missions. In: Proceedings of the 4th international conference on spacecraft formation flying missions & technologies, St-Hubert, Quebec, 18–20 May 2011

Chapter 22
Formation for Atmospheric Science and Technology Demonstration

Jian Guo, Daan Maessen, and Eberhard Gill

Abstract FAST (Formation for Atmospheric Science and Technology demonstration) is a cooperative Dutch-Chinese formation flying mission led by Delft University of Technology (TU Delft) in the Netherlands and Tsinghua University in China. FAST shows the potential of a first international micro-satellite formation flying mission in three distinct fields: technology demonstration, earth science and space education. Here, the FAST mission is presented. The mission scenario consisting of different formation flying stages is described, and the system design of both the space and the ground segments is introduced, with emphasis on Dutch contributions. Some key technical issues related to autonomous formation flying are also addressed.

22.1 Introduction

Space missions consisting of multiple small satellites flying in formation are now often considered as a feasible and logical alternative to the traditional approach of using a large and complex monolithic spacecraft. Important arguments in favour of space missions with Formation Flying (FF) satellites are improved spatial coverage, enhanced functionality, risk mitigation, flexibility, and reduced spacecraft complexity.

An example of the opportunity arising from the usage of small satellites flying in formation is provided in this chapter. The information results from the Formation for Atmospheric Science and Technology demonstration (FAST) mission, a joint project that was initiated in 2007 between Delft University of Technology (TUD) in The Netherlands and Tsinghua University in China [1].

J. Guo (✉) • D. Maessen • E. Gill
Delft University of Technology, Space Systems Engineering, Kluyverweg 1,
Delft 2629 HS, The Netherlands
e-mail: j.guo@tudelft.nl; D.C.Maessen@tudelft.nl; e.k.a.gill@tudelft.nl

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*,
Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_22,

FAST is a mission for the synoptic evaluation of local, regional and global aerosol data and altitude profiles of the cryosphere utilizing two formation flying micro-satellites. The primary mission objectives, which are all equally important, are:

- Demonstrate Autonomous Formation Flying (AFF) by means of various communication architectures, distributed propulsion systems and Micro Systems Technologies (MSTs);
- Characterize atmospheric aerosols, monitor the variation of height profiles in the cryosphere, and correlate these data for improved scientific return; and
- Teach cutting-edge technology, broaden the international view of students and boost skills through the exchange of students and staff.

In this chapter the details of the FAST mission are presented. The rest of the chapter is organized into five parts. The first part describes the mission scenario, highlighting different formation flying phases and operational modes. The second part details the system design of both the space and the ground segments, with emphasis on Dutch contributions. The third part describes scientific applications of the FAST mission, including the characterization of atmospheric aerosols, the seasonal variation of height profiles in the cryosphere, and the correlation between the two. The forth part deals with key technical issues related to autonomous formation flying. Design challenges related to systems engineering, inter-satellite link, distributed computing, and cooperative control are addressed. The fifth part briefly discusses the educational opportunities of the FAST mission.

22.2 Mission Scenario

The baseline orbit for the two satellites, i.e. FAST-D and FAST-T, is a sun synchronous orbit (SSO) at 650 km altitude with a local time of the ascending node (LTAN) of approximately 10:00 h. The satellites are foreseen to be launched as piggyback payloads on a Chinese launcher. Each satellite has been assigned a total velocity change (ΔV) budget of 12 m/s. This includes a 100% contingency to allow for failure of the propulsion system on one of the satellites [2]. The 2.5 years mission is divided into three phases, each with a distinctly different orbital geometry [3]. The rationale for and the orbital geometry during these phases are treated in the following.

22.2.1 Phase 1: LEOP and Initial Commissioning

The first mission phase is the Launch and Early Operations Phase (LEOP), which consists of the launch, orbit insertion, detumbling, acquisition of a safe attitude pointing mode, and establishment of two-way communications with the ground stations. After LEOP, the 1 month long initial commissioning phase starts.

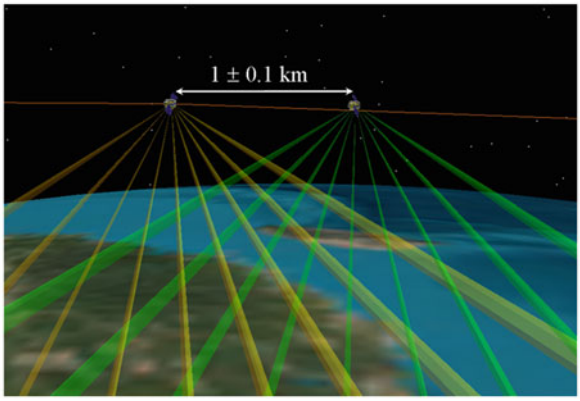


Fig. 22.1 The technology demonstration phase

In this phase the satellites will not fly in formation, thus allowing differential aerodynamic drag to cause a large along-track separation of the two satellites. The current estimate for the along-track separation distance after 1 month is 60 km, requiring a bang-bang formation initialization maneuver costing 0.2 m/s to start the next mission phase.

During the initial commissioning phase the individual satellite subsystems and payloads will be tested and calibrated. The characterization of the attitude control, differential GPS, inter-satellite communication, and propulsion subsystems are of vital importance to allow transitioning to the next mission phase.

22.2.2 Phase 2: Technology Demonstration

During technology demonstration mode, cf. Fig. 22.1, the two satellites will demonstrate the ability to autonomously maintain an along-track separation of 1 ± 0.1 km by means of Autonomous Formation Flying (AFF) using propellant optimized distributed propulsion. This capability requires a propulsion system on each spacecraft as well as an inter-satellite communication link. For solar maximum conditions, it will take the satellites slightly over 1 day to traverse the 200 m control window of this mission phase. To maintain the desired formation geometry, taking into account differential drag only, a ΔV of 2 mm/s is then required to bring the inter-satellite distance back to the other extreme of the control window.

Although the orbital geometry is not optimized for it, some first science observations will be performed during this phase of the mission. Due to the relatively small inter-satellite distance, this phase also lends itself well to perform distributed computing experiments between the two spacecraft and to perform cross-calibration between the instruments on the different spacecraft. The AFF demonstration will be performed for several weeks and will be followed by a transition to the required relative geometry for the science phase.

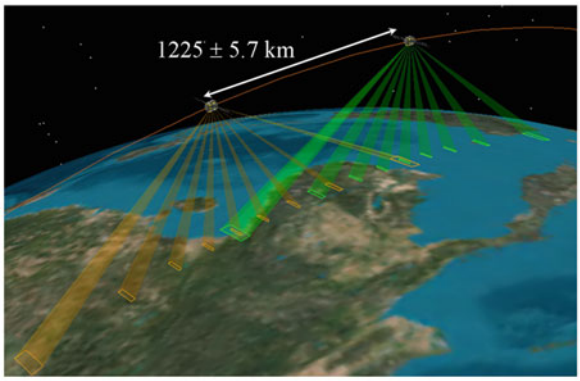


Fig. 22.2 The science observation phase

22.2.3 Phase 3: Science Observation

In this mission phase, the along-track separation between the satellites will be 1,225 ± 5.7 km, as shown in Fig. 22.2. Since natural drift of the formation will require approximately 100 days before this geometry is achieved, preference is given to a controlled maneuver requiring a total ΔV of 2 m/s and taking slightly longer than 2 days. For 1 year of formation maintenance, taking only differential drag into account, approximately 1.2 m/s of ΔV is required and a maneuver needs to be executed once every 9 days.

The orbital geometry of this mission phase lends itself extremely well to perform synoptic and synergetic observations with the spectropolarimeters on both spacecraft. The orbital geometry combined with the nine Earth-looking Fields-Of-Views (FOVs) of the spectropolarimeter results in many simultaneous intersections of the spectropolarimeter FOVs and in several overlapping spectropolarimeter FOVs at the Earth's surface. This allows retrieval of aerosol characteristics at specific altitudes at a single moment in time and it allows making more observations of geolocations from various angles during a single pass (especially near the equator), which is highly desired for aerosol characterization. The ±5.7 km accuracy in the inter-satellite range is driven by the ground pixel size of the spectropolarimeter at 650 km altitude, which is approximately 11.4 km.

