A Method for Accurate Measurement of Heliostat Mirror Orientation

by

Benjamin Dawid Swart

Thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering (Electronic) in the Faculty of Engineering at Stellenbosch University

Supervisor: Mr J. Treurnicht
Co-supervisor: Dr H.A. Engelbrecht

March 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2017

Abstract

Heliostat tracking is a critical component of the solar field of concentrating solar power central receiver systems and can be the source of significant losses in power and profit when it lacks the necessary accuracy. The current industry norm seems to be open-loop tracking using an error model for a tracking error of one milliradian, with error model parameters requiring periodic recalibration. In this thesis a method is developed which is able to determine the orientation of all heliostats in the field simultaneously, providing near real-time feedback to the control system, removing the need for recalibration while maintaining an accurate tracking accuracy.

The method works by determining the positions of three or more distinct points on the heliostat mirror surface. These points form a plane which is then used to calculate the heliostat normal vector. The performance of the method is analysed and quantified using simulations, and experiments are used to validate the theoretical models. Quantification of the system performance is used to determine the initial feasibility of the method.

Results indicate that for the error sources included in the model it is theoretically possible to achieve a tracking error in the order of one milliradian or less. Furthermore, experimental results are found to differ by only 5 % from theoretical results. This is promising and merits further investigation into the method as a possible solution to the current heliostat tracking problem.

Opsomming

Heliostaat volging (die vermoë om die son te volg) is 'n kritiese komponent van die sonveld in gekonsentreerde son-energie sentrale ontvangerstelsels. Indien die nodige akkuraatheid ontbreek, kan dit aansienlike energie– en winsverliese tot gevolg hê. Die huidige industry standaard is oënskynlik opelus-volging met 'n foutmodel wat 'n volgfout van een milliradiaal tot gevolg het, met foutmodel parameters wat periodiese herkalibrasie noodsaak. In hierdie tesis word 'n metode ontwikkel wat die oriëntasie van ál die heliostate in 'n veld tegelykertyd kan bepaal. Dit voorsien byna-intydse terugvoer aan die beheerstelsel wat herkalibrasie onnodig maak terwyl volging steeds akkuraat kan plaasvind.

Die metode werk deur die posisies van drie of meer unieke punte op die heliostaatspieël se oppervlak te bepaal. Hierdie punte vorm 'n vlak wat gebruik kan word om die heliostaat se oriëntasie te bereken. Die doeltreffendheid van die metode word geanaliseer en gekwantifiseer deur middel van simulasies en eksperimente word uitgevoer om die teoretiese modelle te bevestig.

Die resultate wys dat vir die foutbronne wat in die model ingesluit is dit teoreties moontlik is om 'n volgfout in die orde van een milliradiaal of minder te behaal. Daar is ook bevind dat eksperimentele resultate met minder as 5 % van teoretiese resultate verskil. Dit is belowend en verdien verdere ondersoek na hierdie metode as moontlike oplossing vir die huidige probleem met heliostaat volging.

Acknowledgements

I would like to acknowledge and thank the following people:

Mr Johann Treurnicht, my supervisor, for sharing his invaluable insights and experience, and for his gracious assistance with administration, particularly with regards to finance.

Dr Herman Engelbrecht, my co-supervisor, for his invaluable insights in all things signal processing, and his willingness to take on a supervising role halfway through the project.

Ms Mart-Mari Swart, my wife, primarily for her support on this endeavour, and also for her assistance with some of the illustrations.

Everyone else, particularly the people at STERG, who contributed in some way or another.

My Lord and Saviour Jesus Christ for his faithfulness and provision. The fear of the Lord is the beginning of wisdom, and knowledge of the holy is understanding.

Contents

List of Figures

List of Tables

Nomenclature

The conventions, operators, units, variables and acronyms listed here are used throughout the document except where explicitly stated or the context clearly infers otherwise.

Conventions

$\vec{v}$	vector – arrow on top, usually lowercase
P	point coordinates – usually uppercase
$\boldsymbol{M}$	matrix – bold, usually uppercase

Operators

$\vec{v} \cdot \vec{u}$	dot product
$\vec{v} \times \vec{u}$	cross product, or normal multiplication
$\|\|\vec{v}\|\|$	vector size
$\|\theta\|$	absolute value
E$[n]$	expected (mean) value
Var$[n]$	variance

Units

k	thousand
M	million
mm	millimetres
cm	centimetres
m	metres
km	kilometres
ns	nanoseconds
s	seconds
h	hours
mrad	milliradians
rad	radians

$^\circ$	degrees
$^\circ$C	degrees Celsius
mV	millivolt
W	watts
MW	megawatts
GW	gigawatts
kWh	kilowatt-hours
TWh	terawatt-hours
Hz	hertz
kHz	kilohertz
MHz	megahertz
GHz	gigahertz
dB	decibels
dBm	decibel-milliwatts
Ω	ohm
MΩ	megohm
%	percentage
¢	American cent

