Autonomous Relative Navigation for Small Spacecraft

Daan Maessen
Autonomous Relative Navigation for Small Spacecraft

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 2 september 2014 om 10:00 uur

door

Daan Corneel MAESSEN

Ingenieur Luchtvaart en Ruimtevaart
geboren te Tegelen.
Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. E.K.A. Gill

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof. dr. E.K.A. Gill, Technische Universiteit Delft, promotor
Prof. dr. G. Gómez, Universitat de Barcelona, Spain
Prof. dr. K. Schilling, Julius Maximilian University of Würzburg, Germany
Prof. ir. P. Hoogeboom, Technische Universiteit Delft
Prof. ir. B.A.C. Ambrosius, Technische Universiteit Delft
Dr. M. Delpech, Centre National d’Études Spatiales, France
Dr. ir. A.A. Verhagen, Technische Universiteit Delft

Cover design: Proefschriftmaken.nl || Uitgeverij BOXPress
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, ’s-Hertogenbosch

This document has been typeset in \TeX.


Copyright © 2014 by Daan Maessen

All rights reserved. No part of this publication may be reproduced, stored in a re-
trieval system, or transmitted, in any form or by any means, electronic, mechanical,
photocopying, recording, or otherwise, without express permission of the author.

Printed in The Netherlands
Summary

Autonomous Relative Navigation for Small Spacecraft

Daan Maessen

Spacecraft formation flying, i.e. the maintenance of a specific relative motion between two or more spacecraft, is considered to be a new and enabling technology in space engineering. However, it is almost as old as spaceflight itself. It has been performed since the American Gemini missions in the 1960s. However, what sets the "new" formation flying missions apart from the "old" formation flying missions is that the "new" missions rely on a certain amount of autonomy for the spacecraft, enabling them to maintain the formations with little to no human intervention. This autonomy saves cost in ground operations and enables missions that are not feasible when formation control needs to be performed using ground operators.

Spacecraft formations are a subclass of distributed space systems. In these systems, multiple spatially separated platforms are used to fulfill the mission objectives. Performing space missions using spacecraft formations brings several advantages over the more traditional monolithic spacecraft design, but also comes with new challenges. It enables for instance challenging missions where a very high angular resolution is required of the object to be studied or where single-pass interferometric measurements are needed. In addition, missions can be made intrinsically redundant or failure tolerant: If one spacecraft in a formation fails, the mission is not necessarily lost. Spacecraft formations can also be considered for missions where a formation is not strictly necessary but which would require a very complex, and thus costly, monolithic spacecraft (e.g., Envisat). Such a complex spacecraft with many different payloads that all have conflicting requirements could be split into multiple smaller spacecraft that are optimized for one or two payloads. However, formation flying also adds complexity at mission level since multiple spacecraft need to be built, tested, and launched. Once in orbit, these spacecraft also need to be brought and kept in a certain configuration for the entire mission duration. This also requires additional subsystems on each spacecraft in order to determine and maintain the positions and velocities of the spacecraft relative to each other and to avoid spacecraft collisions.

In the last decade, much research has been performed in the field of spacecraft formation flight. However, much of this research has focused on specific missions, the modeling
of the orbital dynamics, or efficient formation control and collision avoidance algorithms. In the field of navigation for formation flying spacecraft, several important aspects still need further investigation. For instance, it is not known how the ranging accuracy, inter-satellite distance, and antenna baseline (i.e., the distance between two antennas on the same spacecraft) impact the accuracy in the estimation of the relative positions and velocities, or 'state', of the spacecraft. If the ranging accuracy is increased by a factor two and the inter-satellite distance is also increased by a factor two, will the accuracy in the estimation of the relative state remain the same or not? This is important to know since spacecraft formations should preferably consist out of relatively small spacecraft (otherwise the total mission costs will quickly become prohibitive) and these small spacecraft are limited in terms of power, mass, and volume. Furthermore, it is also not known if the type of relative orbit has an impact on the estimation of the relative state. Perhaps one type of relative orbit will lead to intrinsically better relative state estimates than another type of relative orbit. Lastly, it is known that a large relative out-of-plane motion of the spacecraft leads to more accurate relative state estimates because of the favorable viewing geometry. In fact, a very small relative out-of-plane motion will even cause the estimation process to fail. However, it has never been quantified how small the relative out-of-plane motion can be before the estimation process will fail. In order to improve the knowledge in the areas discussed in this paragraph, the following research questions have been formulated at the start of the research. These research questions are answered in this thesis.

**Research Question 1** How do the receiver antenna baseline, ranging accuracy, and inter-satellite distance affect the accuracy in the relative state determination for two formation flying spacecraft? Does there exist a transition point between the use of a single antenna and the use of multiple antennas at which the navigation results obtained by using one antenna equal or even outperform the navigation results obtained by using multiple antennas?

**Research Question 2** How are the relative navigation results for a two spacecraft formation influenced by the type of relative motion?

**Research Question 3** How does the magnitude of the relative out-of-plane motion between two spacecraft influence the relative navigation results?

In order for formation flying spacecraft to be able to determine their relative state, some means of navigation is needed on the spacecraft themselves: They must be able to at least measure the inter-satellite range (and preferably also velocity) and, using these measurements, be able to accurately and reliably estimate their relative state using a model of the system dynamics. In (low) Earth orbit, these measurements can be obtained by exchanging measurements of their absolute orbital position that have been obtained using a Global Navigation Satellite System (GNSS) receiver. However, such a strategy will not suffice for all Earth orbits and is not feasible for missions that are not in Earth orbit. Then, the spacecraft need to obtain the measurements in a different manner. This can be achieved using locally generated radio frequency (RF) ranging signals, cameras, or lasers. In this research, it is assumed that RF ranging signals are used, more specifically, Direct Sequence
Spread Spectrum (DSSS) signals. This type of signal is also used for GNSS signals. Therefore, a large technology and knowledge base is present that can be used to implement such signals in a ranging device. In addition, this type of signal can also be used to exchange information between the spacecraft. Typically, the data rate that can be achieved in this manner is not very high, on the order of kbit/s, but it is sufficient to exchange essential data such as measurements, commands, and perhaps even some spacecraft health data. More importantly, it takes away the need for a separate inter-satellite communication system.

The specific implementation of the inter-satellite links (ISLs) depends on many interlinked aspects which cause the implementation to vary from mission to mission: A frequency band has to be selected, but this is limited by regulations and technology. For (very) small satellites, which typically have limited power available, a trade-off is likely to result in the selection of a relatively high carrier frequency. In addition, signal interference has to be prevented, especially for formations with many spacecraft. This can be done by means of multiple access technology, which separates the various signals in the frequency, time, and/or code domain. The choice for the multiple access technology to be used is closely linked to the communication topology. In turn, the choice for the communication topology is driven by the formation control strategy to be implemented and can be centralized, hierarchical, or distributed.

The design of the ranging signal itself is naturally also important as it has a large influence on the obtainable ranging accuracy. For DSSS signals, a pseudo-random noise (PRN) code has to be selected that has low cross-correlation properties. In addition, a signal spreading waveform has to be selected that has a good compromise between (code) ranging accuracy and acquisition robustness. The BOC(1,1) (Binary Offset Carrier) spreading symbol shape is found to offer a good balance between these aspects and is regarded to be a good candidate for use on formation flying spacecraft in the near future. Another, more complex, promising signal shape is Ultra-BOC, which has several advantageous properties for integer ambiguity resolution (IAR).

To enable inter-satellite ranging and to obtain the required accuracy in the relative state estimate, a specific spacecraft design has to be implemented. The inclusion of a Relative Navigation System (RNS) implies that other subsystems need to be adapted in order to enable the correct functioning of the RNS. The use of a RNS also has implications for spacecraft operations such as safe modes and instrument calibration phases.

For the RNS, a transceiver is needed to transmit, receive, and to perform (part of) the processing of the ranging signals. This transceiver will be largely based on GNSS-technologies and needs to implement a very stable frequency source to minimize ranging errors due to clock biases between the various spacecraft in the formation. The transceiver further needs to implement an internal calibration to compensate for bias and drift of the electronics. Active power control, based on predictions of the actual inter-satellite range, is also a very useful property to reduce power consumption for missions where the inter-satellite distance can vary significantly.

To answer the research questions posed at the onset of the research, a dynamic scenario has been set up where two spacecraft fly in formation in a low Earth orbit and
where the state of one of the spacecraft, the ‘deputy’, is estimated by means of a non-linear batch least squares (LSQ) estimator using range measurements between the deputy and the other spacecraft, the ‘chief’. The chief can have one or three receiving antennas with varying baselines. The deputy is modeled as a point mass and has only one transmitting antenna. The distance between the two spacecraft and the ranging accuracy are varied as well. The influence of the type of formation is also studied by comparing results obtained for so-called elliptical and pendulum formation geometries. In an elliptical formation geometry, the deputy describes an elliptical motion with respect to the chief. Here, it has been chosen to have the center of the ellipse coincide with the center of mass of the chief. In a pendulum formation geometry, the deputy is at a constant along-track distance from the chief and has a periodic relative out-of-plane motion with a certain amplitude. To investigate the influence of the relative out-of-plane motion, this is also varied for both formation geometries.

To enable a general treatment of the research questions, various system level simplifications have been adopted. This has resulted in the most basic scenario possible, which still captures the essential parts of the problem to be studied. In this scenario, signal level influences such as frequency, multiple access, ambiguities, biases, signal coding, interference, etc. are not considered since these are very specific properties that vary between missions and during the mission itself. Instead, these contributions are ‘lumped’ together in an assumed value for the standard deviation of the range measurement. Also for the relative orbit dynamics, a very simple model, the Hill-Clohessy-Wiltshire (HCW) equations, has been applied since the objective of the study is not to obtain the most accurate relative navigation results possible, but to determine how various important parameters influence the relative navigation results.

However, even when implementing these simplifications, it is important to understand how a range measurement is obtained and which error sources should be taken into account for a design. For instance, due to the relative drift of the clocks on the satellites, a one-way range measurement such as in GNSS applications is not sufficient to measure the range with good accuracy since the timing information is too inaccurate. Therefore, a so-called dual one-way range measurement needs to be made between each satellite pair, which allows the clock bias to be determined. Various statistical and systematic errors also influence the ranging accuracy, such as signal multipath, integer ambiguities, receiver bias, the ionosphere (if present), phase windup for circularly polarized signals, signal interference, antenna phase center location, system noise temperature, and receiver resolution.

For the estimation of the relative state, a nonlinear LSQ estimator is implemented. An important consideration for the choice for a batch LSQ estimator instead of a sequential filter such as an extended Kalman filter (EKF) is that use of a batch LSQ estimator allows a direct coupling between the estimation results and a so-called observability analysis. Such an analysis indicates in how far the state of a system can be estimated from knowledge of the inputs (i.e., control forces) and/or outputs (i.e., measurements). If the observability is poor, the state of the system, or a subset of the state parameters, cannot be determined with high accuracy or cannot be determined at all. On the other hand, a good observability
indicates that the state of the system can be determined with high accuracy. Thus, an observability analysis can provide insight on the question of how small the relative out-of-plane motion can be to still allow an estimate of the relative state to be performed.

Next to the numerical analysis, a statistical analysis has been developed to obtain insight on the relationship between ranging accuracy, inter-satellite distance, and antenna baseline. It can also be used to obtain a quick first estimate of the navigation accuracy that can be expected for a formation flying mission. General equations are derived for the expected value and variance in the estimation of the position of a stationary point in space (the transmitting antenna) using range measurements, incorporating a variable, normally distributed ranging error, to three different points (the receiving antennas). In the statistical analysis it is also shown that if the receiving antennas are positioned at the corners of a right triangle, the variance in the position estimate of the transmitter is minimized. In addition, to obtain the highest accuracy in the position estimate, the obtained equations show that, as a general rule, the ranging accuracy should be high, the inter-satellite distance should be small, and the antenna baseline should be large. This last conclusion agrees with general knowledge for navigation. The analysis also briefly treats results for two simple formation configurations.

From the numerical analysis it has been found that, for the formation geometries considered, the accuracy in the estimation of the relative state depends linearly on the ranging accuracy and inter-satellite distance, which is as expected and is not considered to be a new result. However, a new result is that the accuracy in the estimation of the relative state does not depend on the antenna baseline: For an elliptical formation geometry, it is very important that the vector from the origin of the local frame to one (of the) antenna(s) in the local frame is perpendicular to the direction of the least observable eigenvector of the system. In addition, the accuracy in the estimate of the relative state scales linearly with the length of this vector. For a pendulum formation geometry, the antenna baseline is only of importance if the observation arc length is less than one orbit. If it is longer, the baseline does not contribute to the accuracy in the relative state estimate.

For a pendulum formation geometry, it is more likely that a single ranging antenna can be used per satellite than in case of an elliptical formation geometry. This is due to a better observability for the pendulum formation geometry and due to a more favorable line-of-sight throughout the orbit. Thus, a pendulum formation geometry lends itself more to the use of small satellites than an elliptical formation geometry.

For the relative navigation, it has been found that the magnitude of the relative out-of-plane motion in itself is not very important for the observability of the relative state and the accuracy of the relative state estimate. Instead, it is the relative out-of-plane angle that drives this. This observation can be exploited in formations consisting out of more than two spacecraft to reduce the number of inter-satellite links: If there is one pair of spacecraft with limited relative out-of-plane motion, the relative navigation between these spacecraft is best done via another spacecraft which has a large relative out-of-plane motion with the other spacecraft.

Lastly, this research has shown that in case of a pendulum formation geometry the relative out-of-plane angle can be several times smaller (3°) than in case of an elliptical for-
mation geometry (14°) without causing a very high probability of the estimation process to fail.

This thesis has researched, using analytical and numerical methods, the impact of ranging accuracy, inter-satellite distance, and antenna geometry for various spacecraft formation flying configurations taking into account adequate technologies for inter-satellite ranging. The results of this work are expected to contribute to establishing and enabling effective and efficient missions for formation flying using small satellites in the future.
Samenvatting

Autonome Relatieve Navigatie voor Kleine Ruimtevaartuigen

Daan Maessen

Het in formatie vliegen van satellieten, dat wil zeggen, de handhaving van een specifieke relatieve beweging tussen twee of meer satellieten, wordt beschouwd als een nieuwe technologie voor de ruimtevaart die nieuwe missies mogelijk maakt. Echter, het in formatie vliegen van satellieten is feitelijk bijna net zo oud als de ruimtevaart zelf. Het wordt namelijk al uitgevoerd sinds de Amerikaanse Gemini missies in de jaren 1960. Echter, de "nieuwe" manier van formatie vliegen onderscheidt zich van de "oude" manier in dat de nieuwe manier gebruik maakt van een zekere mate van autonomie bij de satellieten, waardoor het handhaven van de formatie met weinig tot geen menselijke interventie kan plaatsvinden. Deze autonomie bespaart kosten in grondoperaties en maakt het mogelijk om missies te vliegen die niet uitvoerbaar zouden zijn indien de formatie vanaf de grond gecontroleerd zou moeten worden.

Formaties van satellieten zijn een subklasse van gedistribueerde satellietsystemen. In zulke systemen worden meerdere ruimtelijk verspreide platforms gebruikt om de doelen van de missie te bereiken. Het uitvoeren van missies met behulp van formaties brengt enkele voordelen met zich mee ten opzichte van het gebruik van traditionele monolithische satellieten. Echter, er zitten ook nadelen aan verbonden. Formaties maken het mogelijk om uitdagende missies uit te voeren waarbij voorbeeld een extreem hoge resolutie behaald moet worden of waar 'single-pass' interferometrische metingen nodig zijn. Missies kunnen ook intrinsieke redundant of tolerant voor fouten worden gemaakt: Indien een satelliet in een formatie uitvalt, dan hoeft de missie nog niet geheel verloren te zijn. Formaties kunnen ook gebruikt worden voor missies waar het gebruik van een formatie niet strikt noodzakelijk is, maar waar het gebruik ervan kan voorkomen dat een zeer complexe, en dus dure, satelliet nodig is (bijv. Envisat). Een dergelijk complexe satelliet met vele verschillende instrumenten die allen verschillende, conflicterende, eisen hebben, zou opgesplitst kunnen worden in meerdere kleine satellieten die geoptimaliseerd zijn voor een of twee instrumenten. Echter, het vliegen in formatie voegt ook complexiteit toe op missieniveau, omdat meerdere satellieten moeten worden gebouwd, getest, en gelanceerd. Eenmaal in de ruimte moeten deze satellieten in een bepaalde configuratie
worden gebracht en gehouden voor de duur van de missie. Het vereist ook het gebruik van additionele subsystemen op elke satelliet om de posities en snelheden van de satellieten ten opzichte van elkaar te bepalen, te controleren en om botsingen te vermijden.

In het afgelopen decennium is er veel onderzoek verricht op het gebied van formaties van satellieten. Echter, het onderzoek heeft zich met name toegespitst op specifieke missies, het modeleren van de baanmechanica, of efficiënte technieken om de posities van de satellieten te regelen en om botsingen te voorkomen. Voor de navigatie tussen formatievliegende satellieten zijn er echter nog onderwerpen die verder onderzocht dienen te worden. Het is bijvoorbeeld niet bekend hoe de nauwkeurigheid in de afstandsmeting, de afstand tussen de satellieten, en de afstand tussen de antennes op een satelliet (de basislijn) de nauwkeurigheid in de schatting van de relatieve posities en snelheden, ook wel aangeduid als relatieve toestand, van de satellieten beïnvloedt. Indien de meetnauwkeurigheid wordt verdubbeld en als de afstand tussen de satellieten wordt verdubbeld, blijft de nauwkeurigheid in de schatting van de relatieve toestand dan hetzelfde of niet? Het is belangrijk om dit te weten omdat het gebruik van relatief kleine satellieten de voorkeur heeft voor formaties (anders zouden de kosten al snel te hoog worden) en omdat deze kleine satellieten gelimiteerd zijn in beschikbaar elektrisch vermogen, massa, en volume. Het is verder ook niet bekend of het type van de relatieve beweging van de satellieten de nauwkeurigheid in de schatting van de relatieve toestand beïnvloedt. Wellicht leidt de ene relatieve beweging tot intrinsieke nauwkeurigere schattingen van de relatieve toestand dan een andere relatieve beweging. Ten slotte is het bekend dat een grote beweging van een satelliet t.o.v. het baanvlak van een andere satelliet leidt tot een nauwkeurigere schatting van de relatieve toestand doordat de geometrie van het systeem gunstiger wordt. Echter, een zeer kleine beweging leidt tot het falen van het schattingsproces. Het is nooit gekwantificeerd hoe klein deze beweging kan zijn zonder dat het schattingsproces faalt. Teneinde de kennis te vergroten in de gebieden die in deze paragraaf zijn besproken, zijn aan het begin van het onderzoek de nuvolgende onderzoeksvragen geformuleerd. Deze onderzoeksvragen worden in deze thesis beantwoord.

**Onderzoeksvraag 1** Hoe beïnvloeden de basislijn van de antennes, de meetnauwkeurigheid, en de afstand tussen de satellieten de nauwkeurigheid in de schatting van de relatieve toestand van twee formatievliegende satellieten? Bestaat er een omslagpunt waarbij het gebruik van een enkele antenne resulteert in dezelfde of zelfs betere navigatie-resultaten dan bij het gebruik van meerdere antennes?

**Onderzoeksvraag 2** Op welke wijze worden de navigatie-resultaten voor een formatie bestaande uit twee satellieten beïnvloed door het type van de relatieve beweging tussen de twee satellieten?

**Onderzoeksvraag 3** Op welke wijze beïnvloedt de magnitude van de beweging van een satelliet t.o.v. het baanvlak van een andere satelliet de resultaten van de relatieve navigatie tussen die twee satellieten?
Om formatievliegende satellieten in staat te stellen om hun relatieve toestand te bepalen is er enige vorm van navigatie nodig op de satellieten: Ze moeten minimaal de relatieve afstand (en liefst ook de relatieve snelheid) kunnen meten en, gebruik makend van deze metingen, in staat zijn om een nauwkeurige en betrouwbare schatting van de relatieve toestand te maken met behulp van een model van de systeemdynamica. In een (lage) baan om de Aarde kunnen deze afstandsmetingen worden verkregen door middel van het uitwisselen van metingen van de absolute posities van de satellieten die zijn verkregen door middel van het gebruik van een Global Navigation Satellite System (GNSS) ontvanger op de satellieten. Echter, een dergelijke strategie is niet afdoende voor alle aardbanen en is niet toepasbaar voor missies waarbij de satellieten niet in een baan om de Aarde vliegen. In een dergelijk geval moeten de metingen op een andere manier verkregen worden. Dit kan door middel van lokaal gegenereerde radiofrequentie (RF) signalen, camera's, of lasers. In dit onderzoek wordt er vanuit gegaan dat RF signalen worden gebruikt, meer specifiek, zogenaamde Direct Sequence Spread Spectrum (DSSS) signalen. Dit type signaal wordt ook gebruikt voor GNSS signalen. Daardoor is er een grote technologie-en kennisbasis beschikbaar die kan worden gebruikt voor het implementeren van zulke signalen in een systeem voor afstandsmeting. Verder kan dit type signaal ook gebruikt worden om informatie uit te wisselen tussen de satellieten. De typische datasnelheid die op deze wijze behaald kan worden is niet groot, in de orde van kbit/s, maar dit is voldoende om essentiële data zoals metingen, commando’s, en wellicht zelfs data betreffende de gezondheid van de satelliet uit te wisselen. Nog belangrijker: het neemt de noodzaak voor een apart inter-satelliet communicatiesysteem weg.

De manier waarop de communicatie tussen de satellieten, de "inter-satellite links" (ISLs), wordt geïmplementeerd hangt af van vele aspecten. Hierdoor zal de implementatie van missie tot missie verschillen: Ten eerste moet er een frequentieband worden geselecteerd. Echter, de selectie ervan wordt beperkt door regelgeving en technologie. Voor (zeer) kleine satellieten, die een beperkt elektrisch vermogen beschikbaar hebben, zal een afweging van de voor- en nadelen waarschijnlijk resulteren in de selectie van een signaal met een relatief hoge centrale frequentie. Ook moet interferentie tussen de verschillende signalen worden voorkomen. Dit speelt vooral een rol bij formaties met veel satellieten. Dit kan worden bereikt door middel van meervoudige toegang ("multiple access") technologie waarbij de signalen van elkaar worden gescheiden in het frequentie-, tijds-, en/of code-domein. De keuze voor de te gebruiken multiple access technologie is ook intiem verweven met de communicatie topologie. Op zijn beurt wordt de keuze voor de communicatie topologie bepaald door de manier waarop de communicatie wordt gecontroleerd. Dit kan op een gecentraliseerde, een hiërarchische, of een gedistribueerde manier gebeuren.

Het ontwerp van het meetsignaal zelf is uiteraard ook belangrijk omdat het een grote invloed heeft op de haalbare meetnauwkeurigheid. Voor DSSS signalen moet een pseudo-ruis code worden geselecteerd met lage kruiscorrelatie eigenschappen. Ook moet een verspreidende golfvorm worden geselecteerd met een goed compromis tussen de nauwkeurigheid van de (code-gebaseerde) afstandsmeting en robuustheid van de signaalacquisitie. De BOC(1,1) (Binary Offset Carrier) spreidende symboolvorm biedt een
goede balans tussen deze aspecten en wordt beschouwd als een goede kandidaat voor gebruik door formatievliegende satellieten in de nabije toekomst. Een andere veelbelovende, maar complexere, signaalvorm is Ultra-BOC. Deze heeft enkele voordelige eigenschappen voor de zogenaamde "integer ambiguity resolution" (IAR).

Om afstandsmetingen tussen satellieten mogelijk te maken en om de benodigde nauwkeurigheid in de schatting van de relatieve toestand mogelijk te maken moet een specifieke satellietontwerp worden geïmplementeerd. De toevoeging van een relatief navigatie systeem (RNS) heeft als gevolg dat andere subsystemen moeten worden aangepast om het RNS behoorlijk te laten functioneren. Het gebruik van een RNS heeft ook implicaties voor het operationele gebruik van de satelliet zoals het implementeren van extra veiligheidsmodi en kalibratiefases voor instrumenten.

Voor het RNS is een zendontvanger nodig om de meetsignalen te verzenden, te ontvangen, en om (een deel van) de meetsignalen te verwerken. Zo’n zendontvanger zal grotendeels gebaseerd zijn op GNSS-technologien en moet gebruik maken van een zeer stabiele frequentiebron om meetfouten, veroorzaakt door afwijkingen tussen de klokken op de verschillende satellieten, in de afstandsmeting te minimaliseren. Verder moet de zendontvanger voorzien worden van een interne kalibratie om te compenseren voor de constante afwijking en drift van de elektronica. Actieve controle van de signaalsterkte, gebaseerd op voorspellingen van de actuele afstand tussen de satellieten, is ook een zeer gunstige eigenschap om het stroomverbruik te reduceren voor missies waarbij de afstand tussen de satellieten significant kan variëren.

Om de onderzoeksvragen die aan het begin van de studie gesteld zijn te kunnen beantwoorden is er een dynamisch scenario opgesteld waarin twee satellieten in configuratie vliegen in een lage aardbaan. De toestand van een van de satellieten, genaamd de ‘assistent’, wordt geschat door middel van een niet-lineaire batch kleinste kwadraten (least squares, LSQ) schatter die gebruik maakt van afstandsmetingen tussen de assistent en de andere satelliet, genaamd de ‘baas’. De baas kan een of drie ontvangende antennes hebben waarbij de afstand tussen de antennes kan variëren. De assistent wordt gemodelleerd als een puntmassa en heeft slechts een zendende antenne. De afstand tussen de satellieten en de meetnauwkeurigheid worden ook gevarieerd. De invloed van het type formatie wordt ook bestudeerd door de resultaten te vergelijken die worden behaald voor zogenaamde elliptische en pendulum formatiegeometrieën. In een elliptische formatiegeometrie beschrijft de assistent een elliptische baan ten opzichte van de baas. Er is hier gekozen om het centrum van de ellips samen te laten vallen met het massamiddelpunt van de baas. Bij een pendulum formatiegeometrie heeft de assistent een constante afstand tot de baas in de vliegrichting van de baas evenals een periodieke beweging uit het baanvlak van de baas met een bepaalde amplitude. Om de invloed van de beweging van de assistent uit het baanvlak van de baas te onderzoeken is ook deze beweging gevarieerd voor beide formatiegeometrieën.

Om een algemene behandeling van de onderzoeksvragen mogelijk te maken, zijn er op systeemniveau verscheidene vereenvoudigingen toegepast. Dit heeft geresulteerd in het simpelst mogelijke scenario dat nog steeds de essentiële onderdelen van het probleem bevat. In dit scenario worden signaalinvloeden zoals frequentie, meervoudige toegang,
ambiguïteiten, constante afwijkingen, signaalcodering, signaal interferentie, etc. niet beschouwd omdat deze allen zeer specifieke aspecten zijn die verschillen tussen missies en tijdens de missie zelf. In plaats daarvan worden al deze aspecten samen beschouwd in een aangenomen waarde voor de standaardafwijking van de afstandsmeting. Voor de baanmechanica is ook een zeer simpel model, de Hill-Clohessy-Wiltshire (HCW) vergelijkingen, toegepast. De reden hiervoor is dat het doel van het onderzoek niet het behalen van de meest accurate relatieve navigatie resultaten is, maar om te bepalen hoe enkele belangrijke parameters de relatieve navigatie resultaten beïnvloeden.

Echter, zelfs wanneer deze vereenvoudigingen worden geïmplementeerd blijft het belangrijk om te begrijpen hoe een afstandsmeting verkregen wordt en welke foutbronnen meegenomen moeten worden in een ontwerp. Een voorbeeld is de afwijking van de klokken op de satellieten ten opzichte van elkaar. Doordat deze afwijking continu verandert, is een meting in slechts een richting zoals in GNSS applicaties niet voldoende om de afstand met hoge nauwkeurigheid te meten omdat de tijdsinformatie te onnauwkeurig is. Daarom zijn er tussen elk satelliet-paar altijd metingen in twee richtingen nodig. Dit maakt het mogelijk om de afwijking tussen die twee klokken te schatten. Er zijn ook verschillende statistische en systematische fouten die de meetnauwkeurigheid beïnvloeden, zoals meerwegrefleccties, integer ambiguïteiten, constante afwijking van de ontvanger, de ionosfeer (indien aanwezig), opwinding van de fase van het signaal (voor circulair gepolariseerde signalen), signaal interferentie, de locatie van het fasecentrum van de antenne, de ruistemperatuur van het systeem, en de resolutie van de ontvanger.

Voor de schatting van de relatieve toestand wordt een niet-lineaire batch LSQ schatter gebruikt. Een belangrijke afweging voor de keuze voor de batch LSQ schatter in plaats van een sequentiële filter zoals een uitgebreide Kalman-filter ("extended Kalman filter", EKF) is dat het gebruik van een batch LSQ schatter een directe koppeling tussen het resultaat van de schatting en een zogenaamde waarneembaarheids-analyse ("observability analysis") toestaat. Een dergelijke analyse geeft aan in hoeverre de toestand van een systeem geschat kan worden aan de hand van kennis van de invoer (controle krachten) en/of uitvoer (metingen). Als de waarneembaarheid slecht is, dan kan de toestand van het systeem, of een deelverzameling van de parameters die de toestand van het systeem beschrijven, niet met hoge nauwkeurigheid of in zijn geheel niet bepaald worden. Anderzijds geeft een goede waarneembaarheid aan dat de toestand van het systeem met hoge nauwkeurigheid geschat kan worden. Een analyse van de waarneembaarheid kan dus inzicht bieden op de vraag hoe klein de relatieve beweging uit het baanvlak kan zijn opdat er toch nog een schatting van de relatieve toestand gemaakt kan worden.

Naast deze numerieke analyse is er ook een statistische analyse ontwikkeld om inzicht te verkrijgen in de samenhang tussen nauwkeurigheid van de afstandsmeting, de afstand tussen de satellieten, en de basislijn van de antennes. De statistische analyse kan ook worden gebruikt om een snelle eerste schatting te verkrijgen van de nauwkeurigheid in de navigatie die kan worden verwacht voor een missie met een formatie van satellieten. Hoewel de analyse uitgaat van een statische configuratie kan deze ook worden toegepast in dynamische configuraties. Echter, de analyse zal niet de systeemdynamica meenemen zoals dit wordt gedaan in een schatings-algoritme. Generieke vergelijkingen zijn afgeleid
voor de verwachte waarde en de variantie in de schatting van de positie van een stationair punt in de ruimte (de zendende antenne) gebruikt makend van afstandsmetingen, met een variabele normaal verdeelde meetfout, naar drie verschillende punten (de ontvangende antennes). In de statistische analyse wordt ook aangetoond dat, indien de ontvangende antennes zich op de hoekpunten van een rechte driehoek bevinden, de variantie in de schatting van de positie van de zender geminimaliseerd wordt. De afgeleide vergelijkingen tonen verder aan dat, om de hoogste nauwkeurigheid in de positieschatting te verkrijgen, over het algemeen de meetnauwkeurigheid hoog moet zijn, de afstand tussen de satellieten klein moet zijn, en dat de basislijn tussen de antennes groot moet zijn. Deze laatste conclusie komt overeen met algemene kennis op het gebied van navigatie. De analyse behandelt ook kort resultaten voor twee simpele formatie-configuraties.

De numerieke analyses die zijn uitgevoerd tonen aan dat, voor de beschouwde formatie-geometrieën, de nauwkeurigheid in de schatting van de relatieve toestand lineair afhangt van de meetnauwkeurigheid en de afstand tussen de satellieten. Dit is als verwacht en wordt niet beschouwd als een nieuw resultaat. Echter, een nieuw resultaat is dat de nauwkeurigheid in de schatting van de relatieve toestand niet afhankelijk is van de basislijn tussen de antennes: Voor een elliptische formatie-geometrie is het zeer belangrijk dat de vector van het beginpunt van het lokale referentievlak naar een (van de) antenne(s) in het lokale referentievlak haaks staat op de richting van de slechtst waarneembare eigenvector van het systeem. Verder schaalt de nauwkeurigheid in de schatting van de relatieve toestand evenredig met de lengte van deze vector. Voor een pendulum formatie-geometrie heeft de basislijn tussen de antennes alleen invloed indien de meetperiode kleiner is dan een omwenteling. Indien de meetperiode langer is heeft de basislijn geen invloed meer op de nauwkeurigheid in de schatting van de relatieve toestand.

Voor een pendulum formatie-geometrie is het aannemelijker dat een enkele antenne per satelliet kan worden gebruikt voor de afstandsmeting tussen de satellieten dan voor een elliptische formatie-geometrie. Dit wordt veroorzaakt door een betere waarneembaarheid voor de pendulum formatie-geometrie en door een gunstigere zichtlijn gedurende een omwenteling. Dus, een pendulum formatie-geometrie leent zich beter voor het gebruik van kleine satellieten dan een elliptische formatie-geometrie.

Voor de relatieve navigatie is gebleken dat de magnitude van de relatieve beweging uit het baanvlak op zich niet heel belangrijk is voor de waarneembaarheid van de relatieve toestand en voor de nauwkeurigheid in de schatting van de relatieve toestand. Het is namelijk de relatieve hoek ten opzichte van het baanvlak die hiervoor leidend is. Deze observatie kan gebruikt worden in formaties die uit meer dan twee satellieten bestaan om het aantal signaalverbindingen tussen de satellieten te reduceren: Indien er een satelliet-paar is met een kleine relatieve beweging uit het baanvlak, dan kan de relatieve navigatie tussen deze twee satellieten het beste gedaan worden via een andere satelliet die, ten opzichte van het originele satelliet-paar, een grote relatieve beweging uit het baanvlak heeft.

Tot slot heeft dit onderzoek aangetoond dat in het geval van een pendulum formatie-geometrie de relatieve hoek ten opzichte van het baanvlak enkele malen kleiner kan zijn (3°) dan in het geval van een elliptische formatie-geometrie (14°) zonder dat de kans dat het schattingsproces faalt zeer hoog wordt.
In deze thesis is gebruikmakend van analytische en numerieke methodes en adequate technologieën voor de afstandsmeting tussen twee satellieten, onderzoek gedaan naar de invloed van de meetnauwkeurigheid, de afstand tussen de satellieten, en de geometrie van de antennes voor verscheidene configuraties van formatievliegende satellieten. Verwacht wordt dat de resultaten van dit onderzoek bijdragen aan het creëren en het mogelijk maken van effectieve en efficiënte toekomstige missies met kleine, in formatie vliegende, satellieten.
First and foremost, I want to express my sincere gratitude to my supervisor and promotor, Prof. Dr. Eberhard Gill, for providing me with the opportunity to perform the research that has culminated in this thesis. Furthermore, his support, criticism, encouragement, and suggestions have been of great value during my research.

I also would like to thank TU Delft staff members Jian Guo, Barry Zandbergen, Jasper Bouwmeester, and Brian Gunter for the work I performed with them, which I have done with pleasure.

Of course, a big "thank you" also goes out to my fellow Ph.D. students Rouzbeh Amini, Arash Noroozi, Prem Sundaramoorthy, Rui Sun, Jing Chu, and Steven Engelen for the valuable discussions and for the laughs that we had.

I also would like to thank Prof. Marco D'Errico for providing me the opportunity to write a chapter in his book on distributed space missions for Earth system monitoring. I also would like to thank the co-authors of that chapter, Michel Delpech and Thomas Grelier of CNES, for their valuable contributions.

Last but not least, I want to thank my parents and my lovely wife Jantine, for always being there when I needed you and for putting up with me during the many weekends and holidays which I spent completing my thesis.

Daan Maessen  
's-Hertogenbosch, The Netherlands, February 2014
Table of Contents

Summary ................................................................. i
Samenvatting ............................................................ vii
Acknowledgments ....................................................... xv
Table of Contents ....................................................... xix
List of Tables ........................................................... xxi
List of Figures .......................................................... xxiv
Nomenclature ........................................................... xxv

1 Introduction ........................................................... 1
   1.1 Background ...................................................... 1
      1.1.1 Distributed Space Systems .............................. 2
      1.1.2 Constellations and Formations .......................... 5
      1.1.3 Autonomous Formation Flying ........................... 7
      1.1.4 Potentials and Challenges of Formation Flying ........ 9
      1.1.5 Relative Navigation Needs ............................... 12
   1.2 State of the Art ................................................ 17
      1.2.1 Basic Research .......................................... 17
      1.2.2 Achievable Relative Navigation Accuracy .............. 19
      1.2.3 Opportunities for Knowledge Improvement ............. 24
   1.3 Research Questions .......................................... 24
   1.4 Methodologies ................................................. 26
      1.4.1 Approach ................................................ 26
      1.4.2 Scenario ................................................ 27
   1.5 Thesis Outline ................................................. 28

2 Relative Navigation System Design Considerations ........... 29
   2.1 Relative Navigation System Needs ........................... 29
   2.2 Signal Level Design Considerations ......................... 35
      2.2.1 Frequency Selection .................................... 35
      2.2.2 Multiple Access ....................................... 37
      2.2.3 Ranging Signal Design .................................. 40
      2.2.4 Data Modulation and Bit Errors ....................... 46
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.5 Cramér-Rao Covariance Bound</td>
<td>47</td>
</tr>
<tr>
<td>2.3 Hardware Level Design Considerations</td>
<td>53</td>
</tr>
<tr>
<td>2.3.1 Spacecraft Design</td>
<td>53</td>
</tr>
<tr>
<td>2.3.2 Transceiver Design</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3 Clock Stability</td>
<td>57</td>
</tr>
<tr>
<td>2.4 Summary</td>
<td>61</td>
</tr>
<tr>
<td>3 Relative Navigation Methods and Analysis</td>
<td>65</td>
</tr>
<tr>
<td>3.1 Rationale for System Level Simplifications</td>
<td>65</td>
</tr>
<tr>
<td>3.2 Relative Orbital Dynamics Modeling</td>
<td>67</td>
</tr>
<tr>
<td>3.2.1 Nonlinear Relative Motion Around a Common Primary</td>
<td>68</td>
</tr>
<tr>
<td>3.2.2 Linearized Relative Motion Around a Common Primary</td>
<td>71</td>
</tr>
<tr>
<td>3.3 Observations</td>
<td>74</td>
</tr>
<tr>
<td>3.3.1 Ranging Principles</td>
<td>74</td>
</tr>
<tr>
<td>3.3.2 Code and Carrier Observations</td>
<td>77</td>
</tr>
<tr>
<td>3.3.3 Line-Of-Sight Observations</td>
<td>79</td>
</tr>
<tr>
<td>3.4 Relative State Estimation</td>
<td>80</td>
</tr>
<tr>
<td>3.4.1 Integer Ambiguity Resolution</td>
<td>81</td>
</tr>
<tr>
<td>3.4.2 Estimation Approach</td>
<td>84</td>
</tr>
<tr>
<td>3.4.3 Observability Analysis</td>
<td>90</td>
</tr>
<tr>
<td>3.4.4 (Apparent) Filter Divergence</td>
<td>93</td>
</tr>
<tr>
<td>3.4.5 Statistical Analysis of the Relative Position Determination</td>
<td>101</td>
</tr>
<tr>
<td>3.5 Summary</td>
<td>106</td>
</tr>
<tr>
<td>4 Relative Navigation Simulation and Analysis</td>
<td>109</td>
</tr>
<tr>
<td>4.1 Objectives</td>
<td>109</td>
</tr>
<tr>
<td>4.2 Scenario</td>
<td>110</td>
</tr>
<tr>
<td>4.3 Simulation Geometries</td>
<td>113</td>
</tr>
<tr>
<td>4.3.1 Elliptical Formation</td>
<td>113</td>
</tr>
<tr>
<td>4.3.2 Pendulum Formation</td>
<td>116</td>
</tr>
<tr>
<td>4.4 Analysis</td>
<td>118</td>
</tr>
<tr>
<td>4.4.1 Observability of the Relative States</td>
<td>118</td>
</tr>
<tr>
<td>4.4.2 Relative Position Estimation Results</td>
<td>126</td>
</tr>
<tr>
<td>4.4.3 Apparent Estimator Divergence</td>
<td>136</td>
</tr>
<tr>
<td>4.4.4 Antenna Baseline Influence for the Pendulum Formation</td>
<td>139</td>
</tr>
<tr>
<td>4.4.5 Observability Increase for three Spacecraft</td>
<td>145</td>
</tr>
<tr>
<td>5 Conclusions</td>
<td>153</td>
</tr>
<tr>
<td>5.1 Summary</td>
<td>153</td>
</tr>
<tr>
<td>5.2 Conclusions</td>
<td>156</td>
</tr>
<tr>
<td>5.3 Main Contributions of this Work</td>
<td>162</td>
</tr>
<tr>
<td>5.4 Recommendations for Further Work</td>
<td>166</td>
</tr>
<tr>
<td>5.5 Outlook</td>
<td>167</td>
</tr>
<tr>
<td>Appendix A Derivation of the Expectation and Variance in the Position Estimate of a Stationary Transmitter</td>
<td>169</td>
</tr>
<tr>
<td>A.1 Background Material</td>
<td>169</td>
</tr>
<tr>
<td>A.1.1 Expected Value</td>
<td>169</td>
</tr>
<tr>
<td>A.1.2 Variance</td>
<td>170</td>
</tr>
<tr>
<td>A.1.3 Covariance</td>
<td>170</td>
</tr>
<tr>
<td>A.1.4 Multivariate Taylor Series Expansion</td>
<td>171</td>
</tr>
<tr>
<td>A.2 Derivation</td>
<td>172</td>
</tr>
<tr>
<td>A.2.1 Derivations for $x$</td>
<td>173</td>
</tr>
<tr>
<td>A.2.2 Derivations for $y$</td>
<td>177</td>
</tr>
<tr>
<td>A.2.3 Derivations for $z$</td>
<td>178</td>
</tr>
<tr>
<td>List of Author's Publications</td>
<td>181</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Control ratio</td>
<td>15</td>
</tr>
<tr>
<td>1.2</td>
<td>Inter-satellite ranging technologies</td>
<td>22</td>
</tr>
<tr>
<td>2.1</td>
<td>Recommended frequencies for RF-based inter-satellite crosslinks</td>
<td>35</td>
</tr>
<tr>
<td>2.2</td>
<td>Key effects on frequency of RF ranging signals</td>
<td>37</td>
</tr>
<tr>
<td>2.3</td>
<td>Effect of chipping frequency and pulse shape on the MCRB</td>
<td>51</td>
</tr>
<tr>
<td>2.4</td>
<td>Oscillator comparison</td>
<td>59</td>
</tr>
<tr>
<td>2.5</td>
<td>Allan variance-induced oscillator jitter</td>
<td>60</td>
</tr>
<tr>
<td>4.1</td>
<td>Variations in simulation parameters</td>
<td>111</td>
</tr>
<tr>
<td>4.2</td>
<td>Observability of the relative state in case of a single Rx antenna not located at the origin of the Hill frame</td>
<td>125</td>
</tr>
<tr>
<td>4.3</td>
<td>Empirical relations for the estimation accuracies for a measurement arc length of five orbits</td>
<td>127</td>
</tr>
<tr>
<td>4.4</td>
<td>Mean and maximum relative errors in the empirical relation for the safe ellipse formation</td>
<td>130</td>
</tr>
<tr>
<td>4.5</td>
<td>Mean and maximum relative errors in the empirical relation for the pendulum formation</td>
<td>130</td>
</tr>
<tr>
<td>4.6</td>
<td>Accuracy in the determination of the relative state for an elliptical formation geometry in case of a single Rx antenna not located at the origin of the Hill frame</td>
<td>134</td>
</tr>
<tr>
<td>4.7</td>
<td>Observability results for various formations for $\Delta x^0_0$</td>
<td>148</td>
</tr>
<tr>
<td>4.8</td>
<td>Estimation results for various formations</td>
<td>149</td>
</tr>
</tbody>
</table>
# List of Figures

1.1 Breakdown of all currently active space systems in Earth orbit .......................... 2  
1.2 Qualitative categorization of distributed space systems ................................. 3  
1.3 Examples of distributed space systems .......................................................... 4  
1.4 A typical constellation and a typical formation .............................................. 6  
1.5 Synergy between the data from the payloads on different spacecraft in a formation .............................................................................................................. 7  
1.6 Formation control accuracy versus relative distance ............................................ 14  
1.7 Scaling of important formation flying characteristic length parameters ............. 16  
1.8 Relative navigation options .............................................................................. 20  
2.1 RF-based relative navigation principle .............................................................. 30  
2.2 Dual one-way ranging principle ........................................................................ 34  
2.3 Schematic representation of CDMA, FDMA, and TDMA ................................. 38  
2.4 Communication topologies .............................................................................. 39  
2.5 Example of the correlation of two binary sequences ........................................... 42  
2.6 DSSS modulation ............................................................................................. 44  
2.7 Auto-correlation and amplitude spectra of several ranging codes ..................... 45  
2.8 Effect of bandlimiting on signal autocorrelation .............................................. 50  
2.9 Simplified transceiver block diagram ............................................................... 55  
3.1 Definition of the Hill frame .............................................................................. 69  
3.2 Definition of the Line-Of-Sight angle $\psi$ ......................................................... 79  
3.3 Ultra-BOC signal structure ............................................................................. 83  
3.4 Schematic representation of the relative state estimation process ....................... 95  
3.5 State estimation results for linear and nonlinear reference trajectories ............. 95  
3.6 Graphical representation of the estimator divergence problem .......................... 97  
3.7 Results for various filters and estimators .......................................................... 100  
3.8 Location of the various antennas in the receiver reference frame ....................... 101  
3.9 Definition of the angles in the triangles formed by the Tx and Rx antennas 102
Nomenclature

Abbreviations

ADCS  Attitude Determination and Control System
AFF   Autonomous Formation Flying
ASIC  Application Specific Integrated Circuit
BER   Bit Error Rate
BOC   Binary Offset Carrier
BPSK  Binary Phase Shift Keying
CDHS  Command and Data Handling System
CDMA  Code Division Multiple Access
CLT   Crosslink Transceiver
COM   Center Of Mass
CRB   Cramér-Rao Bound
CRTBP Circular Restricted Three Body Problem
CW    Clohessy-Wiltshire
DGPS  Differential Global Positioning System
DLL   Delay Lock Loop
DSP   Digital Signal Processor
DSSS  Direct Sequence Spread Spectrum
ECI   Earth-Centered Inertial
EKF   Extended Kalman Filter
EPS   Electrical Power Subsystem
FDIR  Fault Detection, Isolation and Recovery
FDMA  Frequency Division Multiple Access
FOV   Field Of View
FPGA  Field Programmable Gate Array
GCO   General Circular Orbit
GNC   Guidance, Navigation & Control
GPS   Global Positioning System
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNSS</td>
<td>Global Navigation Satellite System</td>
</tr>
<tr>
<td>HCW</td>
<td>Hill-Clohessy-Wiltshire</td>
</tr>
<tr>
<td>I</td>
<td>In phase signal component</td>
</tr>
<tr>
<td>IAR</td>
<td>Inter Ambiguity Resolution</td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate Frequency</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>KF</td>
<td>Kalman Filter</td>
</tr>
<tr>
<td>LEO</td>
<td>Low Earth Orbit</td>
</tr>
<tr>
<td>LNA</td>
<td>Low Noise Amplifier</td>
</tr>
<tr>
<td>LOS</td>
<td>Line-of-Sight</td>
</tr>
<tr>
<td>LVLH</td>
<td>Local Vertical Local Horizontal</td>
</tr>
<tr>
<td>LSQ</td>
<td>Least Squares</td>
</tr>
<tr>
<td>MAI</td>
<td>Multiple Access Interference</td>
</tr>
<tr>
<td>MCRB</td>
<td>Modified Cramér-Rao Bound</td>
</tr>
<tr>
<td>MCXO</td>
<td>Microcomputer Compensated Crystal Oscillator</td>
</tr>
<tr>
<td>MEO</td>
<td>Medium Earth Orbit</td>
</tr>
<tr>
<td>max</td>
<td>Maximum</td>
</tr>
<tr>
<td>mtSOF</td>
<td>Modified-truncated Second-Order Filter</td>
</tr>
<tr>
<td>NAVSTAR</td>
<td>Navigation Signal Timing and Ranging</td>
</tr>
<tr>
<td>NCLT</td>
<td>Nanosat Crosslink Transceiver</td>
</tr>
<tr>
<td>NORAD</td>
<td>North American Aerospace Defense Command</td>
</tr>
<tr>
<td>OBC</td>
<td>On-Board Computer</td>
</tr>
<tr>
<td>OCXO</td>
<td>Oven Controlled Crystal Oscillator</td>
</tr>
<tr>
<td>PCO</td>
<td>Projected Circular Orbit</td>
</tr>
<tr>
<td>PLL</td>
<td>Phase Lock Loop</td>
</tr>
<tr>
<td>PPS</td>
<td>Pulse-Per-Second</td>
</tr>
<tr>
<td>PRARE</td>
<td>Precise Range and Range Rate Equipment</td>
</tr>
<tr>
<td>PRN</td>
<td>Pseudo-Random Noise</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectral Density</td>
</tr>
<tr>
<td>PSF</td>
<td>Pulse Shaping Factor</td>
</tr>
<tr>
<td>Q</td>
<td>Quadrature signal component</td>
</tr>
<tr>
<td>QPSK</td>
<td>Quadrature Phase Shift Keying</td>
</tr>
<tr>
<td>RAAN</td>
<td>Right Ascension of the Ascending Node</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RIC</td>
<td>Radial In-track Cross-track</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>RNS</td>
<td>Relative Navigation System</td>
</tr>
<tr>
<td>RSS</td>
<td>Root Sum Square</td>
</tr>
<tr>
<td>RTN</td>
<td>Radial Tangential Normal</td>
</tr>
<tr>
<td>RVD</td>
<td>Rendezvous and Docking</td>
</tr>
<tr>
<td>Rx</td>
<td>Receiver</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>TC</td>
<td>Telecommand</td>
</tr>
</tbody>
</table>
TCS Thermal Control System
TCXO Temperature Compensated Crystal Oscillator
TDMA Time Division Multiple Access
TEC Total Electron Content
TLE Two-Line Elements
TM Telemetry
TPF Terrestrial Planet Finder
TT&C Telemetry, Tracking and Command
Tx Transmitter
UHF Ultra-High Frequency
XO Crystal Oscillator

Roman Symbols
A System matrix
$A_e$ Effective antenna area
a Semi-major axis [m]
B Bandwidth [Hz]
$B_L$ Equivalent noise bandwidth of the tracking loop [Hz]
b Path difference [m]
C Average signal power [W]
$C_r$ Received signal power [W]
$C_t$ Transmitted signal power [W]
$C/N_0$ Carrier-to-noise density ratio [dB-Hz]
Cov[·] Covariance
c PRN code bits vector
c Speed of light in vacuum [m/s]
c_k PRN code bit
d Antenna baseline [m]
$\mathcal{E}$ Energy [Ws]
E Lumped error term
E[·] Expectation
e$_x$, e$_y$, e$_z$ Rotating frame unit vectors
e error
F Probability density function
f Signal frequency [Hz]
f$_c$ Chipping rate [Hz]
f$_{dev}$ Relative frequency deviation [-]
f$_{ref}$ Reference frequency [Hz]
f$_s$ Subcarrier frequency [Hz]
G Gramian matrix
G(f) Pulse shape in the frequency domain
$G_r$ Receive antenna gain
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_t$</td>
<td>Transmit antenna gain</td>
</tr>
<tr>
<td>$\mathbf{g}$</td>
<td>Control law vector function</td>
</tr>
<tr>
<td>$g(t)$</td>
<td>Pulse shape in the time domain</td>
</tr>
<tr>
<td>$\mathbf{H}$</td>
<td>Measurement sensitivity matrix</td>
</tr>
<tr>
<td>$\mathbf{h}$</td>
<td>Measurement vector function</td>
</tr>
<tr>
<td>$h$</td>
<td>Measurement function</td>
</tr>
<tr>
<td>$\mathcal{J}$</td>
<td>Inertial reference frame</td>
</tr>
<tr>
<td>$I$</td>
<td>Identity matrix</td>
</tr>
<tr>
<td>$\mathbf{I}$</td>
<td>Fisher information</td>
</tr>
<tr>
<td>$J_2$</td>
<td>Second-order zonal coefficient of the Earth's gravity field</td>
</tr>
<tr>
<td>$J$</td>
<td>Loss function</td>
</tr>
<tr>
<td>$j$</td>
<td>Number of receiver antennas</td>
</tr>
<tr>
<td>$K$</td>
<td>Integer number of signal cycles</td>
</tr>
<tr>
<td>$k$</td>
<td>Number of iterations</td>
</tr>
<tr>
<td>$\mathcal{L}$</td>
<td>Local reference frame</td>
</tr>
<tr>
<td>$L_i$</td>
<td>Lagrange point identifier</td>
</tr>
<tr>
<td>$l$</td>
<td>Number of measurements</td>
</tr>
<tr>
<td>$m$</td>
<td>Mass [kg]</td>
</tr>
<tr>
<td>$N$</td>
<td>'Large' number</td>
</tr>
<tr>
<td>$N_0$</td>
<td>Noise power spectral density [dBW/Hz]</td>
</tr>
<tr>
<td>$N_c$</td>
<td>PRN code length</td>
</tr>
<tr>
<td>$n(t)$</td>
<td>White Gaussian noise</td>
</tr>
<tr>
<td>$n$</td>
<td>Orbital mean motion [rad/s]</td>
</tr>
<tr>
<td>$\mathbf{P}$</td>
<td>Covariance matrix</td>
</tr>
<tr>
<td>$\mathbf{P}$</td>
<td>Period [s]</td>
</tr>
<tr>
<td>$R(\tau)$</td>
<td>Correlation function</td>
</tr>
<tr>
<td>$\mathbf{R}$</td>
<td>Orbit radius vector</td>
</tr>
<tr>
<td>$R$</td>
<td>Orbit radius [m]</td>
</tr>
<tr>
<td>$r$</td>
<td>Range or relative position vector</td>
</tr>
<tr>
<td>$r$</td>
<td>Range [m]</td>
</tr>
<tr>
<td>$r_e$</td>
<td>Inter-satellite distance at epoch [m]</td>
</tr>
<tr>
<td>$S(f)$</td>
<td>Power spectral density of the signal $s(t)$</td>
</tr>
<tr>
<td>$\mathbf{s}$</td>
<td>Vector representation of the signal $s(t)$</td>
</tr>
<tr>
<td>$s(t)$</td>
<td>Ranging signal in the time domain</td>
</tr>
<tr>
<td>$T$</td>
<td>Signal length [s]</td>
</tr>
<tr>
<td>$T_c$</td>
<td>Chip period [s]</td>
</tr>
<tr>
<td>$T_{\text{pre}}$</td>
<td>Predetection integration time</td>
</tr>
<tr>
<td>$t$</td>
<td>Time [s]</td>
</tr>
<tr>
<td>$\mathbf{U}$</td>
<td>Matrix with the left eigenvectors of $\tilde{\Xi}$</td>
</tr>
<tr>
<td>$\mathbf{u}$</td>
<td>Control input vector</td>
</tr>
<tr>
<td>$\mathbf{V}$</td>
<td>Inertial velocity vector</td>
</tr>
<tr>
<td>$V$</td>
<td>Output voltage [Volts]</td>
</tr>
<tr>
<td>$\text{Var} [\cdot]$</td>
<td>Variance</td>
</tr>
</tbody>
</table>
Relative velocity vector $v$

Weighting matrix $W$

Process noise vector $w$

Weighting factor $w$

Position along the $X$-axis of an inertial frame [m] $X$

State vector $x$

Position along the $x$-axis of a local (rotating) frame [m] $x$

Position along the $Y$-axis of an inertial frame [m] $Y$

Position along the $y$-axis of a local (rotating) frame [m] $y$

Position along the $Z$-axis of an inertial frame [m] $Z$

Position along the $z$-axis of a local (rotating) frame [m] $z$

Angle between two antenna baselines [rad] $\alpha$

Angle in triangle $Rx_0$-$Rx_1$-$Tx$ [rad] $\beta$

Angle in triangle $Rx_0$-$Rx_2$-$Tx$ [rad] $\gamma$

Square of the Gabor bandwidth $\Delta \omega^2$

Phase error [rad] $\delta\varphi$

Measurement noise vector $\varepsilon$

Normally distributed random noise $\varepsilon$

Voltage error [Volts] $\varepsilon_V$

Fraction of signal power remaining after bandlimiting to $B$ [-] $\zeta$

Angular velocity vector $\Theta$

Angular velocity [rad/s] $\dot{\Theta}$

Angle [rad] $\theta$

Condition number $\kappa$

Information matrix $\Lambda$

Wavelength [m] $\lambda$

Gravitational coefficient of the primary body [m$^3$/s$^2$] $\mu$

Normalized measurement sensitivity matrix $\Xi$

Pulse shaping factor $\xi$

Pseudorange [m] $\rho$

Matrix with singular values $\Sigma$

Singular value $\Sigma$

Variance $\sigma^2$

Standard deviation $\sigma$

Ranging accuracy [m] $\sigma_{\rho}$

Accuracy of the a-priori relative position estimate [m] $\sigma_{\text{apr}}$

Signal travel time (delay) [s] $\tau$

Eigenvector matrix $\Upsilon$

Eigenvector $\upsilon$

Eigenvalue $\upsilon$
Φ  State transition matrix
φ  Phase [rad]
ψ  Line-Of-Sight angle [rad]
ω  Angular frequency [rad/s]

Operators

d  Differential operator
∂  Partial differential operator
Δ  Arithmetic difference operator
(·)*  Complex conjugation
(·)  Random error
(·)  Derivative with respect to time
(·)’  Derivative with respect to normalized time
< · >  Average
∥·∥  Euclidian vector norm
(·)T  Matrix transpose
×  Vector cross-product
|·|  Absolute value
•  Vector dot product
(·)̂  Estimated quantity
(·)˜  Modified matrix
(·)−  Predicted variable
(·)+  Updated variable
Chapter 1

Introduction

This chapter provides an introduction to the contents of this thesis. First, an extensive background of the subject area of the thesis, autonomous relative navigation for small spacecraft, is provided. Then, the research questions and the motivation for these questions are treated, followed by the methodology used to answer the research questions. A short description of the individual chapters of this thesis is provided at the end of this chapter.

1.1 Background

In the context of spaceflight, navigation commonly involves the estimation of the position and velocity of a spacecraft expressed in an inertial reference frame. For spacecraft orbiting the Earth, a reference frame that is often used is one whose origin is located at the Earth's center of mass and which has one axis aligned with the vernal equinox and the other axis aligned with the Earth's rotational axis. For disambiguation, spacecraft navigation in such a reference frame is sometimes referred to as absolute navigation. This is needed to distinguish navigation in this frame from relative navigation, where the position and velocity of one or more spacecraft with respect to another spacecraft is estimated. Relative navigation can be advantageous when multiple spacecraft are used in a single mission and when these spacecraft need to be in (relatively) close proximity to each other. Relative navigation for spacecraft practically always concerns either spacecraft formations, where multiple spacecraft need to maintain a predefined motion or position relative to each other, or spacecraft that perform rendezvous and docking (RVD). Depending on the mission needs, spacecraft formations can take many shapes and sizes. They can for instance trail each other (a so-called train), rotate around a common center in an ellipse (elliptical formation), or align and rotate themselves as if part of a single monolithic structure. In this thesis, the spacecraft are always assumed to fly in a spacecraft formation, but the results are also applicable to spacecraft performing RVD. In this chapter, the
three 'hinge points' of the thesis, *formation flying*, *autonomy*, and *relative navigation* are discussed in a general context.