22.3 The System Design

In this section, the design of the FAST system is presented. The system architecture is briefly discussed, followed by information about the science payloads (a spectropolarimeter, an altimeter and a microwave radiometer). The system/subsystem design of FAST-D is specified along with an introduction of FAST-T. The architecture of the ground segment is also described with a focus on scientific data processing and distribution.

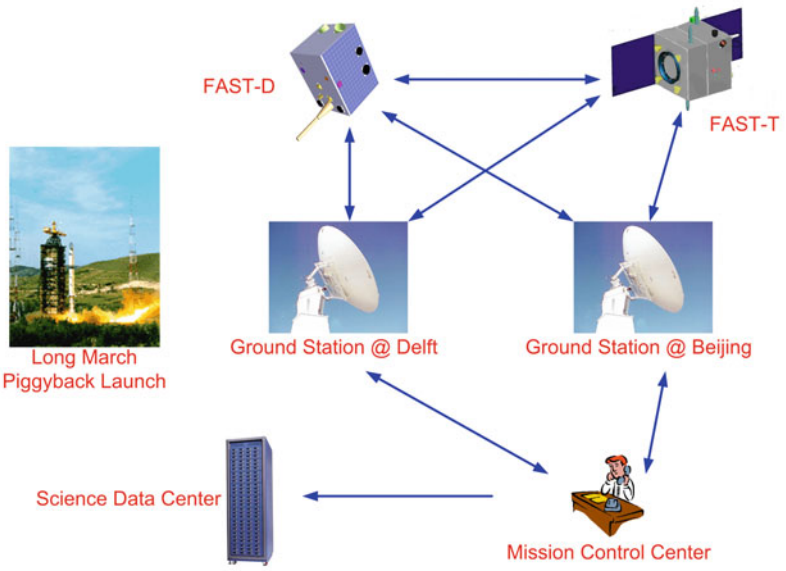


Fig. 22.3 The FAST system architecture

22.3.1 System Architecture

The architecture of the FAST system is schematically shown in Fig. 22.3 [4]. Like many other space systems, the FAST system consists of both space and ground segment. However, due to the feature of formation flying, the space segment is comprised of two spacecraft, i.e. FAST-D and FAST-T. During the procedure of formation flying, i.e. the scientific mode as well as the technology demonstration mode, the two spacecraft will communicate with each other either through a direct Inter-Satellite Link (ISL) or through ground stations. The distributed computing and other technologies will be implemented via the ISL. Details about the elements of this architecture, the ISL and other issues will be discussed in the remaining part of this chapter.

22.3.2 Payloads

In total three scientific payloads will be flown on the FAST mission: a Dutch spectropolarimeter (on both FAST-D and FAST-T spacecraft), the Dutch Stereo Imaging Laser AlTimeter (SILAT, on FAST-D), and the Chinese Dual-Frequency Microwave RadioMeter (DFMRM, on FAST-T).

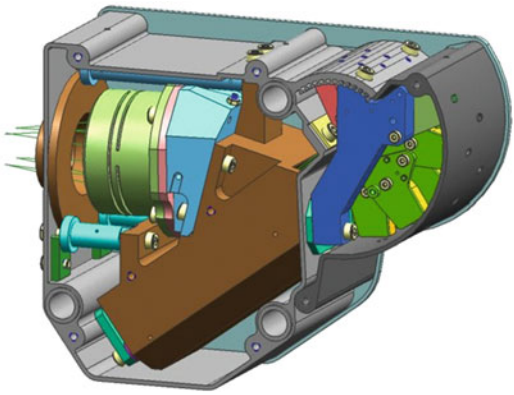


Fig. 22.4 SPEX opto-mechanical design [5]

22.3.2.1 Spectropolarimeter

The candidate for the spectropolarimeter payload is the Spectropolarimeter for Planetary Exploration (SPEX), cf. Fig. 22.4, which is under development by a Dutch consortium consisting of companies and knowledge institutes. Since SPEX is ultimately intended for Mars missions, a special version, i.e. SPEX-FAST, is specifically being designed for the FAST mission. An important modification of SPEX-FAST compared to SPEX-Mars is the change from seven planet-looking FOVs and two limb-looking FOVs to nine planet-looking FOVs. Reason for this change is that SPEX-FAST will focus on characterizing aerosols, leading to a preference for planet-looking FOVs over limb-looking FOVs. Due to SPEX' capability of accurate full linear spectropolarimetry without moving parts or liquid crystals, this spectropolarimeter is an ideal instrument for measuring and characterizing aerosols from space. The current opto-mechanical design of SPEX (without detector and electronics) has a mass of 0.9 kg and dimensions of 15 $\times$ 12 $\times$ 6 cm^3. The complete instrument is expected to have a volume of ~1 litre, a mass of 2 kg, and 2 W maximum power consumption [5].

22.3.2.2 SILAT

SILAT, cf. Fig. 22.5, is a highly integrated payload suite that is composed of three separate instruments contained in a thermally stable SiC frame; an innovative single photon counting Laser AlTimeter (LAT), a High Resolution Camera (HRC) and a Stereoscopic Camera (SCAM) [6]. It is under development by a consortium led by cosine Research. A structural re-design of the instrument has been made for the FAST mission, which results in considerable mass savings as compared to the earlier design for a mission to Jupiter's moon Europa. SILAT contains no moving parts, and all cooling devices are passive; only cold fingers and radiators are used. This increases the reliability of the payload suite, which is essential for space

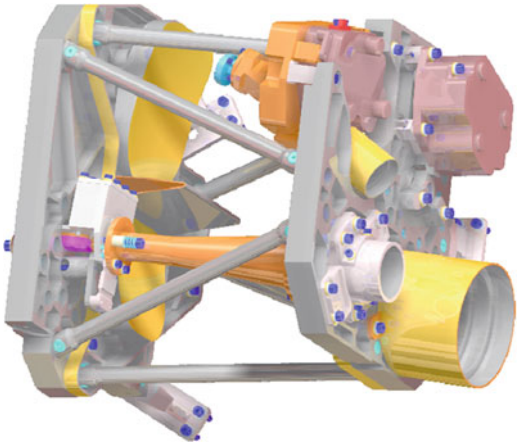


Fig. 22.5 SILAT [6]

missions. The optical filters used on the cameras have been optimized for the spectral response from water.

A breadboard prototype of the LAT has been completed by cosine Research and has been successfully tested both indoor and outdoor in early 2010, demonstrating a distance measurement resolution better than 15 cm [7]. An engineering model of SILAT, consisting of a mixture of space qualified and commercial-off-the-shelf (COTS) parts, is foreseen to undergo flight testing on a TU Delft laboratory aircraft. SILAT has a mass of 5.2 kg and measures $28 \times 28 \times 30$ cm^3. The maximum power consumption of SILAT is ~14 W with the LAT requiring ~12 W.

22.3.2.3 DFMRM

DFMRM is an instrument that measures global luminosity/temperature data at K-band (23.8 GHz) and Ku-band (37 GHz). Currently, DFMRM is under development by the General Establishment of Space Science and Application, Chinese Academy of Sciences (GESSA/CAS). The designed DFMRM is box-shaped with dimensions of $33 \times 27 \times 20$ cm^3, has a mass of no more than 15 kg and an average power consumption of less than 20 W (peak power $<$25 W). There are two observation channels on DFMRM: the primary one is nadir-oriented, and the secondary one points to cold space (off sunlight) for calibration purposes [4].

22.3.3 Spacecraft

The space segment of the FAST system consists of two satellites: FAST-D and FAST-T.

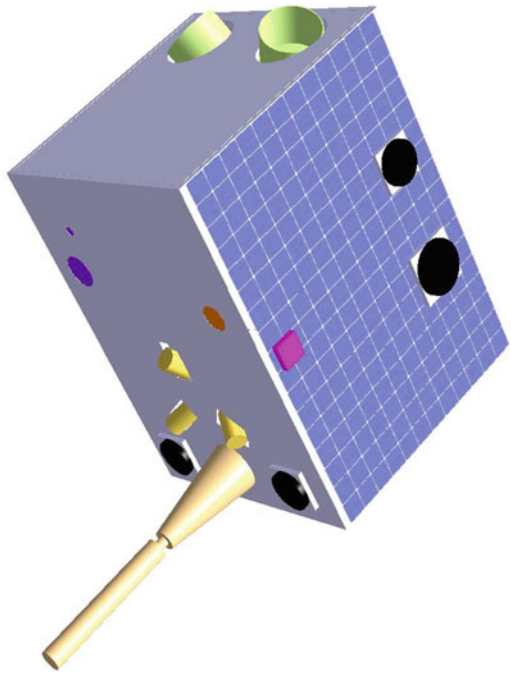


Fig. 22.6 FAST-D spacecraft

22.3.3.1 FAST-D

FAST-D is a 50 kg micro-satellite considered for development by Dutch academia and industries [8]. As shown in Fig. 22.6, FAST-D has a 0.5 × 0.5 × 0.7 m box-shaped structure with three body mounted solar panels on the zenith surface and the two along-track surfaces, respectively. The internal structure consists of multiple trays, which are used both for easy integration and for load carrying. On the nadir surface, the payload detector heads and a high-gain S-band data-downloading antenna are accommodated, as well as low-gain Telemetry and Tele-Command (TTC) antennas. The optical heads of the star tracker are installed on the surface that is parallel to the orbit plane.

FAST-D will be a synthesis and showcase of Dutch achievements on space-related Micro-Systems Technology (MST). Most of the components onboard are developed/under development by Dutch universities or industries, and some of them have been demonstrated in space. For instance, the wireless sun sensor developed by TNO has been flown on the Delfi-C3 nano-satellite [9]. From this aspect, FAST-D will not only be a platform for MST demonstration, but also to some extent a System-of-MicroSystems (SoMS).

The overall design of the FAST-D is schematized in Fig. 22.7 [10]. The subsystem design is briefly introduced as bellows.

Due to the requirements from the payloads, FAST-D's Attitude Determination and Control Subsystem (ADCS) shall be able to provide attitude information with at

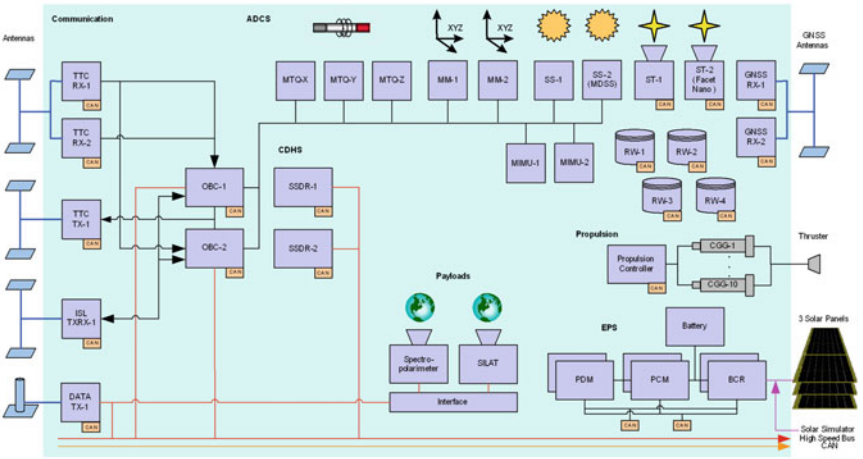


Fig. 22.7 Schematic of the FAST-D spacecraft

least 30 arcseconds accuracy around the various body axes using fine attitude sensors, and shall allow the control of attitude with a minimum pointing accuracy of 300 arcseconds.

The attitude determination is achieved with the use of two redundant 3-axis magnetometers (MM) and two micro sun sensors (SS) for coarse information. For precise determination two miniaturized CMOS star trackers (ST) and two Micro Inertial Measurement Units (MIMU) are used.

The attitude control is performed with reaction wheels (RW) and uses magnetoquers (MTR) for momentum dumping and coarse control. The four micro reaction wheels are mounted in a pyramid configuration maximizing control effectiveness, and the three magnetoquer rods are mounted in an XYZ configuration.

A double-loop Nonlinear Dynamic Inversion (NDI) technique will be adopted, which allows flexible control of a nonlinear system in different flying modes [8]. The double-loop NDI technique avoids strong nonlinear feedback which causes unnecessary actuator saturations, and provides flexibility for multiple operational modes by separating rate and angular control.

Two redundant Global Navigation Satellite System (GNSS) receivers will function as absolute navigation sensors for onboard orbit determination. The onboard orbit determination is further enhanced through uploading precise GNSS ephemerides from ground. The relative position and attitude information will be obtained by exchanging navigation and attitude data via inter-satellite communication.

FAST-D uses three body mounted solar panels as the primary power supply, and two Li-ion battery packages as the secondary power supply. Each solar panel is populated with triple-junction GaAs cells which have a minimum average efficiency of 26.8% at Begin-Of-Life (BOL). In total the power subsystem is able to provide at least an orbit average power of 45 W at BOL, 50% more than required.

The Command and Data Handling Subsystem (CDHS) consists of OnBoard Computers (OBC), data storage, data buses, and the relevant software. There are two OBCs on FAST-D, one for housekeeping and data processing, and the other for AOCS and formation flying. They are also hot redundant for each other in case of failure. The payload data are stored in a radiation tolerant solid state memory with a capacity of 8 GB. The data buses are composed of a high speed bus for payload data transfer and a low speed bus for command/control. The onboard software performs three primary functions: system boot, housekeeping, and AOCS/FF processing.

The communication subsystem, also called Radio Frequency (RF) subsystem, operates at S-band due to the payload data download requirement. The subsystem consists of two parts. One part is the Satellite-Ground Communication Module, which is composed of two hot redundant command receivers, one low power transmitter, one high power transmitter, and associated antennas. The low power transmitter is primarily used for telemetry, and the high power one is primarily for payload data downlink. However, for redundancy on the transmit chain, these two transmitters both can be used as the backup of the other one. The two command receivers share two patch antennas, which are mounted on the nadir and the zenith surfaces of the satellite, respectively. The low power transmitter feeds another two patch antennas. The high power transmitter is connected with a helical antenna.

Another part is the Inter-Satellite Communication Module, which is composed of a transceiver and two patch antennas. During FF, the satellites will exchange information, such as position and attitude, between each other through this link. As payload data are not intended to be exchanged, only a low power transceiver (ISL TXRX) and low-gain antennas are utilized.

Due to the requirements of a piggy-back launch, FAST-D should avoid the utilization of hazardous components, such as high energy propellant, high pressure vessels, etc. Meanwhile, the relatively large ΔV requirements and the volume limitation on the spacecraft also prevent FAST-D from utilizing conventional chemical propulsion techniques. Alternatively, a cold gas generator is used on FAST-D to replace the cold gas tank.

The propulsion subsystem of FAST-D is composed of a cold gas generator, an electronics board, an orbit control thruster, and the associated valves and tubes. It provides for initial launcher injection corrections and formation maintenance. The cold gas generator, developed by TNO and Bradford Engineering, consists of a solid block of chemical material inside a casing [11]. Once ignited, this solid material will decompose into the required gas that leaves the gas generator at ambient temperature. The major advantage of the cold gas generator over the gas tank is that it stores the propellant in solid state. Hence, no large volume, high-pressured tank and associated valves are needed; no risks for leakage; and the mass and volume of the complete propulsion subsystem are both optimized.

Currently nitrogen and oxygen generators are available. For FAST-D the nitrogen generator has been selected, because of its relatively high gas output efficiency

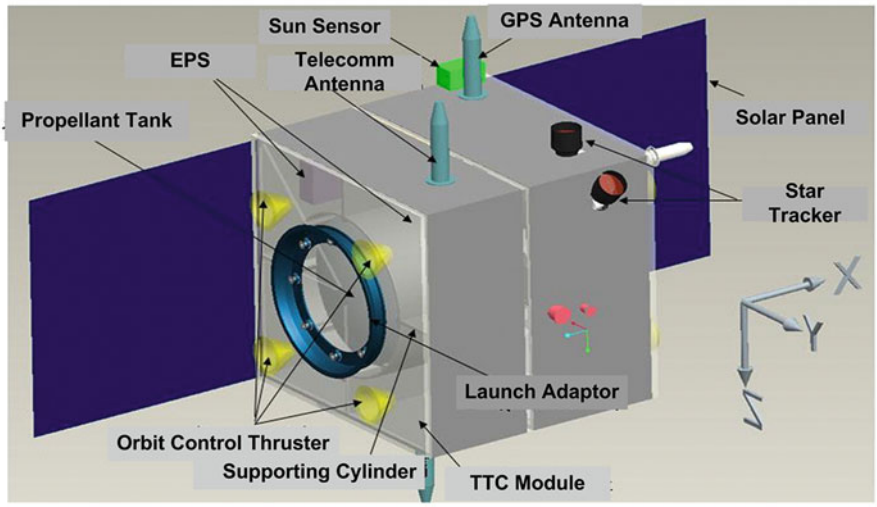


Fig. 22.8 FAST-T spacecraft

(each kg of solid propellant can output 260 liters of gas with the pressure range of 0.1–15 MPa) and its availability for space applications through its space qualification on ESA's PROBA–2 satellite.