### 1.1.1 Distributed Space Systems

Ever since the dawn of the space era in the 1950s up to the present, space systems predominantly consist out of a single spacecraft as this is mostly sufficient to meet the requirements the system has to fulfill. This is illustrated in Figure 1.1 which shows a breakdown of all currently active space systems in Earth orbit into systems consisting of a single spacecraft, a constellation of spacecraft, or a formation of spacecraft. However, as requirements for new missions are becoming harder, or even impossible, to meet with a single, monolithic spacecraft, it is expected that there will be an increase in the number of spacecraft constellations and especially formations relative to the number of single spacecraft systems.

![Figure 1.1](image.png)

*Figure 1.1:* Breakdown of all currently active space systems in Earth orbit into systems with a single spacecraft, a constellation of spacecraft, or a formation of spacecraft. The percentages represent the number of systems of a certain type \(x\) relative to the total number of systems \(y\), equal to 325. Thus, for the single spacecraft systems \(x/y \times 100\% = 64\%\). The information for this figure has been obtained from [UCS 2011].

Constellations and formations are examples of the more general class of distributed space systems, to which also spacecraft swarms and spacecraft that perform RVD can be counted. A distributed space system is a collection of spacecraft which are purposefully located at different locations in space in order to achieve certain mission objectives. The number of spacecraft in a distributed spacecraft system can in essence range from as few as two up to as many as several thousands. Up to date the largest distributed spacecraft system that has actually been deployed is the Iridium constellation, consisting out of 66 active satellites. Although not always necessary, individual spacecraft in a distributed space system can be required to be able to communicate with each other. No commonly agreed upon definitions exist for the various types of distributed space systems. However,
a distinction between these systems can usually be made based on the distance between the spacecraft and the required accuracy with which this distance needs to be controlled, cf. Figure 1.2.

![Figure 1.2: Qualitative categorization of distributed space systems, after [Gill 2008].](image)

Spacecraft that perform RVD require high (centimeter-level) control accuracy and operate at small (meter-level) inter-satellite distance. An example of an advanced form of RVD is the Orbital Express mission, cf. Fig. 1.3(a). Typical RVD missions are flown by spacecraft that resupply a space station, such as the Progress, Soyuz, ATV, HTV, Dragon, and the (decommissioned) Space Shuttle spacecraft.

Spacecraft in a constellation require very relaxed (kilometer-level) control accuracy and operate at large (> 1000 km) inter-satellite distance such as in the aforementioned Iridium constellation, cf. Fig. 1.3(b). Large constellations are primarily used for telecommunication (e.g., Iridium, Globalstar, Orbcomm) and navigation (e.g., GPS, GLONASS, Galileo, Compass), but there are also many examples of smaller communication (e.g., Inmarsat, Thuraya), remote sensing (e.g., Disaster Monitoring Constellation, RapidEye, SAR-Lupe) and meteorologic (e.g., COSMIC, GOES, Meteosat) constellations.

Formations fill the gap between RVD systems and constellations and generally require high to moderate (meter-level) control accuracy and commonly operate at moderate (tens of m to several km) inter-satellite distance. An example of an ambitious formation flying mission is the TPF-I mission, which is designed to detect Earth-like exoplanets by means of nulling interferometry, cf. Fig. 1.3(c). Other examples of spacecraft formations are PRISMA, TerraSAR-X and TanDEM-X, PROBA3, Darwin, MMS, JC2Sat, MAXIM, and Planet Imager.
In Figure 1.2, the spacecraft swarm is semitransparent since no operational spacecraft swarm has been deployed yet. Conceptually, spacecraft swarms mimic swarms found in nature as they consist out of a very large (tens to thousands) number of small, identical, and simple elements whose relative motion is uncontrolled, except for collision avoidance or for ad hoc observations where a certain relative geometry of (some) swarm elements is needed for a limited time duration. An example of a spacecraft swarm is the OLFAR mission, cf. Fig. 1.3(d). Other examples of spacecraft swarms are ANTS and APIES.

To be complete, the class of the so-called fractionated spacecraft is also briefly mentioned as final example of a distributed space system. Unlike constellations or formations, where usually similar spacecraft are spatially distributed, a fractionated spacecraft distributes the functional capabilities of a conventional monolithic spacecraft amongst multiple heterogeneous modules which perform distinct functions and interact through wireless communication links, cf. [Guo et al. 2009]. However, due to the technical immaturity of the enabling technologies for these systems (wireless networking, cluster operation, inter-satellite communication, wireless power transfer, distributed computing) and the lack of a convincing business case, these systems are far from operational and are not considered in Fig. 1.2. In 2013, the most advanced fractionated spacecraft program, System F6, was even canceled [Ferster 2013].
1.1.2 Constellations and Formations

As no clear and universally agreed upon definitions exist for the various types of distributed systems, particularly the classification of a distributed space system as either a formation or a constellation can be rather arbitrary as their inter-satellite distances and control accuracies can be similar. For example, the Afternoon Constellation (also known as the A-Train or the PM Constellation), where the inter-satellite distances are at least tens of km and control accuracies are on the order of km [NASA 2011], can be regarded as a formation but also as a constellation. For this reason, the following definitions of spacecraft formations and constellations are postulated here, which will be used throughout this thesis:

**Definition 1** A spacecraft formation is a distributed space system in which the relative position, velocity, and possibly the relative attitude of the involved spacecraft have to be actively controlled in order to maintain these within preset boundaries, thereby enabling the synergistic use of payloads on board the different spacecraft.

**Definition 2** A spacecraft constellation is a distributed space system in which the absolute positions of the involved spacecraft are controlled such that a shorter revisit time and/or increased coverage of the subject of interest is achieved compared to a space system consisting of a single spacecraft.

Thus, following the above definitions, the Afternoon Constellation is considered a formation in this work.

The difference between a constellation and a formation is made more insightful by comparing the Afternoon Constellation with the Navigation Signal Timing and Ranging (NAVSTAR) Global Positioning System (GPS), which is a typical example of a spacecraft constellation and a specific example of a Global Navigation Satellite System (GNSS). The space segment of the GPS system consists out of 24 active spacecraft in medium Earth orbit (MEO), which are equally divided over six orbital planes, cf. Figure 1.4(a). Each GPS spacecraft transmits a distinct radio frequency (RF) ranging signal that can be acquired and processed by a receiver, allowing an accurate estimation of the position of the receiver if the signals of at least four GPS spacecraft are received. Thus, the minimum number of spacecraft needed for this system, provided that they can all be in view of a receiver at the same time instant, is four. The additional 20 spacecraft in the operational system are thus, in essence, primarily needed for increased coverage, reliability, and, due to the nature of the position estimation process, for increased accuracy (provided that signals from more than four spacecraft can be processed at the same time instance). Increased coverage and reliability both contribute to increased availability.
In contrast to the GPS system, the performance of the Afternoon Constellation, currently consisting of the Earth observation spacecraft Aqua, Aura, CALIPSO, CloudSat, and PARASOL, does rely heavily on the inter-satellite distance between these satellites. The purpose of the Afternoon Constellation is namely to create synergy between the data from the payloads on the different spacecraft. This can only be achieved if the payloads observe the same geolocation at, preferably, the same time instance. Thus, the orbits of the spacecraft have to be very similar and the spacing of the orbit nodes and the times of node crossing have to be kept within certain bounds. Since the payloads on the different spacecraft all have different fields of view (FOV), cf. Figure 1.4(b), the inter-satellite distances and control boxes are different for each spacecraft. Examples of science results that can be obtained in this manner are shown in Figure 1.5. Figure 1.5(a) shows the ash cloud of the Eyjafjallajökull volcano in Iceland, which erupted in 2010, as observed by the MODIS instrument on Aqua (left pane) and by the OMI instrument on Aura (right pane). Figure 1.5(b) shows an image of hurricane Bill as observed by the MODIS instrument with cloud heights from the CALIOP lidar on CALIPSO in 2009. Superimposed on the MODIS image is the polarized reflected sunlight observed by the POLDER instrument on PARASOL. The combination of these different types of information allow scientists to better understand the phenomena studied than when using data from a single instrument.
1.1. BACKGROUND

(a) Ash cloud of the Eyjafjallajökull volcano in Iceland. ©NASA
(b) Hurricane Bill observed by different payloads. ©NASA

Figure 1.5: Examples of the synergy between the data from the payloads on different spacecraft in a formation. Subfigure (a) shows the ash cloud of the Eyjafjallajökull volcano in Iceland as observed by the MODIS instrument on Aqua (left) and by the OMI instrument on Aura (right). Subfigure (b) shows hurricane Bill as seen by the MODIS instrument on Aqua, the POLDER instrument on PARASOL, and the CALIOP lidar on CALIPSO.

1.1.3 Autonomous Formation Flying

Using definition 1 and excluding the rather special case of collocated geostationary satellites (which will be discussed briefly in the next section), there are currently five active formations in space, all of them in Earth orbit. The GRAIL mission, which orbited the Moon, ended in December 2012 when the spacecraft purposefully impacted the Moon, and is listed here for reference:

1. Afternoon Constellation, consisting of the spacecraft Aqua, Aura, CALIPSO, CloudSat, and PARASOL,
2. GRACE, consisting of the spacecraft GRACE-A and GRACE-B ("Tom and Jerry"),
3. Morning Constellation (also known as the AM Constellation), consisting of the spacecraft Terra, Landsat-7, SAC-C, and EO-1,
4. PRISMA, consisting of the spacecraft Mango and Tango (the mission is currently, i.e., May 2013, in the extended / final phase),
5. The spacecraft TerraSAR-X and TanDEM-X,
6. GRAIL, consisting of the spacecraft Ebb and Flow.

Of the listed formations, the formations Landsat-7/EO-1, PRISMA, TerraSAR-X and TanDEM-X, and GRAIL are considered to be autonomous, implying that they can execute a certain sequence of tasks without human intervention. Thus, they do not have to wait for commands to be uploaded from ground to, e.g., respond to a certain event, but can respond to that event via a set of pre-programmed steps. For instance, if the distance between any two spacecraft in the formation becomes too large or too small, a control algorithm computes and commands a series of thruster firings to correct for this. However,
since all the actions that the spacecraft can undertake are pre-programmed, the spacecraft are not truly autonomous in the sense that they can rationalize and make decisions in case of ill-defined problems and/or lacking data. Rather, in this context autonomy must be regarded as an advanced form of automation of a subset of the functions of the formation.

One of the functions that often benefits from a high level of autonomy is the guidance, navigation and control (GNC) function, which is needed to control the formation geometry. The guidance part of this function determines the trajectory that the spacecraft need to take in order to be at a certain position (and/or velocity and/or attitude) with respect to each other at a certain moment in time, the navigation part determines where the spacecraft actually are or will be, and the control part ensures that the correct control forces are applied at the correct times to steer the spacecraft towards the target dictated by the guidance function. For the GNC of a formation, several areas of autonomy can be distinguished:

1. **Relative navigation** - The spacecraft in the formation use measurements and a model of the (relative) spacecraft dynamics to estimate their relative state (e.g., relative positions, relative velocities) on board in real time. No man-made information external to the formation, e.g. from GNSS or ground-based systems, is used to obtain the measurements. Instead, only information generated by the spacecraft in the formation themselves (through e.g. optical, infrared, and RF means) and natural navigation references such as stars, planets, and moons is used. The measurements, e.g. \( z_i \), must be such that they can be modeled as a function of the state components, e.g. \( x_j \), such that \( \frac{\partial z_i}{\partial x_j} \neq 0 \).

2. **Guidance** - Based on the relative navigation results, the spacecraft themselves determine which path they should follow to reach a certain targeted relative state and determine the timing, magnitude, and direction of the control actions needed to achieve this.

3. **Control** - The spacecraft themselves generate and execute the actions needed to actuate the formation control actuators such as to comply with the guidance needs.

The above areas, or functions, provide a convenient step-wise approach to safely initiate fully autonomous formation flying (AFF). For example, after launch and orbit injection the spacecraft can first be brought in a relatively coarse formation using ground commands and ground-based or GPS measurements. Then, one by one, the various functions in the above list can be activated and their functioning can be verified using telemetry downloads. This can be done independent of each other. For instance, the guidance function can be fed with navigation data and a target for the formation geometry, both of which have been generated on ground. It will then generate a sequence of control actions that must be executed for the formation to arrive at the desired relative geometry. This sequence is then not sent to the control actuators, but downloaded to the ground where it is verified. If the sequence is as desired, then the guidance function can be allowed to send control commands to the control actuators directly (once the correct functioning of these has been verified). If one function is not behaving as desired, there is still a possibility to
upload new software. Once the correct independent functioning of the GNC functions has been verified, they can be activated in the listed order, which allows an additional check of the correct functioning of a combination of functions, for instance the relative state measurements and the relative navigation, before the next function, in this case the guidance, is activated.

Autonomous formation flying may be required for several reasons:

1. To maintain tight, e.g., m to cm-level, guidance requirements for demanding science missions.

2. If the relative motion of the spacecraft between consecutive ground station contacts is such that they will drift out of their control box.

3. If the relative motion of the spacecraft between consecutive ground station contacts is such that there is a significant probability of spacecraft collision in case of an anomaly during that time.

4. To significantly reduce operations cost through reduced ground workload.

AFF is typically required for missions with tight control boxes or for (deep space) missions with a long communication delay. Earth observation missions such as the A-Train do usually not require a high level of autonomy since the communication delays are short and the control boxes are large. The first autonomous formation flying mission was the Japanese ETS-VII mission in 1998. The most advanced example of AFF is the Swedish PRISMA technology demonstration mission, which is currently (September 2013) in the extended mission phase (the nominal mission duration was foreseen to be about 300 days). AFF missions that will be launched in the near future are mainly technology demonstration missions and include the missions CanX-4&5, JC2Sat, and PROBA-3. The GRAIL mission has been launched in September 2011 and measures the gravity field of the Moon in high resolution, making it the first spacecraft formation not in Earth orbit [Roncoli and Fujii 2010].

1.1.4 Potentials and Challenges of Formation Flying

The advent and growth in the number of AFF missions can be attributed to increasing cost pressure and increasing scientific demands (pull), but also to the curiosity of engineers (push). Declining budgets for space missions heavily conflict with the desire of scientists to obtain the most detailed information possible or to investigate phenomena that have not been studied yet. This forces mission managers and designers to consider non-standard solutions and disruptive technologies such as formation flying. Typical examples of missions that can, from a cost perspective, only be flown using a spacecraft formation are large sparse aperture telescopes or interferometers; the cost of creating these as a monolithic structure spanning hundreds of meters would be huge or prohibitive. For such missions, formation flying is an enabling technology, although it must be noted that hybrid solutions involving tethers are investigated as well [Rinehart 2006].
Formation flying also offers enhanced science return of operational missions. An example of this is the case of an Earth observation radar mission where small radar receiver spacecraft can be brought in the vicinity of a large radar transceiver spacecraft, thereby enabling multistatic observations and single-pass interferometric measurements. Such an opportunistic scheme has been studied as an addition to several Earth observation missions, such as TerraSAR-L \cite{Zink2003}. In addition, as already demonstrated in the AM and PM Trains, formation flying enables synergy between different (or similar) payloads on different spacecraft and allows synoptic observations of the same geolocation from different angles, all of which enhances the science return. Lastly, payloads with similar functionality on different spacecraft can be cross-calibrated, which adds robustness to the science data.

For geostationary spacecraft, formation flying is already being applied to enable collocation of multiple spacecraft. In this context, collocation implies that two or more geostationary spacecraft have the same argument of longitude. Collocation can be needed to vary the capacity of the orbital slot, which is a scarce commodity in geostationary orbit, according to the current needs and to increase redundancy for the service to a certain geolocation. The probability of collision or mutual interference between the spacecraft is commonly minimized by means of a formation flying strategy called eccentricity/inclination vector separation \cite{Eckstein1989}. Collocation was pioneered by the ASTRA satellites with ASTRA 1A and ASTRA 1B being the first collocated spacecraft in geostationary orbit \cite{SES2013}. Current day, collocation is regularly applied for geostationary satellites.

Formation flying also requires a fundamental change in the way a space mission needs to be designed, managed, and operated. As the space segment now consists out of multiple assets, mission success no longer depends on the functioning of a single spacecraft, but on the functioning of multiple spacecraft. This way, the mission can be designed such that if one spacecraft fails, either mission performance is reduced slightly or a spare spacecraft takes over the task of the failed spacecraft. This, in principle, allows spacecraft to be designed with less functional redundancy and thus less complexity since functional redundancy is guaranteed at the system level. Distributing all payloads needed for a mission over several spacecraft also contributes to a reduction in complexity of the individual spacecraft since the individual spacecraft can be tailored to a single payload whereas the design of a satellite carrying many payloads (e.g., Envisat) is extremely complex due to inevitable conflicting demands of the different payloads. Such a formation can then also act as a kernel to which additional spacecraft with new or improved payloads can be added in the future. In addition, since the spacecraft complexity is intended to be relatively low and the number of payloads per spacecraft small, the spacecraft itself could be small as well, a feature that offers short development time and the possibility to make use of the latest developments in technology.

Yet, next to all these potential benefits, formation flying also introduces new challenges in the way the mission needs to be designed, build, tested, launched, and operated. In the design phase, the system must be designed such that efficient, robust, and safe functioning of the formation as a whole is guaranteed. This leads to the need for additional
specific mission phases such as formation deployment, (re-)establishment, and maintenance. Depending on the inter-satellite distances, it can also require a collision avoidance strategy. Furthermore, if the spacecraft are clamped together when they are inserted into orbit, as in the PRISMA and JC2Sat missions, there is a physical interface needed between the spacecraft. In addition, the individual spacecraft design needs to take into account early operations in the clamped configuration.

Additional testing and verification as compared to a mission with a single spacecraft is also needed. For instance, functional and acceptance tests need to be performed for all spacecraft and qualification tests have to be performed for all non-identical hardware (e.g. in case of dissimilar spacecraft). It is also likely that multiple launches are needed to deploy the complete formation, which adds risk. If multiple launches are needed, there are constraints on the launch opportunities since the spacecraft need to be inserted into roughly the same orbit to avoid the need for significant propulsion system mass on the spacecraft for formation establishment.

For the operations phase, to reduce operator workload and therefore costs, as much tasks as possible must be delegated to and handled autonomously by the formation itself. In the extreme, this requires an inter-satellite ranging and communication system on each spacecraft and can result in a complex networking scheme in case of many (approximately five or more) spacecraft and tight formation control requirements. Even so, in the early operations phase of a mission, during which system performance needs to be calibrated and verified, operator workload is likely to even be higher than in a single spacecraft mission since there are more system elements and functions to be handled.

Further implications of formation flying on the mission and spacecraft design are the need for a propulsion system on at least one of the spacecraft and, in the extreme, a complete GNC system on all spacecraft. This in turn affects all other spacecraft subsystems since it requires power (which is also partially dissipated, impacting spacecraft temperature), data handling, adds mass (and thus inertia which influences the attitude control system), volume, and so on. Formation flying also impacts the layout of the spacecraft since the location of sensors and payloads can be constrained by formation flying requirements. For instance, antennas must be in the field of view of antennas on other spacecraft and thrusters must sometimes be positioned such that the ejected particles do not impinge on sensitive optical surfaces of other spacecraft in the formation. Propellant usage must also be balanced between the spacecraft in the formation to avoid several spacecraft spending much more propellant than other spacecraft, leading to a situation where several spacecraft have spent all their propellant while others still have plenty left. This can cause formation maintenance to become difficult or even impossible and in the worst case, depending on the mission design, a premature end of the mission. Lastly, at end-of-life a formation creates much more space debris than a single spacecraft unless the spacecraft are de-orbited.

The preceding discussion makes clear that there are several advantages and challenges to formation flying and that an advantage in one area can lead to a disadvantage in a different area. For that reason, there will be situations where it is not clear at the start of the
design phase whether a mission should be executed using a formation or with a single spacecraft. In fact, there are examples of missions which have initially been designed as a spacecraft formation, but have later been ‘downscaled’ to a single spacecraft mission. This is usually due to a re-iteration of the mission design, where an improved understanding of the mission characteristics and/or changing requirements lead to a different choice for the mission type. Launcher availability can also be an important factor, e.g. for space telescopes. For space telescopes namely, the amount of light collecting area is a key performance characteristic. Thus, if a launcher is available in which a monolithic design of the observatory fits, preference is usually given to a single launch with the large launcher over multiple launches with a smaller launcher [Coulter 2009]. Lastly, concerns on the maturity of the formation flying technology and a risk-reducing strategy can lead to a more traditional monolithic design with possibly reduced performance.

A good example of such a process is the evolution in the, still ongoing, design of the next-generation X-ray observatory, to be launched in the next decade. At NASA, the mission concept was named Constellation-X and first consisted out of four formation flying spacecraft, which were later merged into a single spacecraft with four telescopes to reduce launch costs [Hornschemeier 2006]. At the same time, ESA and JAXA were jointly studying a mission concept called XEUS, which consisted out of a mirror spacecraft and a detector spacecraft flying in formation [Turner and Hasinger 2007]. In 2008, the Constellation-X and XEUS missions merged into IXO, which was to be a monolithic spacecraft with a deployable telescope structure [Barcons et al. 2011]. Finally, in March 2011, the IXO concept was abandoned and superseded by the ATHENA mission, whose baseline concept currently considers two telescopes in a single spacecraft with no deployable telescope structure [Lumb 2011].

Thus, spacecraft formations offer a number of advantages over monolithic spacecraft missions and even enable new mission types, but they also introduce new challenges to missions. At the present time, these complexities introduce extra development risk since no standard procedures and technologies exist yet to handle these. For that reason, continued research into and in-orbit demonstration of formation flying technologies is needed to actually execute the new types of science missions made possible through formation flying.

1.1.5 Relative Navigation Needs

The difficulty in controlling the relative position between the spacecraft in a formation depends on the required relative position accuracy, the inter-satellite distance, the number of spacecraft, the level of external disturbances, and the available resources (power, mass, volume). For example, maintaining an inter-satellite distance accuracy of 10 m at a separation of 100 m is relatively easy, but doing this at a separation distance of 100 km is much more challenging as the accuracy of the relative navigation typically tends to decrease with increasing inter-satellite distance. Also, achieving 10 m relative positioning accuracy is much easier for a formation consisting of two spacecraft than for a formation consisting of 100 spacecraft since the latter scenario requires a much more complex
command and control dissemination than the former scenario. Very accurate formation keeping in low Earth orbit (LEO) is also much harder than, e.g., in a halo orbit around a Sun-Earth Lagrange point due to the magnitude of external disturbance forces in LEO (such as atmospheric drag and Earth’s complex gravity field).

To minimize propellant usage for formation maintenance, internal and external disturbances to the formation need to be minimized as much as possible. For this reason, LEO orbits are not preferred for these missions. Highly elliptic Earth-centered orbits and halo orbits around the Sun-Earth Lagrange points L1 and L2 however are very attractive environments since gravitational disturbances there are very small and since the formation is always in view of Earth, allowing for continuous contact between ground stations and the formation. In addition, in the future there might be formations orbiting other planets or moons. Formation flying in these locations is complicated through the reduced availability (highly elliptical orbit) or even complete lack (non-Earth orbit) of GNSS signals, which is a very convenient tool for relative navigation in LEO. Therefore, in these environments an autonomous sensor suite is needed to provide omnidirectional navigation capabilities for formation acquisition and maintenance. It is also required for formation safety and highly accurate directional navigation for precise formation maintenance.

Figure 1.6 provides insight into the typical inter-satellite distances and formation control accuracies needed for currently operational missions (A-Train, GRACE, PRISMA, TanDEM-X), missions that are to be launched in the near future (CanX-4&5, JC2Sat-FF, MMS, PROBA-3), and mission concepts which, if not canceled, are to be launched after 2020 (Darwin, Generation-X, GRI, HIRISE, MAX, MAXIM, New Worlds Observer, Stellar Imager, TPF-I). The figure also shows several well-known mission concepts that have already been canceled or for which the design has been frozen (Cross-Scale, FAST, Gemini, PEGASE, Romulus, SIMBOL-X, XEUS). Figure 1.6 clearly reveals a trend in the development of formation flying missions over time, indicated by the curved arrow: The first formation flying missions exhibit large inter-satellite distances of tens to hundreds of km and very relaxed control requirements up to km-level, placing them at the upper right corner of the figure and only allowing for very safe and ‘easy’ train-like missions. As the technology matures, inter-satellite distances can be decreased to roughly 100 m and control requirements tightened to 0.1-1 cm, allowing interferometric and close proximity missions. Even more advanced missions where extremely high angular resolution is needed, to image e.g. exoplanets (NWO) or black holes (MAXIM), require extremely large inter-satellite distances of more than 10,000 km coupled to moderate control accuracies of 0.1-1 m and/or moderate inter-satellite distances of 0.1-1 km coupled to extremely high control accuracies of 10-100 μm.
Figure 1.6: Formation control accuracy versus relative distance for operational and future formation flying missions. The diagonal dashed lines indicate the dynamic range (distance over control ratio). Note that some of the missions depicted in the figure have been canceled or 'shelved'.

The current state-of-the-art in formation flying control performance for LEO orbits is cm-level relative control accuracy at tens of meters inter-satellite separation and has been achieved in the PRISMA mission [Guidotti 2011]. This translates to a distance over control ratio, or dynamic range, of $10^3 – 10^4$, which can be regarded as an indicative performance metric for formation control: The larger this value is, the more challenging the formation control is likely to be. The equation for the dynamic range, which is a unitless metric, is as follows

$$\text{dynamic range} = \frac{\text{relative distance}}{\text{control accuracy}}$$

with the units for the relative distance and the control accuracy being the same. PRISMA’s performance with respect to this metric is several orders of magnitude better than e.g. the A-Train, which achieves a dynamic range of $10^1$ at best (between CloudSat and CALIPSO),
but is still several orders of magnitude away from the performance required for e.g. MAXIM or Stellar Imager, which is around $10^8$, cf. Table 1.1.

Table 1.1: Decrease in control accuracy over distance ratio, or dynamic range, for formation flying missions in time, indicating the need for increasingly advanced formation control systems.

<table>
<thead>
<tr>
<th>Mission</th>
<th>Year (to be) launched</th>
<th>Dynamic range [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRACE</td>
<td>2002</td>
<td>$10^1$</td>
</tr>
<tr>
<td>PRISMA</td>
<td>2010</td>
<td>$10^3 - 10^4$</td>
</tr>
<tr>
<td>MAXIM</td>
<td>2020+</td>
<td>$10^8$</td>
</tr>
</tbody>
</table>

With the foregoing in mind, the relative navigation accuracy needed to enable future formation flying missions may be attempted to be predicted. In the field of formation flying, there are several important characteristic length parameters that drive the design of the formation flying GNC system. These are the inter-satellite distance, the control window, the control accuracy, and the navigation accuracy. These parameters are all interconnected in a manner here referred to as Gill’s scaling law [Gill1 et al. 2010]. This law states that they all differ with approximately a factor of 10 in case no other requirements or constraints on the spacecraft formation exist. The scaling law is exemplified in the following paragraph.

Assume two spacecraft in formation flight that have to be kept, in one axis, at a mean inter-satellite distance $d_{\text{mean}}$. The actual inter-satellite distance is allowed to vary between a maximum and minimum bound that together define the control window. The size of the control window is $d_{\text{cw}}$. Thus, assuming a symmetric control window, the maximum allowed inter-satellite distance $d$ is equal to $d_{\text{mean}} + 1/2 d_{\text{cw}}$. Often, the control window is significantly smaller than the maximum inter-satellite distance and a factor of 10 difference is usually justified for the definition of the control window size, resulting in $d_{\text{cw}} \approx 0.1d$. Thus, when the distance between the two spacecraft is not allowed to be larger than, say, 1,000 m, this implies $d = 1,000$ m and $d_{\text{cw}} = 100$ m. Staying within a certain control window of size $d_{\text{cw}}$ requires controlling the relative motion typically a factor of about 10 better than the size of the window. In the current example, this implies a control accuracy $\sigma_c$ that is 10 times smaller than the control window, or $\sigma_c \approx 0.1 d_{\text{cw}} = 10$ m. The required control accuracy is the summed square of various contributions such as navigation accuracy, thruster misalignments, thruster performance uncertainties, and attitude pointing accuracy. Therefore, estimating that the required navigation accuracy $\sigma_n$ needs to be about 10 times smaller than the required control accuracy is reasonable. This leads to $\sigma_n \approx 0.1 \sigma_c = 1$ m for this specific example, cf. Fig. 1.7.

However, since the navigation accuracy depends on factors like the attitude estimation accuracy, the accuracy of the dynamic model of the relative spacecraft motion, and the accuracy of the inter-satellite distance measurement $\sigma_d$, the foregoing line of reasoning can be taken even one step further than in [Gill1 et al. 2010] by assuming that $\sigma_d \approx 0.1 \sigma_n = 0.1$ m. Thus, following Gill’s scaling law, the required accuracy for the mea-
Maximum spacecraft separation $d$

Control window: $d_{cw} \approx 0.1d$

Control accuracy: $\sigma_c \approx 0.1d_{cw}$

Navigation accuracy: $\sigma_n \approx 0.1\sigma_c$

Figure 1.7: Typical scaling of characteristic length parameters that are of importance in formation flying missions according to Gill’s scaling law (distances not to scale).

surement of the inter-satellite distance in a formation flying mission is typically four orders of magnitude smaller than the inter-satellite distance itself. Yet, Gill’s scaling law assumes a rather relaxed size of the control window. For challenging missions, the control window size is much more restricted, which, due to the other scaling factors remaining as they were, results in much smaller values for the measurement accuracy.

Having estimated that the distance measurement accuracy required to enable future formation flying missions needs to be approximately two orders of magnitude better than the required control accuracy, Fig. 1.6 indicates that the distance measurement accuracy for future formation flying missions with an inter-satellite distance up to 1 km must be equal to or better than 0.1 mm. For missions like MAXIM, which require a very large dynamic range, a very small inter-satellite distance becomes extremely challenging to achieve since this would require measurement accuracies on the order of micrometers or less, which places extremely demanding requirements on all spacecraft subsystems. Therefore, a logical approach to enable such missions, from a formation flying perspective, is to increase the inter-satellite distance, leading to more manageable measurement accuracies. Of course, an analysis has to be performed to verify that the scientific objectives of the mission can be met with a relatively large spacecraft separation.

Naturally, it is desirable to use a single type of metrology system for the relative navigation. However, if the required dynamic range is very large this is not possible with current technology. In that case a metrology chain must be created where a relatively coarse metrology system provides coarse inter-satellite distance(s) and/or angles and, if needed, full sky coverage to enable relative navigation in all directions. This can be required in case the formation consists out of more than two spacecraft, during formation (re-)initialization, and for formation safety. Since this system has to provide the information needed for formation safety, it needs to be robust. In addition, if full sky coverage is
needed, multiple sensors (antennas) typically need to be distributed over the spacecraft. Using the information provided by the coarse metrology system, a relatively coarse formation can be established. When this has been achieved, the next element in the metrology chain is activated, providing the additional information needed to achieve more accurate formation control. This is commonly an optical (laser) system. If this system needs to determine the distance to multiple other spacecraft, then also here multiple sensors are typically needed since the FOV of these sensors is typically relatively small. In a third step, even higher accuracy distance and/or angular metrology can be achieved by means of laser interferometry [Verlaan and Cuylle 2008, Verlaan et al. 2008]. Thus, the total number of sensors needed for the complete metrology system can easily become quite large. An extreme example is presented in [Tien et al. 2004] where a coarse relative navigation system (RNS) design is presented which requires a total of 16 RF antennas on each spacecraft.

1.2 State of the Art

In the previous section, the concept of satellite formation flying has been introduced and discussed with emphasis on the relative navigation aspects. This section will treat the current state of the art in terms of research and results obtained in practice for relative navigation of formation flying satellites. It will also briefly treat the areas where knowledge can and should be improved and which are thus the topic of this research.

1.2.1 Basic Research

In this subsection, the state of the art in the basic research of autonomous relative navigation is introduced. In order to do this, the observability of a system, i.e., to what extent the (unknown) system state\(^1\) can be observed, and thus estimated, is used as a ‘red line’ since this aspect will play an important role in the remainder of this work. In the extreme, if the observability of a system is very bad, the state of the system cannot be estimated, meaning that there is not enough information available to obtain an estimate for all variables in the state vector.

Early work in the field of autonomous navigation by means of inter-satellite measurements has mainly focused on absolute orbit determination. Markley [Markley 1987] was the first to thoroughly analyze the observability of the problem of absolute orbit determination by means of inter-satellite vector measurements without knowing the absolute orbit of one of the satellites. Several other authors, including Herklotz [Herklotz 1987], Psiaki [Psiaki 1999b], Liu and Liu [Liu and Liu 2001], and Yim et al. [Yim et al. 2004] also analyzed the problem of autonomous absolute orbit determination using inter-satellite vector measurements.

\(^1\)In the field of estimation, the ‘state’ of a system is a vector of variables. The combination of these variables, e.g., position \((x, y, z)\), velocity \((v_x, v_y, v_z)\), and acceleration \((a_x, a_y, a_z)\) describes the properties of the system (the system state) at a certain moment in time.
In the 1980s and 1990s, autonomous relative orbit determination however, did not receive much attention in the literature. In the appendix to his paper, Markley determined, as a ‘byproduct’ of the problem he considered, that the relative state of the two satellites is fully observable if the inertial attitude of at least one of the satellites is known and if the relative position vector can be measured. In the last decade, several authors have investigated this topic further. Woffinden and Geller [Woffinden and Geller 2007, Woffinden and Geller 2009a, Woffinden and Geller 2009b] have studied the problem of relative navigation in case of angular measurements only. They proved that with the aid of a calibrated thrust maneuver, observability can be guaranteed for all relative trajectories. Chen and Xu [Chen and Xu 2010] also studied angles-only relative navigation for autonomous rendezvous. They showed that the relative state of a chaser and target spacecraft becomes more observable if an auxiliary spacecraft is used to provide additional angular measurements. Doolittle et al. [Doolittle et al. 2005] found that in case of range and angle measurements, the observability of a two-satellite formation deteriorates when the relative out-of-plane motion is small. Chavez and Lovell [Chavez and Lovell 2004] studied the influence of varying sensor on-off duty cycles on the estimation of the relative state in case of inter-satellite range measurements and coplanar elliptical formations. Holt and Lightsey [Holt and Lightsey 2005] considered relative navigation in a seven-satellite formation in deep space with range measurements between all satellites in the formation. They use the concept of Geometric Dilution Of Precision (GDOP) to explain that particular navigation performances are due to the geometrical distribution of the satellites in the formation. Kang et al. [Kang et al. 2009] studied the observability of a networked satellite constellation, where the satellites perform inter-satellite range measurements and exchange navigation data, and showed that such a system can be completely observable even with incomplete communication and sensor topologies, provided that the two topologies work in harmony with each other. Matko et al. [Matko et al. 2010] showed that when the relative state is expressed in rectilinear coordinates and if the observations are a linear function of the individual states, only the along-track distance and the out-of-plane position or velocity needs to be measured to have full observability. Perea [Perea 2010] studied estimator divergence in the context of formation flying missions and proposed a modified filtering algorithm that is more robust to nonlinearities and which provides good estimates.

Several authors furthermore studied the effect of inter-satellite ranging to augment ‘external’ measurements like those from GPS and ground stations for either absolute or relative orbit determination. Lightsey and Um [Lightsey and Um 2001] studied autonomous rendezvous and docking of a spacecraft with the International Space Station (ISS) by means of, amongst others, pseudolites (GPS-like signal transmitters) mounted on the ISS to augment GPS measurements. Holt and Lightsey [Holt and Lightsey 2008] compared GPS-only relative navigation, transponder-only relative navigation, and GPS plus transponder relative navigation for a tetrahedral formation in a highly elliptical Earth orbit (HEO). In several publications, Huxel and Bishop studied the observability of a multiple spacecraft system in the vicinity of the Moon [Huxel and Bishop 2004, Huxel 2006, Huxel and Bishop 2009]. It was found that an increase in observability is achieved when intersatellite range measurements are used next to measurements from tracking stations. Fur-
thermore, a geometrical explanation was provided for the variability in the observability of the system over time.

Lastly, although not related to a multi-satellite system but containing a detailed observability analysis, Yim et al. [Yim et al. 2000] and Chang et al. [Chang et al. 2009] studied the autonomous absolute orbit determination of interplanetary spacecraft. Yim et al. showed that autonomous absolute orbit navigation is possible by means of radial velocity and line of sight measurements with respect to the Sun. They also showed that by adding one more line of sight measurement from an Earth sensor, the estimation accuracy can be considerably improved. Chang et al. proved through an analytical and numerical observability analysis that Sun line-of-sight vector measurements are sufficient for autonomous navigation of interplanetary spacecraft. Both [Yim et al. 2000] and [Chang et al. 2009] found that a larger semi-major axis in the heliocentric system leads to a poorer observability due to poorer angular information.

### 1.2.2 Achievable Relative Navigation Accuracy

The relative orbit between multiple satellites can be determined in several ways. The most preferable method depends on the mission characteristics and its needs. Figure 1.8 depicts a breakdown of the various options to perform relative navigation. First, a choice needs to be made whether or not to perform the navigation using absolute orbit navigation solutions, using relative orbit navigation solutions, or a combination of both (hybrid). The first method requires the estimation of the absolute orbits of the spacecraft. These are subsequently subtracted from each other, yielding a relative navigation solution. The second method uses inter-satellite measurements only. These are used to estimate the relative orbit directly. A hybrid method, using elements of both previous methods, can also be applied. In fact, it is often necessary to use absolute orbit data as initial estimate for the direct relative navigation method to ensure that the estimator used to compute the relative orbit converges to the correct relative orbit. This is due to the ambiguous nature of the relative navigation problem, i.e., different relative orbits can yield the same measurement data. In such cases, lack of a sufficiently accurate initial estimate for the relative orbit implies that the resulting estimate for the relative orbit cannot be trusted since the estimator can have converged to a relative orbit which is very different from the actual relative orbit.

As shown in Fig. 1.8, absolute orbit determination can be performed using ground station measurements, Two-Line Elements (TLEs) as provided by the North American Aerospace Defense Command (NORAD), the Tracking and Data Relay Satellite System (TDRSS), and GNSS.

Ground station measurements can be performed in various manners. If this is done using RF signals, usually a two-way range or range-rate measurement is used. The classical method employs a transponder on a satellite that receives and transmits a special RF ranging signal from and to a ground station. Such a system can reach an accuracy of 15 m for the range determination of a single satellite [Montenbruck and Gill 2000]. Laser ranging or very specialized equipment such as the Precise Range and Range Rate Equip-
ment (PRARE) system can achieve accuracies at cm-level for the range determination to a single satellite. When using such systems, the relative position of two closely spaced satellites could potentially, again due to cancellation of common errors, be determined with meter-level accuracy.

Probably the least accurate method, but also the cheapest one, is to differentiate the TLEs of the individual satellites. TLEs are a data format where two lines of text are used to list a set of satellite (orbital) parameters that are determined and distributed freely by NORAD. The upside of using TLEs is that no dedicated ground- or space-based functionality needs to be implemented for the mission. The downsides are the irregularity of TLE updates (typically several days) and the poor accuracy. According to [Gill et al. 2010, Kahr et al. 2013], the typical position accuracy that can be achieved for a single satellite using TLE data is on the order of 1 kilometer for all three axes. Due to cancellation of common error terms (atmospheric effects, systematic errors, etc.), the relative position determination of closely spaced satellites using TLEs can have an accuracy of roughly 30 m in the radial and out-of-plane directions and subkilometer accuracy in along-track direction [Kirschner et al. 2001].

**Figure 1.8:** Relative navigation options.
TDRSS is a constellation of six geosynchronous satellites and a ground system which provides tracking and communications support for LEO space vehicles. A TDRS allows relayed two-way range and range rate tracking of satellites and, for satellites equipped with an ultra-stable frequency reference, also precise relayed one-way range rate measurements. In case of two-way measurements, the signals are transmitted from the White Sands ground terminal to the TDRS, where they are coherently forwarded to the spacecraft. The signals are transponded by the user satellite and transmitted back to the TDRS, where they are relayed to the White Sands ground terminal. Two-way range measurements can have an accuracy of 2 m [Montenbruck and Gill 2000].

However, the downside of using these methods to determine the relative positions and velocities of formation flying satellites is that the time between measurements can be rather long, depending, e.g., on the number and the locations of the ground stations. Thus, there can be significant periods of time during which the relative positions of the satellites are unknown or, if dynamic models for orbit prediction are used, known only with inferior accuracy (the accuracy of the predicted orbit typically decreases with increasing elapsed time since the last measurement). In addition, the density of the measurements over one orbit can be poor, which leads to a poor relative state estimate even when the ranging itself is very accurate.

To overcome this drawback, many ground stations would need to be employed, which is a costly solution. For relatively loose formations, where inter-satellite distances are large (many kilometers) and where the control windows are also large, a moderately accurate relative state estimation is not necessarily a problem since there is sufficient time to plan and upload the necessary orbital corrections. Also there is little chance that the satellites will collide, and there is no need for very high navigation accuracy. However, for tight formations or for formations not orbiting the Earth, all the required information needs to be obtained and processed by the satellites themselves since ground contacts can be too infrequent to meet the relative positioning requirements and safety considerations.

In Earth orbit, satellites can be equipped with GNSS receivers, enabling them to determine their absolute orbital positions without the need for ground-based measurements. Equipping the satellites with an inter-satellite communication system allows them to directly exchange their absolute position data, which can be used to compute their relative orbit. However, it is possible to achieve higher accuracy using a method called Differential GPS (DGPS), as will be discussed further in this subsection.

Instead of differencing absolute orbit navigation results, the relative orbit of two satellites can also be obtained from direct measurements between the satellites themselves. These measurements can be performed using RF signals, optical sensors, or by means of DGPS. Table 1.2 lists the typical accuracies that have been obtained in recent years for such systems. Note that the K-band ranging system only measures the change in inter-satellite range, not the absolute range itself and that AVGS (Advanced Video Guidance Sensor) is a camera-based system. All other systems are based on RF ranging.

As already mentioned earlier, when the formation flying satellites are in LEO and equipped with GPS receivers and an inter-satellite communication system, DGPS can be
Table 1.2: Typical accuracies that can be obtained with various inter-satellite ranging technologies from [Dunn et al. 2002], [Delpech et al. 2011], and [D’Amico et al. 2013b, Krieger et al. 2013, Accardo et al. 2013] in [D’Errico 2013]. Note that the K-band ranging system uses signals at 24 GHz and 32 GHz, which are considered in this thesis to fall in the Ka-band.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Accuracy [m]</th>
<th>Sample mission</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-band ranging system</td>
<td>$10 \cdot 10^{-6}$</td>
<td>GRACE</td>
<td>2002</td>
</tr>
<tr>
<td>DGPS (real time)</td>
<td>$5 \cdot 10^{-2}$ (3D r.m.s.)</td>
<td>PRISMA</td>
<td>2010</td>
</tr>
<tr>
<td>DGPS (post-processed)</td>
<td>$1 \cdot 10^{-3}$ ($\sigma$)</td>
<td>TanDEM-X</td>
<td>2010</td>
</tr>
<tr>
<td>FFRF</td>
<td>$9 \cdot 10^{-2}$ (range 20 m, 3D r.m.s.)</td>
<td>PRISMA</td>
<td>2010</td>
</tr>
<tr>
<td></td>
<td>3.5 (range 0.2 – 4 km, 3D r.m.s.)</td>
<td>PRISMA</td>
<td>2010</td>
</tr>
<tr>
<td>AVGS</td>
<td>$7 \cdot 10^{-2}$ ($\sigma$)</td>
<td>Orbital Express</td>
<td>2007</td>
</tr>
</tbody>
</table>

used to obtain a highly accurate estimate of the relative orbit of the satellites. This is achieved through the exchange of raw pseudorange measurements to GPS satellites that are in view of both satellites. Differencing of the pseudoranges measured to the same GPS satellite leads to a new ‘delta-pseudorange’ in which several error terms that were present in the original pseudoranges have been canceled. The ‘delta-pseudoranges’, of which a minimum of seven are needed (three for the relative position, three for the relative velocity, and one for the semi-major axis of the absolute orbit), can be used to compute an accurate estimate of the relative orbit of the formation flying satellites, cf. Table 1.2.

In case DGPS cannot be used (e.g., for missions not in Earth orbit) or does not satisfy the requirements for the mission (e.g., not sufficiently accurate, too constraining for the allowable attitude of the satellite, insufficient availability for elliptical orbits) direct range (rate) measurements can be employed between the formation flying satellites using locally generated RF ranging signals. To achieve a certain level of autonomy for the formation, it is advantageous if the system used for range measurements can also be used to exchange information. Examples of such systems are the FFRF, AFE, CCNT, (N)CLT, LPT, Star Ranger, and IRAS. Detailed information on these systems can be found in [Lestarquit et al. 2006], [Tien et al. 2004], [Bar-Sever et al. 2001], [Stadter et al. 2001], [Winter et al. 2003], [Zenick and Kohlhepp 2000], and [Heckler et al. 2008]. These systems all have in common that they are derived from GPS technology and use GPS-like signals that can be used to measure a range but also to exchange data. Depending on the inter-satellite distance, these systems can provide centimeter to meter-level ranging accuracy and allow for autonomous formation flying even in deep space. Thus, these systems enable advanced formation flying missions such as Darwin and TPF, which need to be deployed at a Lagrange point to minimize disturbances to the formation. Up to now, of the aforementioned systems only the FFRF has actually been used in space in the PRISMA mission.

Other methods that can be employed to measure the inter-satellite range using RF signals are radar systems (either pulsed or continuous wave) and transponder systems. These systems have in common that they use signals that are sent from one satellite to the other satellite and then return to the satellite from which they originated. This results in a so-called ‘two-way’ range measurement instead of the ‘one-way’ range measurement em-
ployed by GPS-like system. Radar systems operate by measuring the time delay (pulsed signal) or phase shift (continuous wave signal) of a signal that is reflected from the target. However, since the strength of a received radar signal is inversely proportional to the fourth power of the range between the transmitter and the target, radar systems need relatively high power signals to perform a measurement. Systems using transponders also use two-way signals for the range measurement, but now the signal sent from the first satellite is acquired by the second satellite and coherently sent back with increased signal power (and different frequency) to the first satellite. This overcomes the drawback of the radar system. Yet, a pure radar system only allows for one-way exchange of information (if the targeted spacecraft can acquire the data that is modulated onto the radar signal). This means that each satellite would have to carry a radar and a receiver in order to allow for two-way communication, which is inefficient. A transponder system does allow for two-way communication since a transponder can store the received signal and can also modulate data onto the signal that is sent back. Yet, such systems have not yet found their way into application on formation flying missions.

Optical systems can also be used to measure inter-satellite ranges. Examples of such systems are cameras and lasers. A scanning laser range finder can be employed in the same manner as a radar and is also limited by similar constraints, cf. [Fehse 2003]. Cameras can also be used to measure ranges by means of pattern evaluation algorithms, cf. [D’Amico et al. 2013a]. A target pattern can be constructed on one of the satellites and by knowing the dimensions of this pattern, a camera image of this pattern can be used to reconstruct the range between the camera and the pattern. For camera-based systems, the light to illuminate the pattern can originate from various sources such as reflected sunlight (which leads to difficulties in adverse illumination conditions), a ring illuminator around the camera, or by using Light Emitting Diodes (LEDs) in the target pattern. An example of a camera-based ranging system is the AVGS, which flew on the Orbital Express mission and achieved a ranging accuracy of 7 cm (1σ), cf. Table 1.2. However, optical ranging systems cannot (yet) be used to exchange information and are therefore less preferable than a GPS-like ranging system for autonomous formations.

Note that in Fig. 1.8, range rate measurements are not explicitly shown. However, they can be obtained via either differentiation of range measurements over time or via Doppler measurements of the RF signal (integrated Doppler count). In addition, line-of-sight (LOS) measurements are also not treated in the figure. LOS measurements can be obtained using a combination of a narrow RF beam and a scanning antenna (two-way RF only), by measuring the time delay or phase shift of a signal received by two antennas separated by a fixed distance, or by measuring the received signal power and combining this with known range and antenna pattern [Fehse 2003].

From the preceding discussion, it is to be expected that from all possible relative navigation options in Fig. 1.8, systems using GNSS-like ranging signals are very likely candidates for future autonomous formation flying missions. Therefore, the work presented in this thesis assumes that such a system is used for the relative navigation measurements.
1.2.3 Opportunities for Knowledge Improvement

Although results obtained thus far are impressive and important, the understanding of the results obtained in some of the theoretical works needs to be improved by performing a detailed study of the observability of the system as function of the relative state of the system. For instance, by systematically varying the type of formation and the formation’s orientation and size, insight can be gained in the manner with which the observability of the system changes as a function of these parameters, which can aid in the design of future formation flying missions. Previous works often considered only one particular formation geometry and could thus not provide this insight. In addition, a detailed study of the system’s observability as a function of sensor suite properties such as sensor accuracy and sensor placement has not yet been performed. Such a study has the potential to facilitate the design of a RNS as design trade-offs between for instance sensor accuracy and sensor placement can be made without extensive simulations. The next section will address the specific research questions that have been formulated to improve the knowledge in these areas. They form the core of this study.

1.3 Research Questions

As discussed in section 1.1, autonomous relative navigation is a key enabling technology for future formation flying missions. However, to keep mission costs limited as much as possible the use of large (> 500 kg) spacecraft has to be avoided if possible. The reasons are that the spacecraft themselves will be relatively expensive and that the cost of launching multiple large spacecraft is very high. Therefore, it is to be expected that future formation flying missions will typically consist out of micro-satellites (10-100 kg) or even nanosatellites (1-10 kg). This is supported by the fact that the capabilities of these relatively small spacecraft have increased significantly during the last decade due to the growing availability of subsystems and payloads that can be supported by these small spacecraft.

As discussed in section 1.2, precise and affordable autonomous formation flying requires that the spacecraft in the formation do not depend on ground-based range measurements. In addition, to widen the general applicability of this research, it is assumed that no use can be made of GNSS systems. Then, a dedicated inter-satellite ranging and communication system has to be used to obtain the required information. For RF-based relative navigation between two spacecraft, it is common to use multiple ranging antennas on at least one of the spacecraft as this provides both range and angular information: Using one antenna will yield only a range \(^2\), using two antennas will yield a range and one angle, and using three (non-aligned) antennas will yield a range and two angles (azimuth and elevation). It is generally known that to improve the accuracy of the relative state (position and velocity) determination of two spacecraft using RF inter-satellite ranging,  

\(^2\)Usage of only one antenna per spacecraft can also yield angular information, but requires knowledge of the distance between the antennas (which can be directly measured), the relative orientation of the antennas, the antenna beam patterns of both antennas, and the strength of the signal upon transmission and reception. With this information, which requires a lot of effort to obtain, a rough estimate can be made of the direction of the signal.
several options are available: Increasing the accuracy of the range determination, increasing the distance between the antennas on the receiver spacecraft, and/or decreasing the distance between the spacecraft themselves. However, mission requirements might prohibit the use of one or more of these options. In addition, for very small spacecraft some of these options can be difficult to achieve. Therefore, foreseeing an increasing use of small satellites for formation flying missions, it is important to know how these three variables influence the relative position determination and what the transition point is between using one or multiple antennas on the spacecraft. The latter consideration is driven by the fact that it is not unlikely that on a small satellite the baseline between multiple antennas can be very small as compared to the range measurement error. Then, the benefit of this baseline is very small as well and a single antenna might provide a comparable performance since, in the limit, as the antenna baseline goes to zero, multiple antennas will lead to measurements as if they were obtained by a single antenna.

**Research Question 1**  How do the receiver antenna baseline, ranging accuracy, and inter-satellite distance affect the accuracy in the relative state determination for two formation flying spacecraft? Does there exist a transition point between the use of a single antenna and the use of multiple antennas at which the navigation results obtained by using one antenna equal or even outperform the navigation results obtained by using multiple antennas?

In Earth orbit, there are several types of relative trajectories for which the relative motion repeats after one orbit (if gravitational and other disturbances are not considered). These are based on energy-matching considerations for the orbits of the spacecraft. Two well-known examples are the pendulum (or helix) and the relative ellipse (or football orbit). For the pendulum orbit, the relative motion between the spacecraft is a simple harmonic oscillation perpendicular to the orbital plane of the reference satellite. For the ellipse however, the relative motion occurs in an ellipse around a reference point and can involve all three spatial dimensions. This difference in the relative motion has an influence on the information obtained by the RNS and thus on the accuracy with which the relative state of the spacecraft can be estimated.

**Research Question 2**  How are the relative navigation results for a two spacecraft formation influenced by the type of relative motion?

For spacecraft formations in Earth orbit employing a dedicated inter-satellite navigation system, it is known that a small relative out-of-plane motion leads to poor relative navigation results. However, little to no effort has yet been put into a quantification of this effect. It is for instance not known how the relative navigation results vary with changing magnitude for the relative out-of-plane motion, whether this effect differs for different types of relative motion, and at what point the relative out-of-plane motion becomes so small that it precludes a reliable estimation of the relative state.

**Research Question 3**  How does the magnitude of the relative out-of-plane motion between two spacecraft influence the relative navigation results?
The three research questions put forth in this section are the subject of investigation for the research presented in this thesis. However, as will become apparent, these questions cannot be answered separately from each other as they are intimately linked. For example, as the type of relative motion influences the information obtained on the relative state of the spacecraft, this will also affect the location of the transition point between the use of one or multiple antennas for the RNS. Therefore, the research questions are treated and answered conjunctly in the remainder of the thesis.

It is emphasized that, although the research focuses on formations in Earth orbit, the results are also (partially) applicable to formations orbiting other planets or even in halo orbits around a Lagrange point as the (linearized) dynamics of the motion in these environments are very similar to the motion in an Earth orbit. The main similarity is the lack of coupling between the relative in-plane and out-of-plane motion of the spacecraft for linearized systems. This is briefly addressed in Chapter 4. Thus, answering research question 3 is of importance for all these environments. The reasons for focusing on Earth orbits are that time periods and distances are easier to grasp for this familiar environment than for, e.g., a formation orbiting Neptune and that, contrary to a halo orbit, the relative motion of the spacecraft can be designed such that it is passively stable (to first order).

Lastly, the research questions have deliberately been formulated such that they refer to a satellite formation consisting out of only two satellites. This has two reasons. Firstly, having more than two satellites in a formation creates many options for the relative motion of the satellites and the (non-)existence of inter-satellite links. Exploring all possible combinations, even with only three satellites, will be a huge effort and is beyond what can be done in the time frame of this research. Secondly, a formation consisting out of two satellites can be regarded as a ‘building block’ for larger formations. Thus, insights gained while studying such a small formation can be readily implemented for larger formations, which is likely to be more efficient than attempting to directly optimize a large formation without this background knowledge.

## 1.4 Methodologies

This section briefly describes the approach and the scenario used in answering the research questions posed in the preceding section.

### 1.4.1 Approach

The approach to answer the above research questions starts with examining the state-of-the-art in RF-based relative navigation for formation flying spacecraft. The methods and hardware used, as well as the inevitable problems in the practical application of these devices must be recognized and understood before the problem can be modeled in a sound and justifiable manner that lends itself for analysis. In addition, a good understanding of the system dynamics and of the tools available for relative state estimation must be acquired.
An analytical analysis of the relative navigation problem is required to understand how the variables in research question 1, antenna baseline, ranging accuracy, and inter-satellite distance, impact the error in the estimation of the relative state of two vehicles for a simple dynamic model. Also, the influence of the angle between the antenna baselines, assuming a maximum number of two baselines, on the estimation needs to be determined. The results of such an analysis can potentially be applied in the design of a RNS.

To understand the influence of the system dynamics on the relative navigation results, which is the subject of research questions 2 and 3, a numerical analysis is needed since an analytical approach becomes impractical after more than a few time steps due to the complexity of the dynamic model.

Verification of the results from the numerical analysis can be achieved by comparing these to the results from the analytical analysis. When the results match, the dynamics have little to no influence on the relative navigation result since the analytical analysis is performed for a simple dynamical setting. If the results differ, the results from the analytical analysis can help in determining how the dynamics of the problem influence the results.

### 1.4.2 Scenario

To investigate the research questions, the following scenario is studied: Two spacecraft in LEO orbit are flying in a passively stable formation. Using observations in the form of RF inter-satellite range measurements and a dynamic model of the relative motion, estimates of the relative state of the two spacecraft are made over a period spanning multiple orbits. By varying the initial conditions of the relative state, different inter-satellite distances and formation geometries are created. RF sensor properties such as antenna baseline and ranging accuracy are also varied. The relative navigation results for all these sub-scenarios are then used to create empirical relationships between the obtained relative navigation accuracy and the four design parameters (formation geometry, inter-satellite distance, antenna baseline, and ranging accuracy). These relationships show how a design parameter influences the relative navigation result and can therefore be useful in the design of a future formation flying mission.

Verification of these results is firstly provided through an analytical analysis. Here, statistical theory is used to derive equations for the 3D position estimate of one object relative to another object. The two objects remain stationary with respect to each other. The object for which the position is to be estimated is represented by a single point for which the position is unknown while the other object is represented by three points of which the positions are known. When the distances from the unknown point to the three known points are known, the relative position of the objects can be computed. When the distances are known exactly, an exact estimate will be the result, but when there is an error in the three distances, there will be an error in the estimate which depends on the error in the distance to the known points, the distance between the known points, and the distance of the unknown point to each of the known points. This is analogous to the estimation of the position of a RF transmitter antenna by measuring its range to three RF receiver anten-
The statistical analysis provides equations for the expected value and variance of the position estimate as a function of the mentioned variables. The equations are compared against the empirical relations found through numerical simulation.

The second means of verification is an observability analysis of the system. This analysis shows how well the relative state can be estimated, if at all, by indicating how well-conditioned the system is. Varying the design parameters changes the observability of the system. This change should match the change in relative navigation accuracy predicted by the empirical relations derived from the estimator results. The observability analysis also allows a study of the eigenvalues and eigenvectors of the system, which provide detailed information on how accurately the components of the relative state vector can be estimated relative to each other. The variation in the eigenvectors and eigenvalues due to a variation in the design parameters provides valuable information needed to answer the research questions.

1.5 Thesis Outline

This thesis is structured into five chapters of which chapters 2, 3, and 4 contain the main contributions of this research. Chapter 1, the current chapter, has introduced the concept of formation flying and autonomous relative navigation and has discussed some general issues involving these subjects. This led to the formulation of three research questions and a methodology to answer these.

Chapter 2 presents an investigation into the technologies used for relative navigation of spacecraft, which is the natural starting point for this research. Advantages and drawbacks in the use of these technologies are explored from a system point of view, i.e. it is investigated how the available technologies influence the design of the complete RNS. This investigation is split into two main parts: signal level and hardware level.

Chapter 3 discusses the estimation of the state of a system. As the system under consideration is a dynamical system, several dynamical models are introduced which describe the relative motion of two objects orbiting a primary body and one of these models is selected to be used during the research. Next, the types of observations that can be fed to the estimator are treated. The chapter concludes with a section devoted to numerical state estimation. The estimator selected for the research is described as well as the method with which the estimator results are analyzed. The section also contains a statistical treatment of the relative state estimation for two stationary objects (i.e., with respect to each other).