The structure of FAST-D is to be manufactured from conventional aluminum alloy including the use of honeycomb panels. The primary structure consists of a launcher separation system, separation panel, avionics stack, payload panel, top panel, three solar panels and one nadir panel.

The overall spacecraft is designed to ensure a balanced thermal environment in all operating, eclipse and safe modes. Due to the utilization of the cold gas generator, the local heating for the propellant tank is not necessary anymore. Therefore, thermal control is achieved primarily by simple passive approaches.

22.3.3.2 FAST-T

FAST-T is a small satellite considered for development at Tsinghua University [4]. As shown in Fig. 22.8, FAST-T is a $0.75 \times 0.75 \times 0.73$ m cube with two deploy-fix solar panels, and the total mass is less than 130 kg.

The primary attitude mode of FAST-T is 3-axis stabilized nadir pointing with a control accuracy of better than 0.1° (3σ) and stability of better than 0.01°/s (3σ). In case of any critical fault, the satellite will automatically enter into a safe mode in which the negative Y axis will point to the Sun with an accuracy of better than 5° and a stability better than 0.05°/s (3σ).

22.3.4 Ground Segment

According to the architecture in Fig. 22.3, the ground segment of FAST consists of two ground stations, one Mission Control Center (MCC), and one Science Data Center (SDC).

In order to increase the window for satellite-ground access, two geographically distributed ground stations are planned: one in Delft, and one in Beijing. Both ground stations are equipped with 3 m paraboloid antennas and have identical functions: sending commands, receiving telemetry and payload data, and acting as backup route for ISL.

All ground operation decisions are made in a single MCC and then sent to the ground station for generating control commands, and all data received by ground stations are also initially analyzed in the MCC. The telemetry data will be further processed here, and the payload data will be sent to the SDC.

22.4 Scientific Applications

The FAST mission will carry several innovative small science payloads that enable research in several areas. These are the characterization of atmospheric aerosols, the seasonal variation of height profiles in the cryosphere, and the correlation between the two.

22.4.1 Aerosol Characterization

Aerosols are suspensions of fine solid or liquid droplets in a gas. The size of these particles can range from about 0.001 to 100 μm. While liquid aerosols will be spherical, solid aerosols will have irregular shapes which depend on their composition and their formation process. Primary types of terrestrial aerosols are oceanic sea salt, desert dust, volcanic ashes and volcanic sulfuric acid and various types of anthropogenic aerosols (e.g. black carbon). Most aerosol particles reside in the lower layers of the Earth's atmosphere, extending up to around 1 km. Aerosol particles can have significant effects on air quality, weather, and the climate. Aerosols affect the climate through scattering and absorption of solar radiation, scattering, absorption and emission of thermal radiation, acting as cloud condensation nuclei, and catalyzing chemical reactions [12].

Characterization of atmospheric aerosols is important since current uncertainties in the aerosol forcing and the total solar irradiance are so large that they preclude meaningful climate model evaluation [13]. Since the degree and direction of polarization of scattered light is very sensitive to the shape and composition of the scattering particle, spectropolarimetry lends itself extremely well to aerosol

characterization. However, this has never been done from space. The first mission that intended to do this was NASA's Glory mission, which failed to reach orbit after liftoff aboard a Taurus XL rocket on 4 March 2011 [14].

Additional significant scientific outputs on aerosol characterization are expected through the configuration of FF. As mentioned in Section 22.2, in the technology demonstration phase, the close FF will allow cross-calibration between payloads on the different spacecraft since they are essentially measuring the same properties at almost the same moment in time. Then in the science observation phase, the distance between the two FF spacecraft will be 1,225 $\pm$ 5.7 km. This orbital geometry lends itself extremely well to perform synoptic and synergetic observations especially with the spectropolarimeters on both spacecraft. The orbital geometry combined with the nine Earth looking FOV of the spectropolarimeter results in many simultaneous intersections of spectropolarimeter FOVs and in several overlapping spectropolarimeter FOVs at the Earth's surface, cf. Fig. 22.2. This allows retrieval of aerosol characteristics at specific altitudes at a single moment in time and it allows making more observations of geolocations from various angles during a single pass (especially near the equator), which is highly desired for aerosol characterization.

22.4.2 Cryosphere and Precipitation

A recent study by Shindell and Faluvegi [15] states that "[. . .] aerosols have had a large role in both global and regional climate change during the entire twentieth century [. . .] Arctic climate is especially sensitive to Northern Hemisphere short-lived pollutants. [. . .] Our calculations suggest that black carbon and tropospheric ozone have contributed [. . .] to Arctic warming since 1890, making them attractive targets for Arctic warming mitigation. In addition, they respond quickly to emissions controls, and reductions have ancillary benefits including improved human and ecosystem health." Thus, recent atmospheric models suggest a strong link between aerosols (pollutants) and Arctic climate. Validation of these results can lead to more stringent emissions restrictions worldwide.

This can, to some extent, be done in the FAST mission due to its unique payload suite. Although the 15 cm height resolution of the LAT is not spectacular, it is sufficient to, in combination with the HRC and SCAM, clearly distinguish between ice/snow and rocks and to monitor seasonal variations in height profiles in the cryosphere. When correlated with the spectropolarimeter data, this information can be used to investigate whether there is evidence for positive correlations between certain types of aerosols and (local) reductions/increases in ice/snow cover [3, 16].

The measurements made by the DFMRM instrument will enable acquisition of the physical parameters related to water present in both liquid and vapor form in clouds. These parameters can be used to construct more accurate precipitation prediction models and will aid the in-depth analysis of the effect of aerosols, characterized by the spectropolarimeter, on precipitation.

22.5 Technological Challenges

The FAST mission will demonstrate FF technology using micro satellites; and will also generate increased science data return through FF. From this point, FF is the core of the whole mission, and many technical challenges are related to it, especially in the case of using micro satellites. In this section, four key technical questions are to be discussed: (1) How to build up a capable and reliable micro-satellite platform based on MST? (2) How do the two satellites exchange information? (3) How do they process the received information? And (4) how do they utilize the processed information?

22.5.1 Systems Engineering of FAST-D

The technical objective of developing FAST-D is to build up a capable and reliable micro-satellite platform based on MST, which will be the first step towards a more compact and flexible SoMS nano-satellite for future multi-mission applications. Therefore, the following system engineering philosophy is adopted for FAST-D [10]:

- Extensively utilize MST components as constituents of the platform;
- Apply redundancy where needed for higher reliability; and
- Allow a certain level of technical risks for low cost and short development time.

22.5.1.1 Strategy of Utilizing MST Components

Unlike some missions where MST components are onboard as payloads for test or demonstration, FAST-D extensively utilizes MST components as constituents of the platform. However, there is an important issue that needs to be considered before utilization, i.e. the maturity of the MST components.

The maturity of MST for space use is still doubted by many developers of conventional or even small spacecraft. This is due to the fact that many MST components are originally developed for terrestrial applications and, therefore, usually cannot fully meet the special requirements (on performance or working environment) from space industries. Moreover, space MST components are still in the infant stage and more developments and demonstrations need to be done in the near future.

The immaturity of space MST also applies to the MST components onboard FAST-D. Currently some of the components are still in the stage of development with relatively low Technology Readiness Levels (TRLs). Considering the current and short-term TRL of MST components, the FAST-D spacecraft is not designed as a highly compact SoMS satellite, but as a mixture of MST and conventional technologies. For example, one of the two star trackers onboard FAST-D is the

newly developed Facet Nano [17], and the other one is the Advanced Stellar Compass (ASC) that has been adopted by a number of missions.

There are several arguments against this mixture strategy. The first argument is about the complexity. Since different types of components are utilized, the workload for interface control is significantly increased. The second argument is the reliability due to relatively low TRL components. However, previous experiences and lessons indicate that micro- and nano-satellite missions should allow a certain level of technical risks for lower cost and shorter development time. Therefore, after a careful trade-off on cost, schedule, complexity, reliability and other factors and, most importantly, keeping in mind the technical objective of the mission, this mixture strategy is eventually chosen for FAST-D.

22.5.1.2 Strategy for Achieving High Reliability

Although a certain level of risks is allowed, realizing as high as possible reliability is still one of the most important technical goals of FAST-D. To achieve this goal, a Single-Point-of-Failure Free (SPOFF) design is performed for critical systems. Here critical systems are defined as the systems that work in the "Safe Mode" and consist of the CDHS, TTC, Electrical Power Subsystem (EPS), and coarse attitude sensors/actuators.

The SPOFF design of critical systems is based on a combined redundancy strategy. Due to mass and power limitations, "functional redundancy" is adopted as the primary approach for redundancy of critical systems. Here "functional redundancy" indicates that the function of a component can be performed (sometimes with decreased efficiency) by another component in case of failure and thus duplication of the hardware is not necessary. One example of "functional redundancy" is the CDHS, where two processors are adopted, one for housekeeping and payload data handling and the other one for attitude determination and control (see Fig. 22.7). If any of these two processors crashes, the other one will take over its responsibilities. Another example is the TTC, where the low-power transmitter will provide the function of payload data downlink in case the dedicated data transmitter fails although the efficiency of the downlink will decrease due to the relatively lower data rate; the data transmitter is also able to send housekeeping data. The "functional redundancy" is also implemented by re-configuring other components. For instance, underactuated control algorithms are provided with the AOCS. Even if two reaction wheels do not work well, the remaining two reaction wheels can still provide the capability of 3-axis stabilization through underactuated attitude control. In addition, "Hardware redundancy" is adopted by FAST-D as the secondary approach for redundancy. The EPS is a good example of "hardware redundancy". Since two sets of Battery Charge Regulator (BCR), Power Conditioning Module (PCM) and Power Distribution Module (PDM) are provided, full redundancy is achieved. Two warm-redundant TTC receivers also ensure SPOFF of the uplink.

In summary, the reliability of FAST-D is achieved by synergistically adopting "functional redundancy" and "hardware redundancy" instead of solely utilizing expensive space-grade components.

22.5.2 Inter-Satellite Information Exchange

The most important issue for a successful FF mission is acquiring state information from other satellites in the formation. Usually, this could be done through direct ISL (RF or optical based). For the FAST mission, however, a mixed strategy for inter-satellite information exchange is proposed.

In the technology demonstration mode, the distance between the two satellites is only 1 km, and each satellite is in the other one's line of sight. Therefore, a direct ISL is utilized as the nominal communication route. To avoid technical complexity, an RF based ISL is chosen, which can achieve a relatively high data rate due to the low free space loss. Therefore it will be beneficial not only to exchange state information, but also to the distributed processing of data. This direct ISL in principle also allows for relative ranging between the satellites using the same signal.

In the science mode, the along-track separation between the satellites will be around 1,225 km, which implies that the two satellites are still in each other's line of sight and the direct RF ISL is still available. However, a lower data rate is expected as the free space loss is much higher.

Except the direct link, a ground-in-the-loop communication route is used as backup [4]. For this type of communication, there are two options as shown in Fig. 22.9: (1) One satellite transmits data to one ground station, then this ground station transmits the data to the second satellite when it flies over; (2) One satellite transmits data to one ground station, then through internet the data are transferred to the other ground station, which is responsible for transmitting the data to the second satellite. In the FAST mission, these two options are mixed adopted subject to the shortest delay time.

22.5.3 Onboard Information Processing

Utilizing onboard information processing has a number of advantages for small satellites. The most important one is to provide near real-time results while avoiding the costs and limitations involved with sending the data down to Earth to be processed.

In order to avoid the primary disadvantage associated with onboard processing, i.e. the errors inherent in the onboard processing cannot be undone for previously computed results, the FAST mission adopts a conservative strategy: The scientific data will be processed onboard and, meanwhile, also stored in the large volume solid state memory. A substantial part of the onboard-processed data will then be validated with the original data in the ground station.

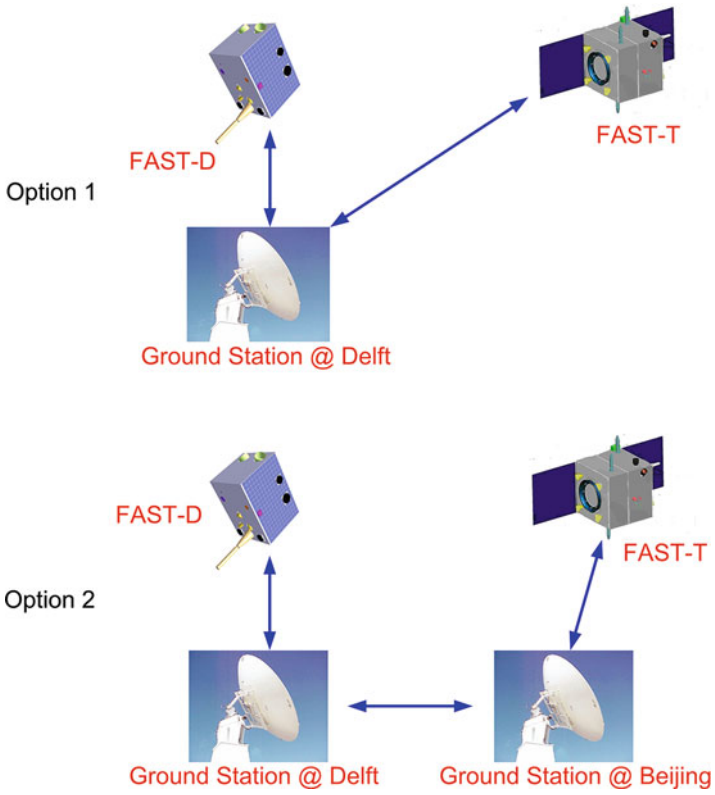


Fig. 22.9 Ground-in-the-Loop communication routes

Considering about the technology readiness, the data processing will be primarily implemented in each satellite's own OBC. The items for onboard processing include data compression and image processing. For example, the volume of data to be collected by SILAT could be dramatically reduced (i.e. to $<1\%$ of the original amount), if the images captured by HRC and SCAM can be identified and the useless images can be deleted onboard.

Besides the conventional "local" computing, also space-based distributed computing will be demonstrated by the FAST mission to explore the possibility of fully utilizing the computing powers of FF small satellites for onboard processing. The concept of space-based distributed computing will utilize the processors on each satellite as a true distributed computer. For the FAST mission this indicates a space-based computer with (at least) two spatially distributed processors.

The types of problems under consideration for such a distributed computing network are typically large, dense linear algebra problems. This is because these operations can be processed using blocked algorithms, which are well suited for distributed and, possibly, heterogeneous systems. An important experiment will be the real-time orbit determination of the FAST formation, because the size of this

problem could be very large (i.e. thousands of parameters using thousands of observations) and beyond the capability of a single processor.

22.5.4 Distributed Utilization of Information

Through inter-satellite communication and onboard "local" or distributed processing, parts of the data received by the two satellites are shared, and each satellite has acquired the other's state such as precise position and attitude. This information will be used to implement different tasks related to FF. Within the FAST mission, two of these tasks are given priority: the propellant optimized distributed propulsion, and the cooperative control.

22.5.4.1 Propellant Optimized Distributed Propulsion

Without formation maintenance, the two satellites in the FAST formation will quickly drift apart due to differential air drag and differential semi-major axis. Since it is vital that the two satellites remain at a specified distance (for the technology demonstration mode this is 1 ± 0.1 km, and for the science mode this is $1{,}225 \pm 7.5$ km) with respect to each other, the inter-satellite distance needs to be periodically adjusted using the propulsion subsystem. As each of the two satellites in the formation is equipped with a propulsion system, there is the possibility to implement propellant optimized distributed propulsion. This means that by virtue of the direct ISL, the two satellites can autonomously decide when and which satellite should fire its thruster(s), thereby minimizing and optionally also balancing the propellant consumption of the two satellites. The two propulsion systems also provide redundancy: even if one does not function, the other one still can perform formation maintenance maneuvers.