Chapter 4 describes the simulations that have been performed for this research and presents the results obtained from these simulations. It also links the results from the statistical analysis to the results obtained with the numerical estimator. The results obtained in this chapter provide the answers to the research questions posed earlier.

Chapter 5 concludes this research through a short summary of the research performed and presents the main conclusions that have been drawn. It also presents several recommendations for further continuation of this research and addresses the impact of this research on the design of future formation flying missions.
Before the research questions formulated in section 1.3 can be answered it is important to understand system aspects and the technology that enables relative navigation between satellites. What are system needs on relative navigation? Which signal design aspects need to be considered? What is the impact on hardware? What can be expected and what cannot? This is especially important for small satellites as these often have limited capabilities. To that end, this chapter deals with the design challenges for an RF-based RNS for formation flying satellites and is divided into four sections. The design challenges are mainly concerned with technology, but regulatory constraints considering frequency usage are also briefly considered. First, a brief discussion is presented on the top-level needs of an RF-based RNS: When is it needed and what information and hardware are needed to create a functional system? The second and third sections deal with in-depth system design considerations. These sections discuss qualitatively which design options there are on signal and hardware level and how these options affect the functioning of the RNS. The fourth section summarizes this chapter.

2.1 Relative Navigation System Needs

As already discussed in the previous chapter, accurate relative navigation for Earth-orbiting formation flying satellites can be achieved using differential GNSS measurements, as is demonstrated in the GRACE, PRISMA, and TanDEM-X missions. 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, which is the case for the PROBA-3 mission. For such missions, an autonomous RNS is needed. 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. Naturally, formation flying missions not in Earth orbit also require an autonomous RNS. Thus, in the following the focus is put on a dedicated RNS only.

Relative navigation deals with the estimation of a set of parameters which sufficiently defines the relative state (e.g., relative position, velocity, acceleration, clock offset) of two platforms at a certain epoch. In this work it is assumed that the measurements from the RNS do not allow for a full absolute orbit determination at a single epoch. Therefore, it is not possible to perform kinematic relative navigation (i.e., without any dynamic model). This type of navigation would be feasible if the RNS would use GNSS to obtain absolute position measurements of the satellites. Thus, as shown in Figure 2.1, physical measurements and a (relative) dynamics model are needed as inputs to estimate the relative state.

It is noted that in orbital dynamics, a relative dynamics model always has to include some information on the absolute orbit of the satellites. The reason for this is that the relative dynamics of the satellites are a result of the absolute dynamics of the satellites. In section 3.2 it is shown how the information on the absolute orbits of the satellites enters the equations for the relative motion. One consequence is that relative navigation can only be performed if the orbital radius of one of the satellites is known (measured or estimated). Thus, to perform relative navigation, information on the relative and on (some elements of) the absolute orbit of the satellites has to be available.

Here, the physical measurements constitute a number of RF range (rate) measurements between a transmitter (Tx) and a receiver (Rx). Figure 2.1 also shows the relative attitude estimate as an input for the relative state estimator since this information may be needed to convert the range (rate) measurements from the range between the Tx and Rx antennas to the range between the centers of mass (COMs). The 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 Tx and Rx antennas if these are present in sufficient quantity, hence the dashed line in Figure 2.1.

Figure 2.1: RF-based relative navigation principle.

At every individual block in Figure 2.1, errors are introduced and the challenge is to reduce their cumulative effect as much as needed to fulfill all requirements for the relative navigation accuracy at a cost that is as low as possible: For the range measurement itself, the generated signal will never be perfect, e.g. due to drift of the frequency source and
bandwidth limitations, the signal that arrives at the receiver will be distorted due to reflections, and noise is introduced in the receiver electronics. No relative dynamics model will ever be perfect and the model used onboard the satellites will have to be simplified in order to reduce the computing power needed to produce a relative state estimate. The relative attitude estimate, if produced using inter-satellite range measurements, will be inaccurate due to ranging errors and uncertainties in the antenna phase center locations and the satellite COM. The type of estimator and its implementation will also affect the accuracy in the relative state estimate.

For relative navigation, it is typical to estimate the position and velocity of one satellite (the ‘deputy’) with respect to another satellite (the ‘chief’) in a local orbital reference frame. The local orbital frame is typically defined with respect to an absolute orbital frame, in which the absolute orbit parameters of the satellites are expressed. The origin of the absolute orbital frame typically coincides with, for orbits around the Earth, the center of the Earth. It is common for the origin of the local orbital reference frame to lie on the same absolute orbit as the chief satellite. Thus, the local orbital frame can be said to ‘co-rotate’ with the chief satellite. Often, the origin of the local orbital frame is also chosen to coincide with the COM of the chief. This is convenient since relative dynamics models are based on the relative motion between the COMs of the satellites.

However, the measurements of the inter-satellite distance, which are needed to enable the estimation of the relative position and velocity, are made between antennas on both spacecraft. These are rarely positioned at the COM of the satellites. Thus, the range measurements between the antennas need to be converted to range measurements between the COMs. It is therefore necessary to accurately know the position of these antennas in the local orbital reference frame. More precisely, the positions of the phase centers of the antennas since the distance between these phase centers is what is actually measured. Knowing the position of the antennas in the local orbital reference frame requires that the orientation of the satellites in the local orbital frame has to be known and that the vectors between the antennas and the COM of the spacecraft are known. This is complicated by the fact that these can vary over the course of a mission when consumables, e.g., propellant, are used. Knowing that there are many error sources in the estimation of the inter-satellite range, accurate relative navigation will typically require knowledge on the position of the COM at mm-level. A similar accuracy is also required for the knowledge of the distance between the antenna phase centers and the COM.

For example, for the chief satellite the orientation of the satellite typically needs to be determined using the inertial attitude and the absolute orbit of the chief satellite. From the absolute orbit, the orientation of the unit vectors of the local orbital frame can be calculated, cf. subsection 3.2.1. For a circular orbit around Earth with a semi-major axis of 7000 km, errors in the knowledge of the position and velocity of the chief satellite of 10 km and 10 m/s, respectively, lead to errors of approximately 0.1° in the orientation of the unit vectors of the local orbital frame. Such errors in the absolute orbit can be expected when using TLE data to determine the satellite’s orbit, cf. [Gill et al. 2010], and are considered to be relatively poor. As the error in the location of the antennas on the chief satellite scales with the cosine of the error angle, the impact on the estimate of the antenna loca-
tion is quite small: 1.5 μm for an antenna located 1 m from the COM. However, the error in the orientation of the local orbital frame also directly impacts the relative position error since the relative position is being determined in an incorrect reference frame. For an inter-satellite distance of 1 km, a 0.1° error in the orientation of one of the axes of the local orbital reference frame results in an intrinsic position estimation error of 1.5 mm on that same axis. As these two errors are considered not to be very large for most applications, it can be concluded that relatively large errors in the determination of the absolute orbit of the chief satellite can be tolerated if there is no need for highly accurate relative navigation results. Reducing this error to negligible values can be achieved when the satellites are equipped with a GPS receiver. Then, the accuracy in the determination of the absolute orbit can easily be increased by three orders of magnitude (using, e.g., DLR’s Phoenix receiver or SSTL’s SGR-05 receiver, which are both very suitable for use on small satellites) as compared to TLE data.

When the orientation of the local orbital frame with respect to the absolute orbital frame is known, the orientation of the satellite in the local orbital frame can easily be determined by differencing the satellite’s inertial attitude with the orientation of the local orbital frame. For small satellites with limited resources, it is expected that the inertial attitude can be determined with a typical accuracy of 0.1° – 1°. This error is thus likely to dominate over the error caused by the non-perfect knowledge of the satellite’s absolute orbital parameters. This differencing can be performed for both satellites, which then also yields the relative attitude of the two satellites. Since the errors in the inertial attitude estimates of both satellites are uncorrelated, it is concluded that the relative attitude can be expected to be determined with an accuracy of 0.14° – 1.4°.

It also needs to be taken into account that other causes can affect the relative state estimation accuracy. For instance, the relative motion between the satellites can be such that there is no line-of-sight between the transmitting and the receiving antenna(s) for some time. This reduces the number of measurements, which has a detrimental effect on the navigation accuracy. Interference with another signal is also a potential source of measurement error and needs to be avoided. Thus, limiting the influence of all potential errors is not an easy task.

Measurement of the range (rate) using RF signals can be performed in several ways, but it always relies on the usage of either one-way or two-way signals. A one-way ranging signal is sent from a transmitter to a receiver on a target, where it is processed to yield a range measurement. A two-way signal is sent from a transmitter to a target where it is either reflected (radar, passive target) or received and re-transmitted (transponder, active target) towards the location of the original transmitter. Then, it is received and processed to yield a range measurement. Since all RF-based relative navigation systems for formation flying missions that have been designed up to now are derived from GPS-technology, which utilizes one-way ranging, only this method of relative navigation is considered here. The principle of two-way ranging will be treated briefly in subsection 3.3.1.

In order to determine the signal travel time τ of the ranging signal, which multiplied with the speed of light c yields the inter-satellite range ρ, information on the time of trans-
mission has to be available at the receiver. In addition, the estimator needs information on
the origin of the signal as there can be many signal sources in the formation. This can all
be provided by the transmitting satellite through a dedicated communication signal, but
it is convenient to include this data in the ranging signal itself. A Direct Sequence Spread
Spectrum (DSSS) signal, as used in GNSS systems, lends itself well for this. However, due to
hardware limitations, there will always exist an initial bias and relative clock drift between
the spacecraft that causes an error in the determination of $\tau$. Due to the multiplication
of the signal travel time with the speed of light, a small difference of for instance $10^{-10}$ s
between the clocks on two satellites already causes a range measurement error of 3 cm.
It depends on the mission requirements how large the clock offset can be. Note that the
drift between two clocks can be made very small using very stable clocks. If the clock-
induced measurement error cannot be reduced sufficiently, the instantaneous clock bias
needs to be estimated to allow accurate ranging. This requires extra information, which
can be acquired only by means of a second ranging signal, effectively making the system a
dual one-way ranging system.

Figure 2.2 provides a basic representation of the principle of dual one-way ranging.
A ranging signal is transmitted from spacecraft A to spacecraft B where it is received and
processed, yielding range measurement $\rho_{AB}$. At the same time instance, a ranging signal is
transmitted from spacecraft B to spacecraft A where it is received and processed, yielding
range measurement $\rho_{BA}$. However, there exists a clock bias $\Delta t$ between the two spacecraft
such that $t_A = t_B + \Delta t$. This leads to $\rho_{AB} = ct + c\Delta t = \rho + c\Delta t$ and $\rho_{BA} = \rho - c\Delta t$. Addition and
subtraction of the two range measurements provides the true range and the clock bias, cf.
Eqs. (2.1) and (2.2). Naturally, this is only true when perfect measurements are assumed
and signal travel times are not explicitly considered. The effect of measurement errors on
these calculations is treated in detail in subsection 3.3.2. Note that this technique requires
that one of the spacecraft sends its measurements to the other spacecraft to enable this
calculation.

$$\rho = \frac{\rho_{AB} + \rho_{BA}}{2} \quad (2.1)$$
$$\Delta t = \frac{\rho_{AB} - \rho_{BA}}{2c} \quad (2.2)$$

To achieve a high ranging accuracy, the ranges $\rho_{AB}$ and $\rho_{BA}$ in Eq. (2.1) should be the
ranges that are obtained after a so-called light time correction has been applied. This
correction is necessary to account for the relative motion between the spacecraft during
the time which the ranging signal needs to traverse the distance between the two space-
craft. Suppose that the range measurement is made at time $t$. Then, the signal has been
transmitted at time $t - \tau$. During the signal travel time, $\tau$, the relative position of the two
spacecraft has changed. Thus, the range measurement, $\rho$, does not represent the instan-
taneous inter-satellite range at time $t$ but a range between spacecraft A (B) at time $t - \tau$
and spacecraft B (A) at time $t$. To correct for this, a light time correction must be applied.
This is a fixed point iteration that uses the Taylor series expansion for the range at time.
\( t - \tau \). Assuming a ranging signal being transmitted from spacecraft A to spacecraft B, this Taylor series expansion can be written as

\[
\rho_{AB}(t - \tau) = \rho_{AB}(t) - \dot{\rho}_{AB}(t) \tau + \frac{1}{2} \ddot{\rho}_{AB}(t) \tau^2 + \mathcal{O}(\tau^3)
\]  
(2.3)

where \( \dot{\cdot} \) denotes differentiation with time. When the positions of the two spacecraft are expressed in a common reference frame as \( r_A(t) \) and \( r_B(t) \) the actual signal travel time can be expressed as

\[
c \cdot \tau = |r_B(t - \tau) - r_A(t)|
\]  
(2.4)

with \( |\cdot| \) denoting the absolute value. The fixed point iteration starts at step zero with the assumption of zero travel time: \( \tau^{(0)} = 0 \). In each iteration step, \( \tau \) is estimated using the following equation [Montenbruck and Gill 2000]:

\[
\tau^{(i+1)} = \frac{1}{c} \left| r_B \left( t - \tau^{(i)} \right) - r_A(t) \right|.
\]  
(2.5)

When the difference between \( \tau^{(i+1)} \) and \( \tau^{(i)} \) is smaller than a certain threshold the iteration is halted, yielding a more accurate estimation for the signal travel time. Note that typically the Taylor expansion of Eq. (2.3) is performed up to first order, so next to the range also the velocity at time \( t \) must be known. This is normally obtained from navigation estimates.

When the formation consists out of two spacecraft, interference between the signals is relatively easy to prevent. However, when the formation consists out of many spacecraft, signal interference becomes a major design challenge since for \( N \) spacecraft, the number of ranging signals, for a fully networked system, is equal to \( N(N-1) \). This quadratic growth in the number of signals leads to the need for an efficient and cost effective multiple access scheme for large formations. The modulation of data on top of the ranging signal also requires some means of bit error correction.

With respect to system hardware, antennas with sufficient field of view but with a gain which is as high as necessary, to reduce the required transmitter power, are needed. As
will be shown later, these are conflicting requirements. Antennas with passive multipath\(^1\) rejection characteristics, e.g., by suppressing signals with a certain polarization, are also highly desirable since multipath is a significant source of measurement errors. To receive and transmit the ranging signals, a transceiver is needed which should require low power and have low mass and small volume. The transceiver also needs to interface with the GNC system and with the command and data handling system (CDHS). For that reason, it is not uncommon that the hardware needed for the transceiver and the hardware needed for the processing of the relative navigation information are contained in a single unit. Lastly, as relative clock bias is an important navigation parameter, a stable on board frequency source (oscillator) is needed to minimize measurement errors.

### 2.2 Signal Level Design Considerations

This section treats in detail the various design considerations for the RNS on signal level. Since the ranging accuracy is one of the variables under study in research question 1, it is important to understand which accuracies are reasonable to assume for this variable. Therefore, in this section emphasis is placed on the design of the ranging code since this strongly affects the achievable ranging accuracy.

#### 2.2.1 Frequency Selection

Since signal bandwidth is a scarce commodity, RF frequency bands are regulated by the International Telecommunication Union (ITU). Table 2.1 lists the frequency bands between 1 and 100 GHz that are recommended for use in RF-based inter-satellite crosslinks.

<table>
<thead>
<tr>
<th>Band</th>
<th>S</th>
<th>Ku</th>
<th>Ka</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.025-2.110</td>
<td>13.75-14.3</td>
<td>22.55-23.55</td>
<td>59-64</td>
</tr>
<tr>
<td>Frequency range [GHz]</td>
<td>2.200-2.290</td>
<td>14.5-15.35</td>
<td>25.25-27.5</td>
<td>65-71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32.3-33.4</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Signal multipath implies that an RF signal arrives at a receiving antenna from more than one propagation route. Thus, the signal has actually 'traveled' multiple paths, hence the term multipath. It is caused by the fact that a radio signal is an electromagnetic wave that propagates in three dimensions and can be bent or reflected. For spacecraft, signal multipath is mainly caused by reflection off surfaces. In space, when a signal is transmitted from a transmitter to a receiver, part of the electromagnetic wave propagates in a straight line to the receiver but other parts of the wave that do not propagate directly towards the receiver can be reflected (causing full reversal of the polarization of the reflected signal) by spacecraft structures in such a way that they are also received at the receiver. Since these signals have traveled a longer path than the direct signal, they have a small delay. They also have less energy (smaller signal amplitude), a phase shift, and, depending on the number of reflections before being received, can have a polarization that is opposite to that of the direct signal, causing a certain degree of interference. These signals are normally undesired and are therefore effectively an error source.
Next to the regulatory issues, frequency selection for the ranging signal is mostly affected by the required ranging accuracy. As the inter-satellite range is estimated by measuring the phase of the ranging signal, which can be either the code phase or the carrier phase, a higher signal frequency in essence enables a more accurate range measurement than a lower signal frequency. For the carrier phase, this is obvious since a phase measurement accuracy of, e.g., 0.1 rad will result in a more accurate range measurement in case of a short wavelength (high frequency) than in case of a long wavelength (low frequency). For the code phase, the obtainable accuracy essentially depends on the bandwidth of the code that is superimposed on the carrier signal and not on the carrier frequency. However, since the bandwidth of the code is always a fraction of the carrier frequency, a higher carrier frequency allows for a wider code bandwidth and thus higher code phase measurement accuracy. Thus, for high ranging accuracy, a high signal frequency is preferred. Frequency selection for the ranging signal is also affected by the frequency used for the Telemetry, Tracking and Command (TT&C) subsystem of the spacecraft. Additional filtering can be required when these frequencies do not have sufficient spectral separation.

With respect to the signal power of the transmitted and received signals, the situation is a bit more complex. Preferably, the amount of signal power that is received, $C_r$, from a signal that is transmitted with signal power $C_t$ is as large as possible. This way, the influence of noise on the signal is relatively small, which leads to a better quality of the received information (e.g., less bit errors, more accurate code ranging). According to the Friis transmission equation, these two variables are related as [Kaplan and Hegarty 2006]

$$C_r = C_t G_t G_r \left( \frac{\lambda}{4\pi r} \right)^2$$  \hspace{1cm} (2.6)

with $\lambda$ the wavelength of the carrier signal, $r$ the distance between the two antennas, and $G_t$ and $G_r$ the gains of the transmitting and receiving antennas, respectively. The last term in Eq. (2.6) is called the free space loss. In a way, this equation is a bit deceptive since it appears that, in order to increase $C_r$, one should increase $\lambda$. Thus, the signal frequency should be reduced. However, the antenna gain is also a function of $\lambda$. In fact, the antenna gain can be expressed as [Kaplan and Hegarty 2006]

$$G = \frac{4\pi A_e}{\lambda^2}$$  \hspace{1cm} (2.7)

with $A_e$ the effective antenna area. Inserting Eq. (2.7) into Eq. (2.6) leads to

$$C_r = C_t A_{e,t} A_{e,r} \frac{\lambda^2}{r^2}.$$  \hspace{1cm} (2.8)

Equation (2.8) shows that the dependency of the signal propagation loss on the carrier frequency is in fact an antenna effect and not a wave propagation effect: If the wavelength of the signal is increased, the effective antenna area needs to be increased to maintain the same received signal power. Since small spacecraft do not allow for large antennas, it is preferred to use high frequency signals in order to obtain high values for $C_r$. Yet, it is emphasized that this does imply that the antenna gain will be high, which leads to narrow
signal beam widths. This means that, in order to benefit from the higher received signal power, the boresights of the antennas on the various spacecraft must be accurately aligned with each other. This can imply an additional complexity for the relative attitude control system of the spacecraft. In addition, if omnidirectional communication is desired for the spacecraft, for reasons of formation safety or simply due to the geometry of the formation, the high gain of the antennas is a drawback since more antennas will be needed to achieve omnidirectional coverage than in case low frequency signals are used.

In case the spacecraft formation is in LEO, the ionosphere causes a ranging error that is inversely proportional to the signal frequency. As the error is also proportional to the total electron content (TEC) that the signal encounters, this error is small for small intersatellite distances. When this error is significant, it can be mitigated by using two ranging signals at different frequencies. By properly weighing and adding the measurements resulting from these signals, the ionospheric contribution can be eliminated from the measurement. Using multiple signals at different frequencies also enables integer ambiguity resolution (IAR), discussed in subsection 3.4.1.

In conclusion, Table 2.2 lists several effects that are influenced by the carrier frequency of a ranging signal. For each effect, a qualitative statement is provided whether a low or a high carrier frequency causes this effect to have a positive influence on the accuracy in the range measurement. The table shows that frequency selection is a trade-off (between the frequencies listed in Table 2.1). The result of this trade-off will highly depend on the mission characteristics, of which some have already been briefly treated. For (very) small satellites, which typically have limited power available, a trade-off is likely to result in the selection of a relatively high carrier frequency.

Table 2.2: Key effects on frequency of RF ranging signals. For each effect, it is indicated whether a low or a high carrier frequency is desired in order for that effect to have a positive influence on the accuracy in the range measurement.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Favors frequency to be</th>
</tr>
</thead>
<tbody>
<tr>
<td>Received signal power</td>
<td>high</td>
</tr>
<tr>
<td>Phase measurement accuracy</td>
<td>high</td>
</tr>
<tr>
<td>Integer ambiguity resolution</td>
<td>low</td>
</tr>
<tr>
<td>Signal multipath</td>
<td>high</td>
</tr>
<tr>
<td>Ionospheric path delay</td>
<td>high</td>
</tr>
</tbody>
</table>

2.2.2 Multiple Access

Accurate inter-satellite ranging requires the use of multiple access technologies to prevent mutual interference. This is true even in the case of only two satellites. The basic forms of multiple access are Frequency Division Multiple Access (FDMA), Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA), cf. Figure 2.3. Hybrid forms of these can also be employed. Usage of FDMA and CDMA allows transmission and re-
ception 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 multiple access interference (MAI) which results in a reduction of the effective $C/N_0$ from a remote transmitter due to signal interference caused by one or more signals from one or more transmitters in the vicinity of the receiver. A special case of MAI is near-far interference, which occurs when a high power signal transmitted by a nearby transmitter completely ‘drowns’ the signal transmitted by an antenna 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 (e.g., 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.

![Figure 2.3: Schematic representation of CDMA (a), FDMA (b), and TDMA (c). The colored boxes represent signals from different sources.](image)

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 propagate, or predict, the signal properties of a certain channel (assuming multiple channels are used) forward in time in case the transmitter assigned to that channel is silent. However, when signal prediction 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 [Stadter et al. 2004]. 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 $\mu$s 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. Variations on the traditional
TDMA method which are deliberately non-synchronized, but still guarantee that all satellites receive each of the others equally after a long period of time are presented by [Tien et al. 2004]. These methods remove the need for synchronization, but suffer from near-far interference (although self-interference is ruled out).

Figure 2.4: Schematic representation of different communication topologies for spacecraft formations. In both the star and the hierarchical topology, an example of a good multiple access method for that topology is provided. The multiple access method is indicated through the signals $s_{i,j}$ where $i$ indicates the signal frequency and $j$ indicates the signal code. As there are many options for the multiple access method of the distributed topology, the multiple access method is not explicitly provided there.

The manner in which the relative positions of the spacecraft in the formation are controlled, cf. Fig. 2.4, also influences the type of multiple access technology that should be used. If the control is centralized, i.e. one spacecraft (the ‘chief’) determines the control actions of all other spacecraft in the formation (the ‘deputies’), a star topology with the chief at the center of the star is a likely scenario. Then, the signals of the chief and the deputies can be spectrally separated by means of FDMA or TDMA and the signals of the deputies can be spectrally separated by means of CDMA. If the formation control is hierarchical, the method presented in [Stadter et al. 2001] is a good choice for the multiple access method. In case of fully distributed formation control, the communication net-
work does not need to be fully connected to allow for control of the formation and thus the spacecraft only need to share information with a selected subset of the spacecraft in the formation. This leads to many options for the multiple access method since the sizes of the subsets and the formation geometry itself now become driving factors.

The choice of multiple access technology is also driven by the mission need. Sometimes half-duplex communication (TDMA) is sufficient, but full-duplex can be required for challenging missions. In addition, if the mission is designed such that additional spacecraft can be added to the formation in the future, the multiple access technology used must be flexible enough to allow for this. This means that the hardware and software of the RNS must be able to process additional signals. For FDMA and CDMA this requires that the transceiver front-end is designed such that there are sufficient channels available on board the original spacecraft to cope with the additional signals transmitted by the new spacecraft in the formation. For FDMA, this also requires that the filter bandwidth of the front-end is wide enough. For TDMA, no hardware modifications are needed to allow for additional signals if only a single channel is used, but the time slots for the individual signals have to remain wide enough to allow accurate range measurements and sufficient data transfer capacity while at the same time they have to be narrow enough to allow for rapid re-acquisition of a signal.

2.2.3 Ranging Signal Design

A mentioned earlier, 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.

The correlation of two signals $s_i(t)$ and $s_j(t)$, which are the same if $i = j$, is mathematically represented as follows. The correlation function $R(\tau)$ for the correlation of a constant power signal $s_i(t)$ with a constant power signal $s_j(t)$ that has a delay equal to $\tau$ is defined as [Kaplan and Hegarty 2006]

$$R(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} s_i^*(t) s_j(t + \tau) \, dt$$

(2.9)

where $T$ is equal to the length of the signal and where $^*$ denotes complex conjugation.
2.2. SIGNAL LEVEL DESIGN CONSIDERATIONS

Since signals have finite lengths, Eq. (2.9) can be reduced to

$$R(\tau) = \frac{1}{T} \int_0^T s_i^*(t) s_j(t + \tau) \, dt$$

(2.10)

for practical applications. If the two signals are very dissimilar, or orthogonal, the result for $R(\tau)$ will always be (close to) zero. However, if the signals are quite similar, the result for $R(\tau)$ will be non-zero for some $\tau$. In fact, if the signals are the same and periodic with period $P$, then $R = 1$ for $\tau = kP$ with $k \in \mathbb{Z}$. This feature is at the heart of ranging using DSSS signals since it allows:

1. detection of signals that are well below the noise floor (known as de-spreading),
2. reception of many ranging signals at the same time on different channels with very little signal interference,
3. accurate alignment of a local code with a received code and with that, accurate ranging.

The principle of correlation is demonstrated in Fig. 2.5. In Fig. 2.5(a) a very short code $s_1(t)$ is autocorrelated with itself and cross-correlated with a different code $s_2(t)$. In Fig. 2.5(b) a longer code $s_3(t)$ is autocorrelated with itself and cross-correlated with a different code $s_4(t)$. The four, randomly chosen, codes are as follows:

1. $s_1(t) = [-1, -1, 1, -1, 1, 1, -1]$
2. $s_2(t) = [-1, 1, -1, 1, 1, -1, -1]$
3. $s_3(t) = [-1, -1, 1, -1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1, -1]$
4. $s_4(t) = [1, 1, -1, -1, 1, 1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1, -1].$

As an example, when code $s_1$ is shifted by $\tau = -2$ symbols, the shifted code is equal to $s_1(t-2) = [1, -1, 1, 1, -1, -1, -1].$ Correlating $s_1(t)$ with $s_1(t-2)$ using Eq. (2.10) results in

$$R(-2) = \frac{1}{7} \sum \left[ (-1 \cdot 1) + (-1 \cdot -1) + (1 \cdot 1) + (-1 \cdot -1) + (1 \cdot -1) + (1 \cdot -1) + (-1 \cdot 1) \right] = -\frac{1}{7}$$

Performing this calculation for various values of $\tau$ leads to the result displayed in Fig. 2.5. In Fig. 2.5(a), it is seen that when the time-shifted version of $s_1$ matches $s_1$ exactly, the correlation is equal to 1. The correlation of $s_1$ and $s_2$ however, is never more than 0.7. Thus, by time-shifting a local replica of the coded signal to be received, other similarly coded signals can be rejected since they do not result in full correlation. However, in practice, situations can arise where the cross-correlation leads to larger values for $R(\tau)$ than the autocorrelation due to different power levels for the incoming signals and poor cross-correlation properties of the codes used. To minimize the probability of that occurring,
longer codes can be used, since these can have much better cross-correlation properties than short codes, as is clear from Fig. 2.5(b). In addition, longer codes allow the reception of weaker signals as they reduce the noise floor (or increase the process gain). Thus, less transmitter power is needed to enable acquisition of the signal at the receiver.

Figure 2.5: Example of the correlation of two binary sequences. In (a), a length-7 binary sequence $s_1$ is correlated with itself and a different length-7 binary sequence $s_2$. In (b), a length-15 binary sequence $s_3$ is correlated with itself and a different length-15 binary sequence $s_4$.

Ranging by means of a DSSS ranging signal occurs in two steps. First, a coarse two-dimensional search over frequency (to find the Doppler shift of the received signal) and time (to align the received signal and the local replica) 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 in phase (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 so-called Costas PLL loop has to be used to extract the data contained in the signal. For code ranging, the phase measurement discussed in subsection 3.3.2 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 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 over one chip period 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. It is used in the GPS C/A and P(Y) codes. To distinguish between the GPS C/A code and the GPS P(Y) code, the notation BPSK-R($f_c/f_{\text{ref}}$) is often used with $f_{\text{ref}}$ denoting a reference frequency of, usually, 1.023 MHz. Thus, the GPS C/A code is modulated using BPSK-R(1) while the GPS P(Y) code is modulated using BPSK-R(10).

In recent years, Binary Offset Carrier (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 intersatellite ranging. A BOC($f_s/f_{\text{ref}}, f_c/f_{\text{ref}}$) signal is created by modulating a sine wave carrier with the product of a PRN spreading code and a square wave subcarrier, with frequency $f_s$, having each binary ±1 values. Thus, BOC(10,5) implies a 10.23 MHz subcarrier frequency and a 5.115 MHz code rate. Note that BPSK-R(1) and BPSK-R(10) can be regarded as BOC(0.5,1) and BOC(5,10), respectively. Figure 2.6 depicts the principles of BPSK-R and BOC modulation.

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, which is the reason why rectangular waveforms are used in GNSS signals. Spreading symbol shapes under consideration for future implementation are Binary Coded Symbols (BCS), Multilevel Coded Spreading Symbols (MCS), Composite BPSK (CBPSK), Sinusoidal Offset Carrier Signals (SOC), and Prolate Spheroidal Wave Functions (PSWF) [Avila-Rodriguez et al. 2007].

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. Figure 2.7. This comes however at the price of larger signal bandwidth. The pulse shape also affects the autocorrelation function. Figure 2.7(a) shows that BOC(1,1) has a much sharper correlation peak than the 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 BPSK-R signals, 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, cf. subsection 2.2.5, it can be proven [Giugno and Luise 2005] 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 $C/N_0$ for a desired code ranging accuracy. 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
Figure 2.6: DSSS modulation using BPSK-R (a) and BOC (b). In the figures, "D" is short for data, "C" is short for spreading code and "SC" is short for subcarrier. For both modulations, the final signal is created by means of BPSK modulation of the carrier with the resulting code. For BPSK-R, the resulting code is created by means of modulo 2 addition of a data signal and a spreading code. For BOC, an additional step is performed. In this step, the product is formed between a square subcarrier and the resultant of the data and the spreading code. This extra step splits the classical BPSK spectrum in two symmetrical components with no remaining power on the carrier frequency.

used, band limiting is of crucial importance to avoid large guard bands or complex filter designs. When TDMA is used, this is of much less concern.

Signal multipath is an important performance limitation for current ranging signal 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.
Software-based methods to reduce multipath-induced errors do exist, but at present these perform very poor for the ultra-short delays encountered in the space environment [Sun et al. 2011]. Therefore, current practice for multipath mitigation in space is extensive calibration on ground: A ranging signal from a precisely known location is sent to a mock-up of one of the spacecraft in the formation. This is preferably done in an anechoic chamber to minimize noise from other RF sources and to minimize multipath contributions from objects other than the mock-up, cf. [Aguttes et al. 2008]. By changing the azimuth and elevation of the mock-up, ranging errors due to multipath can be collected for various relative attitudes. These are stored in a lookup table on board the spacecraft and are subtracted from the actual range measurements. Downsides of this method are the costs of the calibration campaign, especially when the formation geometry is such that ranging signals can come from very different directions, but also the constraint that the outer surface of
the spacecraft cannot be changed once the calibration has been completed. Even then, in orbit re-calibration can still be needed. Indeed, for the PRISMA mission, differences between calibration data obtained on ground and in flight had, in some areas, a magnitude similar to the magnitude of the error itself (cm-level). Improving the on-ground calibration implies a considerable effort for the FFRF and is likely to never match actual flight conditions. Thus, it has been suggested to mitigate multipath-induced biases by other means such as better antenna accommodation on the spacecraft [Delpech et al. 2011].

The PRN code is, as the name suggests, a code designed to mimic white noise but which is actually deterministic. The most famous example is the set of codes devised by Gold [Gold 1967] which are used in the GPS C/A codes. There, the Gold codes have a length of 1023 chips and have very good cross-correlation properties, making it possible to track multiple codes transmitted at the same time and with comparable \( \frac{C}{N_0} \). 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(Y)-code. 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 [Hein et al. 2006].

### 2.2.4 Data Modulation and Bit Errors

Autonomous 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 where one chief 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 [Sun et al. 2010]. 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, 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 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 \( \frac{C}{N_0} \), which eases signal acquisition. How-
ever, when separate data and pilot channels are used, this argument becomes less strong and data rates can be increased somewhat. It is noted that the French FFRF instrument, currently flying on the PRISMA mission, 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 in-phase channel and the navigation and engineering data on the quadrature channel, thereby relaxing the constraint on the width of a data bit (which is still limited by the data demodulation capability of the digital hardware) [Bourga et al. 2002].

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 zeros 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 which are not detailed here. 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 [Borre et al. 2007].

### 2.2.5 Cramér-Rao Covariance Bound

In this subsection it is explored how the signal design influences the theoretical lower bound for the ranging accuracy, known as the Cramér-Rao Bound (CRB). The goal of this subsection is not to derive an optimum ranging signal design, as this depends on many factors and is likely to be different for different missions, but to show how variations in, e.g., pulse shape or chipping rate affect the CRB, which is linked to research question 1.

Assuming coherent downconversion of the received DSSS signal to baseband, the received signal \( s(t) \) at time \( t \) can be modeled after [Giugno and Luise 2005] as

\[
s(t) = \sqrt{2C} \sum_{k=0}^{N_c-1} c_k g(t - kT_c - \tau) + n(t)
\]

(2.11)

where \( C \) denotes the average signal power, \( c_k \in \{-1, 1\} \) are the code bits of the PRN sequence, \( \tau \) is the signal delay and \( n(t) \) is white Gaussian noise with a power spectral density (PSD) of \( N_0 \). The pulse shape is given as \( g(t) \) and has an energy \( E_s \) equal to \( T_c \). The PRN code bits can be collected in a vector \( c \) with length \( N_c \).

The CRB states that the accuracy of an unbiased estimator is bounded by the inverse of the Fisher information. In the current context, the estimation accuracy is equal to the
variance in the signal delay estimation error $\sigma_\tau^2$. Denoting the Fisher information as $I(\tau)$, the CRB is equal to [Giugno and Luise 2005]

$$\sigma_\tau^2 \geq \text{CRB}(\tau) \triangleq (I(\tau))^{-1} = \left[ E_s \left\{ \left( \frac{\partial \ln F(s|\tau)}{\partial \tau} \right)^2 \right\} \right]^{-1}.$$  \hfill (2.12)

In Eq. (2.12), $s$ is a vector representation of the received signal $s(t)$. The probability density function (PDF) $F(s|\tau)$ of $s$ is called the likelihood function of $\tau$, and $E_s \{ \cdot \}$ is the expectation of the enclosed quantity with respect to $s$. Unfortunately, computation of the CRB is difficult in practice due to the necessity of computing $F(s|\tau)$. For that reason, a modified Cramér-Rao Bound (MCRB) can be derived [Giugno and Luise 2005]:

$$\sigma_\tau^2 \geq \text{CRB}(\tau) \geq \text{MCRB}(\tau) \triangleq \left[ E_s \left\{ \left( \frac{\partial \ln F(s|c, \tau)}{\partial \tau} \right)^2 \right\} \right]^{-1}.$$  \hfill (2.13)

Although Eq. (2.13) has the same structure as Eq. (2.12), the analytical solution to it is much easier to derive since the PDF in Eq. (2.13) is an exponential function whose argument is a quadratic form in the difference between the noisy received signal and the transmitted one. Thus, $\ln F(s|c, \tau)$ equals this quadratic form and the expectation in the MCRB is readily derived. For more information, the reader is referred to section 2.4.6 in [Mengali and D’Andrea 1997].

Note that usage of the MCRB can be misleading in practical situations since $\text{CRB}(\tau) \geq \text{MCRB}(\tau)$ and thus a ranging accuracy reached in practice that is in fact close to the theoretical optimum given by the CRB can be relatively far off the theoretical optimum given by MCRB. However, as here the purpose is to highlight how signal design changes affect the achievable accuracy, the MCRB is considered to be an adequate performance measure.

The MCRB can be rewritten as [Giugno and Luise 2005]

$$\text{MCRB}(\tau) = \frac{B_L T_c^2}{2C/N_0} \frac{1}{4\pi^2 \xi}$$  \hfill (2.14)

where $B_L$ is the equivalent noise bandwidth of the estimator used, $C/N_0$ is the carrier-to-noise density ratio of the received signal, and $\xi$ is the pulse shaping factor (PSF), defined as

$$\xi = \frac{\int_{-\infty}^{\infty} f^2 |G(f)|^2 df}{\int_{-\infty}^{\infty} |G(f)|^2 df} = \frac{T_c^2}{\int_{-\infty}^{\infty} |G(f)|^2 df} \int_{-\infty}^{\infty} f^2 |G(f)|^2 df$$  \hfill (2.15)

where $G(f)$ is the Fourier transform of the chip pulse shape $g(t)$ and $f$ is the signal frequency. Note that the last equality in Eq. (2.15) is due to the chip pulse having an energy equal to $T_c$, thus making the integral in the denominator, which represents the signal power $C$, equal to 1 since $E_s = CT_c$. Furthermore, since

$$|G(f)|^2 = S(f)$$  \hfill (2.16)
with $S(f)$ the power spectral density of $s(t)$, Eq. (2.15) can be rewritten as

$$\xi = \frac{T_c^2}{4\pi^2} \int_{-\infty}^{\infty} \omega^2 S(f) \, df$$

(2.17)

with $\omega = 2\pi f$ the angular frequency of the signal. According to [Spilker Jr. 1996], the following equalities hold:

$$\int_{-\infty}^{\infty} \omega^2 S(f) \, df = -\ddot{R}_s(0) = \Delta\omega^2.$$  

(2.18)

The quantity $\Delta\omega^2$ is known as the square of the Gabor bandwidth and $\ddot{R}_s(0)$ is the second derivative with respect to time of the autocorrelation function of the pulse shape at zero delay (i.e., the slope of the DLL or PLL discriminator curve). A large Gabor bandwidth reduces the MCRB and thus contributed to a high ranging accuracy.

Eq. (2.14) indicates several strategies that can be applied to minimize the variance in the range estimation in case the only source of ranging error is Gaussian white noise. The first means is to maximize the signal-to-noise ratio. This can be achieved through small inter-satellite distance, small free-space loss of the signal (i.e., low frequency), high transmission power, high antenna gain, and low-noise receiver hardware. The second means which reduces the MCRB is to reduce the pulse period $T_c$. The third means with which to minimize the variance in the range estimation is to minimize the noise bandwidth of the estimator, commonly a DLL for code measurements and a PLL for phase measurements. Here however, a practical limitation is introduced by the relative dynamics of the transmitting and receiving platforms: Reducing the noise bandwidth makes it harder to keep lock on a signal transmitted by a source that is moving in a manner that results in large accelerations relative to the receiver. For spacecraft formations, relative velocities are generally not larger than several m/s and change gradually if no control force is applied. Thus, for nominal situations, accelerations are small and a small noise bandwidth is feasible. However, control actions can potentially lead to short periods of high relative dynamics and thus high accelerations. When it is important to maintain lock on the signal, the situations with high relative dynamics become driving for the noise bandwidth. The fourth and last means is to maximize the PSF; which will be treated in more detail in the following.

Eq. (2.15) shows that, in theory, the PSF can be made arbitrarily large if the ranging signal has infinite bandwidth. However, in practice, every signal is bandlimited within some band $[-B, B]$, resulting in

\[^2\] It is possible to design an analog DLL or PLL such that it can actively switch between different (partially overlapping) filters, thereby changing the noise bandwidth, when needed. Then, the DLL or PLL can be optimized for different modes, e.g., for ranging accuracy or tracking robustness. For software defined receivers, changing the noise bandwidth can be done very smoothly since the filter properties are completely software-defined. Thus, changing the filter parameters in the software changes the noise bandwidth.
\[
\xi = \frac{T_c^2}{4\pi^2} \int_{-B}^{B} \omega^2 S(f) \, df.
\] (2.19)

Bandlimiting has two negative effects: a loss of signal power and a loss of higher-frequency content which results in a less sharp correlation peak and therefore a less accurate range estimate. Figure 2.8 visualizes this by displaying the autocorrelation functions of BOC(1,1) and BOC(10,1) for infinite bandwidth and for a 20.46 MHz bandwidth \((f_{\text{ref}} = 1.023 \text{ MHz})\). For BOC(1,1), a 20.46 MHz bandwidth is sufficient to capture almost all energy (98.5\%) in the signal and thus its autocorrelation peak is still relatively sharp when compared to the infinite bandwidth signal. For BOC(10,1), the 20.46 MHz bandwidth is not sufficient to capture all the signal energy (only 55.3\%) and thus the amplitude of the autocorrelation function is about half that of the infinite bandwidth BOC(10,1) autocorrelation function (assuming transmission at infinite bandwidth). Furthermore, BOC(10,1) results in many peaks in the autocorrelation function whose relative magnitude decreases slowly for increasing delay. For a noiseless signal this is not a problem, but in the presence of interference and multipath the signal gets distorted, resulting in different magnitudes for the peaks. In such a scenario it is easy for the receiver to lock onto the wrong peak, resulting in a very poor ranging accuracy. Therefore, acquisition of BOC signals with high \(f_s/f_c\) ratio is often performed in a two-step method where first only a single sidelobe of the signal is acquired, resulting in a BPSK-R-like autocorrelation function. Locking on to this signal is relatively easy and reduces the search space, and thus the chance to lock on the wrong peak, for acquisition of the full BOC signal, which is performed in the second step.

![Figure 2.8: Effect of bandlimiting on signal autocorrelation for BOC(1,1) and BOC(10,1).](image)

To show how the chipping rate and the bandwidth affect the PSF, the Gabor bandwidths of following pulse shapes are compared in Table 2.3: BPSK-R(1), BPSK-R(10), BOC(1,1),
BOC(10,1), BOC(10,5), and BOC(10,10). The PSDs of these pulse shapes are \[ \text{Kaplan and Hegarty 2006} \]

\[
S_{\text{BPSK-R}}(f) = T_c \text{sinc}^2(\pi f T_c) \tag{2.20}
\]

\[
S_{\text{BOC}}(f) = T_c \text{sinc}^2(\pi f T_c) \tan^2\left(\frac{\pi f}{2 f_s}ight), \quad \frac{2 f_s}{f_c} = k \text{ even} \tag{2.21}
\]

and the fraction of signal power \( \zeta \) remaining after bandlimiting to \( B \) is

\[
\zeta(B) = \int_{-B}^{B} S(f) \, df. \tag{2.22}
\]

In order to have unit signal power in the determination of the Gabor bandwidth, the PSD of the signal has to be multiplied by \( \zeta^{-1} \) such that \( \zeta^{-1} S(f) \) has unit area over the bandwidth \( B \). Table 2.3 displays \( \Delta \omega^2 \) for different pulse shapes and different bandwidths. When not taking into account the effects of bandlimiting, the design goal would be to maximize the square of the Gabor bandwidth, which is displayed in the third column. However, when taking bandlimiting into account and thus assuming equal signal power levels, which is a better approach, the design goal is to maximize the values in the last column.

**Table 2.3:** Effect of chipping frequency and pulse shape on the MCRB.

<table>
<thead>
<tr>
<th>Pulse shape</th>
<th>One-sided bandwidth</th>
<th>( \frac{\Delta \omega^2}{f^2_{\text{ref}}} ) [-]</th>
<th>( \zeta ) [-]</th>
<th>( \zeta^{-1} \frac{\Delta \omega^2}{f^2_{\text{ref}}} ) [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSK-R(1)</td>
<td></td>
<td>4</td>
<td>0.9028</td>
<td>4.43</td>
</tr>
<tr>
<td>BPSK-R(10)</td>
<td></td>
<td>2.58</td>
<td>0.1978</td>
<td>13.05</td>
</tr>
<tr>
<td>BOC(1,1)</td>
<td></td>
<td>12</td>
<td>0.6446</td>
<td>18.62</td>
</tr>
<tr>
<td>BOC(10,1)</td>
<td></td>
<td>0.03</td>
<td>0.0025</td>
<td>11.20</td>
</tr>
<tr>
<td>BOC(10,5)</td>
<td></td>
<td>0.07</td>
<td>0.0031</td>
<td>23.13</td>
</tr>
<tr>
<td>BOC(10,10)</td>
<td></td>
<td>0.04</td>
<td>0.0016</td>
<td>24.38</td>
</tr>
<tr>
<td>BPSK-R(1)</td>
<td>( B = f_{\text{ref}} )</td>
<td>40</td>
<td>0.9899</td>
<td>40.41</td>
</tr>
<tr>
<td>BPSK-R(10)</td>
<td>( B = f_{\text{ref}} )</td>
<td>400</td>
<td>0.9028</td>
<td>443.07</td>
</tr>
<tr>
<td>BOC(1,1)</td>
<td>( B = 10 f_{\text{ref}} )</td>
<td>120</td>
<td>0.9697</td>
<td>123.75</td>
</tr>
<tr>
<td>BOC(10,1)</td>
<td>( B = 10 f_{\text{ref}} )</td>
<td>1560</td>
<td>0.4480</td>
<td>3482.14</td>
</tr>
<tr>
<td>BOC(10,5)</td>
<td>( B = 10 f_{\text{ref}} )</td>
<td>1400</td>
<td>0.5531</td>
<td>2531.19</td>
</tr>
<tr>
<td>BOC(10,10)</td>
<td>( B = 10 f_{\text{ref}} )</td>
<td>1200</td>
<td>0.6446</td>
<td>1861.62</td>
</tr>
</tbody>
</table>

The results in Table 2.3 show that a high chipping rate leads to a higher ranging accuracy (for the same ratio of \( f_s/f_c \)), but that it is inefficient to have a signal bandwidth smaller than the chipping rate. Furthermore, a large value for the ratio \( f_s/f_c \), resulting in a sharp autocorrelation peak, is beneficial for the ranging accuracy only if the signal bandwidth is large enough to capture the signal lobe with the largest amplitude, which is approximately located at \( f_s \).
Determination of the bandlimited Gabor bandwidth allows calculation of the theoretical minimum signal delay estimation error $\sigma_\tau$ for the pulse shapes and bandwidths presented in Table 2.3. This will be done here for the BPSK-R(1) and BOC(1,1) pulse shapes with $B = f_{\text{ref}}$. First, the foregoing allows rewriting Eq. (2.14) into its bandlimited version as

$$\text{MCRB}_{\text{bandlimited}}(\tau) = \frac{B_L}{2C/N_0 \cdot \xi^{-1} \Delta \omega^2}. \quad (2.23)$$

Taking the square root of the result obtained using Eq. (2.23) yields the theoretical minimum value for $\sigma_\tau$. Multiplying the result for $\sigma_\tau$ with the speed of light yields the code-based ranging accuracy $\sigma_\rho$. In Table 2.3 it is found that for the BPSK(1) pulse shape $\xi^{-1} \Delta \omega^2 / f_{\text{ref}}^2 = 4.43$ and that for the BOC(1,1) pulse shape $\xi^{-1} \Delta \omega^2 / f_{\text{ref}}^2 = 18.62$. Using this and assuming $B_L = 0.1 \text{Hz}$, $C/N_0 = 40 \text{ dB-Hz}$, and $f_{\text{ref}} = 1 \text{MHz}$ leads to $\sigma_{\text{BPSK-R(1)}} \approx 32 \text{ cm}$ and $\sigma_{\text{BOC(1,1)}} \approx 16 \text{ cm}$. Thus, using the BOC(1,1) pulse shape instead of the BPSK-R(1) pulse shape results in a factor two improvement in the ranging accuracy. In addition, the assumed values for the various parameters in Eq. (2.23) lead to code-based ranging accuracies that are in line with the accuracies expected to be needed for formation flying missions. This is important since the values for these parameters are believed to be feasible for small satellites, see for instance [Bourga et al. 2002] for the $C/N_0$. Yet, it has to be kept in mind that the accuracies computed here are optimistic values and that the accuracy obtainable when using a DLL also depends on, e.g., the DLL correlator spacing and the observation time (in case of noncoherent DLL operation\(^3\)).

Considering the foregoing, the BOC(1,1) pulse shape has a set of properties that are very favorable for use on a small satellite: It offers a relatively good code-based ranging accuracy at a relatively small signal bandwidth, which is favorable to reduce hardware design efforts and power consumption since transmission/reception of a signal with a large bandwidth is more complex and requires more power than transmission/reception of a signal with smaller bandwidth\(^4\). In addition, acquisition and tracking of signals with this pulse shape is considered to be more reliable and robust than that of signals with higher subcarrier frequency and chipping rate such as BOC(10,5) and BOC(10,10). These pulse shapes have better ranging accuracy, cf. Table 2.3, but also have several peaks in the autocorrelation function that have an amplitude that is close to that of the central peak, resulting in a high risk that the signal tracking loops can lock on the wrong peak in the presence of multipath and interference, which is highly undesirable. Thus, the BOC(1,1) pulse shape is recommended for use, or as a starting point for an optimized signal, in formations with small spacecraft.

\(^3\)A coherent DLL requires the PLL to be in lock, i.e., it is tracking the carrier phase and the navigation data bits have been removed from the signal. A noncoherent DLL is able to track the code with the navigation data bit present and the PLL is not necessarily in lock. A properly implemented coherent DLL will always yield more accurate results than a noncoherent DLL since it does not suffer from the so-called squaring loss that is present in the noncoherent DLL.

\(^4\)A perfectly rectangular pulse such as used in BPSK-R and BOC implies infinite signal bandwidth; the smaller the bandwidth, the less ideal the pulse shape and the less accurate the range measurement.
2.3 Hardware Level Design Considerations

Application of an RF-based RNS on a satellite places specific demands on the hardware used on the satellite. Naturally, there are requirements on the hardware for the RNS 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.

2.3.1 Spacecraft Design

The RF-based RNS will impact the satellite design with respect to its layout. Protruding elements that are not part of the RNS, e.g. solar panels and TT&C 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 on board autonomy is required, the antennas of the RNS itself should be mounted such that a $4\pi$ 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, as already mentioned in subsection 1.1.5 this can result in a large amount of antennas needed.

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. [Purcell et al. 1998] arrived at a circular beam with roughly constant gain out to $45 - 50^\circ$ off axis and a sharp drop-off at larger angles. Such a beamwidth is not surprising 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. Alternatively, a narrow, high frequency signal (e.g. Ku-band) can be used for very accurate inter-satellite ranging during nominal conditions while a wide, low frequency (e.g. S-band) signal can be used during off-nominal conditions. This can potentially reduce the total number of antennas needed, but this depends greatly on the number of spacecraft in the formation and on the formation geometry.

The satellite platform should also provide a mechanically and thermally stable environment to enable the most accurate measurements. This is put to the extreme in the GRACE mission where the entire satellite design is driven by the need for micrometer-level accuracy ($\sim 10^{-4}$ cycle) in the carrier phase measurements [Dunn et al. 2002]. For accurate navigation, the COM of the spacecraft also has to be known accurately. This implies accurate knowledge before launch and during flight. The latter requires that accurate pro-
pellant location models and/or measurements have to be employed to take into account the shift in the COM due to thruster firings.

In general, the presence of a relative navigation subsystem has one or more consequences for the other traditional spacecraft bus subsystems (ADCS, CDHS, TT&C, Electrical Power Subsystem (EPS), Thermal Control System (TCS), and the spacecraft structure), which implies an increase in mass for all these subsystems. For the ADCS, the RNS adds inertia to the spacecraft and thus larger moments are needed to maintain the correct spacecraft orientation. In addition, next to pointing the payload to a target and pointing the solar panels towards the sun, the ADCS now also has to ensure that the orientation of the spacecraft is such that relative navigation is possible, which complicates the spacecraft layout. The CDHS needs to relay data, telemetry and commands to and from the RNS, which, if not highly integrated with the On-Board Computer (OBC) requires a substantial amount of cables and interface boards. If on the other hand the RNS is highly integrated with the OBC, then extra functionality and computing power needs to be added to the OBC. The EPS must supply (un)regulated power to all active RNS components during nominal operations. This is very likely to include eclipses, requiring additional power generation (solar panel area) and storage (batteries) in sunlit conditions and power delivery by means of batteries during eclipse. For the TT&C subsystem, extra bandwidth is needed to exchange information between the ground station and the RNS and the signal frequency of the TT&C subsystem must have sufficient spectral separation with the ranging signals. Lastly, the spacecraft structure needs to be designed to handle the loads caused by the RNS, which implies that the structure has to be more massive.

The use of a RNS also has implications for spacecraft operations such as safe modes and instrument calibration phases. However, these are considered to be beyond the scope of this study and are therefore not treated.

2.3.2 Transceiver Design

The core of the RF-based RNS is the transceiver. There, the RF signals to be transmitted are generated 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 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 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 spacecraft 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 spacecraft system
integration is high, most, if not all, high-level processing needed for the relative navigation will be performed by the satellite’s GNC system. An example of a transceiver design is depicted in Figure 2.9.

![Diagram of a transceiver design](image)

**Figure 2.9:** Simplified block diagram of a transceiver design for formation flying missions, based on the design presented in [Tien et al. 2004].

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, cf. [Merodio Codinachs and Weigand 2009, Xil 2011].

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 spatially most separated platforms meets the minimum 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 the minimum output power that should be produced to achieve the desired $C/N_0$. This has as added benefit that the efficiency of the transmitter is higher at lower power levels, leading to even more power savings. 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 [Stadter et al. 2001]. Although complicated, implementation of adaptive power control is not unfeasible for a RNS for small satellites; it has been implemented on the (Nanosat) Crosslink Transceiver ((N)CLT), a modular RNS design that in its smallest form is a 4” cube with a mass of 0.86 kg [Stadter et al. 2004]. Alternatively, a less complex design with only a low and a high power setting available can also be considered. This solution has been implemented for the Autonomous Formation Flying Sensor [Tien et al. 2004], which was to fly on the Terrestrial Planet Finder (TPF) mission.

Internal calibration is very important for accurate ranging and is needed to compensate for biases resulting from satellite thermal, electrical, and mechanical variations. Internal calibration 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 Figure 2.9, 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, of which the first is considered to be the favored option for implementation on small spacecraft, are available to deal with this without resorting to TDMA [Purcell et al. 1998]:

- **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.

In the transceiver design presented in 2.9, switches are incorporated in the receive and transmit channels. Thus, this transceiver design is very suitable for TDMA networks.

The design of the RNS should also allow for on-orbit re-calibration to account for dif-
ferent 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 on board and used to correct measurements, it should be possible to update it after in-orbit re-calibration. In addition, the transceiver design should be scalable to meet the demands of various missions (e.g., increased number of channels, additional functionality) and to provide cost-effectiveness.

2.3.3 Clock Stability

As already mentioned in section 2.1, the stability of the on board clock is an important parameter if very accurate navigation solutions are required. It also provides time to the other spacecraft subsystems using pulse-per-second (PPS) signals. However, due to clock drift, the PPS will drift over time. If there are very stringent requirements on the absolute time reference used by the spacecraft in the formation, the maximum error in the PPS can be a more stringent requirement for the clock 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 [Kaplan and Hegarty 2006].

Quartz crystal oscillators (XOs) are nowadays the standard for on board 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 [Kaplan and Hegarty 2006]

\[
V(t) = (V_0 + \epsilon_V(t)) \sin \left( 2\pi f_0 t + \delta \phi(t) \right)
\]  

(2.24)

where \( V_0 \) and \( f_0 \) are the nominal amplitude and frequency, respectively, with corresponding errors \( \epsilon_V(t) \) and \( \delta \phi(t) \). The argument of the sine function in Eq. (2.24) is equal to the instantaneous phase \( \phi(t) \). 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 [Misra and Enge 2006]

\[
f(t) = f_0 + \Delta f + (t - t_0) \dot{f} + \tilde{f}(t)
\]  

(2.25)

where \( \Delta f \) 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

\[
\Delta t(t_1) = \Delta t(t_0) + \frac{\Delta f}{f_0} (t_1 - t_0) + \frac{\dot{f}}{2f_0} (t_1 - t_0)^2 + \int_{t_0}^{t_1} \frac{\tilde{f}(t)}{f_0} \, dt.
\]  

(2.26)

Differencing Eq. (2.26) between two free running clocks results in 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
Eq. (2.26) represent systematic effects while the last term represents random frequency fluctuations, the size of which depends upon environmental effects and aging. They are characterized by their variance and auto-correlation function.

The Allan variance is a commonly used measure of the timekeeping ability of a clock with respect to random processes and is defined as [Vig 2008]

$$\sigma^2_{f_{\text{dev}}} (\tau) = \frac{1}{2} \left\langle (f_{\text{dev},i+1} - f_{\text{dev},i})^2 \right\rangle$$

where $f_{\text{dev}}$ is a measurement of relative frequency deviation ($f_{\text{dev}} = \Delta f / f_0$) of an oscillator averaged over the period $\tau$, $(f_{\text{dev},i+1} - f_{\text{dev},i})$ are the differences between pairs of successive measurements of $f_{\text{dev}}$, and, ideally, $\langle \cdot \rangle$ denotes a time average of an infinite number of $(f_{\text{dev},i+1} - f_{\text{dev},i})^2$. The nature of the random process associated with the measurements of relative frequency deviations changes with the size of the averaging interval $\tau$ and the Allan variance converges for all of them. The square root of the Allan variance is called the Allan deviation $\sigma_{f_{\text{dev}}} (\tau)$. The root mean square (RMS) error of a clock after an interval $\tau$ is approximated as $\tau \cdot \sigma_{f_{\text{dev}}} (\tau)$.

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 oscillator is thermally decoupled from the surrounding temperature, which is typically achieved by elevating the operating temperature of the oscillator 15 to 20 degrees above the highest expected temperature of the surroundings. 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 an 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 [Vig 1992].

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 2.4. Other oscillator frequency influences are drive energy, gravity, shock, vibration, electromagnetic signals physically close to the oscillator, retrace, and hysteresis [Hewlet Packard 1997]. 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 a stable operating point is reached. It may not be apparent that an MCXO or an TCXO 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 2.4, oscillator stability is very sensitive to temperature fluctuations. Since mass provides inertia against temperature changes and is therefore beneficial, miniaturization of oscillators without sacrificing stability is difficult (for a given technology).

Table 2.4: Oscillator comparison. The first four rows are expressed in terms of the fractional frequency error $\Delta f / f_0$ [Vig 2008].