22.5.4.2 Cooperative Control

As mentioned early in last sub-subsection, FAST will perform propellant optimized distributed propulsion, which requires the cooperation of the propulsion subsystems on both satellites because the firing decisions are not intended to be made only based on one satellite's state. Furthermore, the two satellites will observe overlapped areas on the Earth, which indicates that the AOCS of the two satellites need to work cooperatively to perform relative attitude maintenance and reference attitude tracking. The common feature of these problems is: Not only the position and the attitude, but also the behavior of each satellite shall be coordinated and controlled. To solve these problems, cooperative control technology will be investigated and demonstrated by the FAST mission.

Cooperative control is a technology that allows a team (with at least two members) to cooperatively decide where to go, how to behave, and so on. It is not a fairly new term as it has already been studied in the context of a cluster of mobile robots or Unmanned Aerial Vehicles (UAVs), or even for deep space missions [18]. However this technology has not been demonstrated or tested on orbit so far. In general, there are three approaches for cooperative control: centralized, decentralized, or distributed. Considering about technology readiness, mission need (only two satellites in the formation) and many other factors, the first two approaches are expected to be further studied and, eventually, demonstrated through the FAST mission.

22.5.4.3 Multi-Agent Based Framework

To implement the proposed activities, including propellant optimized distributed propulsion and cooperative control, an agent-based framework (see Fig. 22.10) is preliminarily designed under the FAST system architecture (c.f. Fig. 22.3). The reasons for choosing a Multi-Agent System (MAS) are not only from the research aspect, but also due to the following facts:

- Autonomy – both satellites are required to behave at least partially autonomous;
- Local views – neither satellite has a full global view of the system, although both of them have some knowledge of each other;
- Decentralization – due to the limited ground station coverage, there is no one centralized controller for the system during most of the flight time;
- Flexibility – although so far the FAST mission only has two satellites, it actually has an open-system architecture, i.e. more satellites can join the formation later.

As shown in Fig. 22.10, the framework has agents at two levels: spacecraft level (e.g. FAST-D Agent) and function level (e.g. FF Agent). Spacecraft-level agents are in charge of sending/receiving information, local activity planning, command implementation, coordination of local agents, etc. Each function-level agent has several skills. For example, the FF Agent is able to plan the motion of the satellite, estimate current formation parameters, and produce formation control commands. Function-level agents can communicate with other local function-level agents and determine local behaviors.

22.6 Educational Opportunities

Except for scientific application and technology demonstration, the FAST mission also brings various opportunities for education. For example, the technologies to be developed for or used on FAST have been introduced in various courses in TU Delft. During these courses, students are also asked to implement individual and group assignments that have strong links with the FAST mission. The needs of the

Fig. 22.10 MAS-based framework

FAST mission motivated several M.Sc or Ph.D thesis topics that students are working on. In addition, through the FAST mission, students from both Delft and Tsinghua have the opportunity to do internship and thesis work in the other country with different cultures. This exposure can help students to significantly broaden their international view and to boost their skills.

22.7 Conclusions

The concept and detailed design aspects of the Dutch-Chinese FAST small satellites FF mission have been presented. The mission scenario has been described with the various relative orbital geometries of the different mission phases. The

system design of both the space and the ground segments is introduced, with a focus on the Dutch contributions, i.e. the spectropolarimeter, SILAT and the FAST-D spacecraft. Scientific applications of the FAST mission are discussed, with a special interest on improved scientific return due to FF configuration. Key technical challenges, such as systems engineering, ISL, distributed computing and cooperative control, are addressed.

The FAST mission has the potential to validate innovative technologies, such as MST-based sensors and actuators, and to demonstrate autonomous formation flying using low cost spacecraft with various communication architectures and distributed propulsion systems. The FAST mission is also expected to provide important scientific outputs on aerosol characterization and the monitoring of height profiles in the cryosphere. The formation flying will improve scientific return for aerosol characterization by observing the same geolocation from various angles during a single pass. In addition, the FAST mission provides educational opportunities to universities to teach cutting-edge technologies and broaden the international view of students. Therefore, the FAST is a unique FF mission with impact on technology, science and education.

References

1. Maessen DC et al (2008) Preliminary design of the Dutch-Chinese FAST micro-satellite mission. In: Proceedings of small satellite systems and services, Rhodes, Greece, 26–30 May 2008
2. Maessen DC et al (2008) Increasing system performanace and flexibility: distributed computing and routing of data within the FAST formation flying mission. In: Proceedings of 59th international astronautical congress, Glascow, Scotland, 29 Sept–3 Oct 2008
3. Maessen DC et al(2009) Mission design of the Dutch-Chinese FAST micro-satellite mission. In: Proceedings of 7th IAA symposium on small satellites for earth observation, Berlin, Germany, 4–8 May 2009
4. Guo J et al (2009) Status of the FAST mission: micro-satellite formation flying for technology, science and education. In: Proceedings of 60th international astronautical congress, Daejeon, South Korea, 12–16 Oct 2009
5. Snik F, et al (2010) SPEX: the spectropolarimeter for planetary exploration. In: Proceedings of SPIE space telescopes and instrumentation 2010: optical, infrared, and milimeter wave, San Diego, California, USA, June 2010
6. Moon SG et al (2008) Towards a miniaturized photon counting laser altimeter and stereoscopic camera instrument suite for microsatellites. In: Proceedins of 7th IAA symposium on small satellites for earth observation, Berlin, Germany, 4–8 May 2008
7. Moon SG et al (2010) Results and developments of micropayload prototype activities for use in earth observation and planetary missions. In: Proceedings of small satellite systems and services, Funchal, Maderia, 31 May–4 June 2010
8. Maessen DC et al (2009) Conceptual design of the FAST-D formation flying spacecraft. In: Proceedings of 7th IAA symposium on small satellites for earth observation, Berlin, Germany, 4–8 May 2009
9. Ubbels WJ et al (2008) First flight results of the Delfi-C3 satellite mission. In: Proceedings of 22nd annual AIAA/USU conference on small satellites, Logan, Utah, USA, 11–14 Aug 2008

10. Guo J et al (2010) FAST-D: a capable and reliable micro-satellite based on micro system technology. In: Proceedings of small satellite systems and services, Funchal, Maderia, 31 May–4 June 2010
11. van der List M et al Applications for solid propellant cool gas generator technology. In: Proceedings of the 4th international spacecraft propulsion conference, Sardinia, Italy, 2–9 June 2004
12. Gill EKA et al (2010) Atmospheric Aerosol Characterization with the Dutch-Chinese FAST Formation Flying Mission. Acta Astronaut 66:1044–1051
13. Mishchenko MI et al (2007) Accurate Monitoring of Terrestrial Aerosols and Total Solar Irradiance: Introducing the Glory Mission. Bulletin of the American Meteorological Society 88:677–691
14. NASA, Glory – observing the earth's aerosols and solar irradiance, http://glory.gsfc.nasa.gov/. Retrieved 28 Mar 2011
15. Shindell D, Faluvegi G (2009) Climate Response to Regional Radiative Forcing during the Twentieth Century. National Geographic 2:294–300
16. Moon SG et al (2008) A miniaturized laser altimeter and stereo camera for a microsatellite formation mission. In: Proceedings of small satellite systems and services, Rhodes, Greece, 26–30 May 2008
17. van Breukelen E, et al (2009) Facet nano, a modular star tracker concept for highly miniaturized spacecraft. In: Proceedings of 60th international astronautical congress, Daejeon, South Korea, 12–16 Oct 2009
18. Shima T et al (2008) UAV Cooperative Decision and Control: Challenges and Practical Approaches. Society for Industrial Mathematics, Philadelphia

Chapter 23
Future Trend, Potential, Risks

Marco D'Errico, Eberhard Gill, Antonio Moccia, and Rainer Sandau

Abstract During the last decades, the concept of distributed space systems has significantly progressed in terms of space applications, including Earth remote sensing. This chapter is devoted to a critical analysis of the achieved improvements and of the areas of major key issues in order to analyze potential and risks of distributed space systems. A discussion of future activities needed to prepare more advanced distributed space missions is also provided. In particular, payloads and applications are first discussed. Then, guidance, navigation, and control as well as other technological challenges, including modularity and architecture, follow.

Spaceborne radar missions based on Synthetic Aperture Radars (SAR), which used to be flown by large monolithic spacecraft, have shown a great evolution towards the distributed approach. Many studies have been conducted in the last years on systems relying on two (bistatic) or multiple (multistatic) radars, which have led to a great level of maturity especially in the case of bistatic SAR.

M. D'Errico (✉)
Department of Aerospace and Mechanical Engineering, Second University of Naples, Via Roma 29, Aversa (CE) 81031, Italy
e-mail: marco.derrico@unina2.it

E. Gill
Delft University of Technology, Space Systems Engineering, Kluyverweg 1, Delft 2629 HS, The Netherlands
e-mail: e.k.a.gill@tudelft.nl

A. Moccia
Department of Aerospace Engineering, University of Naples "Federico II", P.le Tecchio 80, Napoli 80125, Italy
e-mail: antonio.moccia@unina.it

R. Sandau
German Aerospace Center (DLR), Rutherfordstr. 2, Berlin 12489, Germany
e-mail: rainer.sandau@dlr.de

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8_23,

The bistatic SAR concept has been studied and developed in support to many applications which can exploit either signal phase and amplitude, if the two radars are sufficiently close to one-another (interferometric techniques), or amplitude-only for large separations (baselines). In both fields, bistatic SAR enables applications which are either impossible or strongly limited with standard monostatic approaches. A prominent example is the generation of digital elevation models by SAR interferometry, which is possible with both monostatic and bistatic approaches. It has first been theoretically demonstrated and then practically experienced in space by ESA ERS-1 and -2 tandem operations and by the Shuttle Radar Topography Mission. Recently, this technology was used extensively in the operational TanDEM-X/TerraSAR-X mission where bistatic SAR interferometry achieves unprecedented height accuracy thanks to the possibilities to avoid time decorrelation and to reconstruct the baseline with an extremely high accuracy. In different geometrical configurations, a bistatic SAR can also perform along-track interferometry to derive velocity estimates of targets on ground. In this field and depending on radar frequency and on the desired velocity range to be detected, a solution with a single but partitioned antenna (e.g. TanDEM-X) can be more reliable with respect to two free flying spacecraft. Different applications have also been studied and assessed for large baseline bistatic SAR, mainly in the ambit of the SABRINA mission study: derivation of digital elevation models by methodologies derived from radargrammetry, estimation of velocities by Doppler analysis, estimation of terrain properties (roughness, dielectric constant, moisture), and terrain slope mapping. At the moment, no space mission implementing a large baseline bistatic SAR is scheduled, although such applications could be explored in an extended TanDEM-X mission phase.

The multistatic SAR concept has raised an increasing interest in the last decade. Multistatic applications exploiting phase information can be conceived as a generalization of bistatic ones: SAR tomography is the major example. It allows 3D-imaging of semi-transparent volume scatterers and permits to estimate vertical structures of the imaged scenes (e.g. forests). SAR tomography is also greatly useful for non-transparent scatterers to remove or reduce the layover, one of the classical limitations in radar imagery. Multistatic SAR tomography has the same advantages of bistatic SAR interferometry with respect to a monostatic approach. Multistatic SARs can also be utilized in more advanced modes. An example is the capability of multistatic SAR to overcome the minimum antenna area constrain, one of the major limitations in SAR design. In such cases, multiple free-flying SARs with small antennas contribute to generate a virtual SAR achieving contemporaneously a wide swath and very small geometrical resolutions, contrary to monostatic SARs which trade swath width versus resolutions. An additional advanced possibility is to create very large sparse apertures. Although very promising for future development, multistatic SAR is at a lower technology readiness level as compared to the simpler bistatic concept and its implementation seems questionable in the near future.

The distributed payload concept has also been studied for optical payloads in both the fields of filled or sparse sensor apertures. A filled sensor aperture concept is the one

currently flying on standard monolithic satellites. Its replacement with a distributed sensor has been studied under different approaches. The utilization of still large but lightweight surfaces to implement the primary mirror (thin-film membranes) has been proposed, with the other needed optical modules either connected with tethers or flying in formation. An alternative possibility to replace a large monolithic aperture has also been analyzed: a number of smaller sensors correctly sample the large aperture and maintain a correct geometry based on formation flying. Sparse aperture concept has been studied at a lesser extent. It is worth noting that such studies were mainly conducted at a conceptual level only. Thus, distributed optical sensors appears much less mature than their microwave counterparts.

Another field which benefits significantly from a distributed space architecture is gravimetry. The GRACE mission has demonstrated the great capability to monitor temporal variations of the Earth gravity field thanks to the innovative techniques to tracking at an extremely high accuracy the inter-satellite distance. Such capability strongly enhances the potential of gravimetry missions, opening the way to new applications with respect to classical ones. As an example, gravimetry data can be utilized for climate research thanks to the capabilities to monitor the variations of both ice mass on polar caps and water distributions in water basins.

If a payload is distributed over several spacecraft, their relative geometry plays an essential role in defining its performance. Since the spacecraft relative geometry is determined by orbital dynamics, it could be said that absolute and relative orbital dynamics as well as formation acquisition and keeping and collision avoidance become an essential part of the mission design. Relative dynamics is essential in any aspects of a distributed, formation-based mission: from formation design to its establishment in orbit, from formation control to reconfiguration, when requested, and for navigation. While in early years formation studies have benefited from previous researches on rendezvous and docking, long-term formation flying missions require a highly accurate handling of the relative position and velocity of the involved spacecraft.

Relative motion modeling has reached a high level of maturity in the last 15 years with a strong improvement with respect to the Hill-Clohessy-Wiltshire equations, which constitute a very limited model for formation flying because of the Keplerian motion and circular orbit assumptions. In particular, relative motion modeling, although still open to additional improvements, seems to have reached an acceptable level in LEO for orbits with small eccentricities. Relative motion for highly eccentric orbits requires additional steps to reach an assessed and agreed operational practice. Furthermore, relative motion on higher orbits (MEO, GEO) has only partially been explored. However, these limitations are not expected to pose a risk in the near-term future as most proposed missions/applications are expected to fly on well understood orbits. Nonetheless, it is likely that missions making use of higher altitudes (GEO in particular) could become of great interest in future.

With reference to relative guidance, navigation and control, several control approaches have been thoroughly tackled and demonstrated, considering both continuous and impulsive control. As for navigation, three main technologies have been developed or are under development. GPS-based relative navigation

has already been successfully demonstrated on orbit since the GRACE mission. It has in particular shown that carrier-phase differential techniques strongly improve navigation accuracies which could be obtained with pseudorange measurements. It has been extensively demonstrated that centimeter accuracy on relative position can be achieved in real time, while millimeter accuracy is achievable in post-processing. Those accuracies are achievable provided that a good relative dynamics model is utilized, the mission allows smooth spacecraft operations, for example with respect to thruster activities, and appropriate GPS receivers are used. These performances open new perspectives with respect to autonomous control on one hand and, in addition, enable high accuracy digital elevation models generation by SAR interferometry. For instance, Tandem-X interferometric performance are in fact guaranteed, among other reasons, by this great capability of high precision baseline reconstruction.

Relative positioning can be also provided by other sensors. Relative navigation by locally generated radio frequency ranging signals (Direct Sequence Spread Spectrum signals) has been studied with a technique similar to GNSS systems. Radio frequency relative navigation techniques and hardware have been proved on orbit during the PRISMA mission. Finally, relative navigation can be provided by optical metrology, as demonstrated for close separations formerly during rendezvous and docking missions and later during the PRISMA mission. Optical metrology has also potential developments for large satellite distance. In particular, if different orders of magnitudes of satellite separations are expected during a distributed mission lifetime, a multisensor approach is needed to cover different distance ranges with adequate performance. However, it should be recognized that the great success and wide use of GPS-based navigation are making less attractive other techniques in low Earth orbit, in particular when dealing with microwave scales and having the possibility to apply differential processing. Finally, it is worth underlining that while a strong maturity has been reached in relative navigation, the determination and control of relative attitude need further investigations.