<table>
<thead>
<tr>
<th></th>
<th>TCXO</th>
<th>MCXO</th>
<th>OCXO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy/year* [-]</td>
<td>$2 \times 10^{-6}$</td>
<td>$5 \times 10^{-8}$</td>
<td>$1 \times 10^{-8}$</td>
</tr>
<tr>
<td>Aging/year [-]</td>
<td>$5 \times 10^{-7}$</td>
<td>$2 \times 10^{-8}$</td>
<td>$5 \times 10^{-9}$</td>
</tr>
<tr>
<td>Temperature stability† [-]</td>
<td>$5 \times 10^{-7}$</td>
<td>$3 \times 10^{-8}$</td>
<td>$1 \times 10^{-9}$</td>
</tr>
<tr>
<td>Stability, $\sigma_y(\tau)$ ($\tau = 1s$) [-]</td>
<td>$1 \times 10^{-9}$</td>
<td>$3 \times 10^{-10}$</td>
<td>$1 \times 10^{-12}$</td>
</tr>
<tr>
<td>Volume [cm$^3$]</td>
<td>10</td>
<td>30</td>
<td>20–200</td>
</tr>
<tr>
<td>Warmup time [s]</td>
<td>2 (to $1 \times 10^{-6}$)</td>
<td>2 (to $2 \times 10^{-8}$)</td>
<td>240 (to $1 \times 10^{-8}$)</td>
</tr>
<tr>
<td>Power [W] (at lowest temp.)</td>
<td>0.04</td>
<td>0.04</td>
<td>0.6</td>
</tr>
<tr>
<td>Approximate Price [US$]</td>
<td>10 – 100</td>
<td>&lt; 1,000</td>
<td>200 – 2,000</td>
</tr>
</tbody>
</table>

*including environmental effects
†-55°C to +85°C

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 than an OCXO.

The importance of having a reference XO with a small short-term Allan variance is made apparent in the following example calculation where the overall accuracy of a PLL is determined. In [Kaplan and Hegarty 2006], the following empirical expression is provided for the Allan variance-induced oscillator jitter $\theta_A$ in degrees:
In Eq. (2.28), \( f_L \) is the central frequency of the ranging signal and the value for the constant is 144 or 160 for a second-order or a third-order PLL, respectively. For a ranging signal central frequency of 2 GHz and \( B_L = 50 \) Hz, using the values given in Table 2.4 for the value of \( \sigma_y(\tau) (\tau = 1 \text{s}) \) leads to the results in Table 2.5.

Table 2.5: Allan variance-induced oscillator jitter assuming a ranging signal central frequency of 2 GHz and \( B_L = 50 \) Hz.

<table>
<thead>
<tr>
<th></th>
<th>TCXO</th>
<th>MCXO</th>
<th>OCXO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[deg]</td>
<td>[mm]</td>
<td>[deg]</td>
</tr>
<tr>
<td>Second-order PLL</td>
<td>5.8</td>
<td>2.4</td>
<td>0.58</td>
</tr>
<tr>
<td>Third-order PLL</td>
<td>6.4</td>
<td>2.7</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Clearly, the selection of the oscillator type has a major impact on the Allan variance-induced oscillator jitter. Together with thermal noise \( \sigma_{PLL_t}^2 \), vibration-induced noise \( \sigma_v^2 \), and dynamic stress error \( \theta_e \) (jerk), this jitter contributes to the total PLL accuracy as shown in the following expression from [Kaplan and Hegarty 2006]:

\[
\sigma_{PLL} = \sqrt{\sigma_{PLL_t}^2 + \sigma_v^2 + \theta_A^2 + \frac{\theta_e^2}{3}}. \tag{2.29}
\]

For space-based applications, the dynamic stress error will only play a role during spacecraft maneuvers and the vibration-induced noise can be expected to be very low since many payloads will require this. Then, the main contributers to the PLL accuracy will be the thermal noise and the oscillator jitter:

\[
\sigma_{PLL} \approx \sqrt{\sigma_{PLL_t}^2 + \theta_A^2}. \tag{2.30}
\]

The thermal noise (in degrees) for an arctangent PLL is given by [Kaplan and Hegarty 2006] as

\[
\sigma_{PLL_t} = \frac{360}{2\pi} \sqrt{\frac{B_L}{C/N_0} \left(1 + \frac{1}{2 T_{pre} C/N_0}\right)}. \tag{2.31}
\]

where \( T_{pre} \) is the predetection integration time. Selecting \( B_L = 50 \) Hz, \( T_{pre} = 20 \) ms, and \( C/N_0 = 40 \) dB-Hz yields \( \sigma_{PLL_t} = 4.1 \text{deg} \) or \( \sigma_{PLL_t} = 1.7 \text{mm} \). Comparing this result

\[5\]According to [Grewal et al. 2007], the order of a PLL refers to its capability to track different types of signal dynamics. Most GPS receivers use second- or third-order PLLs. A second-order loop can track a constant rate of phase change (i.e., constant frequency) with zero average phase error and a constant rate of frequency change with a nonzero but constant phase error. A third-order loop can track a constant rate of frequency change with zero average phase error and a constant acceleration of frequency with nonzero but constant phase error. If the application has relatively benign dynamics and low rate of signal frequency change, a second-order PLL is sufficient. On the other hand, in highly dynamic environments (i.e., missiles) a third-order or even higher-order PLL is needed to avoid loss of signal lock due to the high accelerations encountered.

\[
\theta_A = \text{const} \cdot \frac{\sigma_y(\tau) f_L}{B_L}. \tag{2.28}
\]
with the results in Table 2.5 shows that the oscillator jitter can contribute significantly to the overall PLL accuracy. In case of a TCXO, it can even be the dominant error source. Thus, using a very stable oscillator will contribute significantly to maximizing the overall accuracy in the determination of the phase of a ranging signal.

2.4 Summary

In this chapter, the principles of RF-based relative navigation using DSSS signals between two or more spacecraft were introduced and extensively discussed. This is summarized in this section.

To enable accurate relative navigation, not only are accurate range measurements to, preferably, multiple spatially separated sources needed, but also an accurate (relative) dynamics model, accurate information on the relative attitude, and accurate information on the clock bias. This requires the exchange of information (e.g., range measurements, attitude measurements) between the spacecraft. This low rate data can be included in the ranging signal, which is typically a one-way signal, resulting in an efficient system. Ranging accuracy ultimately depends on the signal frequency, which should therefore be as high as possible. However, frequency selection is heavily constrained by regulatory, physical, and technical considerations which all have to be taken into consideration for the design of the RNS.

As spacecraft formations may consist of more than two spacecraft, an efficient multiple access method that allows for accurate navigation is needed. This method can be based on TDMA, CDMA, FDMA, or hybrid forms of these. The formation control architecture (centralized, hierarchical, or distributed), which dictates the networking topology of the formation, and the total number of spacecraft strongly affects the choice for the multiple access method. In addition, the multiple access method must sometimes take into account that additional spacecraft can be added to the formation in the future. There is no 'ultimate' multiple access method that is optimal for all formation types and thus the multiple access method must be tailored to each mission.

The ranging accuracy that can be achieved is ultimately dependent on the ranging signal. Measurement of the fractional signal phase provides the highest ranging accuracy and thus high frequency signals, which have short wavelengths, offer the potential for the most accurate relative navigation. However, measurement of the signal phase does not provide any information on the total distance between the transmitter and the receiver since there typically is an unknown integer number of signal cycles between the transmitter and the receiver. To allow measurement of the total distance, the time of transmission of the received signal has to be known. This can be achieved by modulating the signal with a long periodic ranging code on top of which also data about the transmission time of the ranging code is modulated. By using spread spectrum techniques, a local replica of the ranging code can be (coarsely) aligned in both time and frequency with the received code by means of correlation. By knowing 1) the time of transmission of the received code, 2) how much the local code replica has been shifted in time to align it with the received code,
and 3) the bias between the clocks on the two spacecraft, a relatively coarse measurement of the inter-satellite range can be obtained. Obtaining the bias between the clocks on the two spacecraft requires a range measurement from spacecraft A to spacecraft B and vice versa. Subtraction of these measurements yields the (instantaneous) clock bias.

The data that needs to be exchanged between the platforms is commonly BPSK modulated on top of the ranging code which limits the achievable data rate since a data bit has to be several code periods wide to ease signal acquisition. This constraint can be partially mitigated by modulating the signal using QPSK such that the code is modulated onto the in-phase component and the data is modulated onto the quadrature component. The data exchanged between the spacecraft commonly consist of formation GNC and FDIR data, but can also contain science data. The data rate depends much on the networking topology of the formation and on the formation control accuracy to be achieved: For a loosely controlled formation with few spacecraft a data rate in the order of kbit/s is sufficient whereas for a tightly controlled, fully networked formation with many spacecraft the required data rate can go up to Mbit/s.

The accuracy of the range measurement is increased by fine-tuning the alignment of the codes in a DLL. By also tracking the carrier phase of the signal by means of a PLL, accurate phase measurements can be made. The phase measurement can also be used to aid the DLL. The waveform with which the code is modulated on top of the ranging signal also determines how accurate the range can be measured by the DLL since it affects the shape of the signal’s autocorrelation function. Analysis of the Modified Cramér-Rao Bound shows that waveforms that result in a sharp autocorrelation peak will yield higher ranging accuracy. A wide signal bandwidth also increases the accuracy. Unfortunately, waveforms that have a sharp autocorrelation peak commonly have multiple autocorrelation peaks that can cause the receiver to lock on a wrong peak, and thus a wrong range measurement, in the presence of noise and multipath. Thus, selection of the ranging waveform is a trade-off between ranging accuracy and acquisition robustness.

In the hardware design of the RNS, an important consideration, which is tied to the ranging signal frequency, is the beamwidth of the antennas. To reduce transmitter power, a high antenna gain and thus a small beamwidth is preferred. However, for formation safety reasons, full sky coverage is often desirable, which implies a large beamwidth to reduce the amount of antennas needed. The satellite design should also take into account multipath mitigation or minimization, as this phenomenon causes measurement errors that can easily be on the order of cm and which are difficult to remove from the range measurement.

For the transceiver, low mass and low power are obviously important design drivers. Scalability of the design can make the design more cost effective since it offers the potential of using the design for different missions. Depending on the level of spacecraft system integration, the transceiver unit can also be required to perform formation GNC tasks such as the (pre-)processing of navigation information or even the estimation of relative positions, relative velocities, and clock biases. To obtain high ranging accuracy, a calibration of the transceiver is very important. There should be the possibility to update multipath
correction tables in orbit and hardware-induced ranging errors should be eliminated by applying internal calibration. In addition, self-interference must either be prevented entirely (by selecting a proper multiple access method) or reduced through passive (filtering) or active (destructive interference) means. Power usage can be minimized by applying dynamic power control for the transmitter.

The stability of the on-board clock can also be an important design driver since it affects the ranging accuracy. Since implementing an atomic clock on a small spacecraft is currently not feasible, the most stable frequency source available is an OCXO, which has much better short-term stability, the most important clock parameter for navigation purposes, than other XOs. However, when other ranging errors, such as multipath, are much larger than the clock error, a TCXO or MCXO become viable alternatives provided that their long term stability is sufficient for the mission purposes.
Chapter 3

Relative Navigation Methods and Analysis

This chapter introduces the relative navigation methods and analyses needed to answer the research questions. In the first section, a rationale is provided for the introduction of assumptions that simplify the overall problem. These assumptions are needed to allow a general analysis of the variables in the research questions. The next three sections of the chapter deal with the separate aspects of the estimation of the state of a dynamic system, which is the manner in which (relative) navigation is often performed. These aspects are: the equations describing the relative orbit dynamics, the observations (measurements) of quantities that contain information on the state of the system to be estimated, and the state estimation itself. The last section briefly summarizes this chapter.

3.1 Rationale for System Level Simplifications

In the preceding chapter, a detailed description has been provided of the considerations in the design of a RNS for formation flying spacecraft. Proper management of all these aspects will result in a system design that provides reliable and accurate relative navigation. However, many of these aspects result in behavior of a specific system that cannot be generalized. As the research questions have been posed in a general framework, this has led to the adoption of various system level simplifications. This allows a concise analysis of results obtained for a large variety of system configurations, leading to general conclusions that are applicable to all configurations or to a large subset of these configurations. Without the system level simplifications, conclusions can only be generated for very specific circumstances, which is not the goal of this work. Adoption of the following simplifications results in the most basic scenario possible, which still captures the essential parts of the problem to be studied.
Firstly, it has been decided not to consider a multiple access (MA) scheme since choosing a particular MA scheme enforces certain choices that impact the complete system. For instance, when choosing CDMA as the multiple access scheme, the minimum time it takes to perform a dual one-way range measurement depends on the inter-satellite distance and on the processing speed of the system. Thus, when a minimum time interval is desired between measurements, some assumption on the processing speed has to be made or a detailed simulation or hardware test has to be performed. In addition, as the measurement accuracy also depends on the relative clock drift and the level of signal interference, again some assumptions have to be made or a detailed simulation or hardware test has to be performed. For the signal interference, this implies choosing a signal power ratio and choosing a certain signal coding scheme, as these both influence the level of interference. This all renders the scenario created in such a way far from general. For TDMA and FDMA, similar considerations apply.

Signal multipath, which causes a ranging bias and correlation between measurements, is also not considered since the effect of multipath strongly depends on the geometry of the spacecraft and will, in general, due to changing relative attitudes, not be constant over time. Therefore, this effect is strongly scenario-dependent and does not lend itself for consideration in a general setting.

The signal travel time between the satellites and the time for acquisition and processing is also not considered. Due to the relative motion of the spacecraft during the signal travel time, this quantity should be included in the measurement equation and a light time iteration should be performed to arrive at a proper value for the signal travel time. However, as the error remaining after the iteration is always very small and since the key aspects to be studied are not depending on this, the iteration is not performed. Instead, potential errors in the computed signal travel time are considered to be captured as a small contribution to the measurement noise. If desired, a light time iteration can, without problem, be included in future, more refined models. In addition, the time needed from signal acquisition to integer ambiguity resolution on the carrier phase is also not considered. The reason for this is that it is strongly dependent on the signal design (number of signal carriers, carrier frequencies, code length, etc.) and thus strongly scenario-dependent. Also, when considering integer ambiguity resolution, failure of this process must be taken into account, which is of little relevance to the research questions.

For the estimation of the relative state, a complicated model for the relative motion in the estimator combined with a high fidelity model for the true absolute orbits will yield results that approximate reality best. However, since the objective of this research is not to obtain the most accurate relative navigation results possible, but to determine how various important parameters influence the relative navigation results, there is no need for a high fidelity simulation and estimation of the relative motion. Even atmospheric drag and the second-order zonal coefficient of the Earth's gravity field $J_2$, which have to be accounted for when designing a formation in LEO, can be neglected in this study since they are small, or can be assumed to be small, for the scenarios considered. That the effects caused by these parameters can indeed be assumed small is justified by the following, discussed extensively in chapter 2 of [D'Amico 2010]: As the magnitude of the relative motion caused
by $J_2$ depends mainly on the difference in inclination of the orbits of the two spacecraft and since a small relative out-of-plane motion implies a small difference in inclination, the effect of $J_2$ will be very small for the scenarios with small relative out-of-plane motion that are of special interest here (cf. research question 3) and can therefore be neglected. For example, the growth in differential inclination between two closely separated orbits will be approximately 1 mm/orbit for a differential inclination of $0.001^\circ$. This results in a maximum relative out-of-plane distance between the satellites of 123 m for a semi-major axis of 7028 km. This growth will not be detectable within 5 orbits (the period that is used in the analysis) for ranging accuracies on the order of cm (the highest accuracy used in the analysis). Thus, not including $J_2$ in the dynamic model will not lead to different conclusions than when $J_2$ would be included in the dynamic model. Note that for scenarios with relative out-of-plane distances of 1 km or more and a ranging accuracy of 1 cm, the effect of $J_2$ would become noticeable (or better: observable, cf. subsection 3.4.3). To obtain a small relative acceleration due to differential atmospheric drag, the ballistic coefficients of the two spacecraft must be very similar in magnitude. According to [D’Amico 2010], with proper design, the ballistic coefficients of two spacecraft can be matched up to a few percent, leading to negligible drift in the relative motion. In addition, the effect of differential drag is not related to the geometry of the formation but only on the spacecraft design and the local environment and thus outside the scope of this research.

Thus, due to these simplifications, a one-way ranging scheme is assumed with a Tx antenna on the deputy and one or more Rx antennas on the chief. In the estimator, measurements are assumed to be made instantaneous and the model for the relative orbital dynamics is kept simple by assuming linearized dynamics (Hill-Clohessy-Wiltshire (HCW) equations, cf. subsection 3.2.2) and absence of perturbations. The measurement noise is assumed to be white Gaussian noise. By varying the ranging accuracy with two orders of magnitude (1 cm, 1 dm, and 1 m), several phases in the treatment of the measurements can be assumed. For instance, a ranging accuracy of 1 m can represent a situation with only code measurements while a ranging accuracy of 1 cm can represent a situation where integer ambiguity resolution has been achieved on the carrier phase. A very high measurement frequency is not needed since the formations studied do not exhibit rapid changes in formation geometry and thus a mediocre measurement frequency of 100 measurements per orbit, roughly once per minute, can be taken, which is realistically attainable for any multiple access scheme and transceiver design.

3.2 Relative Orbital Dynamics Modeling

The first two subsections of this section deal with the derivation of the equations of motion that describe the relative orbital dynamics of two spacecraft, a chief and a deputy, orbiting a common primary body. Contrary to absolute orbital dynamics, in which the motion of one or more objects due to gravity is described with respect to a reference frame that can be considered to be inertial (i.e., non-rotating), relative orbital dynamics commonly describes the motion of one or more objects due to gravity with respect to a reference
frame that is rotating (i.e., non-inertial). Often, the non-inertial frame follows a trajectory in the inertial frame that is similar to that of one of the objects under study. The derivation of the equations of motion follows the approach taken in [Alfriend et al. 2010].

3.2.1 Nonlinear Relative Motion Around a Common Primary

The exact modeling of the motion of a body in space is very complex because many forces have to be accounted for. However, using appropriate simplifications, the motion of a body orbiting another body, the primary (e.g., Earth, Mars, Venus), can be described in a (quasi-)inertial coordinate frame centered at the primary. When it is assumed that there are no external or internal forces except gravity, the primary’s mass is much larger than the orbiting body’s mass, and the gravitational force is Newtonian, the Keplerian two-body equations of motion can be written as

\[
\ddot{R} = -\frac{\mu R}{R^3}
\]

with \( R = (X, Y, Z)^T \) the orbit radius vector of the orbiting body expressed in inertial coordinates \( X, Y, \) and \( Z, R = \|R\| \) the orbit radius, \( \|\cdot\| \) the Euclidean vector norm and \( \mu \) the gravitational coefficient of the primary body, here taken to be the Earth. When two bodies, or spacecraft, orbit the primary, equations that describe their relative motion can be developed. Introducing the subscripts “c” and “d” for the chief and the deputy spacecraft, respectively, the range \( r \) between the chief and the deputy can be written as

\[
r = R_d - R_c.
\]

Using Eqs. (3.1) and (3.2), the relative acceleration \( \ddot{r} \) of the deputy with respect to the chief in the inertial frame yields

\[
\ddot{r} = -\frac{\mu (R_c + r)}{\|R_c + r\|^3} + \frac{\mu}{R_c^3} R_c.
\]

with the superscript \( \mathcal{J} \) indicating that \( r \) is expressed in inertial coordinates.

For spacecraft formations, it is often convenient to express the relative motion not in terms of inertial positions and velocities, but in terms of positions and velocities defined in a local co-rotating frame. There is no commonly agreed upon naming convention for this frame (commonly used names are the Local Vertical Local Horizontal (LVLH) frame, the Radial In-track Cross-track (RIC) frame, the Radial Tangential Normal (RTN) frame and the Hill frame). Here, the co-rotating frame is referred to as the Hill frame and its origin coincides with the COM of the chief, cf. Fig. 3.1. Its unit vectors \( e_x, e_y, e_z \) are defined as

\[
e_x = \frac{R_c}{\|R_c\|}, \quad e_y = e_z \times e_x, \quad e_z = \frac{R_c \times V_c}{\|R_c \times V_c\|}
\]

where \( V_c \) is the inertial velocity vector of the chief.
To express the relative acceleration in the Hill frame, which is a rotating frame, the equations that describe the relative acceleration in the inertial frame have to be related to the equations that describe the relative acceleration in the rotating frame. This is accomplished using kinematic theory. A well-known result from kinematic theory states that the time derivative of a vector $\mathbf{r}$ in an inertial reference frame equals the time derivative of the same vector in a rotating reference frame plus the cross-product of that vector with the angular velocity vector $\dot{\mathbf{\theta}}$ of the rotating frame with respect to the inertial frame. This can be written as

$$\frac{d\mathcal{L}\mathbf{r}}{dt} = \frac{d\mathcal{L}\mathbf{r}}{dt} + \dot{\mathbf{\theta}} \times \mathcal{L}\mathbf{r} \quad (3.5)$$

where the superscript $\mathcal{L}$ indicates that $\mathbf{r}$ is expressed in local (Hill) coordinates. Applying Eq. (3.5) twice leads to

$$\mathcal{L}\ddot{\mathbf{r}} = \frac{d\mathcal{L}\dot{\mathbf{r}}}{dt} = \frac{d\mathcal{L}\dot{\mathbf{r}}}{dt} + \dot{\mathbf{\theta}} \times \mathcal{L}\dot{\mathbf{r}} + \dot{\mathbf{\theta}} \times \left(\frac{d\mathcal{L}\dot{\mathbf{r}}}{dt} + \dot{\mathbf{\theta}} \times \mathcal{L}\mathbf{r}\right)$$

$$= \mathcal{L}\ddot{\mathbf{r}} + 2\dot{\mathbf{\theta}} \times \mathcal{L}\dot{\mathbf{r}} + \ddot{\mathbf{\theta}} \times \mathcal{L}\mathbf{r} + \dot{\mathbf{\theta}} \times \left(\dot{\mathbf{\theta}} \times \mathcal{L}\mathbf{r}\right). \quad (3.6)$$

Since the angular velocity vector of any spacecraft orbit is normal to the orbit plane, $\dot{\mathbf{\theta}}$ can be written as

$$\dot{\mathbf{\theta}} = \begin{pmatrix} 0, 0, \dot{\theta}_c \end{pmatrix}^T$$

with $\dot{\theta}_c$ the angular velocity of the chief. Writing the vectors $\mathbf{R}_c$ and $\mathcal{L}\mathbf{r}$ as
\[ R_c = (R_c, 0, 0)^T \]  
\[ \mathbf{r} = (x, y, z)^T \]  

with \( x, y, \) and \( z \) the position coordinates in the local frame provides all the information needed to express the relative motion of the two spacecraft in the rotating frame. Substituting Eq. (3.3) and Eqs. (3.7–3.9) into Eq. (3.6) yields the following (unperturbed) nonlinear equations for the relative motion:

\[
\ddot{x} - 2\dot{\theta}_c \dot{y} - \dot{\theta}_c^2 x = -\frac{\mu (R_c + x)}{[(R_c + x)^2 + y^2 + z^2]^{3/2}} + \frac{\mu}{R_c^2} 
\]  
\[
\ddot{y} + 2\dot{\theta}_c \dot{x} + \dot{\theta}_c^2 y = -\frac{\mu y}{[(R_c + x)^2 + y^2 + z^2]^{3/2}} 
\]  
\[
\ddot{z} = -\frac{\mu z}{[(R_c + x)^2 + y^2 + z^2]^{3/2}}. 
\]

In general, \( R_c \) and \( \theta_c \) are not constant. Thus, two more equations are needed to describe the evolution of these two parameters [Alfriend et al. 2010]

\[
\ddot{R}_c = R_c \dot{\theta}_c^2 - \frac{\mu}{R_c^2} 
\]  
\[
\ddot{\theta}_c = -\frac{2\dot{R}_c \dot{\theta}_c}{R_c} 
\]

and form together with Eqs. (3.10a–3.10c) a ten-dimensional system of nonlinear differential equations. In these equations, the motion in \( x-, y-, \) and \( z- \)direction is fully coupled and thus a motion in one of these directions results in a motion in the other two directions. In many practical situations however, the full nonlinear equations are not needed and can be simplified without sacrifying noticeable accuracy in the modeling of the dynamics. These simplifications can be made when the orbit of the chief is (near) circular and/or when the distance between the chief and the deputy is small.

For a circular chief orbit, the angular velocity and the orbit radius are constant, thus \( \dot{\theta}_c = 0, \dot{\theta}_c = n_c = \text{const.} \), and \( R_c = a_c = \text{const.} \) where \( a_c \) is the semi-major axis of the chief’s orbit and \( n_c \) is the mean motion of the chief, which is defined as \( n_c = (\mu/a_c^3)^{1/2} \). Substituting these simplifications into Eqs. (3.10a–3.10c) leads to the following nonlinear equations for relative motion in case of a circular chief orbit:
3.2. RELATIVE ORBITAL DYNAMICS MODELING

\[ \ddot{x} - 2n_c \dot{y} - n_c^2 x = -\frac{\mu (a_c + x)}{[(a_c + x)^2 + y^2 + z^2]^{\frac{3}{2}}} + \frac{\mu}{a_c^2} \]  
\[ \ddot{y} + 2n_c \dot{x} - n_c^2 y = -\frac{\mu y}{[(a_c + x)^2 + y^2 + z^2]^{\frac{3}{2}}} \]  
\[ \ddot{z} = -\frac{\mu z}{[(a_c + x)^2 + y^2 + z^2]^{\frac{3}{2}}} \]  

(3.13a)  
(3.13b)  
(3.13c)

Note that Eqs. (3.10) and Eqs. (3.13), which describe the nonlinear relative motion of two satellites, are fully dependent on absolute orbit parameters. The orbital radius is the most obvious of these since it is directly visible in the equations (as \( R_c, a_c, \) or \( n_c \)), but all other absolute orbital parameters such as, e.g., inclination and eccentricity, are implicitly present in the equations through the variables \( x, y, \) and \( z \). This is to be expected since the motion of both satellites is purely driven by the gravitational field of the primary body. Thus, the choice for the absolute orbits of the satellites fully defines the relative motion of the satellites (when the satellites are not maneuvering by their own means using, e.g., a propulsion system).

3.2.2 Linearized Relative Motion Around a Common Primary

If the orbit of the deputy in inertial space is only slightly elliptic and inclined with respect to a circular chief’s orbit and if the distance between the chief and the deputy is small, Eqs. (3.13a–3.13c) may be linearized about the origin of the Hill frame. These linearized equations of relative motion are commonly known as the Clohessy-Wiltshire (CW) or Hill-Clohessy-Wiltshire (HCW) equations. The HCW equations can be obtained by expanding the right-hand side of Eqs. (3.13a–3.13c) into a Taylor series about the origin. Retaining first-order terms only, the fractions in the right-hand sides of Eqs. (3.13a–3.13c) then reduce to

\[ \frac{\mu (a_c + x)}{[(a_c + x)^2 + y^2 + z^2]^{\frac{3}{2}}} \approx n_c^2 (2x - a_c) \]  
\[ \frac{\mu y}{[(a_c + x)^2 + y^2 + z^2]^{\frac{3}{2}}} \approx -n_c^2 y \]  
\[ \frac{\mu z}{[(a_c + x)^2 + y^2 + z^2]^{\frac{3}{2}}} \approx -n_c^2 z. \]  

(3.14a)  
(3.14b)  
(3.14c)

Inserting Eqs. (3.14a–3.14c) into Eqs. (3.13a–3.13c), rearranging terms and omitting the subscript "c" leads to
which are the well-known HCW equations. These equations show that, to first order, the out-of-plane motion is a harmonic oscillation that is decoupled from the in-plane motion. Note that the HCW equations, just as the nonlinear equations of subsection 3.2.1 are a function of the absolute orbital parameters of the satellites.

In state-space form, choosing the state vector \( \mathbf{x} = (x, y, z, \dot{x}, \dot{y}, \dot{z})^T \), the HCW equations can be written as

\[
\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \tag{3.16}
\]

with \( \mathbf{A} \) the system matrix, given by

\[
\mathbf{A} = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
3n^2 & 0 & 0 & 0 & 2n & 0 \\
0 & 0 & 0 & -2n & 0 & 0 \\
0 & 0 & -n^2 & 0 & 0 & 0
\end{bmatrix}. \tag{3.17}
\]

The HCW equations can be solved in a straightforward manner using the matrix exponent, leading to

\[
\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) = \Phi(t, t_0)\mathbf{x}(t_0) \tag{3.18}
\]

with \( \Phi(t, t_0) \) the state transition matrix which maps the state at time \( t_0 \) to the state at time \( t \). For \( t_0 = 0 \), the state transition matrix is given by

\[
\Phi(t) = \begin{bmatrix}
4 - 3c_{nt} & 0 & 0 & \frac{s_{nt}}{n} & \frac{2(1 - c_{nt})}{n} & 0 \\
6(s_{nt} - nt) & 1 & 0 & \frac{2(c_{nt} - 1)}{n} & \frac{s_{nt}}{n} & 0 \\
0 & 0 & c_{nt} & 0 & 0 & \frac{s_{nt}}{n} \\
3ns_{nt} & 0 & 0 & c_{nt} & 2s_{nt} & 0 \\
6n(c_{nt} - 1) & 0 & 0 & -2s_{nt} & 4c_{nt} - 3 & 0 \\
0 & 0 & -ns_{nt} & 0 & 0 & c_{nt}
\end{bmatrix}. \tag{3.19}
\]

with \( s_{nt} \) and \( c_{nt} \) denoting \( \sin(nt) \) and \( \cos(nt) \), respectively. In the following, for brevity, the notation \( \{ \cdot \}_k \) is replaced by \( \{ \cdot \}_k \). The second row of the state transition matrix shows that the relative along-track motion is unstable due to the existence of an along-track drift \( y_{\text{drift}}(t) = -(6x_0 + 3\dot{y}_0/n)(nt) \) which causes a linear growth in along-track separation un-
less the following condition is met

\[ \dot{y}_0 = -2nx_0. \] (3.20)

Satisfying this condition yields, up to first order, a relative motion with a period of one orbital revolution. The state transition matrix also shows that the in-plane motion oscillates with a magnitude in along-track direction that is twice the magnitude in radial direction. The in-plane motion can also have a constant offset from the origin of the Hill frame. In radial direction, this constant offset is equal to the differential semi-major axis \( \Delta a \), again up to first order, of the chief and the deputy, thus

\[ a_d - a_c = \Delta a = 4x_0 + 2\frac{\dot{y}_0}{n}. \] (3.21)

In fact, it is this differential semi-major axis that causes along-track drift since a differential semi-major axis implies a differential orbital period, resulting in an along-track drift. For constant \( \Delta a \), the along-track drift per orbit is easily derived. After one orbital period \( P = \frac{2\pi}{n} \), the along-track drift is

\[ y_{\text{drift}}(P) = -(6x_0 + 3\frac{\dot{y}_0}{n})(nP) = -\frac{3}{2} \Delta a(nP) = -3\pi \Delta a. \] (3.22)

Note that if perturbations due to differential atmospheric drag are taken into account, \( \Delta a \) grows linearly with time (in case of constant differential drag), resulting in a quadratic growth of the along-track separation.

Much research has been performed in the development of linearized equations of relative motion for elliptic chief orbits. These are more involved than the HCW equations and are not treated here, but the interested reader is referred to [Alfriend et al. 2010], which describes several theories developed by a.o. Melton, Lawden, Tschauner and Hempel, Carter, and Yamanaka and Ankersen.

Lastly, the following deals briefly with an important link between the relative dynamics in case of a single primary body and the relative dynamics in case of two primary bodies: When the main gravitational forces acting on a body are caused not by one, but by two primaries rotating in a circular Keplerian motion around their barycenter, a special type of relative motion, called the Circular Restricted Three Body Problem (CRTBP), is created, cf. [Roy 2005]. For the CRTBP, there exist geometrical points, known as Lagrange or libration points, where the motion of the body appears to be stationary in a reference frame co-rotating with the two primaries. For such systems there exist five libration points, \( L_i \) \((i = 1, 2, 3, 4, 5)\), of which three are unstable \((L_1 - L_3\), located on the line connecting the two primaries) and two \((L_4 \text{ and } L_5\), forming an equilateral triangle with the two primaries) are stable. For the Sun-Earth/Moon system these points are interesting locations for astrophysical observatories since they provide a thermally stable, low-disturbance environment. Points \( L_1 \) and \( L_2 \), although unstable, are generally preferred over the other three points since they are relatively close to Earth. The instability of the points is mitigated by placing the observatory in a so-called halo orbit around the libration point (the name halo
orbit comes from the fact that the orbit as seen from the biggest primary looks like a ring around the smallest one). This halo orbit is periodically stable with respect to the libration point.

In [Ardaens and D’Amico 2008], a derivation of the linearized equations of relative motion for two spacecraft in the vicinity of a libration point is performed. The structure of these equations is remarkably similar to that of the HCW equations (in fact, when one of the two primaries is given zero mass, the equations reduce exactly to the HCW equations). For instance, the relative out-of-plane motion of the satellites is, just as for the HCW equations, decoupled from the relative in-plane motion. Thus, to some extent, the conclusions that will be drawn in this thesis regarding the magnitude of the relative out-of-plane motion of two formation flying satellites will also be applicable for formations in halo orbits.

3.3 Observations

To allow estimation of the relative state of two spacecraft, observations of quantities containing information on the relative state need to be fed to the estimator. This section describes how these observations, or measurements, are typically obtained and which error sources need to be accounted for in these observations.

3.3.1 Ranging Principles

When using RF signals to directly estimate the distance between two points, two basic measurement types are possible, namely time-of-flight (requires knowledge of the time \(t_T\) 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). Eqs. (3.23) to (3.25) provide the basic relationships for time-of-flight measurement, phase measurement and Doppler measurement respectively [Fehse 2003]

\[
\begin{align*}
  r_T &= c \tau = c \Delta t = c(t_{RX} - t_{TX}) \\
  r_\phi &= \frac{\lambda}{2\pi} \Delta \phi = \frac{\lambda}{2\pi} (\varphi_{RX} - \varphi_{TX}) = \frac{c}{2\pi f} (\varphi_{RX} - \varphi_{TX}) \\
  \dot{r} &= -c \frac{\Delta f}{f_{TX}} = c \frac{f_{TX} - f_{RX}}{f_{TX}}.
\end{align*}
\]

In Eqs. (3.23–3.25), \(c\) denotes the speed of light, \(\tau\) denotes the signal travel time, \(\lambda\) denotes carrier wavelength, and \(\phi\) denotes carrier phase. 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 Eqs. (3.23–3.25) 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 Eqs. (3.23–3.25) as motivated in section 3.1. Phase measurements can be used to construct highly accurate range and angular measurements, which are treated in subsection 3.3.2.

For all ranging methods, the measurements are affected by statistical (noise) and systematic errors. The most important error sources for space-based relative navigation are:

- **Multipath**: This phenomenon is caused by the reception of reflected, delayed, replicas of the desired signal and has already been discussed in section 2.2.3.

- **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, if not tolerable, must be calibrated for. However, as this bias is not constant over time due to, e.g., temperature differences, radiation effects, 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.

- **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. It can cause a bias which needs to be estimated. 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 (partially) mitigate this effect in case of a single frequency ranging system. If measurements from different receiver antennas on the same satellite are differenced, this error cancels.

- **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 [Kim et al. 2006]. 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 below the noise level 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 [Lestarquit et al. 2006].

- **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 [Balanis 2005], [Grewal et al. 2007]. Uncertainties in the location of the antenna with respect to the satellite COM, also due to uncertainties in the location of the COM itself (e.g., due to propellant sloshing, propellant tank depletion), 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 topic is further elaborated on in subsection 3.3.2.

- **System noise temperature**: This parameter consists out of antenna noise and receiver noise and should be minimized to maximize the signal-to-noise ratio. Antenna noise depends mainly on noise radiated by sources in the FOV of the antenna and will be lowest when the antenna beam is directed towards deep space. Receiver noise is a function of thermal noise (which results from random variations in current or voltage caused by the random movement of electrons due to thermal energy) and the receiver front end circuit design. A low noise front end design should utilize a high gain LNA directly behind the antenna feed as this reduces the noise contribution of subsequent stages in the circuit. Also, as in general the noise figure of amplifiers (i.e. the noise added to the signal by the amplifier) increases with in-
creasing signal frequency, a low signal frequency generally leads to a lower system noise temperature.

- **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).

### 3.3.2 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 of code and carrier 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 ($1\sigma$) [Montenbruck and Gill 2000] for code and carrier phase measurements, leading to measurement accuracies of $\sim 1.6$ m for code wavelengths of 100 m and $\sim 2.2$ 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 subsection 3.3.3. For now, it is assumed that this ambiguity is 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, which is typically done by means of dual one-way ranging. Following [Kim and Lee 2009], dual one-way ranging between platforms A and B results in the following fractional phase measurements

\[
\begin{align*}
\varphi^B_A (t_1) &= [\varphi_A (t_1) + \delta \varphi_A (t_1)] - [\varphi_B (t_1 - \tau_1) + \delta \varphi_B (t_1 - \tau_1)] + E_A \\
\varphi^A_B (t_2) &= [\varphi_B (t_2) + \delta \varphi_B (t_2)] - [\varphi_A (t_2 - \tau_2) + \delta \varphi_A (t_2 - \tau_2)] + E_B
\end{align*}
\]

where $\varphi^B_A (t_1)$ and $\varphi^A_B (t_2)$ represent the phase measurements at times $t_1$ and $t_2$ at receivers A and B respectively due to signals transmitted at times $t_1 - \tau_1$ and $t_2 - \tau_2$. 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 $\delta \varphi(t)$ and a lumped error term $E$ including ionospheric delay, signal multipath, and hardware induced noise. The signal travel time is denoted as $\tau$. Addition and subtraction of Eqs. (3.26) provides estimates of the phase difference $\Delta \varphi$ and clock bias $\Delta \delta \varphi$ as follows
\[
\Delta \phi = \frac{\phi^B_A(t_1) + \phi^A_B(t_2)}{2} = \frac{\Delta \phi_1 + \Delta \phi_2}{2} + \frac{\Delta \delta \phi_1 - \Delta \delta \phi_2}{2} + \frac{E}{2} \tag{3.27}
\]
\[
\Delta \delta \phi = \frac{\phi^B_A(t_1) - \phi^A_B(t_2)}{2} = \frac{\Delta \delta \phi_1 + \Delta \delta \phi_2}{2} + \frac{\Delta E}{2} \tag{3.28}
\]
\[
\Delta \phi_1 = \phi_A(t_1) - \phi_B(t_1 - \tau_1) \quad \Delta \phi_2 = \phi_B(t_2) - \phi_A(t_2 - \tau_2)
\]
\[
\Delta \delta \phi_1 = \delta \phi_A(t_1) - \delta \phi_B(t_1 - \tau_1) \quad \Delta \delta \phi_2 = \delta \phi_A(t_2 - \tau_2) - \delta \phi_B(t_2)
\]
\[
E = E_A + E_B \quad \Delta E = E_A - E_B.
\]

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., \( \delta \phi(t - \tau) = \delta \phi(t) \)), which are reasonable assumptions for many spacecraft formations, the above reduces to

\[
\hat{\Delta} \phi = \frac{\phi^B_A(t) + \phi^A_B(t)}{2} = \Delta \phi + \frac{E}{2} \tag{3.29}
\]
\[
\hat{\Delta} \delta \phi = \frac{\phi^B_A(t) - \phi^A_B(t)}{2} = \Delta \delta \phi + \frac{\Delta E}{2}. \tag{3.30}
\]

Note that in the above, no distinction has been made 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 actual pseudorange \( \rho \) between the antennas consists out of the fractional phase as determined using Eq. (3.29) plus \( 2\pi \) times an unknown integer number of signal cycles \( K \). For ambiguous code-based range measurements, \( K \) 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 \( \lambda \) of the signal and dividing by \( 2\pi \) to obtain an estimate for the pseudorange \( \hat{\rho} \):

\[
\hat{\rho} = \frac{\lambda_{\text{code}}}{2\pi} \hat{\Delta} \phi + \lambda_{\text{code}} K. \tag{3.31}
\]

Similarly, the clock bias \( \Delta t \) is estimated as

\[
\hat{\Delta} t = \frac{\lambda_{\text{carrier}}}{2\pi c} \hat{\Delta} \delta \phi. \tag{3.32}
\]

Thus, in practice, the clock bias can be estimated with very little error due to cancellation of common errors (if these are of similar magnitude for the two platforms), but this is not the case for the range estimate. Therefore, to increase the performance of future
relative navigation systems, emphasis should be placed primarily on the range estimate. Mitigation of ranging errors due to clock bias and drift if $t_1 \neq t_2$ is beyond the scope of this work, but a possible method to achieve this is detailed in [Winter et al. 2003].

### 3.3.3 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 on the variation of the gain of the Tx and Rx antennas with the angle from the antenna boresight. The resulting LOS estimate typically has an accuracy of tens of degrees.

However, 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. 3.2 where the ranging signal of a faraway transmitter (dotted light grey arrows) arrives at receiver antennas Rx$_1$ and Rx$_2$. When the path difference $b$ (grey line) and the antenna baseline length $d$ (black arrow) are known, the LOS angle $\psi$ is computed as

$$\psi = \sin^{-1}\left(\frac{b}{d}\right).\quad (3.33)$$

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.

![Figure 3.2: Definition of the Line-Of-Sight angle $\psi$.](image)

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 $b$ 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 angle and range. A long wavelength for the carrier signal can potentially remove the ambiguity problem, but then the phase measurement is relatively inaccurate, cf. subsection 3.3.2, leading to a poor LOS estimate. Methods to solve the integer ambiguity problem are treated in
Next to the carrier wavelength, the accuracy of the estimate of the LOS angle also depends on the antenna baseline \( d \). For a certain LOS angle \( \psi \), increasing \( d \) will lead to a proportional increase in \( b \). However, the accuracy in the measurement of \( b \) will remain the same. This results in an improvement of the accuracy in the estimate for \( \psi \) since the phase measurement error will have a smaller contribution to the total estimate. For instance, when the measurement error is denoted as \( e \), the antenna baseline \( d = d_1 \), and the path difference \( b = b_1 \), the estimate of \( \psi \) can be written as

\[
\hat{\psi}_{\mid d_1 = d} = \sin^{-1}\left( \frac{b_1 + e}{d_1} \right) = \sin^{-1}\left( \frac{b + e}{d} \right) = \sin^{-1}\left( \frac{b + e}{d} \right).
\]  

(3.34)

For the same angle \( \psi \), changing the antenna baseline to \( d_2 = 2d \) implies a path difference of \( b_2 = 2b \) but the same measurement error \( e \), leading to

\[
\hat{\psi}_{\mid d_2 = 2d} = \sin^{-1}\left( \frac{b_2 + e}{d_2} \right) = \sin^{-1}\left( \frac{2b + e}{2d} \right) = \sin^{-1}\left( \frac{b + e}{2d} \right).
\]  

(3.35)

Thus, determination of the LOS angle is more accurate when a large antenna baseline is used. Compared to physically large satellites, this puts physically small satellites at a disadvantage since achieving the same accuracy in the LOS angle estimate with small satellites as for large satellites therefore requires a shorter carrier wavelength and/or a higher accuracy in the carrier phase measurement. Both these options are very undesirable from a systems engineering perspective as they will inevitably lead to a more costly design: Using a shorter wavelength implies a higher frequency, leading to increased free space loss and thus, possibly, the need to use a higher signal transmission power and/or antennas with higher gain and/or a receiver/signal design that allows the detection of lower power signals. Increasing the accuracy in the carrier phase measurement will require more complex electronics, making them larger and consuming more power.

Note that, for a given LOS angle measurement error \( \delta \psi \) and inter-satellite range \( \rho \), the error in the estimation of the position \( \delta r \) of the other satellite scales linearly with the inter-satellite distance since

\[
\delta r = \rho \cdot \delta \psi.
\]  

(3.36)

Thus, for a given LOS angle measurement error, a small inter-satellite distance will yield a smaller position estimation error.

### 3.4 Relative State Estimation

This section introduces the methods used to answer the research questions. The main tools used are a batch least-squares estimator and an observability analysis, which are discussed in subsections 3.4.2 and 3.4.3. To obtain the most accurate estimation results, the inter-satellite range measurements should be as accurate as possible. This requires measurement of the carrier phase, which brings with it the problem of estimating the in-
integer ambiguity in the carrier phase measurement. This is briefly dealt with in the first subsection. However, as the research is not primarily concerned with how the measurements are obtained, it is assumed in the following that it is possible to resolve the ambiguity and therefore the discussion on the integer ambiguity resolution is kept on a relatively high level. The fourth subsection deals with the cases of estimator divergence that were encountered during the research. Especially the so-called apparent estimator divergence [Perea et al. 2007] has strong ties to research questions two and three. The last subsection presents a derivation of the expected value and standard deviation in the estimation of the position of a stationary transmitter. This derivation serves as a theoretical foundation for the results obtained in numerical simulations, to be discussed in the next chapter.

### 3.4.1 Integer Ambiguity Resolution

As already discussed, obtaining the most accurate estimation of the inter-satellite range, LOS angle 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 angle of the remote transmitter) and instrumental offsets, including an integer-cycle ambiguity. As long as the geometry remains fixed in time, the two contributions are indistinguishable and the LOS angle 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. Note that considerable research has been performed on this topic, cf. [Teunissen 1995, Sun et al. 2013].

The first IAR method uses multiple signals with different frequencies, which allow to construct the 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, $\lambda_{WL}$, 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, e.g., ranging signals S1 and S2 with frequencies $f_{S1} = 2.25\text{GHz}$ ($S_1 = 13.3\text{ cm}$) and $f_{S2} = 2.10\text{GHz}$ ($S_2 = 14.3\text{ cm}$) and with phase measurements $\phi_{S1}$ and $\phi_{S2}$ respectively, the resulting wide-lane wavelength $\lambda_{WL}$ and the wide-lane phase measurement $\phi_{WL}$ are

$$\phi_{WL} = w_1\phi_{S1} - w_2\phi_{S2}$$

$$\lambda_{WL} = \frac{c}{w_1f_{S1} - w_2f_{S2}}$$
with the weights $w_1, w_2 \in \mathbb{N}$. Selecting $w_1 = w_2 = 1$ results in $\lambda_{WL} = 2 \text{m} \gg \lambda_{S1}, \lambda_{S2}$. 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 $\sigma_{WL}$ scales as $\sigma_{WL} = \left( (w_1 \sigma_{S1})^2 + (w_2 \sigma_{S2})^2 + \cdots + (w_N \sigma_{SN})^2 \right)^{1/2}$, the magnitude of $w$ should generally be kept small [Jung 2000].

By now choosing the frequencies of the ranging signals in such a way that the wavelength of the different wide-lane combinations gradually reduces ($\lambda_{WL1} > \lambda_{WL2} > \cdots > \lambda_{S}$), the integer ambiguity on the carrier phase of ranging signal $S$ can be resolved in a cascaded approach. For example, assuming the carrier phase integer ambiguity on a distance measurement is to be removed, 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, $\sigma_{\rho,\text{smoothed}}$, is better than half the wavelength of the wide-lane signal with the longest wavelength, i.e. $\sigma_{\rho,\text{smoothed}} < 1/2 \lambda_{WL1}$, 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 ($\sigma_{WL1,\text{smoothed}} < 1/2 \lambda_{WL2}$). 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 RNS, a proper balance needs to be found between the magnitudes of the various signal wavelengths and their ratios (e.g., $\lambda_{WL1}/\lambda_{WL2}$). For instance, in case of a system with a single wide-lane, a relatively large magnitude for $\lambda_{WL}$ results in a higher confidence for the IAR performed on the wide-lane, but results in a large ratio $\lambda_{WL}/\lambda_{S}$, 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, discussed in section 2.2.1, 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 specially designed "ultra-BOC" signal, cf. Fig. 3.3. 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 estimated using code measurements (for the AFF, the code-based ranging accuracy is 0.5 m). Then, the closely spaced tones on one side of the carrier frequency are used to form a wide-lane (WL1) with a wavelength that is much larger than the accuracy of the code measurements (for the AFF, WL1 = 7.5 m). The code measurements can be used to solve the ambiguity on this wide-lane. Now, a second wide-lane (WL2) is formed using the two outermost tones (for the AFF, WL2 = 1 m). 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 [Tien et al. 2004].
The second IAR method makes use of the fact that a precisely known change in the relative geometry results 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, operations effort, 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 spacecraft, but not for formations consisting out of tens of spacecraft. 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. This is however not always required since the FFRF antenna baseline on the PRISMA mission is so small that the wide-lane LOS is already unambiguous. However, in the PRISMA mission there are situations where large multipath biases exist (which occurs when the spacecraft surfaces on which the antennas are mounted are inclined at a large angle with respect to each other) and no valid direct-aiding data is available from the spacecraft GNC system. Then, this strategy does not always result in high enough confidence levels to assure a correct IAR. In that case, it initiates the IAR procedure using a 50° satellite rotation to resolve the ambiguity on the LOS. However, test campaigns performed during that mission have shown that even then IAR can fail for the baselined rotation angle. In such cases, it has been found that a larger rotation angle of e.g. 70° does lead to successful IAR and thus appears to make the motion-aided IAR process more robust [Grellier et al. 2011].
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 these, signal multipath (especially for large LOS angles) and temperature-induced electrical biases (due to temperature effects on components not covered by the internal calibration loop of the instrument) have been found to be the most performance limiting ones for the FFRF sensor [Grelier et al. 2011]. Therefore, future development of RF-based relative navigation sensors should focus on removing these biases as much as possible.

3.4.2 Estimation Approach

In this work, the relative state of two formation flying spacecraft is estimated by means of a dynamic approach: Observations (measurements) that provide information on the system state are obtained from one or more sensors and are fed to an estimator which uses these observations and models of the system dynamics and sensors to compute an estimate of the system state.

Note that kinematic, i.e., without using any dynamic model, and reduced dynamic, i.e., a combination of kinematic and dynamic, estimation approaches are ruled out for the problem under consideration. A purely kinematic approach is not feasible because the measurements obtained at a single epoch are not sufficient to estimate the entire system state and because a priori information is to be used in the estimation process to overcome observability problems for certain formation geometries. The reduced dynamic approach is used to reduce the estimator's susceptibility to dynamic modeling errors (hence the term "reduced dynamic"). It does this, e.g., by means of empirical acceleration parameters, the use of which results in a trajectory that lies in between the trajectories that would be obtained when using purely kinematic or purely dynamic estimators. However, in this research, the dynamic model used in the estimator is a perfect representation of the 'true' system dynamics (the reason for this will be explained in subsection 3.4.4). Therefore, there are no dynamic modeling errors and thus also no need to step away from a purely dynamic estimator.

In general, the system dynamics and the observations are nonlinear functions of the system states. The general state space model for a nonlinear system can be posed as the following set of ordinary differential equations

\[
\dot{x}(t) = f(x_0, u, w, t) \tag{3.39}
\]

\[
z(t) = h(x_0, u, \epsilon, t). \tag{3.40}
\]

where the system equation \( f(x_0, u, w, t) \) is a nonlinear function of the system states \( x_0 \) at time \( t = 0 \), the control inputs \( u \), and a random forcing function known as process noise \( w \), and where the measurements at different time instances are collected in the vector \( z(t) \) which is modeled by the measurement function \( h(x_0, u, \epsilon, t) \) where \( \epsilon \) is a measurement noise vector. If a control law \( g(x_0, t) \) is applied, then the control inputs satisfy \( u = g(x_0, t) \).
Eqs. (3.39–3.40) will be simplified based on the following three assumptions:

1. There are no control inputs, thus \( u = 0 \).
2. The dynamics of the system are known perfectly, thus \( w = 0 \).
3. The dynamics of the system are described by the HCW equations.

The first assumption is valid since no control forces will be applied in the scenarios to be studied in the next chapter. The reason for this is that the application of control forces implies a change in the state of the system. This is undesired since the navigation results will then depend on the timing, magnitude, and direction of the control forces, which results in a less general scenario. Furthermore, control will add control errors which can obscure the observability problem. The second assumption does not fully reflect reality, since the system dynamics cannot be modeled with absolute accuracy, but is applied to simplify the navigation algorithm, which is justified by the reasoning provided in section 3.1. The third assumption does limit the cases that can be studied since the influence of nonlinearities grows for increasing inter-satellite distance and increasing ranging accuracy. Therefore, investigating formations with inter-satellite distances of many kilometers and very high ranging accuracies under the assumption of linear dynamics will lead to misleading results and can therefore not be performed. Fortunately, the inter-satellite distance for many spacecraft formations is below 10 km, cf. Fig. 1.6, and the ranging accuracy for an RF-based RNS is typically not required to be better than cm-level. Thus, this assumption still allows to study cases that are relevant for many spacecraft formations. These simplifications result in the following system model

\[
\dot{x}(t) = Ax(t) \quad (3.41)
\]
\[
z(t) = h(x_0, t) + \varepsilon \quad (3.42)
\]

where the system matrix \( A \) is given by Eq. (3.17), \( x_0 = x(t_0) \), and where the statistics for the measurement noise \( \varepsilon \) are \( \mathbb{E}[\varepsilon] = 0 \) and \( \mathbb{E}[\varepsilon \varepsilon^T] = W^{-1} = \sigma^2 \rho I \), with \( I \) being the identity matrix and \( W \) a weighting matrix. Because \( W \) is diagonal, it is implicitly assumed that there are no cross-correlations between measurements. In practice, some amount of cross-correlation will be present, due to usage of the same receiver for all Rx antennas and due to the fixed antenna baselines. However, some of this correlation will be mitigated by factors such as signal multipath, differential temperature, antenna phase center variations, different cable lengths, and component aging. Even careful design and extensive calibration cannot fully remove these effects. In addition, assuming some amount of cross-correlation implies adding another variable and makes the results sensitive to these assumptions. As cross-correlation is not considered to be a key variable for this study, the foregoing considerations have led to the assumption of zero cross-correlation.

In this work, the objective is to estimate the position and velocity of the deputy relative to the Hill frame, which is collocated with the COM of the chief. Thus, the system state vector \( x \in \mathbb{R}^6 \) is \( x(t) = (r(t), v(t))^T \) where \( r(t) = (x(t), y(t), z(t))^T \) is the relative position
vector and where $\mathbf{v}(t) = (\dot{x}(t), \dot{y}(t), \dot{z}(t))^T$ is the relative velocity vector. This choice for the state vector implies that the mean motion $n$ and with that, the semi-major axis $a$ of the absolute orbit of the chief, must be assumed to be known perfectly since it is an important parameter in the relative dynamics model. This is a simplification of reality since $n$ is never known perfectly. However, if desired, $n$ can be estimated with high accuracy from the inter-satellite range measurements as these measurements are a function of $x(t)$, $y(t)$, and $z(t)$, which in turn are functions of $n$. The reason that $n$ can be estimated with high accuracy lies in the fact that inter-satellite measurements are used to estimate the system state. Due to this, changes in relative radial position and relative along-track velocity will be highly correlated in the estimator (correlation close to -1). This correlation is vital to obtain an accurate estimate for $n$. It also leads to a good 'balancing' between the radial position and along-track velocity estimates (there should be a factor 2$n$ between these parameters). Since the correlation and balancing conditions are both satisfied, the estimate for $n$ (and with that, the semi-major axis $a$) will be very accurate, cf. [Maessen and Gill 2010a, Carpenter and Alfriend 2003, How et al. 2004]. For instance, the error in the estimation of $n$ is easily better than 0.001% and can be as low as 1E-6%, which leads to an error in the estimation of the semi-major axis at cm-level for a semi-major axis of 7028 km (for a pendulum formation with 1 km inter-satellite distance, a ranging accuracy of 1 cm, an antenna baseline of 12 dm, and an initial error in the estimate of the semi-major axis of 100 km). Therefore, including $n$ in the state estimate would increase the time needed for the estimator to converge and would not contribute significantly to answering the research questions.

When estimating the system state, the goal is to find an estimate for the system state $\hat{x}_0$ that minimizes the following weighted cost function $J$

$$J(x_0) = (z - h(x_0))^T W (z - h(x_0))$$

(3.43)

where $(z - h(x_0))$ is a vector of measurement residuals.

The vector function $h(x_0)$ consists out of scalar functions $h_i(x_0, t_i, \tau_i)$ that model the range between a Tx antenna on the deputy, assumed to be a point mass, and a Rx antenna on the chief at time $t_i$ as the Euclidean vector norm of the relative position vector $r_i$ at time $t_i$ due to signals transmitted at time $t_i - \tau_{i,j}$. For $j$ Rx antennas, the variable $\tau_{i,j}$ denotes the signal travel time from the Tx antenna to antenna Rx$_j$ for a signal that is received at time $t_i$. This leads to
This yields a maximum range estimation error of \( \dot{\tau} \) at time \( t \) and power limitations. For static situations, this is not sufficient to estimate the position three Rx antennas for the relative navigation between two satellites. In addition, three Rx the expectation that future formations with small satellites will only rarely use more than \( j \) thus be neglected. Finally, the number of Rx antennas is limited to be either one or three, thus the smallest range measurement standard deviation considered, which is 1 cm, and can \( \tau \) signal travel time introducing significant errors. This simplification is justified by performing a first-order Taylor series expansion around the signal reception time \( t_i \) for the inter-satellite range \( r \) at time \( t_i - \tau_{i,j} \) as \( r \left( t_i - \tau_{i,j} \right) = r(t_i) - \dot{r}(t_i) \tau_{i,j} + O \left( \tau_{i,j}^2 \right) \). To first order, the range estimation error in this approximation is equal to \( \dot{r}(t_i) \tau_{i,j} \). The formation geometries that have been simulated for this research lead to a maximum inter-satellite distance of just under 23 km and a maximum relative velocity of 24 m/s. Dividing the maximum inter-satellite distance by the speed of light \( (3 \times 10^8 \text{ m/s}) \) leads to a maximum signal travel time of \( \tau_{i,j} \approx 8 \times 10^{-5} \text{ s} \). This yields a maximum range estimation error of \( \dot{\tau} \approx 2 \text{ mm} \), which is much smaller than the smallest range measurement standard deviation considered, which is 1 cm, and can thus be neglected. Finally, the number of Rx antennas is limited to be either one or three, thus \( j \in \{1, 3\} \). The choice for studying only systems with either one or three Rx antennas is based on the argument that this is sufficient to answer the research questions but also on the expectation that future formations with small satellites will only rarely use more than three Rx antennas for the relative navigation between two satellites. In addition, three Rx antennas are usually sufficient to estimate the position of another object. For very small satellites, it is not unlikely that only one Rx antenna can be accommodated due to mass and power limitations. For static situations, this is not sufficient to estimate the position

\[
\begin{align*}
\mathbf{h}(x_0, t_i) &= \\
&= 
\begin{pmatrix}
(x(t_1 - \tau_{1,1}) - x_{Rx_1}(t_1), y(t_1 - \tau_{1,1}) - y_{Rx_1}(t_1), z(t_1 - \tau_{1,1}) - z_{Rx_1}(t_1))^T \\
(x(t_1 - \tau_{1,2}) - x_{Rx_2}(t_1), y(t_1 - \tau_{1,2}) - y_{Rx_2}(t_1), z(t_1 - \tau_{1,2}) - z_{Rx_2}(t_1))^T \\
\vdots \\
(x(t_1 - \tau_{1,j}) - x_{Rx_j}(t_1), y(t_1 - \tau_{1,j}) - y_{Rx_j}(t_1), z(t_1 - \tau_{1,j}) - z_{Rx_j}(t_1))^T \\
(x(t_2 - \tau_{2,1}) - x_{Rx_1}(t_2), y(t_2 - \tau_{2,1}) - y_{Rx_1}(t_2), z(t_2 - \tau_{2,1}) - z_{Rx_1}(t_2))^T \\
(x(t_2 - \tau_{2,2}) - x_{Rx_2}(t_2), y(t_2 - \tau_{2,2}) - y_{Rx_2}(t_2), z(t_2 - \tau_{2,2}) - z_{Rx_2}(t_2))^T \\
\vdots \\
(x(t_2 - \tau_{2,j}) - x_{Rx_j}(t_2), y(t_2 - \tau_{2,j}) - y_{Rx_j}(t_2), z(t_2 - \tau_{2,j}) - z_{Rx_j}(t_2))^T \\
\vdots \\
(x(t_i - \tau_{i,1}) - x_{Rx_1}(t_i), y(t_i - \tau_{i,1}) - y_{Rx_1}(t_i), z(t_i - \tau_{i,1}) - z_{Rx_1}(t_i))^T \\
(x(t_i - \tau_{i,2}) - x_{Rx_2}(t_i), y(t_i - \tau_{i,2}) - y_{Rx_2}(t_i), z(t_i - \tau_{i,2}) - z_{Rx_2}(t_i))^T \\
\vdots \\
(x(t_i - \tau_{i,j}) - x_{Rx_j}(t_i), y(t_i - \tau_{i,j}) - y_{Rx_j}(t_i), z(t_i - \tau_{i,j}) - z_{Rx_j}(t_i))^T
\end{pmatrix}
\end{align*}
\]

(3.44)
of another object, but for dynamic situations this is not necessarily the case. It will be interesting to determine if a single Rx antenna can indeed be sufficient for certain scenarios. This results in the following measurement vector function for a single measurement epoch $i$ and three Rx antennas

$$
h(x_0, t) = \left[ \begin{array}{c} \| (x(t) - x_{Rx1}, y(t) - y_{Rx1}, z(t) - z_{Rx1})^T \| \\
\| (x(t) - x_{Rx2}, y(t) - y_{Rx2}, z(t) - z_{Rx2})^T \| \\
\| (x(t) - x_{Rx3}, y(t) - y_{Rx3}, z(t) - z_{Rx3})^T \| \end{array} \right]. \tag{3.45}
$$

It is important to realize that determination of the location of the antennas in the Hill frame, which are assumed to be known, is not trivial. To enable conversion of the measured ranges (and possibly angles) from the body frame to the Hill frame, the orientation of the spacecraft with respect to the Hill frame has to be known. Therefore, not only the spacecraft attitude, but also the orientation of the Hill frame has to be known onboard. The latter can only be known when the absolute orbit of the chief is known since the orientation of the Hill frame is a function of the absolute orbital parameters of the chief’s orbit, cf. Eq. (3.4). This information can be determined by the spacecraft themselves, but requires, in absence of GNSS measurements, a high-fidelity dynamic model and a specific (attitude) sensor suite [Psiaki 1995, Psiaki 1999a, Psiaki 1999b]. As a solution, such information can be provided by means of on ground processing of relevant measurements which are uploaded to the satellites.

In case of near real-time on board estimation of the relative state by means of inter-satellite range measurements, a sequential filter such as the Extended Kalman Filter (EKF) would typically be applied to obtain regular updates of the relative state using smoothed (carrier phase and) code measurements. For this work however, it has been chosen to estimate the relative state at time $t_0 = 0$ by means of a batch least-squares approach and to treat the measurements as zero mean Gaussian noise. An important consideration for the choice for a batch LSQ is that this allows a direct coupling between the estimation results and the observability analysis, to be described in subsection 3.4.3. The simple model for the measurements also removes the need to specifically model the range measurements. Downside is that this approach does not incorporate the successive improvement in ranging accuracy over time that is obtained by smoothing and IAR. Yet, by selecting different ranging accuracies for the cases to be studied, this process can be mimicked to some extent as each ranging accuracy can be regarded as a different step in this process.

Adoption of the batch LSQ approach requires linearization of the measurement equations around a reference state $x^{ref}$. This results in

$$
\Delta z(t) = H(t) \Delta x(t) + \epsilon \tag{3.46}
$$

where $\Delta z(t) = z(t) - h(x^{ref}, t)$, $\Delta x(t) = x(t) - x^{ref}(t)$, and where the measurement sensitivity matrix $H(t)$ contains the partial derivatives of the modeled observations with respect to the individual states evaluated at the reference state, thus $H = \partial h(x, t) / \partial x|_{x=x^{ref}}$. Since the objective is to estimate the relative state at time $t = 0$, Eq. (3.46) is rewritten as
3.4. Relative State Estimation

\[
\Delta z(t) = \begin{pmatrix} \Delta z_0 \\ \Delta z_1 \\ \vdots \\ \Delta z_{l-1} \end{pmatrix} = \begin{pmatrix} H_{t_0} \Phi(t_0, 0) \\ H_{t_1} \Phi(t_1, 0) \\ \vdots \\ H_{t_{l-1}} \Phi(t_{l-1}, 0) \end{pmatrix} \Delta x_0 + \epsilon = \begin{pmatrix} \tilde{H}_{t_0} \\ \tilde{H}_{t_1} \\ \vdots \\ \tilde{H}_{t_{l-1}} \end{pmatrix} \Delta x_0 + \epsilon = \tilde{H}(t) \Delta x_0 + \epsilon
\]

where the modified measurement sensitivity matrix \( \tilde{H}(t) \) now relates the partial derivatives to the relative state at \( t = 0 \) and where \( l \) indicates the total number of measurements.

In addition to state information contained in the measurements and supplemented by the dynamics model, a priori state information can be used. This information does comprise the initial state vector and its covariance information. In that way the estimator is penalized when it deviates too far from the a-priori estimate. This can be accomplished by minimizing the modified loss function \( \tilde{J} \) [Montenbruck and Gill 2000]

\[
\tilde{J} = (\Delta z - \tilde{H}x_0)^T W (\Delta z - \tilde{H}x_0) + (x_0 - x_0^{apr})^T \Lambda (x_0 - x_0^{apr})
\]

where \( x_0^{apr} \) is the a-priori relative state estimate and \( \Lambda \) is the information matrix, which is defined as the inverse of the a-priori covariance matrix \( P^{apr} \). In principle, \( P^{apr} \) is a full matrix that can have been obtained in various manners which depend on the mission characteristics. However, to simplify the case \( P^{apr} \) is assumed to be a diagonal matrix with the a-priori variances of the relative state estimates on the main diagonal. Minimizing the loss function defined in Eq. (3.48) leads to the following least-squares estimate of the relative state at \( t = 0 \)

\[
\hat{x}_0 = x_0^{ref} + P \left( \tilde{H}^T W \Delta z + \Lambda \Delta x_0^{apr} \right)
\]

where an estimate is denoted as \( (\cdot) \) and with the covariance matrix \( P \) equal to

\[
P = \left( \tilde{H}^T W \tilde{H} + \Lambda \right)^{-1}
\]

and \( x_0^{ref} \) is defined as

\[
x_0^{ref} = x_0^{apr} - \Delta x_0^{apr}.
\]

The inclusion of a sufficiently large non-singular information matrix \( \Lambda \) in Eq. (3.50), ensures that \( P^{-1} \) is non-singular and can thus be inverted even if \( \tilde{H}^T W \tilde{H} \) is singular. In the numerical simulations performed for this research, some formation geometries will indeed lead to \( \tilde{H}^T W \tilde{H} \) being (almost) singular. Thus, the information matrix not only serves as a penalizing effect, but also ensures that a solution can always be found. Lastly, as this is a nonlinear estimation problem, \( k \) iterations are needed to find a solution that has sufficiently converged, leading to
\[
\hat{x}^{k+1} = \hat{x}^k + P^k \left( \bar{H}^k W \Delta z^k + \Lambda \left( x_{0}^{\text{apr}} - \hat{x}^k_0 \right) \right).
\] (3.52)

To conclude this subsection, two relative state estimation considerations that have not been considered for this study, but which are very relevant for relative navigation, are briefly highlighted. Firstly, in case of dual one-way ranging, the measurements will be only available to the filter some time after the actual one-way measurements took place since the measurement data needs to be exchanged between the satellites. Thus, for real time on board implementation, this needs to be accounted for in the filter innovation: The system state must be propagated to the time of measurement, taking into account relative dynamics if high accuracy is needed. Secondly, 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 (with respect to the full covariance 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.

### 3.4.3 Observability Analysis

The observability of a dynamic system indicates in how far the state of that system can be estimated from knowledge of the inputs (i.e., control forces) and outputs (i.e., measurements). As control inputs are not considered here, the only information available is that provided by the measurements. Therefore, in this work, observability is defined as

**Definition 3** A system is said to be observable if and only if, for any unknown initial state there exists a finite time such that knowledge of the output over that finite time suffices to determine the initial state.

In other words, a useful measure of observability should provide a sense of the ability to obtain an estimate of the state vector \( x_0 \in \mathbb{R}^6 \) given the time history of the measurement vector \( z \in \mathbb{R}^l \). If an estimate can be obtained for a particular state \( x_0 \), then the system is observable for \( x_0 \). If an estimate can be obtained for all states \( x(t) \) corresponding to all measurement vectors \( z(t) \), then the system is said to be completely observable [Gelb 1974].

For purely linear systems, (dis-)proving observability, which is a global concept, is fairly straightforward since, due to the linearity of the system, the outcome of an observability
test for a single state is the same for all realizable states of that system. For nonlinear systems however, this is not the case. There, the concepts of local and weak observability need to be introduced. A thorough discussion of these concepts is provided in [Hermann and Krener 1977].

As the system under consideration is essentially nonlinear, the most prudent observability measure would be one that takes this nonlinearity into account. However, as the scenarios to be studied in the next chapter are such that nonlinearities are very small, a linear observability analysis is deemed to be sufficiently accurate. This is supported by a comparison between the relative motion resulting from linear and nonlinear dynamic models in subsection 4.4.1. This choice for the observability measure simplifies the analysis considerably since a nonlinear observability analysis requires the computation of Lie derivatives, which is much more involved than the computations required for a linear observability analysis, which are treated next. For the analysis it must be kept in mind however, that the actual system is nonlinear and thus proving linear observability for a particular state does not imply complete observability.

The measure of linear observability to be used can be derived from the definition of observability, which states that the initial state \( x_0 \) must be a unique solution to the measurements contained in the vector \( z \) in order for that state to be observable. This implies that Eq. 3.47, repeated here for easy reference, must have a unique solution.

\[
\Delta z(t) = \begin{pmatrix} \Delta z_0 \\ \Delta z_1 \\ \vdots \\ \Delta z_{l-1} \end{pmatrix} = \begin{pmatrix} H_{l_0} \Phi(t_0,0) \\ H_{t_1} \Phi(t_1,0) \\ \vdots \\ H_{t_{l-1}} \Phi(t_{l-1},0) \end{pmatrix} \Delta x_0 + \varepsilon = \begin{pmatrix} \tilde{H}_{l_0} \\ \tilde{H}_{t_1} \\ \vdots \\ \tilde{H}_{t_{l-1}} \end{pmatrix} \Delta x_0 + \varepsilon = \tilde{H}(t) \Delta x_0 + \varepsilon
\]  

(3.47)

For Eq. 3.47 to have a unique solution, the matrix \( \tilde{H} \) must be invertible and therefore be of rank 6 as the system under consideration is of sixth order. However, this observability criterion provides very little insight into the problem since it only indicates whether the system is observable or not and how many state components are observable (i.e., if the rank is four, only four state components are observable). It does not provide any information on which state components are more observable than others and how the observability of those state components changes over time. Fortunately, this lack of information is easily overcome by using the so-called Gramian \( G \) as the measure for observability. This matrix is defined for a discrete-time system as [Huxel 2006]

\[
G \triangleq \sum_{l=0}^{l-1} (H(t_l) \Phi(t_l,0))^T H(t_l) \Phi(t_l,0) = \tilde{H}^T \tilde{H}.
\]  

(3.53)

If \( G \) is nonsingular, i.e. full rank, the system is observable for that particular system state. Large eigenvalues for \( G \) result in good observability in the directions (eigenvectors) corresponding to those eigenvalues [Dullerud and Paganini 2000]. Furthermore, if one
eigenvalue is significantly larger than the others, this suggests that actually a linear combination of the state components in the eigenvector corresponding to that eigenvalue is observed [Ham and Brown 1983]. Thus, introducing and estimating an additional state component, defined as the linear combination of state components found, should result in a particular good estimate of that state component, as compared to the result for the original state components.

To allow a proper interpretation of its eigenvalues, \( \mathbf{G} \) needs to be normalized [Ham and Brown 1983]. Denoting the first three columns of \( \tilde{\mathbf{H}} \) as \( \tilde{\mathbf{H}}_{r} \) and the remaining three columns of \( \tilde{\mathbf{H}} \) as \( \tilde{\mathbf{H}}_{v} \), the following normalized Gramian \( \tilde{\mathbf{G}} \) is created

\[
\tilde{\mathbf{G}} = \mathbf{Ξ}^{T} \mathbf{Ξ} = \left( \begin{array}{cc} \tilde{\mathbf{H}}_{r} & n \tilde{\mathbf{H}}_{v} \end{array} \right)^{T} \left( \begin{array}{cc} \tilde{\mathbf{H}}_{r} & n \tilde{\mathbf{H}}_{v} \end{array} \right)
\]

(3.54)

where \( \mathbf{Ξ} \) is the normalized measurement sensitivity matrix.

The eigenvalues and eigenvectors of \( \tilde{\mathbf{G}} \) can be found by performing a singular value decomposition (SVD) of \( \mathbf{Ξ} \) such that

\[
\mathbf{Ξ} = \mathbf{U} \mathbf{Σ} \mathbf{Υ}^{T}
\]

(3.55)

where \( \mathbf{U} \) is an orthonormal matrix containing the left singular vectors of \( \mathbf{Ξ} \), \( \mathbf{Σ} = \text{diag}(\Sigma_{1}, \Sigma_{2}, \ldots, \Sigma_{6}) \) with singular values \( \Sigma_{1} > \Sigma_{2} > \ldots > \Sigma_{6} \), and \( \mathbf{Υ} = (\mathbf{v}_{1}, \mathbf{v}_{2}, \ldots, \mathbf{v}_{6}) \) is an orthonormal matrix containing the right singular vectors \( \mathbf{v}_{i} \) of \( \mathbf{Ξ} \), which are equivalent to the eigenvectors of \( \tilde{\mathbf{G}} \). The singular values in \( \mathbf{Σ} \) are the nonnegative square roots of the eigenvalues \( \Sigma_{i} \) of \( \tilde{\mathbf{G}} \). As column \( \mathbf{v}_{1} \) corresponds to the singular value \( \Sigma_{1} \), the eigenvector corresponding to the largest singular value is the direction with highest gain. In other words, eigenvector \( \mathbf{v}_{1} \) is the most observable 'direction' of the system and the largest value in that vector is the most observable state component.

Furthermore, a small value for the condition number \( \kappa \), defined as the ratio of the largest and smallest singular value of a matrix [Kailath et al. 2000], indicates a good accuracy in the estimate (well-conditioned) whereas a large value for \( \kappa \) indicates poor accuracy (ill-conditioned). As the matrix that will actually be inverted is \( \mathbf{G} \) and not \( \tilde{\mathbf{G}} \), \( \kappa \) needs to be determined for the non-normalized matrix \( \tilde{\mathbf{H}} \). Performing the SVD for \( \tilde{\mathbf{H}} \) as

\[
\tilde{\mathbf{H}} = \tilde{\mathbf{U}} \tilde{\mathbf{Σ}} \tilde{\mathbf{Υ}}^{T}
\]

(3.56)

it follows that \( \kappa \) equals

\[
\kappa(\mathbf{Ξ}) = \frac{\Sigma_{1}}{\Sigma_{6}}.
\]

(3.57)

It is noted that an alternative measure of observability, not explored further in this work, can be performed for the current setting. In case of three Rx antennas namely, a \( 3 \times 3 \) matrix can be constructed from the unit vectors pointing from the Rx antennas on the chief to the Tx antenna on the deputy. When the volume defined by these unit vectors is sufficiently small (i.e. the determinant of the \( 3 \times 3 \) matrix is nonzero or nonsingular) an estimate of the relative position can be obtained. Thus, the volume of this matrix could
also serve as a measure of observability. However, such an approach provides only ‘snap-
shots’ of the observability and does not take advantage of the system dynamics, which is a
source of information that can increase the observability.

In addition, due to the limited machine accuracy, a relative position resulting in a de-
terminant close to zero can still lead to an incorrect estimate due to round-off errors. This
is known as estimator divergence and is treated in subsection 3.4.4. There is no method
available that can be used to determine how small the matrix determinant can be in order
for the estimation process to not diverge. This has to be determined by trial and error.
For the condition number, such a method is available. This method is discussed in sub-
section 4.4.3. In the same subsection, also a comparison is made between the condition
number of the Gramian matrix and the determinant of the Gramian matrix in an effort to
determine which metric provides the best indication whether or not estimator divergence
is likely to occur when using a trial and error approach.

### 3.4.4 (Apparent) Filter Divergence

When estimating the state of a system, the objective is commonly to do this such that
an unbiased (i.e. zero mean error) estimate with minimum variance is obtained. The
batch LSQ estimator introduced in subsection 3.4.2 is an example of an estimator that is
meant to accomplish just that. For nonlinear systems however, situations can arise where
the estimator obtains a state estimate where the error in the estimate is larger than the
theoretical error. This discrepancy between the actual and theoretical estimation error
is called divergence. Two different forms of divergence can be distinguished: apparent
divergence and true divergence. In apparent divergence, the error in the state estimate
is significantly larger than the one predicted by theory, but is still bounded. Thus, it can
be regarded as a biased estimate. In true divergence, the error in the state estimate is
unbounded and eventually becomes infinite [Gelb 1974]. This is typically caused by a
poor conditioning of the covariance matrix $P$, which leads to large errors when computing
$P^{-1}$, even when techniques such as QR-factorization are applied in the inversion of $P$ to
minimize numerical errors.

Apparent divergence is typically caused by improper modeling of the system dynam-
ics and/or outputs. For instance, if the dynamics are properly modeled, but when the
measurements are linearized, the batch LSQ described earlier can diverge to a wrong state
estimate for large batch sizes if the error in the a-priori estimate is too large. This is caused
by the linearization of the range measurements at time $t = 0$, which causes second-order
and higher-order terms to be neglected when propagating the range measurements for-
ward (or backward) in time. If the error in the a-priori state estimate is large, these higher-
order terms become significant when the measurements are propagated in time and thus
neglecting these leads to large errors if the propagation is performed for a long enough
time span. In case the a-priori covariances are large enough not to constrain the solution
space too much, a certain combination of a-priori state estimation error and batch size (if
the measurements are made at equally spaced intervals in time, the larger the error, the
smaller the batch size can be and vice versa) will cause the estimator to depart from the
contour of convergence that leads to the correct global minimum and to enter a contour of convergence that leads to an undesired local minimum. In this local minimum, the state components are such that the measurement residuals are small, causing the estimator to reach its convergence criterion, causing it to stop at a solution with a large error in the state estimate but with small covariances. In this work, apparent divergence is defined to occur when the error in the estimate of a state component is more than three times its standard deviation.

During the research, several scenarios that exhibit (apparent) divergence have been encountered. The experience gained from those scenarios has motivated to a large part the choices made for the estimator to be used and for the initial conditions for the scenarios that will be analyzed in detail in chapter 4. In the remainder of this subsection, two examples of apparent divergence encountered during the research are discussed. The first example deals with apparent estimator divergence when estimating the relative state of two formation flying spacecraft. In this case, the apparent divergence is caused by a combination of highly accurate range measurement and the assumption of linear dynamics while in reality the dynamics are nonlinear. The second example presents an investigation into the different behavior of batch-type estimators and sequential estimators for a scenario where the sequential estimators show apparent divergence.

**Apparent divergence due to erroneous modeling of the system dynamics**

For a system with non-linear dynamics that is modeled as a system with linear dynamics, propagation of both trajectories using the same initial conditions leads to differences in the two trajectories that grow over time. In the case of a spacecraft formation in LEO, such a scenario can be created when propagating the reference trajectory using Keplerian dynamics while propagating the estimated trajectory using the HCW equations. When using an estimator to estimate the system state, these differences are manageable if either process noise is introduced to compensate for the neglected nonlinearities or if the accuracy in the observations is of larger magnitude than the neglected nonlinearities. Then, the difference between the reference trajectory and the estimated trajectory is ‘absorbed’ by the process noise and/or the inaccuracy of the observations and the estimator will converge to a somewhat inaccurate, but still valid, estimate. However, during the research, scenarios were encountered where this was not the case, leading to apparent divergence. An example of such a scenario is given here.

The estimation process used in this work is depicted schematically in Fig. 3.4. In the simulations performed, first the ‘true’ relative state at epoch is defined and is subsequently propagated in time using appropriate dynamics. Range measurements are then constructed by calculating the distances between the transmitting and receiving antennas and adding a random, normally distributed, measurement error. In the estimator, the same procedure is followed, but then using an estimate for the relative state and without adding noise to the predicted measurements. The estimator then uses the difference between the ‘true’ measurements and the predicted measurements, the measurement residuals, to improve the estimate of the state vector at $t = 0$ using the presented batch LSQ
algorithm until a certain convergence criterion is satisfied.

In this scenario, first the 'true' reference state was propagated using Keplerian dynamics while the estimator propagated the estimated state using the HCW equations. This led to the results depicted in the top panel of Fig. 3.5. After that, the reference state was propagated using the HCW equations while the estimator again propagated the estimated state using the HCW equations. This led to the results depicted in the bottom panel of Fig. 3.5. At epoch $t = 0$ the reference state and estimated state are identical.

![Figure 3.4: Schematic representation of the relative state estimation process. (The star indicates usage of a (potentially) different dynamic model for propagation.)](image)

Figure 3.4: Schematic representation of the relative state estimation process. (The star indicates usage of a (potentially) different dynamic model for propagation.)

![Figure 3.5: State estimation results for a $1 \times 2 \times 1$ km safe ellipse when propagating the 'true' relative orbit using Keplerian dynamics (top) and when propagating the 'true' relative orbit using the HCW equations (bottom). In both cases, the estimator propagates the estimated relative orbit using the HCW equations. The measurement accuracy is 1 cm.](image)

Figure 3.5: State estimation results for a $1 \times 2 \times 1$ km safe ellipse when propagating the 'true' relative orbit using Keplerian dynamics (top) and when propagating the 'true' relative orbit using the HCW equations (bottom). In both cases, the estimator propagates the estimated relative orbit using the HCW equations. The measurement accuracy is 1 cm.

In Fig. 3.5, the evolution of the estimation error and the corresponding standard deviation for $x_0$ is depicted for a safe elliptical formation, explained in detail in subsection 4.3.1, with dimensions of $1 \times 2 \times 1$ km when using a single Rx antenna on the chief and a single Tx antenna on the deputy. The ranging accuracy is 1 cm, the a-priori estimation error is zero
for all state components and the a-priori standard deviations are 1 m for all relative positions and 1 n m/s for all relative velocities. In the top panel of Fig. 3.5, depicting the results when the reference trajectory is generated by propagating the absolute Keplerian orbits of the spacecraft, a clear spike in the estimation error is visible for a data batch of 0.9 orbits. However, there is no similar spike in the standard deviation. As the difference between the estimation error and the standard deviation is more than three times the standard deviation, it is concluded that apparent estimator divergence has occurred. The lower panel in Fig. 3.5 depicts the results when using the HCW equations to propagate the reference trajectory. There, the result is what is expected with the estimation error staying well within the 3σ bounds.

The cause of this behavior lies in the fact that the HCW equations are linearized while the Keplerian dynamics are not. This leads to small differences in the range between the satellites for both methods. For inter-satellite distances in the order of kilometers, the neglected nonlinearities are on the order of centimeters, leading to discrepancies between the predicted and the 'true' measurements. Because the measurements are very accurate, the measurement weight is fairly large, causing the state estimate to be 'pulled' towards a solution that closely matches the measurements, but which is erroneous since the HCW dynamics differ from the Keplerian dynamics. For larger inter-satellite distances and higher measurement accuracy, this effect gets more and more pronounced, leading to ever worsening results. This effect also appears in case of multiple Rx antennas, irrespective of the antenna baseline. A more accurate representation of the relative motion instead of the HCW equations in the estimator would relieve this problem, but as the goal of this study is not to achieve a relative position estimation result that closely matches real-life situations, it was decided to propagate both the reference trajectory and the estimated trajectory using the HCW equations for the remainder of the study.

**Batch-type estimators versus sequential estimators**

In their paper titled "Nonlinearity in Sensor Fusion: Divergence Issues in EKF, modified truncated SOF, and UKF" [Perea et al. 2007], Perea et al. describe the inability of several filter types, the Extended Kalman Filter (EKF), the modified truncated Second-Order Filter (mtSOF), and the Unscented Kalman Filter (UKF), to converge to the correct estimate for a highly simplified scenario. The scenario deals with estimating the unitless two-dimensional Cartesian state vector \((x_1, x_2)^T = (100, 100)^T\) of a stationary object when multiple range measurements \(r\) and angle measurements \(θ\) are available. The a-priori state estimate is \(\hat{x}_0 = (20, 80)^T\) and the a-priori covariance matrix is \(P_0 = σ^2 I\) with \(σ = 100\). The accuracy of the range measurements is \(σ_r^2 = 2.5 \times 10^{-5}\) square units and the accuracy of the angular measurements is \(σ_θ^2 = 6 \times 10^{-3}\) square radians. Thus, the measurements are a nonlinear function of the state components and the range measurement is much more accurate than the angular measurement. This difference in measurement accuracy has been made intentionally since the authors believe this to be representative for sensor suites on possible future proposed formation flying missions. In addition, it is assumed in the paper that all measurements are perfect, even without random errors. This is done to represent a
3.4. Relative State Estimation

best-case scenario in which measurements are better-than-expected under normal experimental conditions. A graphical representation of the starting conditions for this estimator divergence problem is provided in Fig. 3.6.

![Graphical representation](image)

**Figure 3.6:** Graphical representation of the starting conditions for the estimator divergence problem.

On first glance, it appears that sufficient information is available for all filters to converge to a solution. However, it is shown in the paper that the standard implementations of all three different filters considered are not able to converge to the correct solution. They all stabilize at an error in the state estimate that is much larger than indicated by their covariance matrices, which is a typical case of apparent filter divergence. The cause of this is identified to be an overly fast reduction in the state covariances compared to the accuracy of the state estimates. In the paper, several adaptations to the filters are made and compared. It is concluded that two 'bump-up' EKF strategies and a different formulation of the measurement equations for the UKF yield results with proper convergence properties.

The work presented in [Perea et al. 2007] inspired an investigation into the convergence properties of various LSQ estimators and Kalman filters for this scenario. However, comparing a batch LSQ algorithm with a sequential Kalman filter leads to a dilemma: How to deal with the measurement vector $z$? If multiple observations at different epochs are available, the standard procedure for a batch LSQ is to collect all observations in $z$ and to iterate until the estimator has converged. A Kalman filter would compute one estimate for each epoch using the results (state estimate and covariance matrix) of the previous epoch and the measurements of the current epoch. Thus, assuming measurements have been collected for ten different epochs, the Kalman filter would have 'iterated' ten times to arrive at the estimate for the tenth epoch while the batch LSQ can in essence iterate

---

1A 'bump-up' strategy "[...] tries to avoid the divergence of the filter by reducing the rate of convergence in the direction of the more accurate measurements, and "waits" for convergence in the less accurate directions. [...] a B-EKF [Bump-up-EKF, red.] will artificially increase the noise associated with the accurate measurements, and reduce this increment as a function of the confidence in current state vector estimates. Geometrically, this strategy means using larger and less eccentric confidence areas compared to the EKF."

any number of times. Thus, these results cannot be compared directly due to the different processes used. Fortunately, the current scenario is static and the measurements are perfect. This allows to assume that there is only one measurement epoch and that both the LSQ algorithms and the Kalman filters iterate $k$ times on that epoch. This allows a direct comparison since the information used in each iteration is similar and the number of iterations is the same. In the following, a total of six LSQ and Kalman filter algorithms have been compared:

1. **Batch LSQ**: The formulation of this algorithm is similar to the one presented in subsection 3.4.2, but now the information matrix $\Lambda$ is not used and the state transition matrix $\Phi$ reduces to the identity matrix due to the absence of dynamics. This results in the following algorithm:

   \[
   \hat{x}_k = \hat{x}_{k-1} + P_k H_k^T W (z_k - \hat{z}_k) \tag{3.58}
   \]

   with

   \[
   P_k^{-1} = H_k^T W H_k.
   \]

2. **Batch LSQ with a-priori information**: Now the information matrix is used, leading to the following algorithm:

   \[
   \hat{x}_k = \hat{x}_{k-1} + P_k \left( \Lambda (\hat{x}_0^{\text{apr}} - \hat{x}_{k-1}) + H_k^T W (z_k - \hat{z}_k) \right) \tag{3.59}
   \]

   with

   \[
   P_k^{-1} = \Lambda + H_k^T W H_k.
   \]

3. **Sequential LSQ**: The sequential LSQ differs from the batch LSQ algorithms in that the covariance matrix $P_k$ is computed using the covariance matrix $P_{k-1}$, resulting in $\hat{x}_k$ being computed using Eq. (3.58) but now with

   \[
   P_k^{-1} = P_{k-1}^{-1} + H_k^T W H_k.
   \]

4. **EKF**: The propagation equations (or prediction phase) of the EKF are modeled as

   \[
   \hat{x}_k^- = \hat{x}_{k-1}^+ \\
   P_k^- = P_{k-1}^+
   \]

   where $(\cdot)^-$ denotes a predicted variable and $(\cdot)^+$ denotes an updated variable. The update equations (or innovations phase) are
3.4. Relative State Estimation

\[ K_k = P_{xz,k} P_{zz,k}^{-1} \]
\[ \hat{x}_k^+ = \hat{x}_k^- + K_k (z_k - \hat{z}_k) \]
\[ P_k^+ = P_k^- - P_{xz,k} P_{zz,k}^{-1} P_{xz,k}^T \]

with

\[ \hat{z}_k = h(\hat{x}_k^-) \]
\[ P_{zz,k} = W^{-1} + H_k P_k^- H_k^T \]
\[ P_{xz,k} = P_k^- H_k^T. \]

5. **Batch EKF**: This version of the EKF 'mimics' the behavior of a batch LSQ algorithm by not updating the covariance matrix during the prediction phase. This avoids the apparent divergence problem, but also results in large covariances, which is not desired. The algorithm for the batch EKF is a copy of the algorithm of the EKF except for the propagation equation for \( P_k \), which is now performed as

\[ P_k^- = P_0. \]

6. **Threshold EKF**: To overcome the downside of the batch EKF, a variation of the EKF here referred to as threshold EKF has been developed. This algorithm switches between the batch EKF and the regular EKF once a certain convergence criterion has been met. The convergence criterion used here is

\[ \frac{\| \hat{x}_k^- - \hat{x}_k^+ \|}{\| \hat{x}_k^- \|} < 0.01. \]

Thus, the covariance matrix is not updated unless the state estimate has converged considerably.

For all algorithms described above, the analytical expression for the Jacobian \( H \) is

\[ H = \begin{pmatrix} \frac{\dot{x}_1}{r} & \frac{\dot{x}_2}{r} \\ \frac{-\dot{x}_2}{r^2} & \frac{\dot{x}_1}{r^2} \end{pmatrix} \]

and the weighting matrix \( W = \text{diag}(\sigma_r^2, \sigma_\theta^2) \).

The results for all algorithms are depicted in Fig. 3.7. The left pane of the figure shows how the root-sum square (RSS) error in the state estimate evolves for the different algorithms while the right pane shows how the RSS of the diagonal elements of \( P \) evolves for the different algorithms. The figure provides several important insights:

1. The covariances of the sequential LSQ and the EKF (algorithms 3 and 4 in the preceding list) converge much more quickly than the error in their state estimates and match the behavior of the EKF in [Perea et al. 2007]. Thus, both purely sequential
estimators show apparent divergence. In addition, the curves of these estimators overlap exactly in both panes. This is not surprising as an EKF without dynamics and process noise can be rewritten such that the algorithm is exactly the same as that of a sequential LSQ.

2. All batch-like estimators (algorithms 1, 2, 5, and 6 in the preceding list) converge to an acceptable solution with all having practically the same variances except for the threshold EKF (algorithm 6), which departs from this trend after four iterations, indicating that it has met its convergence criterion and has switched from the batch EKF to the standard EKF.

Figure 3.7: Results for various filters and estimators. Subfigure (a) depicts the dimensionless root-sum square estimation error versus the number of iterations and subfigure (b) depicts the dimensionless root-sum square of the diagonal elements of $P$ versus the number of iterations.

Thus, for this degenerate problem, a standard sequential estimator will diverge whereas a batch estimator will converge to a correct solution. Thus, apparent filter divergence can for this scenario be prevented by using a different type of estimator although it must be noted that pure batch-type estimators are typically not practical for near real-time applications. Thus, out of all estimators considered here, the threshold EKF scheme could be a useful estimator for low-dynamic scenarios with highly accurate measurements.

As a final note, the problem of filter divergence can be circumvented by applying a coordinate transformation such that the measurement equations can be written as a linear function of the state components. For the current scenario, this implies that either the state components have to be transformed from a Cartesian frame to a polar frame or the measurements have to be transformed from the polar frame to a Cartesian frame. In practice, it will likely be preferable to transform the measurements since conversion of the equations of motion to a different reference frame can be non-trivial. In that case, the EKF and sequential LSQ can be shown to converge to the correct state estimate and even outperform the batch estimators.
3.4.5 Statistical Analysis of the Relative Position Determination

Although a numerical treatment is indispensable when analyzing the influence of variations in certain parameters on the accuracy with which the relative state of two non-stationary objects can be estimated, considerable insight can still be gained by analytical means. For instance, when using multiple Rx antennas to estimate the position of a single Tx antenna, it is generally known that the Rx antenna baseline should be as large as possible to obtain the most accurate position estimate of the Tx antenna. But what about the inter-satellite distance and the ranging accuracy? Are the antenna baseline, inter-satellite distance, and the ranging accuracy all equally important or is it for instance more effective to improve the ranging accuracy? And is the influence of the antenna baseline different in different reference frame axes? This information could be derived using numerical simulations, but an analytical analysis provides a firm foundation for the conclusions and it allows other researchers to quickly analyze similar problems.

No reference could be found in the available literature that provides a thorough analytical analysis of the estimation of the position of an object whose distance is measured from multiple locations, as is the case in this study. For that reason, here a statistical analysis is performed of the expected value and variance in the relative position estimation of a stationary transmitting antenna Tx in a reference frame defined by three receiving antennas Rx\(_0\), Rx\(_1\), and Rx\(_2\), cf. Fig. 3.8. The distance between the transmitting and receiving antennas is measured as a pseudorange \(\rho_i\), where \(i = 0, 1, 2\), which consists of the true range \(r_i\) between Tx and Rx\(_i\) and a normally distributed random noise \(\varepsilon\) with zero mean and variance equal to \(\sigma_i^2\). The baselines between Rx\(_0\) and Rx\(_1\) and between Rx\(_0\) and Rx\(_2\) are equal to \(d_1\) and \(d_2\), respectively. The angle between the two baselines is equal to \(\alpha\). Antenna Rx\(_0\) is located at the origin of the reference frame, antenna Rx\(_1\) is located on the x-axis, and antenna Rx\(_2\) is located in the xy-plane. Antenna Tx has coordinates \(x_{\text{Tx}}, y_{\text{Tx}}, z_{\text{Tx}}\).

![Figure 3.8: Location of the various antennas in the receiver reference frame.](image-url)
Figure 3.9: Definition of the angles in the triangles formed by the Tx and Rx antennas.

As \( d_1 \) and \( d_2 \) are known and \( r_0, r_1, \) and \( r_2 \) are measured, two triangles with 'known' side lengths, Rx0-Rx1-Tx and Rx0-Rx2-Tx, can be formed, cf. Fig. 3.9. Using the cosine-law, the following three equations can be constructed for the angles defined in Fig. 3.9:

\[
\begin{align*}
\cos \beta_0 &= \frac{r_2^2 - r_1^2 + d_1^2}{2r_0d_1} = \frac{r_0 \cdot d_1}{\|r_0\| \|d_1\|} \tag{3.60} \\
\cos \gamma_0 &= \frac{r_2^2 - r_0^2 + d_2^2}{2r_0d_2} = \frac{r_0 \cdot d_2}{\|r_0\| \|d_2\|} \tag{3.61} \\
\cos \beta &= \frac{r_2^2 + r_1^2 - d_1^2}{2r_0r_1} = \frac{r_0 \cdot r_1}{\|r_0\| \|r_1\|} \tag{3.62}
\end{align*}
\]

with \( \cdot \) denoting the vector dot product and \( d \) and \( r \) the vectors of the corresponding distances. Equations (3.60–3.62) may be used to derive the following expressions for \( x_{Tx}, y_{Tx}, \) and \( z_{Tx} \):

\[
\begin{align*}
x_{Tx} &= \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \tag{3.63a} \\
y_{Tx} &= \frac{r_0^2(d_1 - d_2 \cos \alpha) + r_1^2d_2 \cos \alpha - r_2^2d_1 + d_1d_2(d_2 - d_1 \cos \alpha)}{2d_1d_2 \sin \alpha} \tag{3.63b} \\
z_{Tx} &= \left( \frac{d_2^2 \sin^2 \alpha \left[ r_0^2(2r_1^2 - r_0^2 + 2d_1^2) + r_1^2(2d_2^2 - r_2^2) - d_1^4 \right] + \left[ -r_0^2(d_1 - d_2 \cos \alpha) + r_1^2d_2 \cos \alpha - r_2^2d_1 + d_1d_2(d_2 - d_1 \cos \alpha) \right]^2}{2d_1d_2 \sin \alpha} \right)^{1/2}. \tag{3.63c}
\end{align*}
\]

Note that \( z_{Tx} \) is in fact a function of \( x_{Tx} \) and \( y_{Tx} \) and can also be written more concisely as \( z_{Tx} = (r_0^2 - (x_{Tx}^2 + y_{Tx}^2))^{1/2} \).

From Eqs. (3.63a–3.63c), it is immediately obvious that for \( \alpha = 0 \) no result can be obtained for \( y_{Tx} \) and \( z_{Tx} \) since this condition leads to all Rx antennas being aligned. In addition, it is concluded that the sign of \( z_{Tx} \) is undefined. Adding a fourth Rx antenna that is not situated in the plane spanned by the other three Rx antennas would remove the ambiguity in the determination of \( z_{Tx} \). In practice however, there will rarely be a situation...
where the sign of $z_{tx}$ cannot be determined, either through a priori navigation information or through restrictions in the field of view of the antennas, and thus this fourth antenna would primarily serve to provide redundant range measurements in this scenario (when dealing with integer ambiguity and multipath effects, measurements made using a fourth antenna would actually be very valuable).

Using Eqs. (3.63a–3.63c), the influence of the ranging error, the inter-satellite range, and the antenna baseline on the expected value and variance in the determination of $x_{tx}$, $y_{tx}$, and $z_{tx}$ can be determined. First however, it is given that $\alpha = 90^\circ$ is the optimum angle (i.e., leading to the smallest position estimation error) between baselines $d_1$ and $d_2$. This is proven in Appendix A, but is also intuitively obvious as this angle leads to an antenna geometry which is furthest from the situation where all receiver antennas are aligned. Setting $\alpha = 90^\circ$ leads to the following simplified equations for the true relative position

$$x_{tx} = \frac{r_0^2 - r_1^2 + d_1^2}{2d_1}$$  \hspace{1cm} (3.64a)

$$y_{tx} = \frac{r_0^2 - r_2^2 + d_2^2}{2d_2}$$  \hspace{1cm} (3.64b)

$$z_{tx} = \frac{1}{2} \left[ -\left( \frac{r_0^4}{d_1^2} + \frac{r_1^4}{d_1^2} - \frac{2r_0^2 r_1^2}{d_1^2} \right) \left( \frac{r_0^4}{d_2^2} + \frac{r_2^4}{d_2^2} - \frac{2r_0^2 r_2^2}{d_2^2} \right) - \left( \frac{r_1^4}{d_1^2} - \frac{r_2^4}{d_2^2} - \frac{2r_0^2 r_1^2}{d_1^2} \right) \right]^{1/2}.$$  \hspace{1cm} (3.64c)

Not surprisingly, due to the two antenna baselines lying exactly on the $x$- and $y$-axis, Eqs. (3.64a–3.64b) have the same structure. To solve either of these equations, only two ranges are needed ($r_0$ and $r_1$ for $x_{tx}$ or $r_0$ and $r_2$ for $y_{tx}$). The expression for $z_{tx}$ consists out of two major terms which can be regarded as the contributions of baselines $d_1$ and $d_2$. In order to solve for $z_{tx}$, all three ranges are needed.

In practice, the range cannot be measured without error and thus only an estimate of $x_{tx}$, $y_{tx}$, and $z_{tx}$ can be made. By replacing the true range $r_i$ in Eqs. (3.64a–3.64c) with the corresponding pseudorange $\rho_i$ and assuming that the pseudorange is normally distributed with mean $r_i$ and variance $\sigma_{\rho_i}^2$, expressions for the expected value $E[\cdot]$ and variance $\text{Var}[\cdot]$ of the relative position estimate can be derived using a second-order multivariate Taylor series expansion around $r_i$. In the derivation, provided in detail in appendix A, zero cross-correlation between measurements from different Rx antennas is assumed. For the expectation and variance of $x_{tx}$ and $y_{tx}$, this leads to

$$E[x_{tx}] = \frac{r_0^2 - r_1^2 + d_1^2 + \sigma_{\rho_0}^2 - \sigma_{\rho_1}^2}{2d_1}$$  \hspace{1cm} (3.65a)

$$E[y_{tx}] = \frac{r_0^2 - r_2^2 + d_2^2 + \sigma_{\rho_0}^2 - \sigma_{\rho_1}^2}{2d_2}$$  \hspace{1cm} (3.65b)
Var \[x_{\text{Tx}}] = \frac{2r_0^2\sigma_{\rho_0}^2 + 2r_1^2\sigma_{\rho_1}^2 + \sigma_{\rho_0}^4 + \sigma_{\rho_1}^4}{2d_1^2} \tag{3.66a}

Var \[y_{\text{Tx}}] = \frac{2r_0^2\sigma_{\rho_0}^2 + 2r_2^2\sigma_{\rho_2}^2 + \sigma_{\rho_0}^4 + \sigma_{\rho_2}^4}{2d_2^2} \tag{3.66b}

Comparing Eqs. (3.64a-3.64b) and Eqs. (3.65a-3.65b) shows that the expected value for the components of the position vector of Tx that are in the plane defined by the receiver antennas (i.e., \(x_{\text{Tx}}\) and \(y_{\text{Tx}}\)) will be very close to the true value if the magnitude of the ranging accuracy is much smaller than the antenna baseline and the inter-satellite distance, thus if \(\sigma_{\rho_i}/d \ll 1\) and \(\sigma_{\rho_i}/r \ll 1\) for all \(i\). These are reasonable conditions since the intersatellite distance is commonly at least tens of meters and typical ranging accuracies are on the order of centimeters [Harr et al. 2008]. It also makes little sense to use an antenna baseline that is smaller than the ranging accuracy since then the benefit of the antenna separation is removed. When assuming that \(\sigma_{\rho_i}/d \ll 1\), \(\sigma_{\rho_i}/r \ll 1\), and \(\sigma_{\rho_i} = \sigma_{\rho}\) for all \(i\), Eqs. (3.65a-3.66b) reduce to

\[
\left[ E[x_{\text{Tx}}] \right]_{\sigma_{\rho} \ll r, d} \approx \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \tag{3.67a}
\]

\[
\left[ E[y_{\text{Tx}}] \right]_{\sigma_{\rho} \ll r, d} \approx \frac{r_0^2 - r_2^2 + d_2^2}{2d_2} \tag{3.67b}
\]

\[
\left[ \text{Var}[x_{\text{Tx}}] \right]_{\sigma_{\rho} \ll r, d} \approx \left( r_0^2 + r_1^2 \right) \frac{\sigma_{\rho}^2}{d_1^2} \tag{3.68a}
\]

\[
\left[ \text{Var}[y_{\text{Tx}}] \right]_{\sigma_{\rho} \ll r, d} \approx \left( r_0^2 + r_2^2 \right) \frac{\sigma_{\rho}^2}{d_2^2} \tag{3.68b}
\]

Equations (3.68a-3.68b) show that the standard deviation in the estimate of \(x_{\text{Tx}}\) and \(y_{\text{Tx}}\) depends linearly on \(r\) and \(\sigma_{\rho}\) and inversely linear on \(d\).

For the out-of-plane component, \(z_{\text{Tx}}\), the equations for the expectation and the variance consist out of many terms covering a wide variety of scenarios. However, the amount of terms is so large that several simplifications, resulting in a non-general scenario, are required to arrive at a result that is easily understood. The scenario selected here is a perfect (i.e., no cross-track motion) along-track formation where the \(z\)-axis of the Rx antenna reference frame points in the along-track direction towards the trailing satellite. For this scenario it can be assumed that all ranges, baselines, and ranging variances are the same, thus \(r_0 = r_1 = r_2 = r\), \(d_1 = d_2 = d\), \(\sigma_{\rho_0} = \sigma_{\rho_1} = \sigma_{\rho_2} = \sigma_{\rho}\). Secondly, it is assumed that the intersatellite range is much larger than the antenna baseline and the ranging accuracy, thus \(r \gg d, \sigma_{\rho}\). This implies that the true value for \(z_{\text{Tx}}\) is approximately \(r\) since under these conditions Eq. (3.64c) reduces to

\[
z_{\text{Tx}} = (r^2 - 1/2d^2)^{1/2} \approx r - 1/4d^2/r + \mathcal{O} \left( d^4/r^3 \right) \approx r + \mathcal{O} \left( d^2/r \right).
\]
This leads to the following expressions for the expectation and variance for $\tau_{\text{Tx}}$ in case of an along-track formation:

$$
E[\tau_{\text{Tx}}] \bigg|_{r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}, r \gg d, \sigma_{\rho}} \approx r \left( 1 - 2 \frac{\sigma_{\rho}^2}{d^2} \right)
$$

(3.69)

$$
\text{Var}[\tau_{\text{Tx}}] \bigg|_{r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}, r \gg d, \sigma_{\rho}} \approx 5r^2 \frac{\sigma_{\rho}^4}{d^4} + \frac{1}{2} \sigma_{\rho}^2.
$$

(3.70)

From Eq. (3.69) it is clear that the expected value for $\tau_{\text{Tx}}$ is always smaller than the true value for $\tau_{\text{Tx}}$. To obtain an expected value for $\tau_{\text{Tx}}$ close to the true value, the ratio $\sigma_{\rho}/d$ must be as small as possible. Note that this condition also holds for $\tau_{\text{Rx}}$ and $\tau_{\text{Rx}}$. Equation (3.70) shows that the standard deviation in the estimate of $\tau_{\text{Tx}}$ depends linearly on $r$ but quadratically on $\sigma_{\rho}$ and $d$. Thus, for an along-track formation, the estimate for $\tau_{\text{Tx}}$ is expected to be much more accurate than the estimates for $\tau_{\text{Rx}}$ and $\tau_{\text{Rx}}$ because it scales with $\left( \frac{\sigma_{\rho}}{d} \right)^2$ instead of $\sigma_{\rho}/d$ and because $\sigma_{\rho}/d \gg 1$.

When reducing the along-track distance, such as in a rendezvous and docking scenario, simulation results show that Eqs. (3.69–3.70) remain valid up to surprisingly small inter-satellite distances. Only when the inter-satellite distance is such that $r < 2d$, the standard deviation and the error in the expectation can no longer be considered to decrease linearly with $r$. For instance, for $r = d$, the expectation and the variance are

$$
E[\tau_{\text{Tx}}] \bigg|_{d_i=d, r_i=r, \sigma_{\rho_i} = \sigma_{\rho}} = \frac{1}{2} \sqrt{2} r \left( 1 - 4 \frac{\sigma_{\rho}^2}{d^2} \right)
$$

(3.71)

$$
\text{Var}[\tau_{\text{Tx}}] \bigg|_{d_i=d, r_i=r, \sigma_{\rho_i} = \sigma_{\rho}} \approx 10.5r^2 \frac{\sigma_{\rho}^4}{d^4} + \sigma_{\rho}^2.
$$

(3.72)

For $\tau_{\text{Tx}}$ approaching zero, which could be the case in a docking or undocking scenario, the error in the expectation as well as the variance increase towards infinity for this case.

For very small satellites, it is not unrealistic that during the design phase of the mission the magnitude of the (achievable) ranging accuracy is found to be close to the (intended) magnitude of the antenna baseline. For again an along-track scenario with $r_i = r$, $d_i = d$, and $\sigma_{\rho_i} = \sigma_{\rho}$ for all $i$, for which $\tau_{\text{Tx}} = \tau_{\text{Rx}} = d/2$ and $\tau_{\text{Tx}} \approx r$, setting $d = 2\sigma_{\rho}$ leads to an error-free expectation for $\tau_{\text{Tx}}$ and $\tau_{\text{Rx}}$ but the expectation for $\tau_{\text{Tx}}$ is off by 50%:

$$
E[\tau_{\text{Tx}}] \bigg|_{d_i=d, r_i=r, \sigma_{\rho_i} = \sigma_{\rho}} = \frac{1}{2} r \left( 1 - 2 \left( \frac{1}{2} \frac{d}{d^2} \right) \right) = \frac{1}{2} r.
$$

(3.73)

For the variances, the result is
Var \[x_{\text{Tx}}\] = Var \[y_{\text{Tx}}\] \bigg|_{d_i=d, r_i=r, \sigma \rho_i=\sigma \rho=1/2d, r \gg d, \sigma \rho} = \\
\frac{2r^2(\frac{1}{2}d)^2 + 2r^2(\frac{1}{2}d)^2 + (\frac{1}{2}d)^4}{2d^2} = \frac{1}{2} r^2 + \frac{1}{8} d^2 \quad (3.74)

Var \[z_{\text{Tx}}\] \bigg|_{d_i=d, r_i=r, \sigma \rho_i=\sigma \rho=1/2d, r \gg d, \sigma \rho} = 5r^2(\frac{1}{2}d)^4 + \frac{1}{2} (\frac{1}{2}d)^2 = \frac{5}{16} r^2 + \frac{1}{8} d^2. \quad (3.75)

Thus, for all axes, the standard deviation in the position estimate is more than half the inter-satellite distance when \(d = 2\sigma \rho\). Thus, for small satellites it is vital to either decrease \(\sigma \rho\) or to maximize \(d\), which could be achieved using a deployable structure.

With the equations provided in this subsection, many (static) formation geometries can be studied. It is even feasible, and not very complex, to use the results obtained here in a dynamic scenario by making all ranges (and possibly also the ranging accuracy) a function of time. Although out of the scope of this study and not taking into account the influence of the estimation algorithm and the dynamic environment, such an approach can be used to obtain a first estimate of the relative position estimation accuracy that can be expected for a scenario with a certain relative motion. Thus, this analysis contributes valuable support of the mission design of formation flying missions.

### 3.5 Summary

This chapter has laid several foundations for the analyses to be presented in chapter 4. Section 3.1 discussed various system level simplifications that have been made for this research. These simplifications have resulted in the most basic scenario possible, which still captures the essential parts of the problem to be studied. In addition, the results of the analyses will be applicable to a large variety of formation flying system configurations.

Section 3.2 discussed the relative orbital dynamics for two satellites whose motion is dictated by the gravitational influence of either one or two primary bodies. Nonlinear equations of relative motion were derived for the case of one primary body. From these equations, the HCW equations of relative motion were derived. The HCW equations are used in the analyses to be presented in chapter 4. Although the research focuses on formations orbiting a single primary body, the Earth, it was briefly shown that the linearized equations of relative motion in case of one primary body are very similar to the linearized equations of relative motion in case of two primary bodies. Thus, the results to be obtained in chapter 4 can, to some extent, even be applied to scenarios with two primary bodies.

Section 3.3 described how the observations, or measurements, that are necessary to enable an estimation of the state of a system are typically obtained and which error sources need to be accounted for in these observations. Since it is assumed in this research that the
3.5. Summary

Satellites are measuring the inter-satellite range using radio signals, the section focused on measurements and measurement errors typically associated with radio signals.

To perform the analyses in chapter 4, a simulation environment incorporating an estimator is needed. Therefore, various aspects regarding the estimation of the (relative) state of a system were treated in section 3.4. For systems relying on radio ranging signals to provide information on the relative state, integer ambiguities arising in the measurements provide a challenge in the estimation process. As the integer ambiguities are considered to be resolved correctly in this research, only a high level discussion is provided on several strategies that can be used to resolve integer ambiguities. The section continued with detailed mathematical descriptions of two important tools used in the analyses performed in this research: the batch least squares estimator with a-priori information and observability analysis.

The last two subsections of section 3.4 dealt with two specific investigations that were related to the work performed during the research. The first investigation, which has only a weak link to the research questions, dealt with the phenomenon of apparent filter divergence in which an estimator converges to a solution which is not the intended solution. In a highly simplified scenario, sequential estimators, batch estimators, and hybrids of the two were analyzed and it was concluded that, for the scenario under study, a standard sequential estimator will diverge whereas a batch or hybrid estimator will converge to a correct solution. This particular investigation provided valuable insight into limitations, and how to handle them, for a proper treatment of the research questions in chapter 4.

The second investigation dealt with a statistical analysis of the relative position determination problem. Although a numerical treatment is needed to answer the research questions, considerable insight on the relation between the inter-satellite range, the antenna baseline, and the ranging accuracy can be gained by analytical means. The analysis yielded general equations for the expected value and variance in the estimation of the position of a point in space (in this case a transmitting antenna) using range measurements, incorporating a variable, normally distributed ranging error, to three different points (receiving antennas). These equations can be used for many static or even dynamic scenarios to obtain a first estimate of the accuracy that is achievable in the position estimate of an object for any desired value for the inter-satellite range, antenna baseline, and ranging accuracy. Thus, these investigations provide valuable support of preliminary mission design phases of formation flying missions using dedicated inter-satellite ranging.
Chapter 4

Relative Navigation Simulation and Analysis

The current chapter describes the results of numerical investigations that have been performed in order to be able to answer the research questions posed in section 1.3. The first three sections describe the objectives of the simulations, the simulation scenario, and the selected spacecraft formation geometries, respectively. The fourth and last section of this chapter presents the results that have been obtained for the simulations and thus provides the answers to the research questions. The fourth section also contains several subsections that describe brief investigations into 'subquestions' that have arisen while conducting the research.

4.1 Objectives

Naturally, the objectives for the simulations to be performed are that an analysis of their results leads to answers to the research questions. For all research questions, this implies that scenarios must be constructed where reliable relative navigation results can be obtained for at least two formation flying spacecraft by means of inter-satellite range measurements. Thus, the scenario has to be set up such that the estimator does not diverge, except for those cases with inherent poor observability. Those compose scenarios with very small relative out-of-plane motion. Furthermore, the observation arc has to be long enough for the estimator to have converged (i.e., the accuracy in the estimate improves smoothly and only slightly for longer observation arcs). In addition, the estimator must not be constrained too much by the a-priori covariances since this will have a large impact on the relative navigation results in case of scenarios that result in poor accuracy in the estimate of the relative state. For instance, when the inter-satellite ranging accuracy in a scenario is 1 m, setting the a-priori covariances for the relative position to 10 cm will yield an estimation process to which the measurements will have a very small contribu-
tion, leading to a result that provides little useful information.

To answer the first research question, the antenna baseline, the inter-satellite range, the ranging accuracy, and the number of antennas needs to be varied for any selected scenario and these have to be dimensioned such that they are reasonable for space applications. To answer the second research question, at least two types of spacecraft formations have to be simulated. These formations should preferably be of a kind that is considered for or is currently used in space missions. To answer the third research question, the amount of relative out-of-plane motion has to be varied within reasonable bounds.

Comparing relative navigation results and drawing conclusions for different scenarios is complex when the results for all state components are to be compared on an individual basis. Therefore, to ease this process, a single figure of merit should be used as much as possible to characterize the results.

4.2 Scenario

To isolate the fundamental properties of interest for this research, a simple setting has been implemented. The rationale behind this is that more advanced aspects add only minor contributions to the findings while they add considerable complexity to the analysis of the results. Some important simplifications have already been discussed in section 3.1.

The scenario to be investigated is as follows. The chief is in a perfectly circular LEO orbit with a semi-major axis of 7028 km, which corresponds to an orbital altitude of 650 km. It is assumed that the only force acting on the satellites is gravity caused by a spherically symmetric Earth. The deputy flies in formation with the chief, has a single Tx antenna and is treated as a point mass. The state vector \((x, y, z, \dot{x}, \dot{y}, \dot{z})^T\) of the deputy with respect to the origin of the Hill frame is to be estimated for the time \(t = 0\). The origin of the Hill frame coincides with the COM of the chief. The information needed for the estimation is obtained from a-priori information and from range measurements between the deputy and the chief using one out of two possible sensor sets. One sensor set consists out of the Tx antenna on the deputy and one Rx antenna on the chief while the other sensor set consists out of the Tx antenna on the deputy and three Rx antennas on the chief. One-way range measurements are performed between the Tx and Rx antenna(s), which implies perfectly synchronized satellite clocks with negligible relative drift.

To investigate the influence of Rx antenna baseline, ranging accuracy, and inter-satellite distance on the navigation solution, these variables are varied in magnitude. Each of these variables is given three different values, which differ one order of magnitude with respect to each other, cf. Table 4.1. The inter-satellite distance is scaled by varying the inter-satellite distance at the epoch \(t = 0\), denoted as \(r_e\).

For the sensor set with three Rx antennas, the location of the Rx antennas on the chief is chosen such that a homogeneous distribution over the satellite body, assumed to be cubic, is achieved, cf. Fig. 4.1. The plane spanned by the three antennas does not coincide with any plane in the Hill frame, thus ensuring that the observability of the system is not positively or negatively 'biased' towards any particular state component. The selected
chief side lengths lead to antenna baselines of \((150)^{1/2}\) cm \(\approx\) 12 cm, \((150)^{1/2}\) dm \(\approx\) 12 dm, and \((150)^{1/2}\) m \(\approx\) 12 m, respectively. It is assumed that the spacecraft body does not obstruct the field of view of the antennas. Naturally, a spacecraft with side lengths of 10 m is anything but small, but it is necessary to include this length in the simulations to obtain decent information on the trend in the observability and the accuracy in the estimation of the relative state when varying the antenna baseline. In addition, such a large antenna baseline can potentially be created when a spacecraft is equipped with antennas on deployable structures such as booms or deployable solar arrays. In case of a sensor set with a single Rx antenna, the chief is treated as a point mass. Thus, the position of the Rx antenna coincides with the origin of the Hill frame. This case can be considered as a three Rx antenna case with zero baseline, but with only one observation per measurement epoch instead of three.