Distributed systems clearly need novel approaches to system design, development and operation. In particular, with an increasing number of identical (or similar) coordinated spacecraft replacing monolithic systems, modularity becomes a crucial aspect to reduce costs. It also enables rapid insertion of new technologies in space, adaptation of a system to long period project evolution, reduction of qualification issues (multiplication of the same qualified module in lieu of development of a new module). Modularity is greatly beneficial because it allows fast assembly by off-the-shelf components, which is essential when fast reconfiguration capabilities are required. A modular approach also allows one to simplify the design process, breaking it into a number of simpler, concurrent processes. Finally, a modular system exhibits a graceful degradation and it can be easier repaired or upgraded by substituting only required modules.

If on the one hand, repetitive modularity is widespread as a means to achieve high performance utilizing low performance devices by simply increasing their number, hierarchical modularity is still an area to be explored. A hierarchical system is in fact rather different because it works by assembling larger entities

from subordinate entities and by interconnecting them with proper busses. Thus, hierarchical modularity can be applied by conceiving the system as an assembly of different sub-systems, in turn obtained as a combination of several sub-modules on different hierarchical levels. Hierarchical modularity can be greatly beneficial to distributed systems because of its embedded potential to distribute the intelligence throughout the system. An architecture building on modular components and subsystems, which are in turn intelligent to perform self-monitoring and control functions, allows a system with different levels of internal autonomy to be integrated and makes available more resources for the on-board computer to interact and collaborate with nearby platforms. Such effort is aided by two on-going developments: availability of an increasing number of miniaturized and space qualified microelectronics devices (including microcontrollers) and development of a variety of data busses (including space qualified CAN bus) which are particularly suited for modular system implementations.

Autonomy is itself an important feature for any upcoming space mission to reduce mission cost. If autonomy can be still limited for two spacecraft systems operating at reasonable distances, it becomes a fundamental need for distributed systems integrating a large number of spacecraft. Several approaches have been analyzed for formation organization, such as leader follower, virtual structure, and behavioral. The latter appears both better suited for multi spacecraft systems and more flexible and inherently adaptable to an evolutionary environment. In this ambit multi agent systems seem promising because they are consistent with de-centralized problems of a society of interacting, autonomous components. Nonetheless, it must be considered that such technology is not mature in an orbital, operational scenario: it has only been partially tested in laboratories and needs further developments even at a theoretical level. When considering autonomous organization and operation of a distributed system, an additional concern must be analyzed. The need for an adequate information flow among the platforms for both formation control and payload operation has experienced only limited analysis. Nonetheless, this aspect is critical for formation implementation due to limited telecommunication resources (data rate, power, etc.) and geometrical conditions (relative visibility).

The ground segment is also an issue to be considered seriously. The classical approach of ground segments assigns the specific tasks (S/C monitoring & control; P/L data reception & archives; P/L data products & distribution) to specific ground facilities communicating through complicated protocols. As the number of space-craft increases, a change in the operations paradigm is necessary to avoid a concomitant increase in the operational cost, which is inevitable with a simple downscaling. A qualitative change combining the different tasks and facilities in networks with new features is therefore needed. Key words for new ground systems are: open systems, automation, “internet” technology, multi-session operations, ground station networks, enhanced on-board autonomy.

From a systems architecture perspective, distributed space systems can be implemented not only as formations, but as constellations, trains, swarms, and fractionated satellites. Constellations provide a large coverage which is typically

of global nature combined with low control accuracy. Their implementation and operations is well understood at least for Medium Earth Orbits (MEO) such as the existing GPS NavStar system and the upcoming Galileo system. Formations, which have been thoroughly herein analyzed, implement a distributed space architecture with typically a regional coverage requiring a high control accuracy. An efficient alternative to formations are trains, where payload is distributed on various satellites which essentially follow a joint orbit, such as in the A-train. Trains offer flexibility in terms of the programmatic organization, as even various countries can contribute to a single train, decoupling development efforts and allowing for risk mitigation. The coordination effort is moderate as compared to formations. However, the functional adequacy of a train may be much more restricted to certain applications than that of a formation. In other words, overall result of a train is just the sum of contributions of each single component and it is in charge of the user to develop integrated applications. In the case of a formation, each element furnishes, in general, limited outputs, but the integrated system offers higher order performance, not achievable differently.

Beyond the classical space architectures, two new approaches are currently being more and more studied: swarms and fractionated spacecraft. Swarms represent a multitude of identical, typically highly miniaturized spacecraft with limited functionality. A swarm architecture may be used e.g. for multi-point sensing. The technology readiness level of swarms in most of its aspects is however still very low. In contrast, fractionation of spacecraft can be regarded as the opposite to a swarm in that its elements are heterogeneous and the functionality of a single spacecraft is broken apart and distributed among the different elements in the architecture which make up the entire (distributed) spacecraft. Fractionation may allow for new approaches towards reliability and modularity including enhanced responsiveness and decreased vulnerability. Its benefits and challenges are currently controversially discussed. Both swarms and fractionated spacecraft are innovative concepts in its infancy which may demonstrate their value only in the coming two decades.

Distributed space systems pose a number of system challenges. The increased threat and awareness of space debris render a proper treatment of collision risk and end-of-life operations a crucial aspect. Access to space for distributed space systems may benefit from a reduced-risk multiple launch approach at the expense of an increased propulsion and operations effort to acquire the nominal relative configuration. Legal and regulatory aspects are becoming more and more relevant and may represent challenges for distributed space systems, such as frequency allocation and coordination for complex space architectures.

Finally, although the distributed system concept can be in principle applied to both large and small spacecraft, it is naturally coupled with small satellites for a number of reasons: system cost, easy and rapid replacement of a failed satellite, possibility to gradually update flying technologies (which is generally an issue for large space systems). Generally, small satellite missions are already powerful tools to flexibly react on information requirements with space borne solutions. They can be conducted relatively quickly and at a limited cost and provide increased

opportunity for access to space. The spacecraft bus and instruments can be based either on optimized off-the-shelf systems, with little or no requirements for new technology, or on new high-technology systems. Thus a new class of advanced small satellites, including autonomously operating "intelligent" satellites may be created, opening new fields of application using also distributed space systems for scientific purposes as well as operational, public and commercial services. The approach to use available off-the-shelf technology to build a small satellite system (bus and payload) for the intended remote sensing purpose results in low-cost and quick response time. In this context, the restrictions or limitations of small satellites in comparison to large complex satellites are to be considered: limited platform capacity for using instruments with high power consumption or high data rate requirements; size limitations when large microwave antennas or long monolithic telescopes are necessary; restricted options for instrument combinations on a single satellite platform because of the limited size and power capabilities of small satellites. For these reasons small satellite missions supplement conventional Earth observation missions.

The alternative approach is taking full advantage of the ongoing technology developments leading to further miniaturization of engineering components, development of micro-technologies for sensors, instruments, and spacecraft bus components. At the extreme end of the miniaturization, the integration of micro-electromechanical systems (MEMS) with microelectronics for data processing, signal conditioning, power conditioning, and communications leads to the concept of application specific integrated micro-instruments (ASIM). These micro- and nano-technologies have led to the concepts of nano- and pico-satellites, constructed by stacking wafer-scale ASIMs together with solar cells and antennas on the exterior surface, enabling the concept of space sensor webs.

Further milestones in the small satellite, low-cost distributed space systems developments are the availability and improvement of small launchers, the development of small ground station networks connected with rapid and cost-effective data distribution methods, and cost-effective management and quality assurance procedures.

In conclusion, distributed systems integrating a small number of spacecraft have reached a well assessed maturity in many fields, primarily bistatic SAR and gravimetry. For these systems, technology maturity has also been consistently demonstrated by already flown missions.

New research and technology developments and proof-of-concept experiments are however needed to prepare future missions integrating a large number of spacecraft, which appear as an exciting challenging adventure from the engineering point of view. Whereas, from the application point of view, further analyses are required to better: (a) assess the potential of distributed systems, identify (b) new products for the user community and (c) application areas with major benefits from a distributed system concept. Increased communication and cooperation between scientists, data users, and engineers would represent an important step forward to facilitate a mutual understanding of application needs and system and technology capabilities.

Index

M. D'Errico (ed.), *Distributed Space Missions for Earth System Monitoring*, Space Technology Library 31, DOI 10.1007/978-1-4614-4541-8,