To enable the estimation of the relative state, the location of the Rx antennas in the Hill frame needs to be known. This implies that the attitude of the chief in the Hill frame has to be known. As in general the attitude of a spacecraft is determined in an inertial frame, e.g. the Earth-Centered Inertial (ECI) frame, this knowledge has to be converted to the Hill frame. As conversion from the ECI frame to the Hill frame requires the inertial radius vector and velocity vector of the chief to be known, cf. Eq. (3.4), the absolute orbit of the chief must be known. This information can be uploaded from ground using ground-based estimates of the absolute orbit of the chief by means of e.g. ground-station ranging or laser ranging. Alternatively, the spacecraft can estimate their absolute orbits themselves if they are equipped with the necessary sensors and algorithms [Psiaki 1995, Psiaki 1999a]. Either way, the estimate of the absolute orbit is never exact and therefore the location of the Rx antennas on the chief is generally known with limited accuracy. Here however, it is
assumed that the positions of the Rx antennas in the Hill frame are known perfectly.

The ranging accuracy $\sigma_\rho$ is assumed to be 0.01 m, 0.1 m, or 1 m. As signal multipath and IAR are not considered, these accuracies can be considered to represent the following cases.

1. A ranging accuracy of 1 m can be regarded as a situation where only code measurements are available.

2. A ranging accuracy of 1 dm can be regarded as a situation where code measurements are available that have been smoothed using carrier phase measurements.

3. A ranging accuracy of 1 cm can be regarded as a situation where accurate and unambiguous carrier phase measurements are available.

If signal multipath and IAR were to be modeled, then the first two cases would represent cases with large biases due to signal multipath and/or integer ambiguities on code measurements (case 1) or on carrier phase measurements (case 2), but then a ranging bias on the order of meters would have to be included in the range measurement model for case 1 and a ranging bias on the order of decimeters would have to be included in the range measurement model for case 2. The biases would then have to change for changing relative attitude to account for multipath effects and they would have to vary between simulation runs to account for different, possibly erroneous, solutions in the IAR (resulting in ranging biases of integer-valued carrier cycles).

All Rx antennas are assumed to be connected to a single receiver, which introduces some degree of cross-correlation between the range measurements resulting from processing the signals received by different Rx antennas. This cross-correlation adds information and is thus beneficial in the relative state estimation process. However, as already briefly discussed in section 3.4.2 and repeated here, various effects are also present that reduce the cross-correlation such as multipath, differential temperature, antenna phase center variations, different cable lengths, and component aging. Even careful design and extensive calibration cannot fully remove these effects. In addition, assuming some amount of cross-correlation implies adding another variable and makes the results sensitive to these assumptions. As cross-correlation is not considered to be a key variable for this study, the foregoing considerations have led to the assumption of zero cross-correlation between range measurements.

By processing the signals received on each receiver channel (one channel corresponds to one Rx antenna), 100 range measurements are made per orbit, roughly one per minute, for a maximum of five orbits and the estimator performs state estimations for measurement arc lengths of 0.1 to 5 orbits with steps of 0.1 orbits. The measurements are generated by first computing the true range and then adding a zero-mean Gaussian noise. This results in measurements that are uncorrelated between different Rx antennas and agrees with the assumption used in the derivation of the expected values and variances of the relative position in subsection 3.4.5. The frequency of the measurements only affects the results for the relative navigation in a statistical manner, i.e. more measurements per time
interval leads to a higher accuracy in the estimate for a certain time period. It does however not influence how quickly the estimator converges to a stable state estimate. One hundred measurements per orbit has been chosen since it provides a reasonable density of measurement points along the orbit and since it limits computational burden. This frequency of inter-satellite measurements is also considered to be reasonable, taking into account computational costs and multiple access, for implementation on real-life formations consisting of up to ten spacecraft.

The error in the a-priori state is set equal to $+0.001r_e$ in all relative position components and $+0.001nr_e$ in all relative velocity components for all formation geometries studied, resulting in $\Delta x^0 = 1 \cdot 10^{-3} (r_e, r_e, r_e, nr_e, nr_e, nr_e)^T$. The a-priori estimation error has been deliberately kept small to prevent the estimator from diverging to an ambiguous solution for the relative state estimate for the elliptical formation geometries in case of long measurement arcs. This occurs for both sensor sets at the same measurement arc length, irrespective of the measurement accuracy, antenna baseline or inter-satellite distance and is due to the nonlinearity of the problem, which causes the existence of multiple local minima. This is a drawback of using a batch estimation process instead of a sequential filter such as an EKF where the reference state is regularly updated, thus reducing linearization errors. The a-priori standard deviations are set to $0.1r_e$ in relative position, in the following denoted as $\sigma_{\text{appr}}$, and $0.1nr_e$ in relative velocity. This implies poor a-priori knowledge, which can be the case directly after orbit injection by the launcher or after a long period without measurements, potentially caused by a spacecraft having been in safe mode.

To investigate the influence of the out-of-plane motion on the observability of the relative state, the maximum cross-track distance in the formation geometries studied, $z_{\text{max}}$, is varied between 0 and $r_e$.

### 4.3 Simulation Geometries

Two formation geometries have been selected for the study. These are the elliptical formation and the pendulum formation. Reasons for choosing these formation geometries are that they exhibit bounded 1:1 commensurate motion (i.e., the orbital periods of the chief and deputy match) in case of HCW dynamics and that they have various practical uses for e.g. Earth observation, spacecraft inspection, and spacecraft RVD.

#### 4.3.1 Elliptical Formation

In terms of absolute orbital elements, elliptical formation geometries can be created by placing the deputy in an orbit that is similar to that of the chief, but which has a small eccentricity. This leads to a coplanar relative motion in the shape of an ellipse with a maximum negative radial inter-satellite distance at the periapsis of the deputy’s orbit and a maximum positive radial inter-satellite distance at the apoapsis of the deputy’s orbit. This motion is periodic with a period equal to the orbital period of the chief. By creating a differential argument of perigee between the orbits of the deputy and the chief, the center
of the relative ellipse can be located anywhere on the $y$-axis of the Hill frame. For this work however, it has been chosen to let the center of the ellipse coincide with the center of the Hill frame. By creating a differential true anomaly, the phasing of the relative orbit can be adjusted as well. The relative motion can be made three-dimensional by rotating the plane of the deputy’s orbit to create a differential inclination and/or a differential right ascension of the ascending node (RAAN). As discussed before, this relative out-of-plane motion is a harmonic oscillation with a period equal to the orbital period of the chief.

Mathematically, an elliptical formation geometry can be constructed as follows. If Eq. (3.20), $\dot{y}_0 = -2n\dot{x}_0$, is satisfied, the homogeneous solution to the HCW equations can be written as [Alfriend et al. 2010]

$$x(t) = n^{-1}\sqrt{\dot{x}_0^2 + n^2\dot{x}_0^2} \sin\left(nt + \tan^{-1}\left(n\frac{x_0}{\dot{x}_0}\right)\right)$$

$$y(t) = [y_0 - 2n^{-1}\dot{x}_0] + 2n^{-1}\sqrt{\dot{x}_0^2 + n^2\dot{x}_0^2} \cos\left(nt + \tan^{-1}\left(n\frac{x_0}{\dot{x}_0}\right)\right)$$

$$z(t) = n^{-1}\sqrt{\dot{z}_0^2 + n^2\dot{z}_0^2} \sin\left(nt + \tan^{-1}\left(n\frac{z_0}{\dot{z}_0}\right)\right).$$

This set of equations describes an elliptic cylinder (i.e., a cylinder with an ellipse as base) whose base is parallel to the $xy$-plane and whose center is located at $(0, y_0 - 2\dot{x}_0/n, 0)^T$. When the phases of the in-plane and out-of-plane motion match, the equation

$$\tan^{-1}\left(n\frac{x_0}{\dot{x}_0}\right) = \tan^{-1}\left(n\frac{z_0}{\dot{z}_0}\right)$$

holds and the three-dimensional relative orbit is a section of the elliptic cylinder.

Various well-known elliptical formation geometries can be constructed. It is for instance possible to create a circular relative orbit. This general circular orbit (GCO) is created when $x^2(t) + [y(t) - (y_0 - 2n^{-1}\dot{x}_0)]^2 + z^2(t) = \text{const.}$ and requires that

$$\frac{x_0}{\dot{x}_0} = \frac{z_0}{\dot{z}_0}$$

$$\sqrt{\dot{z}_0^2 + n^2\dot{z}_0^2} = \sqrt{3} \sqrt{\dot{x}_0^2 + n^2\dot{x}_0^2}.$$ 

The resulting three-dimensional circle has a radius equal to $2n^{-1}\left[\dot{x}_0^2 + n^2\dot{x}_0^2\right]^{1/2}$.

It is also feasible to create a relative orbit whose projection on the $yz$-plane is a circle, thus the equation $y^2(t) + z^2(t) = \text{const.}$ has to be satisfied. This relative trajectory is referred to as a projected circular orbit (PCO) and results when
4.3. Simulation Geometries

\[ \frac{x_0}{\dot{x}_0} = \frac{z_0}{\dot{z}_0} \quad (4.4a) \]
\[ \sqrt{\dot{z}_0^2 + n^2 z_0^2} = 2 \sqrt{\dot{x}_0^2 + n^2 x_0^2}. \quad (4.4b) \]

Starting with a coplanar ellipse, a PCO is created by rotating the coplanar ellipse around the \( y \)-axis of the Hill frame with an angle that results in the creation of a circular motion in the \( yz \)-plane.

Since the inter-satellite distance between the chief and the deputy remains constant in a GCO, this formation can be advantageous for interferometric observations where a constant baseline between the spacecraft is important. A PCO can be used for sparse aperture sensing, such as considered for the canceled TechSat-21 mission [Martin and Stallard 1999].

A different elliptical formation geometry can be created when the in-plane and out-of-plane motion are 90 degrees out of phase, resulting in [Alfriend et al. 2010]

\[ \tan^{-1}\left( n \frac{z_0}{\dot{z}_0} \right) = \pi/2 + \tan^{-1}\left( n \frac{x_0}{\dot{x}_0} \right). \quad (4.5) \]

This elliptical formation geometry is commonly referred to as a safe ellipse or as eccentricity/inclination vector separation and has originally been developed for the collocation of geostationary satellites [Eckstein et al. 1989]. It is regarded to be a passively safe relative orbit since the radial and along-track separation between the chief and the deputy never become zero at the same time and thus the spacecraft cannot collide with each other provided that the magnitude of the out-of-plane motion is sufficiently large. This collision avoidance based on nonzero radial and out-of-plane separation is preferable over collision avoidance based on nonzero along-track separation. The reason is that, to first order, the relative motion in radial and cross-track direction is purely periodic while the relative motion in along-track direction is in practice quasi-periodic due to secular drift caused by a nonzero differential semi-major axis. Since achieving a nonzero differential semi-major axis is very challenging in a LEO environment, guaranteeing a non-zero along-track separation is difficult and therefore not advisable for a passive collision avoidance scheme. The safe elliptical formation geometry is extensively discussed in [D’Amico 2010]. There, the formulation using parallel eccentricity and inclination vectors is also used.

Starting with a coplanar ellipse, a safe ellipse is created by rotating the coplanar ellipse around the \( x \)-axis of the Hill frame, resulting in a linear motion in the \( yz \)-plane. An example of a safe elliptical formation is presented in Fig. 4.2.

As described in [Maessen and Gill 2010b], there is no major difference in the relative navigation results between a PCO and a safe ellipse for the same inter-satellite distance. In addition, as the amount of relative out-of-plane motion in a GCO and PCO is tightly coupled to the amount of in-plane motion, these formation geometries do not lend themselves well to study the influence of the amount of out-of-plane motion on the observabil-
Figure 4.2: Example of a safe elliptical formation geometry in three dimensions (bold lines) and its two-dimensional projections (thin lines) on the various planes. The position of the chief is indicated by a cross and the position of the deputy at $t = 0$ is indicated by a circle. The size of the ellipse is $2 \times 1$ km and the amplitude of the relative out-of-plane motion is 1 km.

ity of the out-of-plane states since a change in out-of-plane motion also forces a change in inter-satellite distance, which is not desired. For these reasons, it has been decided to not consider the PCO and GCO in detail for this study and to focus on the safe ellipse as being representative for the family of elliptical formation geometries.

The state vector at $t = 0$ for the safe ellipse is provided in Eq. (4.6):

$$x_0 = \begin{pmatrix} r_e & 0 & 0 & -2nr_e & [-nr_e, 0] \end{pmatrix}^T.$$  

(4.6)

The out-of-plane velocity $\dot{z}_0$ is varied within the range $[-nr_e, 0]$, which results in a maximum out-of-plane amplitude $z_{\text{max}}$ equal to $r_e$. Of course, the term safe ellipse does not hold for small values of $z_{\text{max}}$.

4.3.2 Pendulum Formation

When the absolute orbits of both spacecraft are circular, a spacecraft formation can be created by means of a small difference in the argument of latitude, which is the sum of the argument of perigee and the mean anomaly. This results in a so-called along-track formation since the spacecraft are only separated in along-track direction. However, this formation can be regarded as a sub-case of the more general case where there is also differential inclination and/or RAAN. Then, there is a constant along-track separation and a relative motion only in the out-of-plane direction. Again, this motion is a harmonic oscillation with a period equal to the orbital period of the chief. This formation geometry is
known as a *pendulum* formation. A feature of pendulum formations that is advantageous for Earth observation missions is that a proper sizing of the amplitude of the out-of-plane motion results in overlapping ground tracks for the two spacecraft. These will thus be able to observe the same geolocation at the same angle and at almost the same instance in time, which allows for synergy between the observations performed with each individual spacecraft. Such a pendulum formation is commonly referred to as an in-track formation [Sabol et al. 2001] and is applied in the A-Train mission. An example of a pendulum formation is presented in Fig. 4.3.

![Figure 4.3: Example of a pendulum formation geometry in three dimensions (bold lines) and its two-dimensional projections (thin lines) on the various planes. The position of the chief is indicated by a cross and the position of the deputy at $t = 0$ is indicated by a circle. The along-track separation and the amplitude of the relative out-of-plane motion are both 1 km.](image)

For the pendulum formation to exhibit 1:1 commensurate motion, again Eq. (3.20), $\dot{y}_0 = -2n x_0$, has to be satisfied. Maintaining a constant along-track distance requires that $\dot{x}_0^2 + n^2 x_0^2 = 0$, cf. Eq. (4.1b). This condition leads to $x(t) = 0$, thus there is no relative motion in radial direction. In addition, this condition can only be satisfied when $x_0 = x_0 = 0$. Since $\dot{y}_0 = -2n x_0$, this also results in $y_0 = 0$. Essentially, the deputy can be leading or trailing the chief, but this is of little concern here since it does not influence the analysis in any way. It has been chosen to have the deputy trail behind the deputy, leading to a negative value for the along-track distance. The resulting state vector at $t = 0$ for the pendulum formation is provided in Eq. (4.7):

$$x_0 = \begin{pmatrix} 0 & -r_e & 0 & 0 & [nr_e, 0] \end{pmatrix}^T.$$  

(4.7)
4.4 Analysis

In this section the results of various simulations executed for the scenarios presented earlier in this chapter are presented and analyzed in detail. The first subsection treats in detail how the observability of the system as a whole and of the individual state variables depends on the various factors. The second subsection provides results and empirical relations for the accuracy in the relative navigation as a function of the inter-satellite range, antenna baseline, ranging accuracy, and formation geometry. The third subsection explores the behavior of the estimator when the magnitude of the relative out-of-plane motion of the two satellites is reduced. These three subsections provide all the information necessary to answer the research questions. The fourth and fifth subsection deal with investigations that have been conducted to further explore several interesting topics that have arisen while conducting the research.

In all subsections, the maximum observations arc length considered is five orbits and many results are described for this observation arc length. The reason for this is that the estimation results are stable and very repetitive for this observation arc length. Thus, the estimation process itself is regarded to have converged sufficiently at this observation arc length. Yet, some investigations still require averaging of the results over many (one thousand or ten thousand) simulator runs to enable the construction of accurate trendlines, such as those provided in subsection 4.4.4.

Furthermore, since the only type of estimator divergence that can occur for this estimator is apparent divergence, cf. subsection 3.4.4, the term ‘apparent estimator divergence’ is often abbreviated to ‘estimator divergence’ or simply ‘divergence’.

4.4.1 Observability of the Relative States

For both formation geometries studied, the change in the condition number $\kappa$ and in the eigenvalues $\upsilon_i$ and eigenvectors $v_i$ of the Gramian matrix $\tilde{G}$ for changing design parameters provides a valuable insight in the problem at hand. Since the system has been linearized, the state that is effectively estimated at each iteration is $\Delta x_0^k$ with $k$ indicating the iteration step. Since $\Delta x_0^k \neq \Delta x_0^{k-1}$, the observability changes with each iteration step. The observability analyses for the various scenarios have been performed for $x_0 + \Delta x_0^0$, the relative state that is estimated in the first iteration step.

As already discussed in subsection 3.4.3, the eigenvectors and eigenvalues of the Gramian matrix indicate the relative observability of the various state components. The most observable ‘direction’ is the eigenvector corresponding to the largest eigenvalue and the least observable ‘direction’ is the eigenvector corresponding to the smallest eigenvalue. In the ideal case, each state component is chosen such that it corresponds to a system state that has a physical meaning (in case of absolute orbit determination these ‘natural’ states would be the semi-major axis, eccentricity, etc.). Then, all eigenvectors would have a single value close to 1 with all other values close to zero. Thus, as an example, for a system with three state components and state vector $(x_1, x_2, x_3)^T$ the eigenvectors could be
\( \mathbf{v}_1 \approx (1, 0, 0)^T, \mathbf{v}_2 \approx (0, 1, 0)^T, \) and \( \mathbf{v}_3 \approx (0, 0, 1)^T. \) Since \( \mathbf{v}_1 > \mathbf{v}_2 > \mathbf{v}_3, \) the most observable direction is \( \mathbf{v}_1 \) and the least observable direction is \( \mathbf{v}_3. \) There would also be almost no cross-correlation in the estimation of the state components. This implies that the estimate of \( x_1, \) the most observable state component in this example, will have a variance that is, e.g., 10 times smaller than the variance of \( x_3, \) the least observable state component, if \( \frac{v_1}{v_3} = 10. \) On the other hand, if the eigenvectors have significant contributions from multiple state components and if the eigenvalues have similar magnitudes, significant cross-correlation is present and the foregoing \( \frac{\sigma^2_{x_3}}{\sigma^2_{x_1}} = 10 \) when \( \frac{v_1}{v_3} = 10 \) no longer applies. This makes it more difficult to understand the behavior of the system.

**Eigenvector directions**

Figure 4.4 provides an example of how the directions of the six eigenvectors of the Gramian matrix \( \tilde{G} \) change with observation arc length. The depicted results have been obtained for a \( 1 \times 2 \times 1 \) km elliptical formation geometry with an antenna baseline of \( d \approx 12 \) dm. The figure shows that up to roughly one orbit, the direction of the eigenvectors can vary dramatically. However, the direction of the eigenvectors stabilizes for longer observation arc lengths. This is due to the fact that the relative motion of the spacecraft repeats after every orbit and thus little extra information is added after more than one orbit. The figure also shows that, in general, the eigenvectors of the system are not directed to a single state component, but rather to a subset of all state components. In addition, the direction of most eigenvectors changes considerably if the geometry of the formation is changed, thus the eigenvectors depicted in Fig. 4.4 are by no means representative for all formation geometries, except for \( \mathbf{v}_1. \) The direction of that eigenvector does not change when the formation geometry is altered and thus represents a special combination of state components that is significant for this system. What this combination represents is briefly discussed later in this subsection.

Figure 4.4 also partially shows how accurate the various state components will be estimated at this iteration step for a certain observation arc length: The main components of eigenvector \( \mathbf{v}_1, x_0, \) and \( y_0, \) will be estimated with relatively high accuracy while the main components of eigenvector \( \mathbf{v}_6, z_0, y_0, \) and \( x_0, \) will be estimated with relatively low accuracy. However, since a state component can be a significant component in multiple eigenvectors (caused by cross-correlation), for instance \( x_0 \) in \( \mathbf{v}_2 \) and in \( \mathbf{v}_6, \) the ultimate accuracy with which each state component can be estimated depends on the relative magnitude of all eigenvalues since, due to the inversion of the matrix \( \Lambda + \tilde{H}^T W \tilde{H}, \) it is the result of a superposition of all eigenvector/eigenvalue combinations.

The plots in Fig. 4.5 show how the direction of \( \mathbf{v}_6, \) the least observable eigenvector, changes due to an increasing value for the ratio \( z_{\text{max}}/r_e, \) the amplitude of the relative out-of-plane motion normalized by the inter-satellite distance at epoch. Figures 4.5(a) and 4.5(b) show how the results for the safe ellipse and for the pendulum formation, respectively. The plots have been obtained for a scenario with a measurement arc length of five orbits, \( r_e = 1 \) km, and three Rx antennas with \( d \approx 12 \) dm. These results are very similar for all other scenarios investigated.
Figure 4.4: Example of the change in the direction of the eigenvectors with changing observation arc length for $\Delta x^0_0$. The depicted results have been obtained for a $1 \times 2 \times 1$ km elliptical formation geometry with antenna baseline $d \approx 12$ dm. Note that for clarity the absolute values of the eigenvector components are depicted.

For both formation geometries, Figs. 4.5(a) and 4.5(b) show that $\nu_6$ is directed mainly towards an out-of-plane state component for values of $z_{\text{max}}/r_e$ up to and including $z_{\text{max}}/r_e = 1$. However, as $z_{\text{max}}/r_e$ increases, the vector rotates in such a way that other state components, especially the along-track separation $y$, become more and more prominent. Thus, due to the increased out-of-plane motion, or better, the increased ratio $z_{\text{max}}/r_e$, the observability of the out-of-plane state components improves while the observability of the other state components deteriorates relative to the observability of the out-of-plane state components. By increasing $z_{\text{max}}/r_e$ even more, $y$ becomes the dominant direction in $\nu_6$ and thus the least observable state component. This change in the direction of the eigenvectors also occurs for other eigenvectors. Thus, the out-of-plane state components will become the dominant directions in other, more observable, eigenvectors. Not shown here is that for the range of $z_{\text{max}}/r_e$ considered, the largest component in the one but least observable direction, $\nu_5$, is also an out-of-plane state component.

The sudden change in the direction of $\nu_6$ in Fig. 4.5(b) around $z_{\text{max}}/r_e = 0.2$ from $z$ to $\dot{z}$ also occurs for $\nu_5$ but then from $\dot{z}$ to $z$. In fact, the shape of the plot for $\nu_5$ is very similar to the shape of the plot for $\nu_6$, but then the curves for $z$ and $\dot{z}$ are similar to the curves for $\dot{z}$ and $z$ for $\nu_6$, respectively. In addition, the ratio $\nu_5/\nu_6$, which is never very large, cf. Fig. 4.6(a), is at a minimum (1.002) for $z_{\text{max}}/r_e = 0.2$. This change in the direction of eigenvectors $\nu_5$ and $\nu_6$ and the minimum value for the ratio $\nu_5/\nu_6$ also consistently
occurs for the other pendulum formation scenarios around \( z_{\text{max}}/r_e = 0.2 \). It is therefore concluded that for the pendulum formation geometry the level of observability for the out-of-plane state components 'flips' around \( z_{\text{max}}/r_e = 0.2 \), causing the vector which was \( \nu_6 \) for \( z_{\text{max}}/r_e > 0.2 \) to become \( \nu_5 \) for \( z_{\text{max}}/r_e < 0.2 \). It is unclear why this 'flip' occurs for \( z_{\text{max}}/r_e = 0.2 \). However, since the eigenvalues of the out-of-plane state components are always very similar, this change in relative observability has little influence on the observability of the system as a whole.

**Eigenvalues**

The plots in Fig. 4.6 provide insight in the absolute and relative magnitude of the eigenvalues of the system. Again, the plots have been obtained for a scenario with a measurement arc length of five orbits, \( r_e = 1 \text{ km} \), and \( d \approx 12 \text{ dm} \).

In Figs. 4.6(a) and 4.6(b), the magnitudes of eigenvalues \( \nu_1, \nu_2, \nu_3, \) and \( \nu_4 \) are shown to be independent on \( z_{\text{max}}/r_e \), which is expected due to the lack of coupling between the in-plane and out-of-plane motion, while the magnitudes of eigenvalues \( \nu_5 \) and \( \nu_6 \) are clearly dependent on \( z_{\text{max}}/r_e \). The reason that eigenvalues \( \nu_1 \) to \( \nu_4 \) can be expected to be independent of the amount of relative out-of-plane motion is that their dominant components, cf. Fig. 4.4, are always in-plane state components.

With decreasing \( z_{\text{max}}/r_e \), the ratio \( \nu_4/\nu_5 \) increases and is coupled to a rotation of \( \nu_5 \) and \( \nu_6 \) towards the out-of-plane state components. This results in a poor conditioning of the Gramian matrix as two of the six eigenvalues have a small value and thus will both have a large contribution in numerical errors when inverting the matrix \( \Lambda + \tilde{H}^T W \tilde{H} \). For the safe ellipse, the ratio \( \nu_4/\nu_5 \) needs to be smaller than \( 10^2 \), which occurs approximately for \( z_{\text{max}}/r_e > 0.2 \), to influence the final covariance matrix such that estimator divergence has a low probability of occurrence (note that no relation has been found between this observation and the 'flip' in the observability of the out-of-plane state components which also occurs around \( z_{\text{max}}/r_e = 0.2 \) as discussed on the previous page). For the pendulum
formation, the ratio $\nu_4/\nu_5$ needs to be smaller than $10^4$, which occurs approximately for $z_{\text{max}}/r_e > 0.01$, for estimator divergence to have a low probability of occurrence. As the ratio $\nu_4/\nu_5$ increases, the conditioning of the matrix $\Lambda + \tilde{H}^T W \tilde{H}$ becomes worse and thus the probability of estimator divergence increases. For scenarios with one Rx antenna, the ratio $\nu_4/\nu_5$ scales with $r_e/z_{\text{max}}$ while for scenarios with three Rx antennas it scales with $r_e/(d z_{\text{max}})$, which indicates the influence of angular information on the estimation. The probability of estimator divergence as a function of $z_{\text{max}}/r_e$ is explored in more detail further in this subsection and in subsection 4.4.3.

The foregoing indicates that for actual space missions, in case the system design of the formation has already been fixed at a (too) small amplitude for the relative out-of-plane motion, the only ways to ensure that the matrix $\Lambda + \tilde{H}^T W \tilde{H}$ is well-conditioned are to reduce the a-priori covariances of the out-of-plane states or to remove these from the relative state estimate, resulting in a reduced-state estimate. For the first option, the result will be that the accuracy in the estimate for the out-of-plane states will not improve beyond the improved a-priori covariance. Naturally, this requires that the a-priori information can be improved upon, which might not be feasible. If not, only the second option remains, which is only an option if the relative out-of-plane motion is of little importance both from a mission point of view as well as from a formation safety point of view.

Figures 4.6(a) and 4.6(b) also show that $\nu_1$ is always much larger than the other eigenvalues. In subsection 3.4.3, it was already stated that this indicates that the sensors are observing an unspecified state, which is equal to the linear combination of the states in that vector. In [Maessen and Gill 2010a], it is shown that this is indeed the case and that the most observable direction of the system, $(0.89, 0.01, 0.00, 0.01, 0.45, 0.00)^T$, is in fact the relative semi-major axis (the difference between the semi-major axes of the two spacecraft orbits) which is equal to $4x + 2\dot{y}/n$ [Alfriend et al. 2010]. A likely cause for the relative semi-major axis to be a 'natural' direction for this system is the fact that it has a large influence on the in-plane relative motion: If it is nonzero, there will be an along-track drift and a constant offset in radial direction, cf. subsection 3.2.2.
Condition number

Another finding from the observability analysis is that for all cases where estimator divergence can occur, the condition number of $\tilde{H}$ following from the observability analysis is at most three orders of magnitude larger than for the cases where estimator divergence does not occur, which is a relatively small difference considering that the condition number can be in the order of $10^{12}$ for certain scenarios. More importantly, the magnitude of the condition number is always smaller than required by e.g. MATLAB® for the matrix to be considered ill-conditioned and therefore highly susceptible to numerical errors when inverted.

This is shown in Fig. 4.7 where the observability of the elliptical and pendulum formations is depicted for all Rx antenna baselines considered, for all values of $r_e$, and for $z_{\text{max}} = r_e$ and $z_{\text{max}} = 0$. For the most observable formation geometry, $z_{\text{max}} = r_e$, the condition number for all cases lies between $10^4$ and $10^{10}$ for the entire observation arc with the very high condition numbers occurring for a single Rx antenna and measurement arc lengths shorter than half an orbit. For the least observable formation geometry, $z_{\text{max}} = 0$, the condition number for all cases lies between $10^4$ and $10^{12}$ for the entire observation arc. Using the default SVD-based rank computation performed by MATLAB®, the upper and lower limits for rank-deficiency of $\tilde{H}$, based on a machine accuracy of $2^{-52}$ at the number 1.0 for double precision metrics, lie between $\kappa = 1.2 \cdot 10^{15}$ after 10 measurements and $\kappa = 1.4 \cdot 10^{12}$ after 500 measurements. Since $\tilde{H}$ is still far from being considered rank-deficient at the observed condition numbers, the decrease in observability for $z_{\text{max}} = 0$ would normally not be a particular reason for concern. However, simulation results prove the opposite: When attempting to estimate the relative state using the iterative batch LSQ estimator described earlier, the estimator shows divergence for the ratios of $z_{\text{max}}/r_e$ mentioned earlier.

The reason that the condition number of $\tilde{H}$ does not provide sufficient information to predict estimator divergence for this investigation is twofold: Firstly, the estimator uses an iterative scheme in which the error in the state estimate is expected to decrease in each iteration step. Thus, iteration step $k$ starts with a new state estimate, which is the result of iteration step $k-1$. Therefore, for each iteration step the observability of the system is different and a system that is observable at step $k = 0$ can become unobservable after several iteration steps. Secondly, the condition number only takes into account the largest and the smallest eigenvalue of the system. As has been shown in Fig. 4.6, eigenvalue $\nu_5$ has almost the same value as eigenvalue $\nu_6$, and thus also contributes significantly to the conditioning of the matrix $\Lambda + \tilde{H}^T W \tilde{H}$.

To take these two factors better into account, the following two recommendations are made: Since an iterative nonlinear estimator is commonly intended to converge to a solution that is in the limit (i.e., after an infinite number of iterations) equal to the true relative state, it is considered to be good practice to determine the observability of the true relative state. If this state is poorly observable, then estimator divergence at some stage of the iteration is very likely when attempting to estimate that state. In an actual real-life implementation this is of course not feasible since the true relative state is unknown, but during the mission design phase this can be done for the desired relative state. The sec-
ond recommendation is to use a different ‘single value’ observability metric, such as one that takes into account all eigenvalues of the system. This would be the determinant of the Gramian matrix. Since the determinant of a matrix is equal to the multiplication of its eigenvalues [Chen 1999] it will thus also provide information on the variation of $v_5$, which is of importance here since its value is close to that of $v_6$ and thus contributes considerably to the numerical conditioning of the Gramian matrix. However, a downside of using the matrix determinant is that there is no minimum value below which a matrix can be considered rank deficient. In addition, for the scenarios studied in this research, the results from subsection 4.4.3 show that this metric does not provide more insight in the probability of estimator divergence as a function of $z_{\text{max}}/r_e$ than the condition number.

**Probability of apparent estimator divergence**

Even though it has been concluded that the condition number of $\tilde{H}$ is not a very effective observability criterion for the system under study, it still provides a number of useful insights. Firstly, Fig. 4.7 shows that a distinction must be made between ‘short’ (< 1 orbit) measurement arcs and ‘long’ measurement arcs. In the former case namely, a larger antenna baseline always leads to a significantly smaller condition number and therefore pos-

![Figure 4.7: Condition numbers of $\tilde{H}$ for $\Delta x_0^0$ in case of (a) a safe ellipse and (b) a pendulum formation with $z_{\text{max}} = r_e$ and in case of (c) a safe ellipse and (d) a pendulum formation with $z_{\text{max}} = 0$. The top, middle, and lower plots in each figure depict the results for $r_e = 100$ m, $r_e = 1$ km, and $r_e = 10$ km, respectively.](image-url)
Table 4.2: Selected results for the observability of the relative state in case of a single Rx antenna. The condition number is provided for various locations of the Rx antenna in the Hill frame. The formation geometry is an ellipse with \( z_{\text{max}}/r_e = 1 \). The observation arc length is five orbits.

<table>
<thead>
<tr>
<th>Antenna location [m]</th>
<th>Condition number (/10^5) [-]</th>
<th>( r_e = 0.1) km</th>
<th>( r_e = 1) km</th>
<th>( r_e = 10) km</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x ) ( y ) ( z )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 \ 0 \ 0</td>
<td>7.25</td>
<td>7.25</td>
<td>7.25</td>
<td></td>
</tr>
<tr>
<td>-10 \ 0 \ 0</td>
<td>7.29</td>
<td>7.25</td>
<td>7.25</td>
<td></td>
</tr>
<tr>
<td>10 \ 0 \ 0</td>
<td>7.14</td>
<td>7.24</td>
<td>7.25</td>
<td></td>
</tr>
<tr>
<td>0 \ -10 \ 0</td>
<td>12.95</td>
<td>7.65</td>
<td>7.29</td>
<td></td>
</tr>
<tr>
<td>0 \ 10 \ 0</td>
<td>4.67</td>
<td>6.88</td>
<td>7.21</td>
<td></td>
</tr>
<tr>
<td>0 \ 0 \ -10</td>
<td>3.98</td>
<td>6.72</td>
<td>7.19</td>
<td></td>
</tr>
<tr>
<td>0 \ 0 \ 10</td>
<td>14.56</td>
<td>7.85</td>
<td>7.30</td>
<td></td>
</tr>
</tbody>
</table>

ivitively influences the observability of the problem for both formation geometries. Thus, for short measurements arcs, a larger antenna baseline will result in a smaller probability of estimator divergence. Secondly, Fig. 4.7 shows a clear distinction in observability for large and small out-of-plane motion. For scenarios where \( z_{\text{max}} = r_e \), Figs. 4.7(a) and 4.7(b) indicate that for 'long' measurement arcs, the antenna baseline has little influence on the probability of estimator divergence. When \( z_{\text{max}} = 0 \) however, Figs. 4.7(c) and 4.7(d) show that a large antenna baseline does significantly improve the observability for 'long' measurement arcs, although this effect diminishes when the inter-satellite distance increases. The probability of apparent estimator divergence is treated in more detail in subsection 4.4.3.

Influence of the location of the Rx antenna(s)

For the scenarios with a single Rx antenna, not locating the antenna at coordinates (0,0,0) in the Hill frame, but giving it a fixed offset in \( x \)-, \( y \)-, or \( z \)-direction leads to slightly different results for the observability of the relative state for the pendulum and the elliptical formation geometry. For the pendulum formation geometry, the change in Rx antenna placement does not have a significant influence on the observability of the relative state. The reason for this is detailed in subsection 4.4.2. For the elliptical formation geometry however, there is a noticeable change in the observability of the relative state when changing the antenna location. This is shown in Table 4.2 where the condition number of the Gramian is provided for various locations for the Rx antenna. Especially large offsets in \( y \)- and/or \( z \)-direction have a noticeable impact on the observability of the system. In this example, this is the case when \( r_e = 100 \) m. More general, this is the case if the difference between the inter-satellite and the antenna offset is considerably less than two orders of magnitude. This effect leads to interesting results for the relative position determination, which will be treated in detail at the end of subsection 4.4.2.
Chapter 4. Relative Navigation Simulation and Analysis

4.4.2 Relative Position Estimation Results

As already discussed in section 4.1, it is desirable to represent the relative navigation results using a single figure of merit. As the relative in-plane and out-of-plane motion are decoupled, the figure of merit should at least contain one in-plane state component and one out-of-plane state component. In addition, it is preferable for the figure of merit to have a physical meaning that is intuitive to relate to. For these reasons, the figure of merit has been chosen to be $\sigma_{\text{RSS}}$, the three-dimensional root-sum square (RSS) of the standard deviations of the relative position estimates. It provides sufficient information on the convergence of the total state estimate and provides an intuitive understanding of how accurate the state estimate actually is. It is defined as

$$\sigma_{\text{RSS}} = \sqrt{\sigma_{x_0}^2 + \sigma_{y_0}^2 + \sigma_{z_0}^2}. \quad (4.8)$$

**Empirical relationships**

Empirical relationships that describe the accuracy in the estimation of the relative state for the two formation geometries under study are provided in Table 4.3 for a measurement arc length of five orbits. To obtain these results, the maximum out-of-plane motion $z_{\text{max}}$ has been varied between four different values: 0, 0.01$r_e$, 0.1$r_e$, and $r_e$. The empirical relations have been obtained by means of a least squares fit to the simulation data. The relations thus obtained have subsequently been slightly modified to yield equations in which the contributions of the various parameters are explicitly present (e.g., the number of measurements, the number of Rx antennas) and where the relation between the parameters is physically meaningful (e.g., linear, quadratic). For instance, if the least squares fit yields an expression where $\sigma_{\text{RSS}}$ is almost proportional to $r_e/z_{\text{max}}$, then the relation is modified such that $\sigma_{\text{RSS}} \propto r_e/z_{\text{max}}$, even if this would imply an equation with a slightly reduced fit to the data. The mean and maximum relative errors (%) in the empirical relations for the two formation geometries are provided in Table 4.4 and in Table 4.5.

Table 4.3 shows that for both formation geometries and for small ratios of $z_{\text{max}}/r_e$, estimator divergence is likely to occur: With decreasing magnitude of $z_{\text{max}}$, the estimation becomes more and more erratic and at a certain point starts to show divergence for the out-of-plane state components. The probability that this occurs generally increases with decreasing $z_{\text{max}}/r_e$. Thus, when considering for instance the safe ellipse, estimator divergence occurs regularly for $z_{\text{max}}/r_e = 0.1$, but is expected to occur much more often for $z_{\text{max}}/r_e = 0.001$. Estimator divergence has been observed to occur for all possible measurement arc lengths considered. In addition, it occurs for a single Rx antenna as well as for three Rx antennas on the chief with the important difference that the antenna baseline $d$ influences the probability of estimator divergence, as already discussed in subsection 4.4.1.

In case of a single Rx antenna located at the origin of the Hill frame, Table 4.3 shows the following. For an elliptical formation geometry, the information provided by the measurements is of such poor quality that $\sigma_{\text{RSS}}$ hardly improves beyond $\sigma_{\text{apr}}$ for $z_{\text{max}} = r_e$. This implies that $\sigma_{\text{RSS}}$ is not, or only very weakly, a function of $r_e$ and $\sigma_{\rho}$ for this particular
Table 4.3: Empirical relations for the estimation accuracies for a measurement arc length of five orbits (i.e., 500 observations per receiver channel).

<table>
<thead>
<tr>
<th>Formation</th>
<th>$z_{\text{max}}/r_e$</th>
<th>Single Rx antenna</th>
<th>Three Rx antennas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safe ellipse</td>
<td>1</td>
<td>$\sigma_{\text{RSS}} \approx 0.95\sigma_{\text{apr}}$</td>
<td>$\sigma_{\text{RSS}} \approx 0.95\sigma_{\text{apr}}(1 - e^{-1.3\sigma_{\rho}/d})$</td>
</tr>
<tr>
<td></td>
<td>$\leq 0.1$</td>
<td>divergence likely</td>
<td>divergence likely</td>
</tr>
<tr>
<td>Pendulum</td>
<td>[0.01, 1]</td>
<td>$\sigma_{\text{RSS}} \approx \frac{3}{\sqrt{500}} \frac{r_e\sigma_{\rho}}{z_{\text{max}}}$</td>
<td>$\sigma_{\text{RSS}} \approx \frac{3}{\sqrt{3\sqrt{500}}} \frac{r_e\sigma_{\rho}}{z_{\text{max}}}$</td>
</tr>
<tr>
<td></td>
<td>$&lt; 0.01$</td>
<td>divergence likely</td>
<td>divergence likely</td>
</tr>
</tbody>
</table>

Scenario. For the pendulum formation the result is very different: Now the measurements obtained using a single Rx antenna do provide sufficient information to increase the accuracy in the estimation well beyond $\sigma_{\text{apr}}$. Because $\sigma_{\text{RSS}} = \sqrt{3}\sigma_{\text{apr}}$, it is evident that for both formation geometries, the information obtained using a single Rx antenna on the chief does allow for an improvement in the relative state estimate beyond that of the a-priori relative state estimate. However, in case of an elliptical formation the improvement is rather small and capped around $\sigma_{\text{apr}}$. Thus, for that formation geometry at least some state components cannot be estimated accurately and are therefore poorly observable. However, when the Rx antenna is not required to be located at the origin of the Hill frame, as it is now, the results become very different. This is discussed at the end of this subsection.

In case of three Rx antennas on the chief, the following results have been obtained. For a pendulum formation with a relatively large out-of-plane motion, Table 4.3 shows that there is no dependency in the estimation accuracy on the antenna baseline $d$ but that there is a dependency on the out-of-plane motion in the form of $z_{\text{max}}$. Thus, three Rx antennas provide no geometrical advantage over a single Rx antenna in case of a long measurement arc and significant out-of-plane motion. The reason for this is explained later in this subsection. A minor advantage comes from the increased number of measurements per measurement epoch (three instead of one), leading to a factor $\sqrt{3}$ improvement in the estimation accuracy. In addition, the accuracy in the estimation scales linearly with $r_e/z_{\text{max}}$. This is explained by the observation, see subsection 4.4.1, that the out-of-plane states ($z$ and $\dot{z}$) are the least observable directions for this formation geometry and thus a motion that results in a large magnitude for these directions is needed to obtain an accurate estimate of the out-of-plane components. The equation for the safe ellipse in case of three Rx antennas will be explained in detail in the next part of this subsection.

Important to note is that for $r_e = 100$ m and $z_{\text{max}}$ between 1 m and 10 m, the empirical relations for the pendulum formation do not hold if $d \approx 12$ m. This is due to the out-of-plane motion being on the same order of magnitude as the antenna baseline. Thus, the large antenna baseline in this case does provide a geometrical advantage and offers a very good observation geometry that significantly improves the estimation accuracy, which can be exploited, e.g., for rendezvous and docking operations. This result is further investigated in subsection 4.4.4.
Implications for the system design of a formation

The nonlinear empirical relationship in Table 4.3 for the safe ellipse and three Rx antennas is influenced by the information matrix $\Lambda$ that is used in the estimator. Figure 4.8 shows that the empirical exponential function closely follows the simulation data for all three Rx antenna baselines in case $z_{\text{max}} = r_e = 100$ m. Because the information matrix $\Lambda$ is used in the estimator, the slope of the estimation result gradually reduces for increasing $\sigma^d$ due to the damping influence of the a-priori information as it starts to dominate the information obtained from the measurements. For large values of $\sigma^d$, the result settles around $\sigma^{\text{apr}}$ and is thus no longer better than what can be achieved using a single Rx antenna. This is understood by the fact that now the benefit of having multiple antennas, obtaining angular information, is almost completely negated since the measurement error will often be (much) larger than the antenna baseline. Thus, practically, such a system design cannot be expected to be much better than a single antenna in providing angular information. This is confirmed by the estimation results in Table 4.3.

The exponential relation connects the two extreme conditions satisfactorily since it reduces to $\sigma_{\text{RSS}} \approx \sigma^{\text{apr}}$ in case of large values for $\sigma^d$ while it reduces to $\sigma_{\text{RSS}} \approx (r_e \sigma^d) / (8d)$ for small values for $\sigma^d$. The latter result is achieved by means of a first-order Taylor series expansion for the exponential function and using $\sigma^{\text{apr}} = 0.1 r_e$. For small values of $\sigma^d$ however, there exists a linear relationship between $\sigma_{\text{RSS}}$ and $\sigma^{\text{apr}}$, $\sigma^d$, and $d$. Then, the linear approximation $\sigma_{\text{RSS}} \approx (r_e \sigma^d) / (9d)$ provides a better fit for the simulation data than the exponential function, cf. Fig. 4.8. The linear approximation holds approximately up to $\sigma^d = 0.25$ for this particular choice of the a-priori covariances. A more accurate a-priori knowledge causes a deviation of $\sigma_{\text{RSS}}$ from a linear relation with $\sigma^d$ at even smaller values for $\sigma^d$. Thus, a proper system design should ensure that $\sigma^d \leq 0.25$ in order to obtain good angular information for an elliptical formation geometry.

For a single range measurement, using the results from subsection 3.4.5 the statistical result for $\sigma_{\text{RSS}}$ is

$$\sigma_{\text{RSS,statistical}} = \sqrt{\sigma_{x_{\text{Tx}}}^2 + \sigma_{y_{\text{Tx}}}^2 + \sigma_{z_{\text{Tx}}}^2} = \sqrt{\frac{2r^2 \sigma_{\rho}^2}{d^2} + \frac{2r^2 \sigma_{\rho}^2}{d^2} + \frac{5r^2 \sigma_{\rho}^4}{d^4}} \approx \frac{2r \sigma_{\rho}}{d} \quad (4.9)$$

if $d_1 = d_2 = d$, $\sigma_{\rho_0} = \sigma_{\rho_1} = \sigma_{\rho_2} = \sigma_{\rho}$, $r_0 = r_1 = r_2 = r$, and $\sigma_{\rho} \ll d$. In case of $l$ uncorrelated measurements, $\sigma_{\text{RSS}}$ is reduced by the factor $l^{-1/2}$. For a measurement arc consisting of 500 measurements, this leads to $\sigma_{\text{RSS,statistical}} \approx (2r \sigma_{\rho}) / (\sqrt{500d}) \approx (r \sigma_{\rho}) / (11d)$. This result is in close agreement with the empirically deduced result for the elliptical formation geometry when using three Rx antennas and small $\sigma^d$, which is $\sigma_{\text{RSS}} \approx (r \sigma_{\rho}) / (9d)$. Thus, for elliptical formation geometries, Eq. (4.9) can be used in the initial design phase of a RNS as discussed here to obtain a rough estimate of the achievable RSS relative position accuracy.
Figure 4.8: RSS standard deviation in the relative position estimation for the safe ellipse for a measurement arc length of five orbits, $z_{\text{max}} = r_e = 100$ m, and varying ranging accuracy. In the top, middle, and bottom plots, the antenna baseline is approximately 12 cm, 12 dm, and 12 m, respectively. Note the different scales for the axes.

Naturally, the empirical relations that have been obtained for the safe ellipse and the pendulum formation in Table 4.3 are not perfect fits to the simulation results. Table 4.4 and Table 4.5 display the mean and maximum relative error of the empirical relations with the simulation results for the two types of spacecraft formations. For each scenario, the error values have been obtained by subtracting the average value for $\sigma_{RSS}$ that resulted after 10,000 Monte Carlo runs from the corresponding empirical relations given in Table 4.3. The value thus obtained was subsequently divided by the average value for $\sigma_{RSS}$ to obtain the relative error.

For the safe ellipse, the table displays only results for $z_{\text{max}} = r_e$ while for the pendulum formation the table displays results over the range $z_{\text{max}}/r_e = [0.01, 1]$. For the safe ellipse, the empirical relation fits the simulation results very well, which is partially due to the fact that the results have only been obtained for $z_{\text{max}} = r_e$. For the pendulum formation, the fit is less good. This has multiple causes. Firstly, the error in the empirical relation is generally $< 10\%$ for small to moderate values of $z_{\text{max}}/r_e$, but quickly increases for $z_{\text{max}}/r_e > 0.5$. Secondly, the general trend in the results is broken when $z_{\text{max}}/d < 0.1$, which happens for small values of $r_e$ and large values of $d$ and/or $\sigma_\rho$. This is the reason for the lack of numbers in the second and third column of the bottom row in Table 4.5 and is discussed in detail in subsection 4.4.4. Thirdly, the function $\sigma_{RSS} = 4\sigma_\rho/(\sqrt{3}\sqrt{500})\cdot(z_{\text{max}}/r_e)^{-0.9}$ has been found to provide a better fit to the data than the selected empirical function (the mean and maximum relative errors for that function are shown between brackets in Table 4.5), but since no explanation has been found for the power of -0.9 in the function this equation has not been implemented.

The last piece of information that Table 4.3 provides is that it is not so much the magnitude of the relative out-of-plane motion that is important for convergence of the rela-
Table 4.4: Mean and maximum relative errors (%) in the empirical relation for the safe ellipse formation for $z_{\text{max}} = r_e$ and three Rx antennas.

<table>
<thead>
<tr>
<th>$d$</th>
<th>$r_e = 0.1$ km</th>
<th>$r_e = 1$ km</th>
<th>$r_e = 10$ km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>max.</td>
<td>mean</td>
</tr>
<tr>
<td>$(150)^{1/2}$ cm</td>
<td>3</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>$(150)^{1/2}$ dm</td>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>$(150)^{1/2}$ m</td>
<td>6</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 4.5: Mean and maximum relative errors (%) in the empirical relation for the pendulum formation for $z_{\text{max}} / r_e = [0.01, 1]$ and three Rx antennas. Between brackets the results for the equation $\sigma_{\text{RSS}} = 4\sigma_{\rho} / (\sqrt{3})\sqrt{500} \cdot (z_{\text{max}} / r_e)^{-0.9}$ are shown.

<table>
<thead>
<tr>
<th>$d$</th>
<th>$r_e = 0.1$ km</th>
<th>$r_e = 1$ km</th>
<th>$r_e = 10$ km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>max.</td>
<td>mean</td>
</tr>
<tr>
<td>$(150)^{1/2}$ cm</td>
<td>15 (12)</td>
<td>37 (24)</td>
<td>15 (11)</td>
</tr>
<tr>
<td>$(150)^{1/2}$ dm</td>
<td>20 (12)</td>
<td>46 (23)</td>
<td>16 (11)</td>
</tr>
<tr>
<td>$(150)^{1/2}$ m</td>
<td>-</td>
<td>-</td>
<td>19 (11)</td>
</tr>
</tbody>
</table>

tive state estimation (and thus for the reliability of the relative navigation) but the angle $z_{\text{max}} / r_e$, which confirms the findings from the observability analysis in subsection 4.4.1. Denoting the maximum along-track distance between the chief and the deputy as $y_{\text{max}}$ and knowing that $r_e = y_{\text{max}}$ for the pendulum formation and that $r_e = 0.5y_{\text{max}}$ for the safe ellipse, the angle $z_{\text{max}} / r_e$ can be converted to the ‘out-of-plane’ angle $z_{\text{max}} / y_{\text{max}}$ for both formation geometries. This out-of-plane angle provides a better geometrical insight into the issues concerning the relative out-of-plane motion than the angle $z_{\text{max}} / r_e$.

**Explanation of the obtained results**

The estimation results from Table 4.3 are explained by performing an observability analysis for a state estimate that is equal to the true relative state $x_0$, which implies $\Delta x_0^k = 0$. In other words, there is no error in the a-priori state estimate. This is the state that the batch LSQ algorithm will ideally converge to when starting from the initial relative state estimate and will be very close to the relative state estimate that has been obtained in the one but last ($k - 1$) iteration. Thus, an observability analysis of the true relative state will yield results that are very close to performing an observability analysis of the relative state estimate that has been obtained in iteration $k - 1$. This is necessary since every run of the batch LSQ algorithm will yield slightly different results and thus the relative state estimate obtained in iteration $k - 1$ will differ between runs, which complicates an analysis. Performing the observability analysis for $x_0$ removes this complication. For this relative state, the rotation of eigenvector $\psi_6$ due to a change in $z_{\text{max}} / r_e$ is very similar to the result depicted earlier in Figs. 4.5(a) and 4.5(b). Thus, the out-of-plane states are still the least observable states. The variation in the magnitude of the eigenvalues is very similar to Fig.
4.4. Analysis

4.6(b) for the pendulum formation, but notably different to Figure 4.6(a) for the safe ellipse. Fig. 4.9 displays the variation in the magnitude of the eigenvalues for the safe ellipse and the pendulum for \( \Delta x_0^k = 0 \). It shows that for the safe ellipse, \( v_6 \) is now not dependent on \( z_{\text{max}} \) and orders of magnitude smaller than the other eigenvalues, except for very small values for \( z_{\text{max}}/r_e \).

**Figure 4.9:** Change in the magnitude of the eigenvalues of the Gramian for \( \Delta x_0^k = 0 \) in case of (a) a safe ellipse and (b) a pendulum formation as a function of \( z_{\text{max}}/r_e \) for a measurement arc length of five orbits, \( r_e = 1 \) km, and \( d \approx 12 \) dm.

Furthermore, for the safe ellipse, the value of \( v_6 \) is a function of \( r_e \) and \( d \): It decreases two orders of magnitude for one order of magnitude increase in \( r_e \) or a one order of magnitude increase in \( d \). Due to the very small value of \( v_6 \), the most prominent state components in eigenvector \( v_6 \) are estimated with very poor accuracy and the relative magnitude of the standard deviation in their estimate is the same as their relative magnitude in vector \( v_6 \). These state components also dominate the overall accuracy in the relative state estimate for the safe ellipse. This is the reason that the antenna baseline \( d \) is an important parameter in the empirical relation for the safe ellipse. Since for the pendulum formation, the value of \( v_6 \) is not a function of \( d \) and since the magnitude of \( v_6 \) is comparable to that of the other eigenvalues, the antenna baseline does not influence the accuracy in the relative state estimate for that formation geometry.

Figure 4.10(a) shows that for the elliptical formation geometry and \( \Delta x_0^k = 0 \), an order of magnitude increase in the antenna baseline leads to an order of magnitude decrease in the condition number for all measurement arc lengths, which is again caused by a two orders of magnitude increase in the magnitude of \( v_6 \). It is recalled that the condition number is determined for the non-normalized matrices while the eigenvalues are determined for the normalized matrices, thus a change in the condition number does not imply the same change in the eigenvalues. For the elliptical formation geometry, the scenario with a single Rx antenna is not observable for \( \Delta x_0^k = 0 \) due to an extremely small value for \( v_6 \) and is therefore not shown.

For the pendulum formation however, a large antenna baseline positively influences the estimation result only for ‘short’ measurement arcs, cf. Fig. 4.10(b). For measurement arcs shorter than one orbit, a larger antenna baseline results in a significantly smaller
condition number and thus a higher accuracy in the estimation. Thus, if there is a time-critical relative navigation accuracy requirement for the pendulum formation, a design using multiple Rx antennas with a large baseline is preferable over a single Rx antenna design. Figure 4.10(b) also shows that, in case of a three Rx antenna design, it is possible to save power by switching to a single Rx antenna for the relative navigation after a measurement arc length of at least one orbit, provided that a $3^{1/2}$ decrease in relative navigation accuracy is acceptable, cf. Table 4.3. The factor $3^{1/2}$ difference between one and three Rx antennas does not show in the condition number since this factor affects all eigenvalues and does therefore not affect the condition number. If the determinant of the Gramian would have been used as an indication of observability however, this difference would have been clear without investigating the eigenvalues.

Note that the shape of the curves in Fig. 4.10(b) depends on the magnitude of $z_{\text{max}}$ while the magnitude of the condition number depends on $r_e$. Thus, if $r_e = 10\text{km}$ and $z_{\text{max}} = 100\text{m}$, plotting the condition number versus the measurement arc length will result in a shape just as in Fig. 4.10(b), but the magnitude of the condition number will be two orders of magnitude larger, leading to a decrease in estimation accuracy. The similarity in shape is apparent from Fig. 4.10(d) for the scenario that $z_{\text{max}} = 100\text{m}$. Figure 4.10(d) also
clearly demonstrates that for the pendulum formation a larger out-of-plane motion leads to improved estimation accuracy while Fig. 4.10(c) shows that this is hardly the case for the safe ellipse, which is consistent with the results shown in Table 4.3.

The increase in the condition number for more than one orbit, cf. Fig. 4.10, implies a decrease in the numerical stability of the system and therefore a larger probability of estimator divergence, but it also implies an increased accuracy in the estimate. The reason for this follows from physics: As the magnitudes of the eigenvalues of the Gramian indicate how well a certain state component can be estimated, it is expected that these magnitudes increase with increasing observation arc length since more and more information on the relative trajectory of the spacecraft is obtained and thus the estimate of the relative trajectory should become more accurate. However, after one orbit, the relative motion repeats, thus adding little extra information. The most observable state component however will benefit more from the extra information than a less observable state component, and thus $\nu_1$ will grow faster than $\nu_6$, causing the growth in condition number for more than one orbit. It also causes the standard deviation in the estimate for the most observable state component to decrease faster with time than the standard deviation in the estimate for the least observable state component. Thus, the growth in condition number here actually represents an increase in the accuracy of the relative state estimate rather than a decrease. Of course, this result can only be obtained when the dynamic model used in the estimator is a perfect representation of the actual dynamic environment. When more realistic dynamic models are used and if some of the dynamics are not modeled in the estimator, which would be the case for a space mission, the accuracy of the estimate cannot improve beyond the accuracy of the modeled dynamics and thus the accuracy in the estimate will not improve anymore after a sufficiently long observation arc length.

Influence of the location of the Rx antenna(s)

In subsection 4.4.1 it was determined that for scenarios with a single Rx antenna, not placing this antenna at the origin of the Hill frame can influence the observability of the system. This is especially true for an offset in $y$- and/or $z$-direction in case of an elliptical formation geometry. Thus, changing the location of the Rx antenna can potentially improve the accuracy in the estimate of the relative state. This has been investigated for both formation geometries. For the pendulum formation, this results in no significant improvement in the estimation accuracy. This is as expected since the results presented earlier in this subsection already showed that the antenna baseline does not influence the estimation accuracy for a pendulum formation geometry. Thus, placing a single Rx antenna not at the COM of the chief, which can also be regarded as a baseline, should also not influence the results. For the elliptical formation geometry however, the location of the Rx antenna is of major importance. This is shown in Table 4.6, which depicts the RSS accuracy in the estimation of the relative position for a $1 \times 2 \times 1$ km elliptical formation for various antenna locations.

From Table 4.6 it is clear that especially an offset in $y$- and/or $z$-direction has a large impact on the accuracy in the estimate of the system state. The results in Table 4.6 also show that the results for a displacement in $y$- and/or $z$-direction have a trend which is very
Table 4.6: Selected results for the accuracy in the determination of the relative state for an elliptical formation geometry in case of a single Rx antenna. The RSS accuracy in meters of the relative position estimate, obtained by averaging the results for 1,000 Monte Carlo estimator runs, is provided for various locations of the Rx antenna in the Hill frame and for various ranging accuracies. The formation geometry is an ellipse with \( z_{\text{max}} / r_e = 1 \) and \( r_e = 1,000 \) m. The observation arc length is five orbits. Note that \( \sigma_{\text{appr}} = 100 \) m.

<table>
<thead>
<tr>
<th>Antenna location</th>
<th>( x ) [m]</th>
<th>( y ) [m]</th>
<th>( z ) [m]</th>
<th>( \sigma_{\rho} = 1 , \text{cm} )</th>
<th>( \sigma_{\rho} = 1 , \text{dm} )</th>
<th>( \sigma_{\rho} = 1 , \text{m} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9.51 \times 10^1</td>
<td>9.51 \times 10^1</td>
<td>9.49 \times 10^1</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8.88 \times 10^1</td>
<td>9.51 \times 10^1</td>
<td>9.49 \times 10^1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.76 \times 10^1</td>
<td>9.51 \times 10^1</td>
<td>9.48 \times 10^1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.65 \times 10^1</td>
<td>6.97 \times 10^1</td>
<td>8.89 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>1.43 \times 10^1</td>
<td>8.38 \times 10^1</td>
<td>9.47 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.45 \times 10^0</td>
<td>1.45 \times 10^1</td>
<td>7.95 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1.45 \times 10^{-1}</td>
<td>1.45 \times 10^0</td>
<td>1.43 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>9.59 \times 10^0</td>
<td>6.43 \times 10^1</td>
<td>9.44 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9.64 \times 10^{-1}</td>
<td>9.49 \times 10^0</td>
<td>6.73 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>9.64 \times 10^{-2}</td>
<td>9.63 \times 10^{-1}</td>
<td>9.58 \times 10^0</td>
</tr>
<tr>
<td>0</td>
<td>-0.06</td>
<td>0.08</td>
<td>0</td>
<td>7.99 \times 10^0</td>
<td>5.83 \times 10^1</td>
<td>9.41 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>-0.55</td>
<td>0.83</td>
<td>0</td>
<td>8.02 \times 10^{-1}</td>
<td>7.93 \times 10^0</td>
<td>6.09 \times 10^1</td>
</tr>
<tr>
<td>0</td>
<td>-5.55</td>
<td>8.32</td>
<td>0</td>
<td>8.02 \times 10^{-2}</td>
<td>8.02 \times 10^{-1}</td>
<td>7.99 \times 10^0</td>
</tr>
</tbody>
</table>

Similar to the trend observed in case of three Rx antennas, cf. Fig. 4.8: The accuracy in the relative state estimation scales linearly with antenna location and ranging accuracy provided that the ranging accuracy is substantially smaller than the distance of the antenna to the center of the Hill frame. If not, the trend becomes nonlinear and approaches \( \sigma_{\text{appr}} \).

In addition, for the linear regime of this trend, the RSS position estimate for an antenna displacement in \( z \)-direction is a factor 1.5 more accurate than the RSS position estimate for the same antenna displacement in \( y \)-direction. Positioning the single Rx antenna on a vector where the ratio between the \( y \)- and \( z \)-components of the vector is \(-1.5\), resulting in the unit vector \( (0, -\frac{4}{13}, \frac{9}{13})^T \), leads to the highest estimation accuracy achievable with a single Rx antenna. This is displayed in the last three rows of Table 4.6.

The large improvement in estimation accuracy due to an antenna offset in \( y \)- and/or \( z \)-direction is in part due to a better observation geometry since the Rx antenna is not positioned in the plane of the relative orbit of the deputy anymore. However, somewhat contrary to what might be expected, the unit vector that results in the optimal location of the Rx antenna, \( (0, -\frac{4}{13}, \frac{9}{13})^T \), is not the unit vector that is perpendicular to the plane of the relative motion. That is the vector \( (0, -\frac{1}{5}, \frac{4}{5})^T \), for which the ratio between the \( y \)- and \( z \)-components is \(-2\). The optimal unit vector on which the single Rx antenna has to be located is in fact the vector which points, in \( \mathbb{R}^3 \), in the same direction as eigenvector \( v_6 \). This leads to a large increase in the magnitude of eigenvalue \( \upsilon_6 \) since the sensitivity...
of the measurement in that direction is greatly improved, resulting in a more accurate overall state estimate since it is less dominated by the major components of eigenvector $\mathbf{v}_6$. It is noted that if the out-of-plane angle $z_{\text{max}} / y_{\text{max}}$ changes, the direction of eigenvector $\mathbf{v}_6$ will also change, cf. Fig. 4.5(a), and thus the Rx antenna location should be adapted accordingly.

Another interesting observation from Table 4.6 is that the optimal placement of a single Rx antenna results in a state estimation accuracy that is better than the state estimation accuracy attainable with 3 Rx antennas for the same distance of the antennas to the COM of the chief. For instance, for a single Rx antenna at $(0, -\sqrt{8/52} \text{ m}, \sqrt{18/52} \text{ m})^T$, which yields a distance of $0.5\sqrt{2} \text{ m}$ with respect to the COM of the chief, the estimation result for the safe ellipse with $z_{\text{max}} = r_e$, $r_e = 1,000 \text{ m}$ and $\sigma_\rho = 1 \text{ dm}$ is on average $\sigma_{\text{RSS}} \approx 11.1 \text{ m}$. For three Rx antennas with standard layout and a baseline of $d \approx 12 \text{ dm}$, which yields for all Rx antennas also a distance of $0.5\sqrt{2} \text{ m}$ with respect to the COM of the chief, the estimation result is on average $\sigma_{\text{RSS}} \approx 9.2 \text{ m}$. This is better than what can be achieved with a single Rx antenna, but when multiplied with $3^{1/2}$ to compensate for the increased number of measurements, the result becomes $\sigma_{\text{RSS}} \approx 16.0 \text{ m}$. This result is clearly worse than the optimal single Rx antenna scenario. Thus, the single Rx antenna scenario with optimal antenna placement clearly outperforms the scenario with three Rx antennas at the standard locations. The reason for this is that the optimal antenna placement yields a relatively large improvement in the accuracy of the estimates for $z_0$ and $y_0$. These state components both contribute the most to the magnitude of $\sigma_{\text{RSS}}$ since $x_0$ can already be estimated with high accuracy. Merely increasing the distance between the Rx antenna(s) and the COM of the spacecraft will not result in such a large improvement in the estimation accuracy for $z_0$ and $y_0$ since the increase in observability of these state components is suboptimal.

For three Rx antenna spanning an equilateral triangle, the ideal antenna locations are as follows: Two Rx antennas have to be located on the line $y = -2/3z$ but one antenna has to have a positive $y$-coordinate while the other antenna has to have the same but negative value for the $y$-coordinate. The third Rx antenna has to be located on the $x$-axis of the Hill frame such that an equilateral triangle is formed. If the antenna baseline is required to be $\sqrt{150} \text{ dm}$ and if the distance of all antennas to the center of the satellite has to be equal to $0.5\sqrt{2} \text{ m}$, a configuration that is very similar to the standard antenna layout, the ideal antenna locations are $(-1/4\sqrt{2} \text{ m}, -\sqrt{12/104} \text{ m}, \sqrt{27/104} \text{ m})^T$, $(-1/4\sqrt{2} \text{ m}, \sqrt{12/104} \text{ m}, -\sqrt{27/104} \text{ m})^T$, and $(1/2\sqrt{2} \text{ m}, 0, 0)^T$. This antenna configuration requires either a rotation of the cubic-shaped chief satellite as displayed in Fig. 4.1 around an appropriate vector or a different geometry for the chief. This optimal configuration leads to $\sigma_{\text{RSS}} \approx 9.1 \text{ m}$, which is 2% better than the standard layout. It is noted that a translation of the equilateral triangle formed by the three Rx antennas along the $x$-axis of the Hill frame has no noticeable effect on the accuracy of the relative state estimate.

Therefore, it is concluded that for the safe ellipse the placement of the Rx antennas with respect to the origin of the Hill frame is very important: The direction and length of the vector(s) from the origin of the Hill frame to the Rx antenna(s) has a major contribution to the accuracy in the relative state estimate, which is actually more important than the
antenna baseline. If positioned at the proper location, a single Rx antenna on the chief will yield sufficient information to produce an accurate relative state estimate. Thus, relative navigation using a single Rx antenna on the chief is feasible for both the pendulum and the safe elliptical formation geometries. However, these conclusions are only valid for scenarios where reasonably accurate a-priori information is available, for which the out-of-plane angle of the formation is large enough to have a very low probability of estimator divergence and for which the signals between the Tx and Rx antennas are not blocked by the satellite structure.

4.4.3 Apparent Estimator Divergence

This subsection deals briefly with the probability of apparent estimator divergence as a function of $z_{\text{max}}/r_e$. To determine this probability, a large amount of estimator runs have been performed for various scenarios and the percentage of runs that led to apparent estimator divergence has been determined. The estimator is said to have diverged when the error in the estimate of a state component is more than three times the standard deviation of that state component.

In Fig. 4.11, the probability of apparent estimator divergence is displayed as a function of $z_{\text{max}}/r_e$ for various scenarios for an observation arc length of five orbits. Recognizing that, for the formation geometries studied, the ratio $z_{\text{max}}/r_e$ can be regarded as an ‘out-of-plane’ angle $z_{\text{max}}/y_{\text{max}}$, this figure shows that the probability of estimator divergence is very low and stable around 0–2% for large out-of-plane angles, but increases sharply once this angle drops below a certain threshold. In addition, the probability of estimator divergence does not necessarily increase with decreasing out-of-plane angle, cf. e.g. Fig. 4.11(f).

Figure 4.11 also shows the influence of the antenna baseline and the inter-satellite distance on the probability of estimator divergence. In general, a large antenna baseline leads to a lower probability of estimator divergence. For some scenarios even, the probability of estimator divergence remains very low throughout the entire range of $z_{\text{max}}/r_e$ considered. A small inter-satellite distance also generally leads to a lower probability of estimator divergence. The influence of the ranging accuracy on the other hand is difficult to assess: For some scenarios a high ranging accuracy leads to a decrease in the probability of estimator divergence while for other scenarios the opposite is true. Why this is the case is not understood and should be researched further.

Careful examination of Fig. 4.11 also reveals that the curve for the probability of estimator divergence does not change when the variables $r_e$, $d$, and $\sigma_\rho$ are all multiplied by the same scalar. For example, the curve for $d \approx 12$ cm in the upper pane in Fig. 4.11(a) is identical to the curve for $d \approx 12$ dm in the middle pane in Fig. 4.11(c) which in turn is identical to the curve for $d \approx 12$ m in the bottom pane in Fig. 4.11(e). When only a single Rx antenna is present, a similar effect is observed: The curve for one Rx antenna in the upper pane in Fig. 4.11(a) is identical to the curve for one Rx antenna in the middle pane in Fig. 4.11(c) which in turn is identical to the curve for one Rx antenna in the bottom pane in Fig. 4.11(e). Thus, the linearized treatment of the system also affects the proba-
Figure 4.11: Probability of apparent estimator divergence as a function of $z_{\text{max}}/r_e$ for an observation arc length of five orbits and different antenna baselines. Subfigures (a), (c), and (e) display the results for a pendulum formation geometry while subfigures (b), (d), and (f) display the results for an elliptical formation geometry. In each subfigure, the top, middle, and bottom panes display the results for $r_e = 100$ m, $r_e = 1$ km, and $r_e = 10$ km, respectively. All data points are the result of 1,000 Monte Carlo simulator runs.
bility of estimator divergence: When all variables are scaled by the same magnitude, the observability of the system is not affected. For example, if $r_e = 0.1\text{km}$, $d \approx 12\text{dm}$, and $\sigma_\rho = 0.1\text{m}$, the probability of estimator divergence will be identical to a scenario where $r_e = 1\text{km}$, $d \approx 12\text{m}$, and $\sigma_\rho = 1\text{m}$.

Although the foregoing observations are interesting, the exact probability of apparent estimator divergence is actually not very important when estimating a system state. More important is whether or not it is higher than a certain threshold: If it is above the threshold, the estimation process is not reliable and the system state should not be estimated with this combination of estimator type and system models (system dynamics and sensors) since the result cannot be trusted to be correct. As already mentioned, for this estimator and for large out-of-plane angles, the probability of estimator divergence is stable at 0–2%. Thus, a reasonable threshold above which the estimator results should no longer be trusted to be correct is a probability of estimator divergence of 2%.

Comparing the results for the pendulum and elliptical formation geometries in Fig. 4.11 leads to the conclusion that, in general, reliable estimation of the elliptical formation geometry requires a significantly larger out-of-plane angle $z_{\text{max}}/y_{\text{max}}$ than reliable estimation of the pendulum formation geometry: For the pendulum formation geometry and most scenarios considered here, this angle needs to be larger than 0.6° ($z_{\text{max}}/y_{\text{max}} > 0.01$). To account for all scenarios, the angle needs to be larger than 2.9° ($z_{\text{max}}/y_{\text{max}} > 0.05$). For the elliptical formation geometry and most scenarios considered here, the angle needs to be larger than 8.5° ($z_{\text{max}}/y_{\text{max}} > 0.15$). To account for all scenarios, the angle even needs to be larger than 14° ($z_{\text{max}}/y_{\text{max}} > 0.25$).

To assess whether estimator divergence can be predicted accurately, the condition number and the determinant of the matrix that is actually being inverted, $\Lambda + \tilde{H}^T W \tilde{H}$, have been determined. This has been done for both $\Delta x_0^0$ and $\Delta x_k^0 = 0$. Figures 4.12 and 4.13 provide sample results for these parameters for scenarios with $r_e = 1000\text{m}$ and $\sigma_\rho = 1\text{dm}$. The probability of estimator divergence for these scenarios is depicted in the middle panes of Figs. 4.11(c) and 4.11(d) for the pendulum formation and the elliptical formation, respectively. The condition number $\kappa$ has been chosen since it provides information on the numerical conditioning of $\Lambda + \tilde{H}^T W \tilde{H}$. For the current setting, if $\kappa > 10^{15}$ the matrix $\Lambda + \tilde{H}^T W \tilde{H}$ is considered to be rank-deficient and thus the estimation algorithm will diverge. The condition number also provides information on the magnitude of $\nu_6$, which is known to vary with varying $z_{\text{max}}/r_e$, with respect to $\nu_1$, which is known to be independent of $z_{\text{max}}/r_e$. The matrix determinant provides information on possible variations in the other eigenvalues that can indicate a change in observability of the system that is not ‘captured’ by the condition number. The matrix determinant is also used in [Giorgi 2011] to evaluate the influence of the number of antenna baselines on the ambiguity resolution in GNSS carrier phase-based attitude determination.

Comparing the subfigures of Fig. 4.12 with the middle pane of Fig. 4.11(c) shows that the curves of Fig. 4.12(a) for $z_{\text{max}}/r_e < 0.01$, where $\kappa > 10^{15}$ for three of the four curves and thus estimator divergence is very probably, are remarkably similar to the curves in the middle pane of Fig. 4.11(c). This is also true for Fig. 4.12(c), but less so for Figs. 4.12(b)
and 4.12(d). This applies to all pendulum formation scenarios in Fig. 4.11: In general, the magnitude of the condition number for $\Delta x_0^0$ provides a good indication for the probability of estimator divergence although it usually does not match so well to the actual variation in probability as in the example shown here. The magnitude of the matrix determinant is a less useful metric since it scales with the number of Rx antennas (measurements) and therefore yields a more pessimistic result for the single Rx antenna scenario than what is actually the case. For $\Delta x_k^0 = 0$, the condition number and matrix determinant can also be used to make a statement on the reliability of the estimation process, but not as well as for $\Delta x_0^0$.

The curves in the subfigures of Fig. 4.13 on the other hand do not resemble the curves in the middle pane of Fig. 4.11(d) at all. Except for Fig. 4.13(a), the subfigures of Fig. 4.13 provide little to no useful information on the probability of estimator divergence for the elliptical formation geometry. Thus, again, the condition number for $\Delta x_0^0$ is the most accurate metric to provide quick information on the reliability of the estimation process. However, it fails to predict the sudden increase, or 'bump', in the probability of estimator divergence for elliptical formation geometries around $z_{\text{max}}/r_e = 0.1$, cf. Figs. 4.11(b), 4.11(d), and 4.11(f). This 'bump' is also not predicted by other metrics such as the matrix determinant, the magnitude of eigenvalue $v_5$ or $v_6$, or the ratios $v_4/v_5$ or $v_5/v_6$. More research into this is recommended to determine whether this 'bump' is caused by the choice of estimator or if this is a fundamental property for the elliptical formation geometry.

### 4.4.4 Antenna Baseline Influence for the Pendulum Formation

As already briefly indicated in subsection 4.4.2, for small values of $r_e$ and $z_{\text{max}}$, it has been observed that the empirical relation for $\sigma_{\text{RSS}}$ in Table 4.3 does not hold for the pendulum formation geometry in case of a large antenna baseline ($d \approx 12$ m). In this subsection, this observation is further investigated by analyzing the state estimates that result for scenarios where $r_e$ is varied between 1 m, 10 m, 100 m, and 1000 m ($\sigma_\rho = 1$ cm) and for scenarios where $\sigma_\rho$ is varied between 1 mm, 1 cm, 1 dm, and 1 m ($r_e = 10$ m). For all scenarios, the setup of the simulations is the same as described in section 4.2 and $\sigma_{\text{RSS}}$ is determined for an observation arc length of five orbits. Due to the poor observability of the system for small $z_{\text{max}}/r_e$ and the resulting spread in the accuracy of the state estimates, all data points in the figures in this subsection have been generated by averaging the results for 10,000 Monte Carlo simulation runs. Note that these scenarios can result in extreme situations such as an inter-satellite distance of 1 m for a spacecraft with a 12 m antenna baseline, which will only occur during formation deployment or during RVD. As RVD maneuvers are typically executed with minimal relative out-of-plane motion, the pendulum formation scenarios with small inter-satellite distance and large relative out-of-plane motion, i.e., $r_e \leq 10$ m and $z_{\text{max}}/r_e > 0.1$, treated here should be considered to be hypothetical scenarios with little relevance to actual space missions.

Figure 4.14 displays the results for $\sigma_{\text{RSS}}$ for four different inter-satellite distances in case $\sigma_\rho = 1$ cm: $r_e = 1$ km, $r_e = 100$ m, $r_e = 10$ m, and $r_e = 1$ m. As the estimation process is not reliable for $z_{\text{max}}/r_e < 0.01$, as determined in subsection 4.4.3, only results for the range
Chapter 4. Relative Navigation Simulation and Analysis

Figure 4.12: Condition number and matrix determinant of $\Lambda + \tilde{H}^T W \tilde{H}$ for a pendulum formation geometry with $d \approx 12$ dm, $r_e = 1000$ m, and $\sigma_p = 1$ dm. Subfigures (a) and (b) provide the condition number for $\Delta x^0_0$ and $\Delta x^k_0 = 0$, respectively. Subfigures (c) and (d) provide the matrix determinant for $\Delta x^0_0$ and $\Delta x^k_0 = 0$, respectively.

$z_{max}/r_e = [0.01, 1]$ are provided.

Figure 4.14(a) displays the trend in $\sigma_{RSS}$ when $r_e = 1$ km. This trend is clearly linear for all antenna baselines for the range of $z_{max}/r_e$ considered and depicts graphically the estimation results reported in Table 4.3 for the pendulum formation geometry. For increasing ratio $z_{max}/r_e$, the result for $\sigma_{RSS}$ improves. This is a direct result of the increased observability of the out-of-plane state components, as treated in subsection 4.4.1. In addition, as already reported in Table 4.3, the estimation accuracy when using a single Rx antenna is a factor $3^{1/2}$ worse than when using three Rx antennas, for which the results are exactly equal for almost all values of $z_{max}/r_e$ considered.

However, Fig. 4.14(a) already shows that the curve for $d \approx 12$ m departs slightly from the linear trend when $z_{max} \approx 10$ m $\approx d$. This is even more clear in Fig. 4.14(b), which also shows that the curve for $d \approx 12$ dm starts to depart from the linear trend at $z_{max} \approx d$. In
4.4. **Analysis**

Figure 4.13: Condition number and matrix determinant of $\Lambda + \tilde{H}^T W \tilde{H}$ for an elliptical formation geometry with $d \approx 12 \text{ dm}$, $r_e = 1000 \text{ m}$, and $\sigma_\rho = 1 \text{ dm}$. Subfigures (a) and (b) provide the condition number for $\Delta x^0_0$ and $\Delta x^k_0 = 0$, respectively. Subfigures (c) and (d) provide the matrix determinant for $\Delta x^0_0$ and $\Delta x^k_0 = 0$, respectively.

In fact, the results for $d \approx 12 \text{ dm}$ in Fig. 4.14(b) are exactly equal to the results for $d \approx 12 \text{ m}$ in Fig. 4.14(a). This trend is continued in Figs. 4.14(c) and 4.14(d) and can be summarized as follows: The curve for an antenna baseline $d$ at inter-satellite distance $r_e$ is exactly equal to the curve for an antenna baseline $10d$ at inter-satellite distance $10r_e$. This commonality is explained by the fact that for those cases the LOS angles from the Rx antennas on the chief to the Tx antenna on the deputy are identical, resulting in identical observation geometries and therefore identical information content. Thus, for long observation arc lengths, the accuracy in the relative state estimate $\sigma_{\text{RSS}}$ in case of a pendulum formation is in fact a function of the antenna baseline $d$. However, the influence of the antenna baseline $d$ is negligible unless it is equal to or larger than the amplitude of the out-of-plane motion ($d \gtrsim z_{\text{max}}$).
Concluding, once the amplitude of the out-of-plane motion becomes smaller than the antenna baseline, the observation geometry changes in such a way that the accuracy in the relative state estimate is no longer a function of the amplitude of the out-of-plane motion. Thus, decreasing this amplitude even further will no longer affect the accuracy in the relative navigation. It does however still influence the observability of the system and should therefore preferably be as large as possible in order to obtain the most reliable state estimates.

Reducing the inter-satellite distance even further, cf. Figs. 4.14(c) and 4.14(d), leads to several additional observations. Firstly, once the inter-satellite distance is smaller than the antenna baseline, \( r_e \approx d \), the accuracy of the state estimate is no longer a function of \( z_{\text{max}} \), \( r_e \) or \( d \). The approximate relation for \( \sigma_{\text{RSS}} \), in case of three Rx antennas on the chief, is now \( \sigma_{\text{RSS}} \approx 3\sigma_\rho / \sqrt{l} \) with \( l \) being the number of measurements.

Figure 4.14(d) also shows that even the curve for a single Rx antenna can deviate from the linear trend. However, this behavior is tied to the accuracy of the a-priori information and is not fundamental to the system. The accuracy of the a-priori information is for
this scenario 0.1 m and is, since \(x_0\) and \(y_0\) can still be estimated with good accuracy, approximately the worst case for the RSS accuracy in the state estimate (since the accuracy in the estimate for \(z_0\) will remain equal to \(\sigma^{apr}\) if it is poorly observable). Thus, the linear trend for the single Rx antenna in Fig. 4.14(d) cannot be sustained for small values of \(z_{\text{max}}/r_e\) since it would then yield a result for \(\sigma_{\text{RSS}}\) that would be considerably larger than \(\sigma^{apr}\). Since the measurement accuracy is high, the estimate for the out-of-plane states can still be improved upon, which is why the curve does not converge towards \(\sigma^{apr}\). When the measurement accuracy is decreased to \(\sigma_{\rho} = 1\) m, the accuracy in the state estimate does converge to \(\sigma^{apr}\) for small values of \(z_{\text{max}}/r_e\), cf. Fig.4.15(d). For completeness, it is noted that even though the estimate for the out-of-plane state components can be improved upon in Fig. 4.14(d), implying that these state components are observable, not using \(\Lambda\) in the estimation process will lead to a very high likelihood of apparent estimator divergence and is therefore not advisable.

When \(r_e\) is kept constant at 10 m, but \(\sigma_{\rho}\) is varied between 1 mm, 1 cm, 1 dm, and 1 m, the results displayed in Fig. 4.15 are obtained. It clearly shows that \(\sigma_{\text{RSS}}\) scales linearly with \(\sigma_{\rho}\), as already shown in subsection 4.4.2. In addition, as the measurement accuracy deteriorates, the influence of the a-priori information on the relative state estimate increases. As can be expected, this is most noticeable for scenarios with small values for \(z_{\text{max}}/r_e\) since these scenarios are poorly observable. Also here it is evident that a large antenna baseline has a positive influence on the accuracy in the relative state estimate for small values of \(z_{\text{max}}/r_e\). However, for values of \(z_{\text{max}}/r_e\) approaching 1, the benefit of the large antenna baseline decreases or even completely vanishes.

Summarizing, for the pendulum formation, if the amplitude of the out-of-plane motion \(z_{\text{max}}\) of the deputy is smaller than or equal to the antenna baseline \(d\) on the chief, the RSS accuracy in the relative position estimation \(\sigma_{\text{RSS}}\) is no longer a function of the out-of-plane angle \(z_{\text{max}}/y_{\text{max}}\) of the formation, but only of the inter-satellite distance \(r_e\), the antenna baseline \(d\), and the ranging accuracy \(\sigma_{\rho}\). A large antenna baseline is beneficial under these conditions as this results in the highest attainable accuracy in the relative state estimate, which improves for decreasing inter-satellite distance. Thus

\[
\sigma_{\text{RSS,pendulum}} \propto \frac{r_e\sigma_{\rho}}{d\sqrt{l}} \quad \text{if} \quad z_{\text{max}} \lesssim d. \tag{4.10}
\]

with \(l\) being the number of measurements.

However, when \(r_e \lesssim d\), \(\sigma_{\text{RSS}}\) is no longer a function of \(r_e\) and \(d\). Then, the approximate relation for \(\sigma_{\text{RSS}}\) is

\[
\sigma_{\text{RSS,pendulum}} \approx 3\sigma_{\rho}/\sqrt{l} \quad \text{if} \quad r_e \lesssim d. \tag{4.11}
\]

Thus, the results obtained in this subsection lead to the conclusion that for pendulum formations with small out-of-plane amplitudes and small inter-satellite distances, which can be considered to be RV(D) scenarios, a large antenna baseline can positively influence the accuracy in the relative navigation if the amplitude of the relative out-of-plane motion is smaller than the antenna baseline. Therefore, a spacecraft with large dimensions is at
an advantage compared to a spacecraft with small dimensions since the former allows for a larger antenna baseline without the need for deployable structures.

In addition, in a RV(D) scenario, where the inter-satellite distance is commonly decreased in a step-wise manner and where the relative out-of-plane motion is small, having a large antenna baseline will be advantageous since this will yield the highest attainable accuracy in the relative navigation during most of the approach phase. However, the accuracy in $\sigma_{\text{RSS}}$ cannot improve beyond $3\sigma_{\rho}/\sqrt{I}$. It therefore depends on which accuracy is desired at which stage of the RV(D) maneuver to decide which antenna baseline and ranging accuracy should be used and thus whether a small spacecraft is adequate. Of course, it must be kept in mind that in real situations the accuracy in the relative state estimate will not continue to increase indefinitely when more measurements are made, as is suggested by Eq. (4.11) (and by Eq. (4.10)). Due to various modeling and systematic errors, there will exist a lower limit for the accuracy that can be obtained, no matter how many measurements are made.
Lastly, it is noted that the relative accuracy in the relative state estimate for a different number of Rx antennas and/or different Rx antenna baselines for the scenarios treated in this subsection can also be determined with moderate accuracy by computing the matrix determinant of \( \Lambda + \tilde{H}^T W \tilde{H} \) for \( \Delta x_k^0 = 0 \) for the different scenarios. This has been verified for multiple scenarios and can also be concluded from the similarity between Figs. 4.14(a) and 4.12(d) and knowing that the accuracy in the relative state determination scales linearly with \( \sigma_\rho \). Thus, the shapes of the curves in Fig. 4.14(a) do not change when \( \sigma_\rho \) is changed to 1 dm.

### 4.4.5 Observability Increase for three Spacecraft

In the preceding subsections, it has been concluded that a small magnitude for the out-of-plane angle of the deputy relative to the chief leads to a poorly observable system with a high probability of estimator divergence. However, it was also observed that an increase in the magnitude of the relative out-of-plane angle increases the observability of the system. With this knowledge, it is of interest to investigate whether adding an additional spacecraft to a formation where the deputy has a small out-of-plane angle can increase the observability of that deputy. Naturally, this will have a non-negligible impact on the mission and spacecraft complexity and on mission costs and should thus only be considered if it is a mission enabler and if it is less costly than other solutions, e.g. using relative GNSS measurements for a mission in LEO. For the investigation, the following scenario has been set up.

A two-spacecraft pendulum formation with the same characteristics as used earlier is expanded to a three-spacecraft formation by adding another deputy. The original deputy, referred to as deputy 1, has a small out-of-plane angle, \( z_{\text{max}} = 0.001 r_e \), with respect to the chief. For this relative orbit, estimator divergence has a very high probability of occurrence. Another deputy, referred to as deputy 2, placed in an orbit resulting in a large relative out-of-plane angle, is added to the formation with the sole purpose of increasing the observability of the state vector of the original deputy, cf. Fig. 4.10(d). It is noted that a similar scenario could have been set up for the elliptical formation geometry, but with the complication that this would have required the use of multiple Rx antennas on the chief. This is obvious from the estimation accuracy results presented in Table 4.3 where it is clear that the observability for the safe ellipse in case of a single Rx antenna on the chief is extremely poor, even if \( z_{\text{max}} / r_e = 1 \). Nevertheless, the conclusions that will be drawn for the pendulum formation in this subsection will also be applicable to a safe ellipse provided that the observability of the added deputy is good.

For the inter-satellite range measurements to be made, there now exist three options, of which one can be split into two sub-options:

1. Range measurements between the chief and deputy 1 and range measurements between the chief and deputy 2.

2. Range measurements between the two deputies and
   
   (a) range measurements between the chief and deputy 1.
(b) range measurements between the chief and deputy 2.

3. Range measurements between all three spacecraft.

For option 1, there is no coupling between the state estimates for deputy 1 and deputy 2 since the measurements for deputy 1 and deputy 2 are independent. Therefore, changing the relative orbit of one of the deputies will not impact the observability of the state vector of the other deputy and thus this option is not expected to increase the observability of a deputy with small out-of-plane angle. For option 3, the system is fully connected (all possible inter-satellite links are established) and will thus contain the best information possible, which should lead to the best observable system and thus the most accurate overall state estimate. However, to decrease the amount of resources needed for the relative navigation, a range measurement between the chief and a deputy can be removed. This can be either the link with deputy 2, resulting in option 2a, or the link with deputy 1, resulting in option 2b. As will be shown, this choice has a large impact on the relative navigation result. The various inter-satellite measurement options are displayed in Fig. 4.16.

![Figure 4.16](image_url)

**Figure 4.16:** Various inter-satellite ranging options in case of two deputies. In option 1, there are range measurements between the chief and both deputies. In option 2, there are range measurements between the chief and one deputy and between the deputies themselves. In option 2a, the chief measures the distance to deputy 1 while in option 2b, the chief measures the distance to deputy 2. In option 3, there are range measurements between all three spacecraft in the formation. The options have been color-coded to ease interpretation of the figure.

The state vector of deputy 1 at $t = 0$ is denoted as $\mathbf{x}_{0,1}$ and the state vector of deputy 2 at $t = 0$ is denoted as $\mathbf{x}_{0,2}$. In the scenario under investigation, the full state vectors are

\[
\mathbf{x}_{0,1} = (0, -r_e, 0, 0, 0, -0.001 n r_e)^T \quad (4.12)
\]
\[
\mathbf{x}_{0,2} = (0, -2r_e, 0, 0, 0, -n r_e)^T . \quad (4.13)
\]

Thus, the relative motion of deputy 2 with respect to deputy 1 is nearly equal to the two-spacecraft formation scenarios investigated earlier where the observability of the system was best. The state vector to be estimated is $\mathbf{x}_0 = (\mathbf{x}_{0,1}, \mathbf{x}_{0,2})^T$. Observability analyses have been conducted for $\Delta \mathbf{x}_0^0$ and relative state estimates have been produced for all inter-
4.4. Analysis

Satellite ranging options using the same a-priori information as earlier for an observation arc length of five orbits. In all scenarios, the chief is equipped with a single Rx antenna, the inter-satellite distance at epoch $t = 0$ is 1 km, and the accuracy of all range measurements is 1 cm. The results are displayed in Tables 4.7 and 4.8. In the tables, the two-spacecraft formation is referred to as option 0. In addition, for brevity, the subscript 0 is omitted in the following.

A difficulty in comparing the observability of deputy 1 for formations with three spacecraft is that the eigenvectors of the system will generally have contributions from state vector components belonging to both deputy 1 and deputy 2. Thus, determining the observability only for deputy 1 by means of the eigenvalues of the eigenvectors whose main component is a state component of deputy 1 will yield a result that inevitably will have contributions from the state components of deputy 2. A first step to overcome this problem is to determine whether there are eigenvalues in the system that are much smaller than all other eigenvalues. If that is not the case, there is only a small probability of estimator divergence and thus the out-of-plane state components of deputy 1 can be determined accurately, which is already a huge gain compared to the original two-spacecraft formation. When multiple options have a satisfactory range of eigenvalues, a more detailed investigation of the eigenvectors is needed. A more practical approach however is to simply perform the estimation process to determine which option can estimate the state of deputy 1 the most accurate.

For all options, the matrix $\tilde{H}$ is full rank for $\Delta x^0_0$ and thus, normally, estimator divergence would not be expected. However, as will be shown, it occurs for options 0, 1, and 2a. Table 4.7 displays the results of the observability analysis for all options with two columns for each option. For each option, the left column displays the magnitude of the eigenvalues of the system in descending order while the right column displays the state component(s) with the largest absolute value in the corresponding eigenvector.

The first observation that can be made when investigating Table 4.7 is that options 0, 1, and 2a all have two eigenvalues that are significantly smaller than the other eigenvalues. The corresponding eigenvectors are all directed towards out-of-plane state components. For option 1, the two smallest eigenvalues correspond to the out-of-plane state components of deputy 1 and are exactly equal to the two smallest eigenvalues for option 0. In fact, all eigenvalues and eigenvectors for option 1 that correspond to a state component of deputy 1 are exactly equal to the eigenvalues and eigenvectors for option 0. In addition, the eigenvalues and eigenvectors for deputy 2 in option 1 are exactly equal to those obtained for a two-spacecraft pendulum formation with $r_e = 2$ km and $z_{\text{max}} = 0.5 r_e$. Lastly, for option 1, all eigenvectors of the system are either directed only towards state components belonging to deputy 1 or only towards state components belonging to deputy 2, clearly demonstrating a complete lack of coupling between the information obtained on deputy 1 and deputy 2. Therefore, as expected, option 1 does not lead to an improvement in the observability of deputy 1.

For option 2a, an interesting result is obtained. There are namely four eigenvectors that do not have a single state component that is clearly larger than all other state components,
Table 4.7: Observability results for two-spacecraft (option 0) and three-spacecraft (options 1-3) formations. Each option is represented by two columns. The left column displays the eigenvalues of the system in descending order. The right column displays the largest state component in the eigenvector corresponding to the eigenvalue to its left. The observability analysis is performed for $\Delta x^0_0$.

<table>
<thead>
<tr>
<th>Option 0</th>
<th>Option 1</th>
<th>Option 2a</th>
<th>Option 2b</th>
<th>Option 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$7.5 \cdot 10^6$</td>
<td>$7.5 \cdot 10^6$</td>
<td>$1.6 \cdot 10^7$</td>
<td>$1.5 \cdot 10^7$</td>
<td>$1.8 \cdot 10^7$</td>
</tr>
<tr>
<td>$1.6 \cdot 10^3$</td>
<td>$6.8 \cdot 10^6$</td>
<td>$2.6 \cdot 10^6$</td>
<td>$2.4 \cdot 10^6$</td>
<td>$7.1 \cdot 10^6$</td>
</tr>
<tr>
<td>$2.0 \cdot 10^2$</td>
<td>$1.6 \cdot 10^3$</td>
<td>$3.5 \cdot 10^3$</td>
<td>$3.4 \cdot 10^3$</td>
<td>$4.0 \cdot 10^3$</td>
</tr>
<tr>
<td>$7.9 \cdot 10^1$</td>
<td>$1.5 \cdot 10^3$</td>
<td>$5.6 \cdot 10^2$</td>
<td>$5.3 \cdot 10^2$</td>
<td>$1.5 \cdot 10^3$</td>
</tr>
<tr>
<td>$4.9 \cdot 10^{-5}$</td>
<td>$2.0 \cdot 10^2$</td>
<td>$3.7 \cdot 10^2$</td>
<td>$3.5 \cdot 10^2$</td>
<td>$4.2 \cdot 10^2$</td>
</tr>
<tr>
<td>$4.8 \cdot 10^{-5}$</td>
<td>$1.7 \cdot 10^2$</td>
<td>$1.9 \cdot 10^2$</td>
<td>$2.0 \cdot 10^2$</td>
<td>$2.3 \cdot 10^2$</td>
</tr>
<tr>
<td>$7.9 \cdot 10^1$</td>
<td>$8.5 \cdot 10^1$</td>
<td>$9.2 \cdot 10^1$</td>
<td>$9.2 \cdot 10^1$</td>
<td>$1.8 \cdot 10^2$</td>
</tr>
<tr>
<td>$7.7 \cdot 10^1$</td>
<td>$8.4 \cdot 10^1$</td>
<td>$7.8 \cdot 10^1$</td>
<td>$1.0 \cdot 10^2$</td>
<td>$1.0 \cdot 10^2$</td>
</tr>
<tr>
<td>$1.2 \cdot 10^1$</td>
<td>$6.1 \cdot 10^1$</td>
<td>$5.7 \cdot 10^1$</td>
<td>$9.2 \cdot 10^1$</td>
<td>$9.2 \cdot 10^1$</td>
</tr>
<tr>
<td>$1.2 \cdot 10^1$</td>
<td>$2.5 \cdot 10^1$</td>
<td>$2.7 \cdot 10^1$</td>
<td>$7.4 \cdot 10^1$</td>
<td>$7.4 \cdot 10^1$</td>
</tr>
<tr>
<td>$4.9 \cdot 10^{-5}$</td>
<td>$2.4 \cdot 10^{-5}$</td>
<td>$5.7 \cdot 10^0$</td>
<td>$9.4 \cdot 10^0$</td>
<td>$9.4 \cdot 10^0$</td>
</tr>
<tr>
<td>$4.8 \cdot 10^{-5}$</td>
<td>$2.4 \cdot 10^{-5}$</td>
<td>$5.3 \cdot 10^0$</td>
<td>$5.7 \cdot 10^0$</td>
<td>$5.7 \cdot 10^0$</td>
</tr>
</tbody>
</table>

but two. These two state components do not have exactly the same magnitude, but their difference is very small. Interestingly, the two smallest eigenvectors of the system are only directed to the two state components listed in Table 4.7; all other state components in those two eigenvectors have zero value. The two smallest eigenvectors both point in a direction that is for 50% towards an out-of-plane state component for deputy 1 and for 50% towards the same out-of-plane state component for deputy 2 ($z_1$ and $z_2$ or $\dot{z}_1$ and $\dot{z}_2$). Thus, when estimating the complete state vector, it is expected that the result for those state components will be highly correlated. Further on in this subsection, it is shown that this is indeed the case. Considering the observability of deputy 1, the observability of the out-of-plane state components remains very poor and estimator divergence is very likely for this option. Thus, it is concluded that a formation such as in option 2a will not improve the observability of deputy 1.

Compared to options 0, 1, and 2a, options 2b and 3 show a large improvement in the magnitude of the two smallest eigenvalues. These are now of a magnitude that follows the trend of the other eigenvalues of the system, except for the two largest eigenvectors, which are the relative semi-major axes of deputy 1 and deputy 2. Thus, it is immediately obvious that the observability of deputy 1 has increased for both options. Therefore, it is concluded that in order to increase the observability of deputy 1, range measurements between the chief and deputy 2 and between the two deputies are required. A range measurement between the chief and deputy 1 is not necessary. Which of these two options leads to a more accurate estimate for deputy 1 is evident from the results displayed in Table 4.8, but since the eigenvalues for option 3 are all larger than those for option 2, it can already be expected that the result for option 3 will be better than for option 2.
Table 4.8: Estimation results for the two-spacecraft and three-spacecraft formations. Entries with an asterisk (*) have a high probability of divergence and their values change for each run of the estimator.

<table>
<thead>
<tr>
<th></th>
<th>Option 0</th>
<th>Option 1</th>
<th>Option 2a</th>
<th>Option 2b</th>
<th>Option 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{x_1}$ [mm]</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>$\sigma_{y_1}$ [mm]</td>
<td>1.2</td>
<td>1.2</td>
<td>1.3</td>
<td>1.9</td>
<td>0.9</td>
</tr>
<tr>
<td>$\sigma_{z_1}$ [mm]</td>
<td>$\sim 200^*$</td>
<td>$\sim 200^*$</td>
<td>$\sim 200^*$</td>
<td>3.1</td>
<td>3.1</td>
</tr>
<tr>
<td>$\sigma_{\dot{x}_1}$ [$\mu$m/s]</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>$\sigma_{\dot{y}_1}$ [$\mu$m/s]</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>1.2</td>
<td>0.6</td>
</tr>
<tr>
<td>$\sigma_{\dot{z}_1}$ [$\mu$m/s]</td>
<td>$\sim 200^*$</td>
<td>$\sim 200^*$</td>
<td>$\sim 200^*$</td>
<td>3.3</td>
<td>2.4</td>
</tr>
<tr>
<td>$\sigma_{x_2}$ [mm]</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>$\sigma_{y_2}$ [mm]</td>
<td>1.3</td>
<td>1.9</td>
<td>1.3</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{z_2}$ [mm]</td>
<td>2.7</td>
<td>$= \sigma_{z_1}^*$</td>
<td>2.7</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\dot{x}_2}$ [$\mu$m/s]</td>
<td>0.3</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>$\sigma_{\dot{y}_2}$ [$\mu$m/s]</td>
<td>0.8</td>
<td>1.1</td>
<td>0.8</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\dot{z}_2}$ [$\mu$m/s]</td>
<td>2.9</td>
<td>$= \sigma_{z_1}^*$</td>
<td>2.9</td>
<td>2.5</td>
<td></td>
</tr>
</tbody>
</table>

The accuracy in the estimation is displayed for all options in Table 4.8. It is noted that in the table, no effort has been made to compensate for the different number of measurements for the different options. This, next to the different inter-satellite link options, also affects the accuracy in the relative state estimates, but is considered to be of secondary importance in this analysis.

For option 1, the estimation results confirm the observations from the observability analysis: The accuracy in the estimate for deputy 1 is the same as in option 0 and the out-of-plane state components are still prone to apparent divergence. The state vector components of deputy 2 on the other hand can all be estimated accurately due to the relatively large ratio $z_{\text{max}} / r_e$ for that spacecraft, although it is clear that the estimate for the out-of-plane state components is much less accurate than the estimate for the in-plane state components.

For option 2a, the estimation result is worse than for option 1 since now also the estimate for the out-of-plane state components of deputy 2 can diverge. In addition, the accuracy in the estimate of the in-plane state components for deputy 2 is not as good as for option 1, which is a result of the indirect range measurement to deputy 2. Lastly, the estimation result for the out-of-plane state components for deputy 1 and deputy 2 is almost identical. That is, the accuracy in the estimate of $z_1 (\dot{z}_1)$ is identical to the accuracy in the estimate of $z_2 (\dot{z}_2)$ and the error in the estimate of $z_1 (\dot{z}_1)$ is almost identical to the error in the estimate of $z_2 (\dot{z}_2)$. This is a direct result of the two smallest eigenvectors of the system pointing equally towards two out-of-plane state components, as observed during the observability analysis.

For options 2b and 3, hardly any apparent divergence is observed anymore. Even when the out-of-plane motion is completely absent for deputy 1, the estimator converges for all state components. For option 2b, compared to the three previous formation options, the
accuracy in the estimate for the in-plane state components of deputy 1 has degraded considerably, which is due to the indirect measurement of the range of deputy 1. The accuracy in the estimate for deputy 2 is the same as in option 1 and thus the range measurement between the two deputies does not add information that improves the estimate for deputy 2. For option 3, the accuracy in the estimate for all state components is as good or better than for all the other formation options. This is due to the system being fully connected, resulting in more complete information on the formation geometry and in 1.5 times more measurements being available than for options 2a and 2b. For both option 2b and option 3, the accuracy in the estimate of the out-of-plane state components is much less accurate than the estimate of the in-plane state components.

Concluding, adding a spacecraft to a two-spacecraft formation for which the relative state is poorly observable due to a very small relative out-of-plane angle of the deputy can increase the observability of that deputy. However, in order to achieve this, range measurements between the chief and the second deputy and between the two deputies are required to be made. Having range measurements between all spacecraft increases the accuracy in the state estimate, but is not required unless the accuracy in the estimate of the in-plane state components of the original deputy is not allowed to deteriorate compared to the original situation.

Next to the extensive analysis performed when $r_e = 2$ km and $z_{\text{max}} = 0.5r_e$ for deputy 2 with respect to the chief, a sensitivity analysis has been performed to obtain a qualitative understanding on how the accuracy in the estimation of the relative state of the three-spacecraft system changes when the relative motion of deputy 2 with respect to the chief is changed. This has been done for options 2b and 3 only.

When $r_e$ is not 2 km, but 1.1 km for deputy 2, there is a relatively small along-track separation between deputy 1 and deputy 2. In case of option 2b, this results in a significant decrease in the accuracy of the estimates for the in-plane state components of deputy 1. However, for the out-of-plane state components of deputy 1, the accuracy increases by almost a factor two, which can be explained by the increase of the relative out-of-plane angle. For deputy 2, the accuracy for the in-plane state components is hardly affected while the accuracy for the out-of-plane state components is also almost doubled. Changing to option 3 'fixes' the poor result for the in-plane state components of deputy 1. For both options 2b and 3, increasing the along-track distance of deputy 2 to 3 km leads to a deterioration in the estimate for the out-of-plane state components while the result for the in-plane state components is hardly affected.

For both options 2b and 3, when increasing the out-of-plane motion of deputy 2 such that $z_{\text{max}} = 2$ km and thus $z_{\text{max}} / r_e = 1$, the accuracy in the estimate of the out-of-plane state components of both deputies increases by almost a factor two. For the in-plane state components, the accuracy in their estimates decreases slightly. Increasing $z_{\text{max}} / r_e$ even further leads to a further improvement in the accuracy for the out-of-plane state components, but also to a further decrease in the accuracy of the estimates of the in-plane state components. This is especially true for option 2b. For option 3, due to the system being fully connected, the decrease in accuracy for the in-plane state components is sig-
nificantly smaller than for option 2b. For the out-of-plane state components, the rate of improvement in the estimation accuracy decreases for increasing $z_{\text{max}}/r_e$, while the rate of deterioration in the estimation accuracy for the in-plane state components increases.

Thus, for a certain formation geometry and network connectivity, there will be an optimal relative motion of deputy 2 with respect to the chief and deputy 1 where the RSS of the accuracy in the estimate of the state vector of deputy 1, or of e.g. the complete state vector, is minimal.
Chapter 5

Conclusions

The main goal of this research was to obtain insight in the key factors that drive the obtainable accuracy in the estimate of the relative position and velocity of two formation flying spacecraft that rely on locally generated ranging signals to obtain the data necessary to perform autonomous relative navigation. The underlying idea is that this insight can subsequently be implemented to aid the design of the relative navigation system (RNS) of future formation flying spacecraft in order to arrive at a more efficient, less costly, design in a structured manner that is supported by theory. Another goal of this research, which is tightly linked to the major goal, was to determine to what extent small spacecraft, with small dimensions and a minimalistic relative navigation sensor suite, can perform autonomous relative navigation. To make the findings applicable to general formations not necessarily in Earth orbit, where Global Navigation Satellite System (GNSS) measurements can potentially be used to provide the necessary observations, the spacecraft in this research perform inter-satellite range measurements using locally generated radiofrequency (RF) signals.

The following sections will summarize the foregoing chapters, list the main conclusions that have been drawn from the research performed, and provide recommendations and an outlook for future work in this field.

5.1 Summary

In order to research a certain subject, it is important to understand the role of that subject in a larger context. To that end, the first chapter of this thesis introduced the concept of spacecraft formation flying and discussed how it compares to other distributed space systems. Next, the spacecraft formation itself was treated in detail. The importance of autonomy in the operations of spacecraft formations, with a clear focus on Guidance, Navigation, and Control (GNC) aspects, was discussed. This was followed by an overview of major potential benefits of spacecraft formations over monolithic space systems. How-
ever, spacecraft formations also pose new challenges ranging from mission design up to mission operations. These challenges need to be tackled in order to fully exploit the potential benefits of spacecraft formations. This research was performed to support this effort. By analyzing the required performance of the RNS for past, present, and future (proposals for) formation flying missions, a clear trend of increasingly demanding relative navigation accuracy requirements over time could be discerned. This has been quantified by introducing the concept of dynamic range, defined as the ratio of the distance between the formation flying satellites over the required formation control accuracy.

Following this introduction, the three main research questions for this research were introduced and motivated. Two of these questions focused on how major design-driven aspects of the chosen RF-based RNS such as ranging accuracy and antenna baseline together with the mission-driven aspects of inter-satellite distance and relative motion influence the accuracy in the estimation of the relative position and velocity of two formation flying spacecraft. This knowledge can be used in the initial design of the RNS for future formation flying missions. In addition, the sub-question was raised whether there exist conditions for which usage of only a single antenna per spacecraft would yield a relative navigation performance comparable to or better than usage of multiple antennas on the spacecraft. Answering this sub-question is important due to the increasing use of simple, low cost spacecraft such as Cubesats. The third research question addressed the issue of estimator divergence, resulting in an erroneous state estimate, in case of a small relative out-of-plane motion between two formation flying spacecraft. Although this phenomenon is known amongst formation flying experts, only qualitative knowledge on this phenomenon existed. No numerical or theoretical results were available that allowed to determine the minimum out-of-plane motion that is required to have a small probability of (apparent) estimator divergence. Therefore, one of the goals of this research was to provide this knowledge.

In order to be able to answer the research questions, all aspects of the RNS have to be properly understood. To that end, chapter 2 presented the findings of extensive research that has been performed into the main technologies needed to facilitate autonomous relative navigation. First, it was discussed which information needs to be present and processed on the spacecraft in order to be able to produce a relative state estimate. Then, assuming that an RF-based RNS would be used, the basic means by which inter-satellite range measurements could be obtained were developed.

The major part of chapter 2 was devoted to a discussion of the major drivers for the design of the inter-satellite ranging signal, which can also act as the communication link between the two spacecraft. Considerations on the frequency of the carrier signal were presented. Since the spacecraft in a formation are regularly transmitting signals, the use of multiple access technology, which separates the signals from different spacecraft in time, code, and/or frequency, is needed. The choice for the multiple access technology to be used depends heavily on the formation design. As such, there exists no single ‘best’ multiple access technology for spacecraft formations. The design features of the ranging signal itself, assumed to be a direct sequence spread spectrum (DSSS) signal, were extensively
covered and an analysis of the Cramér-Rao Bound (CRB) was used to show which signal features are beneficial for an implementation in inter-satellite ranging signals.

The last part of chapter 2 dealt with a discussion of the design considerations at hardware level. The design of the spacecraft has to take into account antenna placement and antenna beamwidth to enable continuous tracking and a high signal-to-noise ratio of the signal(s) transmitted by the other spacecraft. In addition, traditional subsystems of the spacecraft (CDHS, AOCS, etc.) have to interface with the RNS, which requires an additional engineering effort. The design of the transceiver and its main components was also addressed, focusing on design solutions to reduce power consumption and increase the accuracy of the range measurement. Lastly, the importance of clock stability was addressed and several common crystal oscillator designs were compared.

Chapter 3 extensively discussed the relative navigation methods and analyses used during this research. The first section of that chapter explained and argued for the need to introduce several system level simplifications in order to keep the research activities generally manageable and focus on the basic principles driving the performance of such a system. The second section dealt with the derivation of the equations used in the research to model the relative motion of two spacecraft around a primary body such as the Earth. In this section it was also mentioned that the dynamic model used in this research is similar to the dynamic model for the relative motion of two spacecraft orbiting a Lagrange point: Both dynamic models show, to first order, a decoupling of the relative out-of-plane motion from the relative in-plane motion. Thus, the results obtained in this research considering the magnitude of the relative out-of-plane motion can also be qualitatively applied for formation flying missions at Lagrange points. The third section of chapter 3 discussed how range, range rate, and line-of-sight observations can be obtained from one-way range measurements and which sources can introduce errors in the measurements.

The fourth and last section of chapter 3 discussed the relative state estimation. The need for integer ambiguity resolution of carrier-phase measurements was explained and some methods to achieve this were introduced. Then, the estimator used in this research, a batch least-squares estimator with a-priori information, and the major analysis tool, the observability analysis, were introduced and discussed. The last two subsections dealt with two issues that arose while conducting the research. The first of these was apparent filter divergence and it was shown that the type of estimator used can determine whether or not this will occur. The second issue was that no literature could be found that showed how the mean and variance in the estimation of the position of a stationary point by means of range measurements to three other points was affected by the measurement accuracy and the geometry of the problem. Since knowledge of this was likely to be beneficial in answering the research questions, this was determined by means of a statistical analysis.

Chapter 4 presented the major results that have been obtained during this research by means of extensive numerical simulations. The chapter provided the objectives for the simulations and the general scenario with all the assumptions and variables used for the simulations. The general scenario was a formation of two spacecraft in low Earth orbit. One spacecraft was referred to as the chief spacecraft and its position coincided with the
center of a local coordinate frame called the Hill frame with axis pointing in radial, along-track, and cross-track directions. The other spacecraft was called the deputy spacecraft. The state (i.e., position and velocity) of the deputy relative to that of the chief was estimated using a batch least-squares algorithm. The information needed for the estimation was provided by means of noisy range measurements between a transmitting antenna on the deputy and either one or three receiving antennas on the chief spacecraft. A-priori knowledge on the relative state of the deputy was assumed to be available as well. The positions of the receiving antennas in the Hill frame were assumed to be known perfectly and constant with time.

Subsequently, the formation geometries selected for the simulations were introduced and discussed. The geometries considered were the pendulum, where one satellite trails the other satellite and exhibits a periodic out-of-plane motion relative to the orbital plane of other satellite, and the ‘safe’ ellipse, where one satellite describes an elliptical motion with respect to the other satellite. These geometries were selected because they exhibit bounded 1:1 commensurate motion (i.e., the orbital periods of the spacecraft match, resulting in a periodic relative motion that is stable over time) in case of linearized dynamics and because they have various practical uses for e.g. Earth observation, spacecraft inspection, and spacecraft rendezvous and docking (RVD).

The last section of the chapter contained the major results of this research and extensive discussions of these results. It was shown how the antenna baseline, ranging accuracy, and inter-satellite distance affect the accuracy in the relative state estimate. In addition, clear differences between the pendulum and elliptical formation geometries were found. It was also found that for the elliptical formation geometry the location of the antenna(s) in the local reference frame is of major importance for the accuracy in the relative state estimate. Lastly, it was determined that it is not so much the magnitude of the relative out-of-plane motion, but the out-of-plane angle of the formation that determines when estimator divergence has a high probability of occurrence.

The last section in chapter 4 also dealt with two questions that arose while conducting the research. The first of these was tied to the observation that for the pendulum formation geometry, certain scenarios clearly deviated from the general trend for the accuracy in the relative state estimate as a function of the variables under study. The second of these was whether adding another spacecraft to a formation with a small relative out-of-plane angle could substantially improve the estimate of the relative state of the original two-spacecraft formation.

5.2 Conclusions

In this section, the conclusions that have been drawn from the research are listed. First, the conclusions concerning the three research questions are provided. Then, additional conclusions are given. Section 5.3 will focus on the main contributions of this work.

It is stressed that the answers to the research questions depend on several simplifying assumptions that have been applied in order to focus on the fundamental aspects of the
5.2. **Conclusions**

Research subject. In addition, these assumptions allow a generalized treatment of the subject, leading to conclusions that are valid for a wide range of formation flying scenarios. The assumptions have been addressed in sections 3.1 and 4.2. Important assumptions are:

- all measurement biases are removed and the standard deviation in the measurement noise is considered constant,
- the relative motion dynamics can be described by the Hill-Clohessy-Wiltshire (HCW) equations,
- reasonably accurate a-priori information on the relative state at \( t = 0 \) is available,
- the position of the antennas in the Hill frame is known perfectly and the antennas have a \( 4\pi \) steradian field of view.

As the research has revealed that research questions 1 and 2 are tightly linked, they are answered jointly. For easy reference, the research questions formulated in section 1.3 are repeated here:

**Research Question 1** How do the receiver antenna baseline, ranging accuracy, and inter-satellite distance affect the accuracy in the relative state determination for two formation flying spacecraft? Does there exist a transition point between the use of a single antenna and the use of multiple antennas at which the navigation results obtained by using one antenna equal or even outperform the navigation results obtained by using multiple antennas?

**Research Question 2** How are the relative navigation results for a two spacecraft formation influenced by the type of relative motion?

**Research Question 3** How does the magnitude of the relative out-of-plane motion between two spacecraft influence the relative navigation results?

The answers to the first two research questions, starting with the results obtained for the pendulum formation geometry, are as follows:

1. For a pendulum formation geometry, the accuracy in the estimate of the relative state is inversely proportional to the inter-satellite distance and proportional to the measurement accuracy. Thus, a high accuracy in the relative state estimate requires a high measurement accuracy and/or a small inter-satellite distance.

2. For a pendulum formation geometry, for observation arcs longer than one orbit, the antenna baseline has no effect on the accuracy in the estimate of the relative state. The reason for this is that the magnitude of the eigenvalues of this system are not dependent on this parameter for such long observation arcs. For that reason, for long observation arcs, the only advantage of having more than one antenna on the chief is that the accuracy of the state estimate improves with the square root of the
number of antennas because they result in more measurements in the same time interval. Thus, in case of e.g. three Rx antennas, it is possible to save power by switching to a single Rx antenna for the relative navigation after a measurement arc length of at least one orbit, provided that a decrease by $\sqrt{3}$ in relative navigation accuracy is acceptable.

3. For a pendulum formation geometry, for observation arcs shorter than one orbit, an increasing antenna baseline has a positive effect on the accuracy in the estimate of the relative state. Therefore, if there is a time-critical relative navigation accuracy requirement for the pendulum formation geometry, a design using multiple antennas with a large antenna baseline is preferable.

4. For an elliptical formation geometry, the placement of the Rx antennas with respect to the origin of the Hill frame is very important: The direction and length of the vector(s) from the origin of the Hill frame to the Rx antenna(s) has a major contribution to the accuracy in the relative state estimate, which is actually more important than the antenna baseline. The reason for this is that the magnitude of the smallest eigenvalue of this system is a function of both the length and the direction of the vector(s) connecting the origin of the Hill frame and the position of the antenna(s). Positioning an antenna on a vector that has, in $\mathbb{R}^3$, the same direction as the eigenvector corresponding to the smallest eigenvalue of the system increases the information obtained on that vector, resulting in a larger value for the corresponding eigenvalue, which in turn yields a more accurate overall state estimate. An increase in antenna baseline on the other hand will not necessarily always lead to a more accurate overall state estimate: If the increased antenna baseline is accompanied by a rotation of the vectors connecting the origin of the Hill frame and the position of the antennas, this rotation can result in a less favorable orientation of those vectors. The combined effect might then lead to a less accurate relative state estimate even though the antenna baseline has increased.

5. For an elliptical formation geometry, for scenarios with a single antenna on the chief, the distance between the antenna and the origin of the Hill frame must be more than twice the ranging accuracy in order for the accuracy in the relative state estimate to be inversely proportional to the inter-satellite distance and proportional to the measurement accuracy. If it is smaller, the accuracy in the relative state estimate becomes also a function of the accuracy in the a-priori information. In the limit, for very poor ranging accuracy and very small antenna baselines, the accuracy in the relative state estimate is almost equal to the accuracy in the a-priori information. The reason is that the angular information, obtained by placing the antenna not in the origin of the Hill frame, is lost since the measurement error is comparable to or larger than the baseline thus created.

6. For an elliptical formation geometry, the only advantage of having more than one antenna on the chief is that the accuracy of the state estimate improves at most
with the square root of the number of antennas (assuming that each antenna is connected to a dedicated receiver channel) because they result in more measurements in the same time interval. However, since there is only one optimum position for an antenna, in practice the improvement will be smaller. In case of three antennas in an equilateral triangle with the distance of all antennas to the origin of the Hill frame the same, the improvement is at most a factor 1.4 instead of 1.7.

For research question 3, the answers are:

1. The magnitude of the relative out-of-plane motion itself does not drive the accuracy of the relative state estimate or the observability of the relative state. In fact, it is the relative out-of-plane angle of the formation (the amplitude of the out-of-plane motion divided by the amplitude of the along-track motion) that drives this.

2. For the pendulum formation geometry, the accuracy in the relative state estimate is proportional to the magnitude of the relative out-of-plane angle: The larger this angle is, the higher the accuracy. For the elliptical formation geometry, the accuracy in the relative state estimate does not depend on the magnitude of the relative out-of-plane angle.

3. The out-of-plane angle at which the estimation process is stable, i.e., has a low probability of apparent divergence, differs for the two formation geometries studied. For the pendulum formation geometry and most scenarios considered, this angle needs to be larger than $0.6^\circ$. To account for all scenarios, the angle needs to be larger than $2.9^\circ$. For the elliptical formation geometry and most scenarios considered, the angle needs to be larger than $8.5^\circ$. To account for all scenarios, the angle even needs to be larger than $14^\circ$.

Next to these main conclusions of the research, the work performed during the research has lead to the following additional results:

1. Of the signal pulse shapes studied in subsection 2.2.5 of this thesis, BOC(1,1) is considered to offer a good mix of achievable ranging accuracy and ease of acquisition at relatively narrow signal bandwidths. This is favorable for application on small satellites since these typically have limited on-board resources. Other signal design aspects such as the PRN code, data modulation, data frames, and coding are considered to be of lesser importance.

2. In subsection 3.4.4, a simple static scenario with nonlinear measurements was used to show that the apparent divergence of a certain estimator or filter can be mitigated by switching to a different type of estimator or filter. Alternatively, applying a coordinate transformation such that the measurement equations are a linear function of the state components was also shown to remove the apparent divergence and is also a viable option.
3. The statistical analysis of the relative position determination of a stationary transmitting antenna in a coordinate frame defined by three receiving antennas, performed in subsection 3.4.5, revealed the following: To obtain the smallest error in the position estimate, the receiving antennas have to be arranged in a right triangle (a triangle in which one angle is a right, i.e. 90°, angle). In addition, the standard deviation in the estimate for the position coordinates corresponding to the plane in which the receiving antennas are located (x and y) depends linearly on the range and the measurement accuracy and inversely linear on the baseline between the receiving antennas. The standard deviation in the estimate for the remaining position coordinate (z) depends linearly on the range, quadratically on the measurement accuracy, and inversely quadratic on the baseline between the receiving antennas. Thus, the estimate for z will generally be more accurate than the estimate of x and y, provided that the ranging accuracy is much smaller than the baseline between the receiving antennas.

4. The observability analysis performed in subsection 4.4.1 revealed that by measuring the inter-satellite distance between two spacecraft, in fact the relative semi-major axis of the formation is observed and can be estimated with high accuracy.

5. The observability analysis performed in subsection 4.4.1 also revealed that, due to the decoupling of the in-plane and out-of-plane state components in the system dynamics, the out-of-plane state components are the least observable state components unless the out-of-plane angle of the formation is made large (tens of degrees).

6. The results obtained in subsection 4.4.4 lead to the conclusion that for pendulum formations with small out-of-plane amplitudes and small inter-satellite distances, a large antenna baseline can positively influence the accuracy in the relative navigation if the amplitude (i.e., z_{max}) of the relative out-of-plane motion of the deputy spacecraft is smaller than the antenna baseline on the chief spacecraft. Therefore, a spacecraft with small dimensions is at a disadvantage compared to a spacecraft with large dimensions since the latter allows for a larger antenna baseline without the need for deployable structures. In addition, in a RV(D) scenario, where the inter-satellite distance is commonly decreased in a step-wise manner and where the relative out-of-plane motion is small, having a large antenna baseline will be advantageous since this will yield the highest attainable accuracy in the relative navigation during most of the approach phase. However, when the inter-satellite distance becomes smaller than the antenna baseline, which is an extreme scenario, the accuracy in the relative position estimate becomes only a function of the ranging accuracy and the number of measurements. It therefore depends on which accuracy is desired at which stage of the RV(D) maneuver to decide which antenna baseline and ranging accuracy should be used and thus whether a small spacecraft is adequate.

7. As shown in subsection 4.4.5, adding an additional spacecraft to a formation where the deputy has a small relative out-of-plane angle relative to the chief, which implies poor observability for the out-of-plane state components of the deputy, can
substantially improve the observability of the out-of-plane state components of the original deputy. This requires that the relative out-of-plane angle of the second deputy with respect to the first deputy is large and that range measurements between the chief and the second deputy and between the original deputy and the second deputy are performed. Additional range measurements between the chief and the original deputy are not required, but will improve the accuracy in the overall state estimate. Lastly, it was shown that for a certain formation geometry and network connectivity, there will be an optimal relative motion of the second deputy with respect to the chief and the original deputy which leads to the highest accuracy in the state estimate.

Lastly, several strategies have been devised during this research to cope with poor observability due to a small out-of-plane angle of the formation. These strategies can be considered for implementation on formations with small out-of-plane angles and are:

1. **Use different sensors.** Sensors with which the along-track distance and the out-of-plane position or velocity can be directly measured are preferable to all other sensors since measuring these state components is sufficient to guarantee complete observability for the system as shown in [Matko et al. 2010]. Otherwise, a sensor set that uses one sensor to directly measure the line-of-sight angles (azimuth and elevation) between the satellites with sufficient accuracy, e.g. a high performance camera such as used in the PRISMA mission or a phased array antenna, and another (RF) sensor to determine the range between the satellites is preferable. Nevertheless, availability, cost, mass, and experience are important considerations in the selection of alternative sensors and can be reasons to stick to a dedicated RF-based design.

2. **Use a dynamic model that includes second-order effects.** In contrast to linearized models, second-order models maintain the link that exists between the in-plane and out-of-plane motion in a formation, as treated in subsection 3.2.1. However, since the link disappears upon linearization, it is weak and might not provide much benefit during actual implementation. A higher fidelity linear model that includes disturbances like differential \( J_2 \) and differential drag is not likely to improve the relative navigation performance much since differential drag is an in-plane effect and thus does not add information on the out-of-plane motion. For the relative inclinations considered in this research, the effect of differential \( J_2 \), which causes an angle between two orbital planes due to a differential RAAN to slowly grow over time, is too small to be detectable within a few orbits for ranging accuracies on the order of cm.

3. **Change of basis.** If the smallest eigenvalue is much smaller than the other eigenvalues and if it has significant contributions from multiple state components, these state components will be poorly observable. In that case, a change of basis such that the linear combination of the state components in the eigenvector corresponding to the smallest eigenvalue becomes a new state variable will reduce the number of poorly observable states in the new basis to one. This can be implemented by
choosing the new basis to be spanned by the eigenvectors of the Gramian. This will reduce the cross-correlation in the estimate and will lead to a more accurate estimation result although there will still be one state that is poorly observable. Thus, the estimation process still has a large probability of divergence if the observability of the newly created but poorly observable state component is very bad. Therefore, using mathematics to rewrite the problem will not alleviate the fundamental issue of poor observability of the system; it will only reduce the number of poorly observable state components from two to one. For completeness, it is noted that using the so-called relative orbital elements to describe the relative motion of the spacecraft, which is effectively a change of basis, will not alleviate the problem of poor observability since the in-plane and out-of-plane motion are still decoupled. The fundamental reason for this is that the equations of relative motion for the relative orbital elements are still derived from the HCW equations.

4. (Temporarily) increase the (relative) magnitude of the relative out-of-plane motion. This can be achieved by adjusting the amplitude of the relative out-of-plane motion and/or the inter-satellite distance. This requires thruster firings and therefore undesired propellant consumption, but it is the only solution to increase the observability of the out-of-plane state components that does not require the use of additional or other sensors or a higher fidelity model for the relative motion (see the preceding points). Once a stable relative navigation solution has been achieved in the adjusted configuration, the formation can be brought back to the original geometry but now with a much improved accuracy in the estimate of the out-of-plane state components. However, also this solution does not remove the origin of the problem. In addition, this procedure has to be repeated many times over the course of the mission. The reason is that once the original geometry has been re-acquired the accuracy in the estimate of the out-of-plane state components will quickly reduce over time since these are now again poorly observable. Thus, this solution is not likely to be very beneficial in practical applications.

5.3 Main Contributions of this Work

The research performed has led to several new results and insights in the field of relative navigation for formation flying spacecraft. These new insights are summarized in this section.

The research has shown, for the formation geometries considered, that the accuracy in the estimation of the relative state for formation flying spacecraft depends linearly on the ranging accuracy and inversely linearly on the inter-satellite distance, which is as expected and is not considered to be a new result. However, an important new result is that the accuracy in the estimation of the relative state does not depend on the antenna baseline for observation arc lengths of more than one orbit. For an elliptical formation geometry, it is very important that the vector from the origin of the local frame to one (of the) antenna(s)
in the local frame is perpendicular to the direction of the least observable eigenvector of the system. In addition, the accuracy in the estimate of the relative state scales inversely linear with the length of this vector. Yet, the antenna position in the local frame has to be known with high accuracy, which is challenging since it requires accurate knowledge of the satellite orbit and accurate knowledge of the satellite attitude. For small satellites, which are typically constrained in the accuracy with which the orbit and attitude can be estimated when compared to larger spacecraft, this presents a challenge. For a pendulum formation geometry, the antenna baseline is only of importance if the observation arc length is less than one orbit. If it is longer, the baseline does not contribute to the accuracy in the relative state estimate. Thus, theoretically, it is possible to perform relative navigation between two spacecraft using only one ranging antenna per spacecraft. Therefore, autonomous formation flying with small spacecraft that are constrained in terms of mass, power, and volume is feasible provided that they can support a transceiver and one ranging antenna. Yet, more antennas can facilitate more range measurements, which in turn can increase the accuracy in the relative state estimate purely due to statistics. Moreover, engineering considerations such as line-of-sight and redundancy can also create the need for more than one ranging antenna per spacecraft.

There is not a clear transition point between using one or multiple ranging antennas. This depends very much on the constraints imposed by the mission. For a pendulum formation geometry, it is more likely that a single ranging antenna can be used per satellite than in case of an elliptical formation geometry. This is due to a better observability for the pendulum formation geometry and due to a more favorable line-of-sight throughout the orbit. Naturally, if the formation consists out of more than two satellites, the formation geometry quickly becomes such that multiple antennas are needed per spacecraft to enable lines-of-sight to the other spacecraft. Lastly, if the relative attitude of the satellites needs to be determined with an accuracy that is higher than what can be achieved by differencing the absolute attitudes of the satellites, it is unavoidable to use at least three antennas per satellite. Thus, usage of a single ranging antenna per satellite is feasible, which implies that very small satellites can potentially be used, but this will be restricted to very specific applications.

For the relative navigation, a new result is that the magnitude of the relative out-of-plane motion in itself is not very important for the accuracy of the relative state estimate. Instead, for pendulum formation geometries, it is the relative out-of-plane angle that drives the accuracy in the relative state estimate. For the elliptical formation geometry, the accuracy in the relative state estimate does not depend on the relative out-of-plane motion.

It is already known that a very small relative out-of-plane motion causes a high probability of estimator divergence. This research has provided a first attempt to quantify this probability as a function of the out-of-plane angle. In addition, this research has shown that in case of a pendulum formation geometry the relative out-of-plane angle can be several times smaller than in case of an elliptical formation geometry without causing the probability of estimator divergence to increase significantly.
The observation that the relative out-of-plane angle is an important parameter in the estimation of the relative state has led to the realization that this can be exploited in formations consisting of more than two spacecraft to reduce the number of inter-satellite links: If there is one pair of spacecraft with limited relative out-of-plane motion, the relative navigation between these spacecraft is best done via another spacecraft which has a large relative out-of-plane motion with the other spacecraft.

An analytical statistical treatment of the relative navigation problem has been introduced. Although this treatment assumes a general and static scenario and does not account for biases in the range measurements, it provides a framework that can be adapted for use in very specific and in dynamic scenarios. Thus, it allows for quick ‘rule of thumb’ estimates of the achievable relative navigation accuracy. In subsection 4.4.2 it was shown that the setting used in the analytical treatment already yields results that compare well with the results obtained for the elliptical formation geometry. Thus, it can be used in the early design phase of a safe ellipse formation geometry to get a first rough estimate of the accuracy achievable in the estimate of the relative state.

For the first time, the probability of apparent estimator divergence as a function of formation geometry, relative out-of-plane angle, antenna baseline, ranging accuracy, and inter-satellite distance has been performed. This revealed that for large relative out-of-plane angles this probability varies between 0-2%. As the relative out-of-plane angle is reduced, the probability suddenly increases sharply, leading to the conclusion that for small relative out-of-plane angles, the estimation process is unreliable. For the safe ellipse formation geometry, much larger relative out-of-plane angles are needed than for the pendulum formation geometry (14° versus 2.9°) in order for the estimation process to be reliable. In addition, in general, both a large antenna baseline and a small inter-satellite distance lead to a lower probability of estimator divergence. The influence of the ranging accuracy on the probability of apparent estimator divergence is more complex and not well understood. Lastly, when the variables \( \sigma \rho \), \( d \), and \( r_e \) are all multiplied by the same scalar, the probability of apparent estimator divergence does not change, which verifies the linearity of the system.

The work performed has also led to several conclusions that will benefit and contribute to the design of efficient formation flying missions. This is of special importance if the formations consist of small satellites, which have limited resources compared to typical satellites with masses of hundreds of kilograms.

Firstly, the power consumption of the RNS can be minimized through several means. A simple means to achieve this, if feasible within the mission requirements, is to use a low signal frequency and small inter-satellite distance to reduce the free space loss of the ranging signal. In addition, a high antenna gain will also reduce the required signal power, but only if a good alignment of the boresights of the antennas on the different spacecraft can be achieved. Next to that, the pulse shape of the ranging signal must be chosen such that a sharp correlation peak is obtained in the DLL. This will allow a comparable code-based ranging accuracy at a lower signal-to-noise ratio than in case of a less optimal pulse shape. Lastly, active signal power control by the transceiver, based on predictions of the
actual inter-satellite range, is very useful to reduce power consumption for missions where the inter-satellite distance can vary significantly.

Secondly, from a navigation point of view, a pendulum formation geometry is much more preferable than an elliptical formation geometry and should therefore be adopted when possible. The reasons for this preference are as follows:

1. For the pendulum formation geometry, apparent estimator divergence occurs at much smaller relative out-of-plane angles than for the safe ellipse (2.9° versus 14°). Thus, the pendulum formation geometry offers more freedom in the selection of the relative out-of-plane angle.

2. The pendulum formation geometry leads to a higher accuracy in the estimate of the relative state than the safe ellipse formation geometry for comparable scenarios (e.g., \( r_e/z_{max} = 1, \sigma_\rho/d \ll 1 \)).

3. For the pendulum formation geometry there is a reduced dependence on the number of antennas and antenna position(s) on the chief spacecraft:
   (a) The accuracy of the relative state estimate does not depend on the antenna baseline (for an observation arc length of more than one orbit). Thus, a single antenna on a physically small chief spacecraft will yield a good navigation result, irrespective of the ranging accuracy. For the safe ellipse, this is not necessarily the case.
   (b) The along-track distance of the deputy is constant (to first order) and the relative motion is in the \( yz \)-plane of the Hill frame only. This implies that the antenna(s) on the chief can be placed such that the spacecraft structure does not obstruct the line-of-sight between the antenna(s) on the chief and the antenna on the deputy. For a safe ellipse where the chief is positioned inside the ellipse, maintaining a good line-of-sight between the antenna(s) on the chief and on the deputy is much harder. This requires the use of, e.g., two low gain antennas on opposite sides of the chief that hand over tracking of the deputy once the signal-to-noise ratio for one antenna drops below a certain threshold. Alternatively, a rotation of the chief around its axis allows a single antenna to continuously track the deputy. Another possibility is the placement of one hemispherical antenna on a long boom to minimize signal interference from the spacecraft structure during most of the orbit. Obviously, these solutions are much more involved than in case of a pendulum formation geometry.
   (c) If needed to reduce the power consumption of an RNS with three antennas on the chief, the number of active channels can be reduced to one after an observation arc length of more than one orbit if a \( \sqrt{3} \) reduction in the accuracy of the relative state estimate is acceptable. For a safe ellipse formation geometry, it is unlikely that this will be feasible.

4. Contrary to the safe ellipse formation geometry, for the pendulum formation geometry the ratio of the ranging accuracy over antenna baseline does not need to be very
small ($\sigma_{\rho}/d \leq 0.25$) in order for the accuracy of the relative state estimate to be independent on the a priori information. This means that, e.g., a lower ranging signal strength, lower antenna gain, and/or more ‘noisy’ electronics are acceptable for the same antenna baseline (if acceptable for the mission). Thus, a less complex or less costly RNS system can be used.

Thirdly, if a safe ellipse formation geometry must be used for a mission, it is extremely important to place the Rx antenna(s) on the chief in the optimal position(s) (on a vector with, in $\mathbb{R}^3$, the same direction as the least observable eigenvector of the system). This can yield a significant improvement in the accuracy of the relative state estimate, which is due mostly to an improvement in the accuracy for the estimates of the along-track and out-of-plane positions.

5.4 Recommendations for Further Work

The research performed has led to several interesting conclusions that can be of benefit in the design of future formation flying missions using small spacecraft. However, it is recommended to further study several of these topics since this can yield additional and valuable knowledge.

The first recommendation is to study in more detail the probability of apparent estimator divergence as a function of the ranging accuracy, the inter-satellite range, the antenna baseline, and the relative motion (e.g., formation geometry and out-of-plane angle). Especially the sudden increase and subsequent decrease in the probability of apparent estimator divergence for the elliptical formation geometry around an out-of-plane angle of $3^\circ$ is not understood. In addition, the variation in the probability of apparent estimator divergence as a function of the out-of-plane angle for different measurement accuracy, inter-satellite range, and antenna baseline shows results which should be analyzed further. The linear observability analysis applied during the research fails to explain these phenomena. A nonlinear observability analysis might be able to provide the answers.

The second recommendation is to perform a more extensive sensitivity analysis for the three-spacecraft formation. This will lead to an improved understanding of the observability of the relative state for formations consisting out of more than two spacecraft. This can aid in the design of an efficient RNS for future formations with three or more spacecraft.

The third recommendation is to research how reliable integer ambiguity resolution and multipath mitigation can be best achieved on a small spacecraft. Although these topics have hardly been dealt with in this thesis, they are important when accurate range measurements need to be obtained. Especially for small spacecraft in elliptical formation geometries, accurate range measurements are important for accurate estimation of the relative state.
5.5 Outlook

The research performed for this thesis has shown that autonomous relative navigation using locally generated RF ranging signals is feasible for small spacecraft in low Earth orbit. As demonstrated in the PRISMA mission, considered to be the most advanced autonomous formation flying mission to date, this can be accomplished with centimeter to meter-level accuracy at inter-satellite distances of tens to thousands of meters. Yet, the two spacecraft used in this mission still have considerable mass, dimensions, and power (e.g., "Mango’s" central body measures $0.75 \times 0.75 \times 0.82$ m, has a mass of 140 kg and has 400 W power available). The masses of the relative navigation electronics boxes employed on both spacecraft are 7 kg and 9 kg and their power consumption is between 23 W and 30 W. Thus, to obtain the quoted performance, considerable resources are still needed.

For very small satellites, in the order of 1-10 kg, such as for the DelFFi or CanX-4&5 missions, only very limited resources are available as compared to the resources available to the satellites in the PRISMA mission. Restrictions in mass, power, physical dimensions, and mission budget will inevitably lead to less accurate relative navigation if this is performed using an RF-based system (for CanX-4&5, the relative navigation will be performed using carrier differential GPS). Nevertheless, if relatively coarse navigation accuracy can be accepted, autonomous relative navigation using locally generated RF ranging signals is feasible with these very small satellites. Provided that the satellites are equipped with sufficiently accurate propulsion and attitude control systems, formation flight in this manner is feasible.

In subsection 1.1.5, it was shown that for many of the currently proposed formation flying missions there is a trend towards increasing dynamic range (i.e., high control accuracy and large inter-satellite distance). In addition, many of the proposed missions will be much more complex than the missions that have been flown up to date. Two causes for that are the number of satellites involved, often more than two, and the need for very accurate relative attitude determination and control. These factors result in extreme demands on the metrology system to be used and are likely to result in a metrology chain where different subsystems are used for coarse, fine, and very fine relative navigation. For such systems, it is likely that RF systems will be used for the coarse and possibly also fine navigation while a laser-based interferometric system is likely to be used for the fine and very fine navigation. This calls for satellites with sufficient resources (volume, power) to support multiple navigation subsystems and all other bus and payload systems. Provided that the payload is relatively compact, such missions might be performed using satellites similar to PRISMA’s Mango satellite. It is however not unlikely that even satellites with such properties are insufficient to fulfill the mission needs. Thus, it will be a challenge, although not necessarily unfeasible, to implement such small satellites for these demanding missions.

For very small satellites it is foreseen that these will be utilized, at least in the near future, for niche applications that require dense sampling in space and/or time and which do not require very high navigation accuracy and that do not require large payloads. Alternatively, very small satellites could also be used in a configuration where tens of small
satellites form a synthetic aperture to collect measurements (e.g., radiowaves) which are subsequently sent towards a relatively large collector satellite. This satellite can send the raw measurements to a ground station for post-processing or can even perform (some of) the post-processing itself and send the result to a ground station. The collector spacecraft could be used as a central GNC hub that determines the (relative) positions of all the small satellites and which sends thruster firing commands to those satellites to maintain the desired formation geometry. Lastly, very small satellites might be utilized to inspect large satellites for, e.g., external damage. However, in order to do this with RF signals, the target needs to be cooperative. Current spacecraft do not have this capability and thus such usage will depend on future developments. It is more likely that such inspector satellites will use a vision-based system for the relative navigation since this does not require a cooperative target and since visual images need to be acquired anyway.

For very small satellites, the potential for use in future formation flying missions seems limited to niche applications. However, for these satellites it is not uncommon to take more development risk (i.e, smaller engineering margins) than in traditional satellite designs, which implies that more functional mass and/or power is available than what would be expected using a more traditional engineering approach. In addition, such satellites typically employ more advanced electronics than traditional satellites. This again results in more functional mass and enables more complex operations and the use of more advanced signal structures. In addition, when many satellites are present in a mission, there is the potential to use the inter-satellite network to improve the relative navigation accuracy and overall redundancy. Lastly, as shown in this research, an accurate antenna positioning can significantly improve the relative navigation accuracy. This can be further enhanced by employing antennas on deployable structures.

Thus, RF-based autonomous relative navigation using (very) small satellites is feasible. The attainable navigation accuracies are high enough to enable formation flight, even with very small satellites. Smart signal structures and transceiver designs exist and can even be improved for future missions. For the near future, small satellites are able to support the systems needed to achieve cm-level relative navigation accuracy. For very small satellites, the achievable relative navigation accuracy will more likely be at m-level. Depending on the specific mission parameters, small satellites can be used for challenging formation flying missions in the near future.
Appendix A

Derivation of the Expectation and Variance in the Position Estimate of a Stationary Transmitter

In this appendix a derivation is presented of the expectation and variance in the position estimate of a stationary transmitting antenna in a reference frame defined by the positions of three receiving antennas. The first section will provide the necessary background material while the derivation will be conducted in the second section.

A.1 Background Material

In this section, equations are provided that yield the expected value \( E[\cdot] \), the variance \( \text{Var}[\cdot] \), and the covariance \( \text{Cov}[\cdot] \) for basic functions of two normally distributed continuous random variables \( X \) and \( Y \). The equations have been obtained from [Mood et al. 1974]. Random variable \( X \) has a mean \( \mu_X \) and variance \( \sigma_X^2 \) while random variable \( Y \) has a mean \( \mu_Y \) and variance \( \sigma_Y^2 \). The scalars \( a \) and \( b \) are sometimes added to the equations to make the results more general. In the last subsection, the equation for a multivariate Taylor series expansion is provided, which is needed for the derivation in section A.2.

A.1.1 Expected Value

For the summation or subtraction of \( X \) and \( Y \) the result for the expected value is fairly straightforward:

\[
E[aX \pm bY] = E[aX] \pm E[bY] = aE[X] \pm bE[Y] = a\mu_X \pm b\mu_Y. \tag{A.1}
\]

When \( X \) and \( Y \) are multiplied, the result for the expected value is:
APPENDIX A. DERIVATION OF THE EXPECTATION AND VARIANCE IN THE POSITION ESTIMATE OF A STATIONARY TRANSMITTER

\[ E[(aX) (bY)] = abE[X]E[Y] + \text{Cov}[X, Y]. \]  \hspace{1cm} (A.2)

If \( X = Y \) and \( a = b = 1 \), Eq. (A.2) reduces to

\[ E[X^2] = (E[X])^2 + \text{Var}[X] = \mu_X^2 + \sigma_X^2. \]  \hspace{1cm} (A.3)

A.1.2 Variance

The variance of a continuous random variable \( X \) is defined as

\[ \text{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2. \]  \hspace{1cm} (A.4)

For the summation or subtraction of \( X \) and \( Y \) the result for the variance is:

\[ \text{Var}[aX \pm bY] = \text{Var}[aX] + \text{Var}[bY] \pm 2\text{Cov}[aX, bY] \]
\[ = a^2\text{Var}[X] + b^2\text{Var}[Y] \pm 2ab\text{Cov}[X, Y]. \]  \hspace{1cm} (A.5)

When \( X \) and \( Y \) are multiplied, the result for the variance is rather involved:

\[ \text{Var}[XY] = (E[Y])^2\text{Var}[X] + (E[X])^2\text{Var}[Y] + 2E[X]E[Y]\text{Cov}[X, Y] - \text{Cov}[X, Y]^2 \]
\[ + E[(X - E[X])^2 (Y - E[Y])] + 2E[Y]E[(X - E[X])^2 (Y - E[Y])] \]
\[ + 2E[X]E[(X - E[X])(Y - E[Y])^2]. \]  \hspace{1cm} (A.6)

If \( X = Y \), Eq. (A.6) yields after some work

\[ \text{Var}[X^2] = 4(E[X]^2)^2\text{Var}[X] - (\text{Var}[X])^2 + E[(X - E[X])^4] + 4E[X]E[(X - E[X])^3] \]
\[ = 4(E[X])^2\text{Var}[X] + 2(\text{Var}[X])^2. \]  \hspace{1cm} (A.7)

If \( X \) and \( Y \) are independent, implying \( \text{Cov}[X, Y] = 0 \), Eq. (A.6) yields

\[ \text{Var}[XY] = (E[Y])^2\text{Var}[X] + (E[X])^2\text{Var}[Y] + \text{Var}[X]\text{Var}[Y]. \]  \hspace{1cm} (A.8)

A.1.3 Covariance

The covariance of two continuous random variables \( X \) and \( Y \) is defined as

\[ \text{Cov}[X, Y] = E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])] = E[(X - \mu_X)(Y - \mu_Y)]. \]  \hspace{1cm} (A.9)
Thus, if \( X = Y \) then \( \text{Cov}[X, Y] = \text{Cov}[X, X] = \text{Var}[X] \). Multiplication of \( X \) and \( Y \) with scalars \( a \) and \( b \) yields

\[
\text{Cov}[aX, bY] = ab\text{Cov}[X, Y]
\]

while addition of \( X \) and \( Y \) with scalars \( a \) and \( b \) yields

\[
\text{Cov}[X + a, Y + b] = \text{Cov}[X, Y].
\]

### A.1.4 Multivariate Taylor Series Expansion

For complex functions with multiple random variables, it can be difficult to derive the expected value and the variance of the function. In such cases, these can be obtained by means of a multivariate Taylor series expansion of the function: First, the function is written as a Taylor series and then the expected value or the variance of the series is computed. Often, a second order expansion is sufficient for a good approximation of the expected value and the variance.

For a function consisting out of a single variable \( x \), the well-known Taylor series expansion around a point \( a \) is

\[
f(x) = f(a) + (x - a) \frac{\partial f(x)}{\partial x} \bigg|_a + \frac{(x - a)^2}{2!} \frac{\partial^2 f(x)}{\partial x^2} \bigg|_a + \frac{(x - a)^3}{3!} \frac{\partial^3 f(x)}{\partial x^3} \bigg|_a + \ldots
\]

which can be written in short notation as

\[
f(x) = \sum_{n=0}^{\infty} \frac{(x - a)^n}{n!} \frac{\partial^n f(x)}{\partial x^n} \bigg|_a.
\]

For a function consisting out of two independent variables \( x \) and \( y \), the Taylor series expansion around points \( a \) and \( b \) is [Bayin 2006]

\[
f(x, y) = f(a, b) + (x - a) \frac{\partial f(x, y)}{\partial x} \bigg|_{a,b} + (y - b) \frac{\partial f(x, y)}{\partial y} \bigg|_{a,b} + \frac{1}{2!} \left\{ (x - a)^2 \frac{\partial^2 f(x, y)}{\partial x^2} \bigg|_{a,b} + 2(x - a)(y - b) \frac{\partial^2 f(x, y)}{\partial x \partial y} \bigg|_{a,b} \right\} + \frac{1}{3!} \left\{ (x - a)^3 \frac{\partial^3 f(x, y)}{\partial x^3} \bigg|_{a,b} \right\} + 3(x - a)^2(y - b) \frac{\partial^3 f(x, y)}{\partial x^2 \partial y} \bigg|_{a,b} + 3(x - a)(y - b)^2 \frac{\partial^3 f(x, y)}{\partial x \partial y^2} \bigg|_{a,b} + (y - b)^3 \frac{\partial^3 f(x, y)}{\partial y^3} \bigg|_{a,b} + \ldots
\]
Appendix A. Derivation of the Expectation and Variance in the Position Estimate of a Stationary Transmitter

In the more general case of \( n \) independent variables \( x_1, x_2, ..., x_n \) the Taylor series expansion around points \( a_1, a_2, ..., a_n \) is written in short notation as

\[
f(x_1, x_2, \ldots, x_n) = \sum_{k=0}^{\infty} \frac{1}{k!} \left\{ \sum_{i=1}^{n} (x_i - a_i) \frac{\partial}{\partial x_i} \right\}^k f(x_1, x_2, \ldots, x_n) \bigg|_{a_1, a_2, \ldots, a_n}.
\] (A.15)

A.2 Derivation

In subsection 3.4.5, the following set of equations

\[
x_{\text{Tx}} = \frac{r_0^2 - r_1^2 + d_1^2}{2d_1}
\] (A.16a)

\[
y_{\text{Tx}} = \frac{r_0^2 \left( d_1 - d_2 \cos \alpha \right) + r_1^2 d_2 \cos \alpha - r_2^2 d_1 + d_1 d_2 \left( d_2 - d_1 \cos \alpha \right)}{2d_1 d_2 \sin \alpha}
\] (A.16b)

\[
z_{\text{Tx}} = \frac{\left( d_2^2 \sin^2 \alpha \left[ r_0^2 \left( 2r_1^2 - r_0^2 + 2d_1^2 \right) + r_1^2 \left( 2d_1^2 - r_2^2 \right) - d_1^2 \right] + \right.}{2d_1 d_2 \sin \alpha} \left. \right)^{1/2} \frac{\left. -\left[ r_0^2 \left( d_1 - d_2 \cos \alpha \right) + r_1^2 d_2 \cos \alpha - r_2^2 d_1 + d_1 d_2 \left( d_2 - d_1 \cos \alpha \right) \right]^2 \right]}{2d_1 d_2 \sin \alpha}
\] (A.16c)

was derived for the exact position of a transmitting antenna \( \text{Tx} \) in a reference frame defined by the positions of three receiving antennas \( \text{Rx}_0, \text{Rx}_1, \text{Rx}_2 \), cf. Fig. A.1. The variables \( r_i \) in Eq. (A.16) are the distances between \( \text{Tx} \) and \( \text{Rx}_i \) where \( i = 0, 1, 2 \). In addition, it was given that \( z_{\text{Tx}} = \left[ r_0^2 - (x_{\text{Tx}}^2 + y_{\text{Tx}}^2) \right]^{1/2} \).

Figure A.1: Location of the various antennas in the receiver reference frame.
By means of the equations presented in section A.1 it is possible to derive the expected value and variance for \(x_{\text{Tx}}, y_{\text{Tx}},\) and \(z_{\text{Tx}}\) when replacing the true range \(r_i\) in Eqs. (A.16a–A.16c) with the pseudorange \(\rho_i\). This pseudorange is assumed to be a normally distributed random variable with mean \(r_i\) and variance \(\sigma^2_{\rho_i}\). For \(x_{\text{Tx}}\) and \(y_{\text{Tx}}\), the expected value and variance can be worked out by hand. Due to the complexity of the equation for \(z_{\text{Tx}}\), the symbolic math program Maple\textsuperscript{®} was used to determine its expected value and variance. In the following derivations, zero cross-correlation between measurements from different Rx antennas is assumed.

### A.2.1 Derivations for \(x\)

For \(x_{\text{Tx}}\) and \(y_{\text{Tx}}\), the expected value and variance can be determined exactly and thus no Taylor series expansion is strictly required. However, for \(x_{\text{Tx}}\) the expected value and variance are derived by means of a Taylor series expansion and by means of an exact derivation. The reason for this is twofold. Firstly, it is very instructive to derive an equation in two different manners. Firstly, it will show that the Taylor series expansion and the exact derivation can lead to exactly the same result. Secondly, the complete Taylor series expansion for \(z_{\text{Tx}}\) is very lengthy and therefore not shown completely in the corresponding subsection. Since it is very instructive to show how the Taylor series expansion can be used to approximate expected values and variances of complex functions, it was decided to show this for the moderately complex function \(x_{\text{Tx}}\), which makes it easier to understand the individual steps.

Performing the Taylor series expansion up to second order of \(x_{\text{Tx}}\) around the mean values \(r_0\) and \(r_1\) of the random variables \(\rho_0\) and \(\rho_1\) results in

\[
x_{\text{Tx}} = \frac{\rho_0^2 - \rho_1^2 + d_1^2}{2d_1} \approx \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} + (\rho_0 - r_0) \frac{\partial}{\partial r_0} \left( \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right) + (\rho_1 - r_1) \frac{\partial}{\partial r_1} \left( \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right) \\
+ \frac{1}{2!} (\rho_0 - r_0)^2 \frac{\partial^2}{\partial r_0^2} + 2(\rho_0 - r_0)(\rho_1 - r_1) \frac{\partial^2}{\partial r_0 \partial r_1} + (\rho_1 - r_1)^2 \frac{\partial^2}{\partial r_1^2} \left( \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right).
\]

(A.17)

The first and second order partial derivatives are

\[
\begin{align*}
\frac{\partial}{\partial r_0} \left( \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right) &= \frac{r_0}{d_1} \\
\frac{\partial}{\partial r_1} \left( \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right) &= -\frac{r_1}{d_1} \\
\frac{\partial^2}{\partial r_0 \partial r_1} \left( \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right) &= 0.
\end{align*}
\]
Therefore

\[ x_{\text{Tx}} \approx \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} + (\rho_0 - r_0) \frac{r_0}{d_1} - (\rho_1 - r_1) \frac{r_1}{d_1} + \frac{1}{2d_1} \left[ (\rho_0 - r_0)^2 - (\rho_1 - r_1)^2 \right]. \]  

(A.18)

Taking the expected value of both sides of the equation leads to

\[
E[x_{\text{Tx}}] \approx E \left[ \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} + (\rho_0 - r_0) \frac{r_0}{d_1} - (\rho_1 - r_1) \frac{r_1}{d_1} + \frac{1}{2d_1} \left[ (\rho_0 - r_0)^2 - (\rho_1 - r_1)^2 \right] \right] \\
= E \left[ \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} \right] + E \left[ (\rho_0 - r_0) \frac{r_0}{d_1} \right] - E \left[ (\rho_1 - r_1) \frac{r_1}{d_1} \right] + \frac{1}{2d_1} E \left[ (\rho_0 - r_0)^2 \right] \\
= \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} + \frac{r_0}{d_1} E[(\rho_0 - r_0)] - \frac{r_1}{d_1} E[(\rho_1 - r_1)] + \frac{1}{2d_1} \text{Var}[\rho_0] - \frac{1}{2d_1} \text{Var}[\rho_1] \\
= \{E[\rho_i - r_i] = E[\rho_i] - E[r_i] = r_i - r_i = 0\} \\
= \frac{r_0^2 - r_1^2 + d_1^2 + \sigma_{\rho_0}^2 - \sigma_{\rho_1}^2}{2d_1}.
\]

(A.19)

The expected value could also have been determined in the following manner

\[
E[x_{\text{Tx}}] = \frac{E[r_0^2 - r_1^2 + d_1^2]}{2d_1} \\
= \frac{1}{2d_1} E[\rho_0^2] + \frac{1}{2d_1} E[-\rho_1^2] + \frac{1}{2d_1} E[d_1^2] \\
= \frac{1}{2d_1} \left( r_0^2 + \sigma_{\rho_0}^2 \right) - \frac{1}{2d_1} \left( r_1^2 + \sigma_{\rho_1}^2 \right) + \frac{d_1^2}{2d_1} \\
= \frac{r_0^2 - r_1^2 + d_1^2 + \sigma_{\rho_0}^2 - \sigma_{\rho_1}^2}{2d_1}.
\]

(A.20)

which shows that for \(x_{\text{Tx}}\) the Taylor expansion leads to the same result as the exact derivation.

Using Eqs. (A.16a) and (A.20), the error \(e\) in the expected value of \(x_{\text{Tx}}\) can be determined to be equal to
\[ e_{x_{1x}} = x_{1x} - E[x_{1x}] = \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} - \frac{r_0^2 - r_1^2 + d_1^2 + \sigma_{\rho_0}^2 - \sigma_{\rho_1}^2}{2d_1} \]  
\[ = \frac{\sigma_{\rho_1}^2 - \sigma_{\rho_0}^2}{2d_1}. \]  

(A.21)

Thus, to minimize the error in the estimation of \( x_{1x} \), \( d_1 \) has to be maximized, the difference between \( \sigma_{\rho_0} \) and \( \sigma_{\rho_1} \) has to be minimized, and/or the magnitudes of \( \sigma_{\rho_0} \) and \( \sigma_{\rho_1} \) have to be minimized.

In a similar manner, re-using the results from the second order Taylor series expansion derived earlier, the variance of \( x_{1x} \) can be determined to be as follows.

\[
\text{Var}[x_{1x}] \approx \text{Var}\left[ \frac{r_0^2 - r_1^2 + d_1^2}{2d_1} + (\rho_0 - r_0) \frac{r_0}{d_1} - (\rho_1 - r_1) \frac{r_1}{d_1} + \frac{1}{2d_1} \left( (\rho_0 - r_0)^2 - (\rho_1 - r_1)^2 \right) \right]
\]
\[= \frac{1}{d_1^2} \text{Var}\left[ r_0 \rho_0 - r_1 \rho_1 + \frac{1}{2} (\rho_0 - r_0)^2 - \frac{1}{2} (\rho_1 - r_1)^2 \right]
\]
\[= \frac{1}{d_1^2} \text{Var}\left[ \frac{1}{2} \rho_0^2 - \frac{1}{2} \rho_1^2 + \frac{1}{2} r_0^2 - \frac{1}{2} r_1^2 \right]
\]
\[= \frac{\text{Var}[\rho_0^2] + \text{Var}[\rho_1^2] - 2\text{Cov}[\rho_0^2, \rho_1^2]}{4d_1^2}. \]

(A.22)

By using Eq. (A.7) and by assuming \( \text{Cov}[\rho_0^2, \rho_1^2] = 0 \) the above equation can be written as

\[
\text{Var}[x_{1x}] \approx \frac{2r_0^2 \sigma_{\rho_0}^2 + 2r_1^2 \sigma_{\rho_1}^2 + \sigma_{\rho_0}^4 + \sigma_{\rho_1}^4}{2d_1^2}. \]

(A.23)

Just as for the expected value, the variance could also have been determined using an exact derivation which leads to the same result as the derivation using the Taylor series expansion:
\[ \text{Var}[x_{TX}] = \text{Var} \left[ \frac{\rho_0^2 - \rho_1^2 + d_1^2}{2d_1} \right] \]

\[ = \frac{1}{4d_1^2} \left[ \text{Var}[\rho_0^2] + \text{Var}[\rho_1^2] + \text{Var}[d_1^2] - 2\text{Cov}[\rho_0^2, \rho_1^2] + 2\text{Cov}[\rho_0^2, d_1^2] - 2\text{Cov}[\rho_1^2, d_1^2] \right] \]

\[ = \frac{1}{4d_1^2} \left[ 4r_0^2 \sigma_0^2 + 2\sigma_0^4 + 4r_1^2 \sigma_1^2 + 2\sigma_1^4 \right] \]

\[ = \frac{2r_0^2 \sigma_0^2 + 2r_1^2 \sigma_1^2 + \sigma_0^4 + \sigma_1^4}{2d_1^2}. \]

\[ (A.24) \]

Equation (A.24) shows that in order to minimize the variance in the estimate of \( x_{TX} \), \( d_1 \) has to be maximized and \( r_0, r_1, \sigma_{\rho_0}, \text{ and } \sigma_{\rho_1} \) have to be minimized.

Lastly, for completeness, it is now shown that it is in fact feasible to determine the variance of \( x_{TX} \) for \( \text{Cov}[\rho_0^2, \rho_1^2] \neq 0 \) when the correlation between \( \rho_0^2 \) and \( \rho_1^2 \), \( \text{Cor}[\rho_0^2, \rho_1^2] \), is known or assumed. Since by definition \(-1 < \text{Cor}[\rho_0^2, \rho_1^2] < 1\) and since the correlation is defined as \( \text{Cor}[X, Y] = \text{Cov}[X, Y] / (\sigma_X \sigma_Y) \), assuming \( \text{Cor}[\rho_0^2, \rho_1^2] = a, a \in (-1, 1) \) leads to \( \text{Cov}[\rho_0^2, \rho_1^2] = a(\text{Var}[\rho_0^2] \text{Var}[\rho_1^2])^{1/2} \). This can be worked out to

\[ \text{Cov}[\rho_0^2, \rho_1^2] = a \sqrt{\text{Var}[\rho_0^2] \text{Var}[\rho_1^2]} \]

\[ = a \sqrt{(4r_0^2 \sigma_0^2 + 2\sigma_0^4)(4r_1^2 \sigma_1^2 + 2\sigma_1^4)} \]

\[ = 2a \sigma_{\rho_0} \sigma_{\rho_1} \sqrt{(2r_0^2 + \sigma_0^2)(2r_1^2 + \sigma_1^2)}. \]

\[ (A.25) \]

If \( r_i \gg \sigma_i \) then Eq. (A.25) can be approximated as

\[ \text{Cov}[\rho_0^2, \rho_1^2] \approx 4ar_0r_1 \sigma_{\rho_0} \sigma_{\rho_1} \]

\[ (A.26) \]

and Eq. (A.22) would reduce to

\[ \text{Var}[x_{TX}] \approx \frac{r_0^2 \sigma_{\rho_0}^2 + r_1^2 \sigma_{\rho_1}^2 - 2ar_0r_1 \sigma_{\rho_0} \sigma_{\rho_1}}{d_1^2}. \]

\[ (A.27) \]
A.2.2 Derivations for $y$

Using Eq. (A.16b) and replacing $r_i$ with $\rho_i$, the expected value of $y_{\text{TX}}$ is derived as follows

$$
E[y_{\text{TX}}] = E\left[ \frac{r_0^2(d_1 - d_2 \cos \alpha) + r_1^2 d_2 \cos \alpha - r_2^2 d_1 + d_1 d_2 (d_2 - d_1 \cos \alpha)}{2d_1 d_2 \sin \alpha} \right]
$$

$$
= \frac{1}{2d_1 d_2 \sin \alpha} \left[ \left( d_1 - d_2 \cos \alpha \right) E[\rho_0^2] + d_2 \cos \alpha E[\rho_1^2] - d_1 E[\rho_2^2]
+ d_1 d_2 \left( d_2 - d_1 \cos \alpha \right) \right]
$$

$$
= \frac{1}{2d_1 d_2 \sin \alpha} \left[ \left( d_1 - d_2 \cos \alpha \right) \left( r_0^2 + \sigma_{\rho_0}^2 \right) + d_2 \cos \alpha \left( r_1^2 + \sigma_{\rho_1}^2 \right) - d_1 \left( r_2^2 + \sigma_{\rho_2}^2 \right)
+ d_1 d_2 \left( d_2 - d_1 \cos \alpha \right) \right]
$$

$$
= \frac{1}{2d_1 d_2 \sin \alpha} \left[ r_0^2(d_1 - d_2 \cos \alpha) + r_1^2 d_2 \cos \alpha - r_2^2 d_1 + d_1 d_2 (d_2 - d_1 \cos \alpha)
+ \sigma_{\rho_0}^2(d_1 - d_2 \cos \alpha) + \sigma_{\rho_1}^2 d_2 \cos \alpha - \sigma_{\rho_2}^2 d_1 \right].
\tag{A.28}
$$

The result for the variance is

$$
\text{Var}[y_{\text{TX}}] = \text{Var}\left[ \frac{r_0^2(d_1 - d_2 \cos \alpha) + r_1^2 d_2 \cos \alpha - r_2^2 d_1 + d_1 d_2 (d_2 - d_1 \cos \alpha)}{2d_1 d_2 \sin \alpha} \right]
$$

$$
= \frac{1}{4d_1^2 d_2^2 \sin^2 \alpha} \left[ \left( d_1 - d_2 \cos \alpha \right)^2 \text{Var}[\rho_0^2] + d_2^2 \cos^2 \alpha \text{Var}[\rho_1^2] + d_1^2 \text{Var}[\rho_2^2] \right]
$$

$$
= \frac{1}{4d_1^2 d_2^2 \sin^2 \alpha} \left[ \left( d_1 - d_2 \cos \alpha \right)^2 \left( 4r_0^2 \sigma_{\rho_0}^2 + 2\sigma_{\rho_0}^4 \right) + d_2^2 \cos^2 \alpha \left( 4r_1^2 \sigma_{\rho_1}^2 + 2\sigma_{\rho_1}^4 \right)
+ d_1^2 \left( 4r_2^2 \sigma_{\rho_2}^2 + 2\sigma_{\rho_2}^4 \right) \right]
$$

$$
= \frac{1}{2d_1^2 d_2^2 \sin^2 \alpha} \left[ 2r_0^2 \sigma_{\rho_0}^2 \left( d_1 - d_2 \cos \alpha \right)^2 + 2r_1^2 \sigma_{\rho_1}^2 d_2^2 \cos^2 \alpha + 2r_2^2 \sigma_{\rho_2}^2 d_1^2
+ \sigma_{\rho_0}^4 \left( d_1 - d_2 \cos \alpha \right)^2 + \sigma_{\rho_1}^4 d_2^2 \cos^2 \alpha + \sigma_{\rho_2}^4 d_1^2 \right].
\tag{A.29}
$$

The results for the expected value and the variance show that the angle $\alpha$ between the baselines $d_1$ and $d_2$ has an influence on the estimate of the position of the transmitter. Since a small error in the estimate and a small variance in the estimate are desirable, it is of importance to known whether $\alpha$ should be large or small.

Using Eqs. (A.16b) and (A.28), the error $e$ in the expected value of $y_{\text{TX}}$ can be determined to be equal to
$e_{y_{Tx}} = y_{Tx} - E[y_{Tx}]$

$$= \frac{1}{2d_1d_2\sin\alpha} \left\{ \left[ r_0^2\left( d_1 - d_2\cos\alpha \right) + r_1^2d_2\cos\alpha - r_2^2d_1 + d_1d_2\left( d_2 - d_1\cos\alpha \right) \right] - \\
\left[ r_0^2\left( d_1 - d_2\cos\alpha \right) + r_1^2d_2\cos\alpha - r_2^2d_1 + d_1d_2\left( d_2 - d_1\cos\alpha \right) \right] \\
+ \sigma_{\rho_0}^2\left( d_1 - d_2\cos\alpha \right) + \sigma_{\rho_1}^2d_2\cos\alpha - \sigma_{\rho_2}^2d_1 \right\}$$  

$$= \frac{d_1\left( \sigma_{\rho_2}^2 - \sigma_{\rho_0}^2 \right) + d_2\cos\alpha\left( \sigma_{\rho_0}^2 - \sigma_{\rho_1}^2 \right)}{2d_1d_2\sin\alpha}. \quad (A.30)$$

From Eq. (A.30) it is clear that the error in the estimate is minimized when the denominator is maximized, which occurs when $\alpha = 90^\circ$ (in the limit, $\alpha = 0^\circ$ leads to an error of either $+\infty$ or $-\infty$, depending on the values of $d_1$, $d_2$, $\sigma_{\rho_0}$, $\sigma_{\rho_1}$, and $\sigma_{\rho_2}$). This is also intuitively obvious as this angle leads to an antenna geometry which is furthest from the situation where all receiver antennas are aligned. Setting $\alpha = 90^\circ$ also leads to the smallest value for the variance in the estimate of $y_{Tx}$, which is easily observed from Eq. (A.29).

Setting $\alpha = 90^\circ$ leads to the following simplified expressions for the expected value and the variance of $y_{Tx}$:

$$E[y_{Tx}] \bigg|_{\alpha=90^\circ} = \frac{r_0^2 - r_2^2 + d_2^2 + \sigma_{\rho_0}^2 - \sigma_{\rho_2}^2}{2d_2} \quad (A.31)$$

and

$$\text{Var}[y_{Tx}] \bigg|_{\alpha=90^\circ} = \frac{2r_0^2\sigma_{\rho_0}^2 + 2r_2^2\sigma_{\rho_2}^2 + \sigma_{\rho_0}^4 + \sigma_{\rho_2}^4}{2d_2^2}. \quad (A.32)$$

Not surprisingly, due to baseline $d_2$ now lying exactly on the $y$-axis of the reference frame, the equations for $y_{Tx}$ are now identical to the equations for $x_{Tx}$ except that the subscript "1" in the equations for $x_{Tx}$ is now replaced by the subscript "2".

### A.2.3 Derivations for $z$

In subsection A.2.2, it was shown that $\alpha = 90^\circ$ leads to the smallest error and to the smallest variance in the estimate of $y_{Tx}$. Since $z_{Tx} = [r_0^2 - (x_{Tx}^2 + y_{Tx}^2)]^{1/2}$, it is fairly easy to determine for which value of $\alpha$ leads to the smallest error in the estimate of $z_{Tx}$. Since the error in the estimate of $z_{Tx}^2$ will be minimal when the error in the estimate of $z_{Tx}$ is also minimal, the derivation can also be performed for the error in the estimate of $z_{Tx}^2$, which simplifies the derivation:
\[ e_{z_{\text{Tx}}} = z_{\text{Tx}}^2 - E\left[ z_{\text{Tx}}^2 \right] \]
\[ = r_0^2 - x_{\text{Tx}}^2 - y_{\text{Tx}}^2 - E\left[ \rho_0^2 \right] + E\left[ x_{\text{Tx}}^2 \right] + E\left[ y_{\text{Tx}}^2 \right] \]
\[ = \left( r_0^2 - E\left[ \rho_0^2 \right] \right) + \left( E\left[ x_{\text{Tx}}^2 \right] - x_{\text{Tx}}^2 \right) + \left( E\left[ y_{\text{Tx}}^2 \right] - y_{\text{Tx}}^2 \right). \tag{A.33} \]

Minimization of the bracketed terms in the above equation will lead to the smallest error for \( z_{\text{Tx}}^2 \) and thus also for \( z_{\text{Tx}} \). Since \( \{E\left[ y_{\text{Tx}}^2 \right] - y_{\text{Tx}}^2\} \) is the only term that depends on \( \alpha \) and since \( \alpha = 90^\circ \) leads to the smallest error in the estimate of \( y_{\text{Tx}} \), it is concluded that the error in \( z_{\text{Tx}} \) is minimized when \( \alpha = 90^\circ \).

For the variance, knowing that there is zero covariance between \( r_0^2 \), \( x_{\text{Tx}}^2 \), and \( y_{\text{Tx}}^2 \), a similar reasoning can be applied:

\[ \text{Var}\left[ z_{\text{Tx}}^2 \right] = \text{Var}\left[ r_0^2 - x_{\text{Tx}}^2 - y_{\text{Tx}}^2 \right] = \text{Var}\left[ r_0^2 \right] + \text{Var}\left[ x_{\text{Tx}}^2 \right] + \text{Var}\left[ y_{\text{Tx}}^2 \right]. \tag{A.34} \]

As \( \text{Var}\left[ y_{\text{Tx}}^2 \right] \) is minimized when \( \alpha = 90^\circ \), the variance of \( z_{\text{Tx}}^2 \) and thus the variance of \( z_{\text{Tx}} \) is minimized when \( \alpha = 90^\circ \).

Setting \( \alpha = 90^\circ \) leads to the following, slightly simplified, expression for \( z_{\text{Tx}} \):

\[ z_{\text{Tx}} \bigg|_{\alpha=90^\circ} = \frac{1}{2} \sqrt{-\left( \frac{\rho_0^4}{d_1^2} + \frac{\rho_1^4}{d_1^2} - \frac{2\rho_0^2\rho_1^2}{d_1^2} - 2\rho_1^2 + d_1^2 \right) - \left( \frac{\rho_0^4}{d_2^2} + \frac{\rho_2^4}{d_2^2} - \frac{2\rho_0^2\rho_2^2}{d_2^2} - 2\rho_2^2 + d_2^2 \right)}. \tag{A.35} \]

Unfortunately, calculation of the expected value and the variance of this expression leads to extremely long expressions that provide little insight. For that reason, after the complete results have been obtained with Maple\textsuperscript{\textregistered}, they have been simplified by imposing the hypothetical situation that \( r_0 = r_1 = r_2 = r \), \( d_1 = d_2 = d \), and \( \sigma_{\rho_0} = \sigma_{\rho_1} = \sigma_{\rho_2} = \sigma \), which is a simplification that is not far from reality for many formation flying scenarios, especially for along-track formations. In this situation, for which the true \( z \)-coordinate of the transmitter is \( z_{\text{Tx}} = 0.5(4r^2 - 2d^2)^{1/2} \), the expected value reduces to

\[ \text{E}\left[ z_{\text{Tx}} \right] \bigg|_{r=r,d=d,\sigma_{\rho_1}=\sigma_{\rho}} \approx \frac{1}{2} \sqrt{4r^2 - 2d^2} \frac{4r^4 \left( d^2 - 2\sigma_\rho^2 \right) + r^2 d^2 \left( 5\sigma_\rho^2 - 4d^2 \right) + d^4 \left( d^2 - \sigma_\rho^2 \right)}{d^2(d^2 - 2r^2)^2}. \tag{A.36} \]

and the variance is
APPENDIX A. DERIVATION OF THE EXPECTATION AND VARIANCE IN THE POSITION ESTIMATE OF A STATIONARY TRANSMITTER

\[
\text{Var}[z_{\text{Tx}}]_{\mid r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}} \approx \sigma_{\rho}^2 \left( 80r^8 \sigma_{\rho}^2 + 8r^6 d^2 \left( d^2 - 11 \sigma_{\rho}^2 \right) + r^4 d^4 \left( 34 \sigma_{\rho}^2 - 8d^2 \right) + 2r^2 d^6 \left( d^2 - 3 \sigma_{\rho}^2 \right) + d^8 \sigma_{\rho}^2 \right) \over 2d^4 \left( d^2 - 2r^2 \right)^3.
\]

(A.37)

The error in the estimate for \( z_{\text{Tx}} \) is now

\[
e_{z_{\text{Tx}}} \bigg|_{r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}} = \left\{ z_{\text{Tx}} - \mathbb{E}[z_{\text{Tx}}] \right\}_{r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}} = -4r^4 \left( d^2 - 2 \sigma_{\rho}^2 \right) + r^2 d^2 \left( 5 \sigma_{\rho}^2 - 4d^2 \right) + d^4 \left( d^2 - \sigma_{\rho}^2 \right) \over d^2 \left( d^2 - 2r^2 \right)^2.
\]

(A.38)

When further assuming that \( r \gg d, \sigma_{\rho} \), the result reduces to

\[
\mathbb{E}[z_{\text{Tx}}]_{\mid r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}, r \gg d, \sigma_{\rho}} \approx r \left( 1 - 2 \frac{\sigma_{\rho}^2}{d^2} \right).
\]

(A.39)

and

\[
\text{Var}[z_{\text{Tx}}]_{\mid r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}, r \gg d, \sigma_{\rho}} \approx 5r^2 \sigma_{\rho}^4 + \frac{1}{2} \sigma_{\rho}^2.
\]

(A.40)

The error in the estimate for \( z_{\text{Tx}} \) reduces to

\[
e_{z_{\text{Tx}}} \bigg|_{r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}, r \gg d, \sigma_{\rho}} = \left\{ z_{\text{Tx}} - \mathbb{E}[z_{\text{Tx}}] \right\}_{r_i=r, d_i=d, \sigma_{\rho_i} = \sigma_{\rho}, r \gg d, \sigma_{\rho}} = r - r \left( 1 - 2 \frac{\sigma_{\rho}^2}{d^2} \right) = 2r \frac{\sigma_{\rho}^2}{d^2}.
\]

(A.41)

Thus, for this scenario, the error and the variance in the estimate of \( z_{\text{Tx}} \) can be minimized by minimizing \( r \) and \( \sigma_{\rho} \) and by maximizing \( d \).
List of Author’s Publications

Below is a list of relevant publications published in the context of the thesis activities.

**Book Chapters**


**International Journals**


APPENDIX A. DERIVATION OF THE EXPECTATION AND VARIANCE IN THE POSITION ESTIMATE OF A STATIONARY TRANSMITTER


*International Conferences*


Daan Maessen was born in Tegelen, The Netherlands, on September 23, 1981. He attended secondary school (the Dutch VWO) from 1993 until 2000 at the St. Thomascollege in Venlo. From 2000 to 2007 he studied Aerospace Engineering at the Delft University of Technology. During his studies he actively participated in the Delft Aerospace Rocket Engineers (Delft), an amateur student rocket society. During this time he designed, built, and launched several small rockets. For six months he conducted an internship at Dutch Space in Leiden on the conceptual design of an orbit transfer vehicle for the VEGA rocket. In 2007 he graduated cum laude with a Master’s thesis on the development of a generic inflatable de-orbit device for cubesats. From 2007 to 2011 he worked at the Space Systems Engineering chair of the Delft University of Technology as a Ph.D. student in the field of satellite formation flying under the supervision of Prof. Dr. Eberhard Gill. From 2012 to July 2013, he has been employed by the company Science [&] Technology Corp. for which he worked at Moog Bradford in Heerle, The Netherlands, on the design and development of electric propulsion subsystems. From December 2013 on, he has been employed by Stirling Cryogenics B.V. in Son, The Netherlands, on the optimization of cryogenic cooling equipment. He has over 20 scientific publications including conference proceedings, peer-reviewed journal articles, and book chapters.
Bibliography


http://hdl.handle.net/2014/10401.