Variables

θ_z	solar zenith angle	[rad]
α	solar altitude angle	[rad]
A	solar azimuth angle	[rad]
$\vec{S}$	solar vector	[m]
T	transmitter position	[m]
R	receiver position	[m]
d	distance	[m]
Δ_{ij}	distance difference $(d_i - d_j)$	[m]
λ	wavelength	[m]
f	frequency	[Hz]
f_0	oscillator frequency	[Hz]
f_s	sampling frequency	[Hz]
ω	angular frequency	[rad/s]
θ	initial phase	[rad]
θ_{ij}	phase difference $(\theta_i - \theta_j)$	[rad]
γ	phase variation due to noise	[rad]
Θ	phase altered by noise	[rad]
Φ	phase shift from oscillator to transmission	[rad]
ϕ	phase shift from transmission to reception	[rad]
t	time	[s]
T	period	[s]

a	amplitude	[$\sim$]
VF	velocity factor	[$\sim$]
s	signal	[$\sim$]
n	noise	[$\sim$]
$\vec{s}$	sampled signal values	[$\sim$]
S	number of samples	[$\sim$]
$\hat{s}$	Hilbert transform of signal	[$\sim$]
$\hat{n}$	Hilbert transform of noise	[$\sim$]
N	number of transmitters	[$\sim$]
D	number of dimensions	[$\sim$]
B	equivalent normalised bandwidth	[$\sim$]
SPC	samples per cycle	[$\sim$]
SNR	signal-to-noise ratio	[$\sim$]
d_R	inter-receiver distance	[m]
ϵ_ϕ	phase difference error	[rad]
ϵ_Δ	distance difference error	[m]
ϵ_R	position error	[m]
ϵ_{Rr}	normalised position error	[$\sim$]
ϵ_{Rz}	position error z-axis factor	[$\sim$]
ϵ_{Ts}	tracking error scale	[$\sim$]
ϵ_T	tracking error / receiver plane normal error	[rad]
Γ	reflection coefficient	[$\sim$]
Z	impedance	[Ω]
s_{ij}	scattering parameter	[$\sim$]

Acronyms

AAF	anti-aliasing filter
ADC	analog-to-digital converter
AOA	angle-of-arrival
AWGN	additive white Gaussian noise
BCS	beam characterisation system
BPF	band pass filter
BW	bandwidth
CCD	charge-coupled device
CRS	central receiver system
CRSES	Centre for Renewable and Sustainable Energy Studies
CSIR	Council for Scientific and Industrial Research
CSP	concentrating solar power

DC	direct current
DP	dot product
DPNC	dot product with noise compensation
DFT	discrete Fourier transform
DIR	digital image radiometer
DNI	direct normal irradiation
FDM	frequency division multiplexing
FFT	fast Fourier transform
GPS	Global Positioning System
HT	Hilbert transform
IFFT	inverse fast Fourier transform
LCoE	levelised cost of electricity
LED	light emitting diode
LoS	line of sight
LLSQ	linear least squares
LPR	local-positioning radar
MAP	maximum *a posteriori* probability
MLE	maximum likelihood estimator
NLLSQ	non-linear least squares
PCB	printed circuit board
PSA	Plataforma Solar de Almerýa
PSD	power spectral density
PV	photovoltaic
REIPPP	renewable energy independent power producer procurement
REN21	Renewable Energy Policy Network for the 21st Century
RF	radio frequency
RMS	root mean square
RSS	received-signal strength
RTOF	round-trip time-of-flight
RX	receiver
SGA	sliding Goertzel algorithm
SNR	signal-to-noise ratio
SPA	Solar Position Algorithm
SPC	samples per cycle
STERG	Solar Thermal Energy Research Group
TDM	time division multiplexing
TDOA	time-difference-of-arrival
TOA	time-of-arrival
TX	transmitter

Chapter 1

Introduction

"Energy is central to nearly every major challenge and opportunity the world faces today" [1]. This sentiment is shared by all of the 193 Member States of the United Nations who unanimously adopted the UN's historic new Sustainable Development Agenda in late 2015. One of the goals of this agenda is "to ensure access to affordable, reliable, sustainable and modern energy for all" by 2030.

The work in this thesis takes place at Stellenbosch University in the context of the Solar Thermal Energy Research Group (STERG), which is affiliated with the Centre for Renewable and Sustainable Energy Studies (CRSES), the national academic hub for renewable and sustainable energy. The focus of STERG is on concentrating solar power (CSP), in particular central receiver systems (CRSs). A critical component of a CRS is the heliostat field which focuses the incident solar irradiation onto the central receiver. Each heliostat is required to track the movement of the sun with high accuracy throughout the day in order to maintain focus of the reflected beam at the desired location. Improper tracking causes less solar irradiation to be reflected onto the receiver, leading to losses in power output and profit.

The current industry norm for heliostat tracking seems to be open-loop tracking using an error model for a tracking error of one milliradian [2]. Error model parameters require periodic recalibration which can take two to three weeks for an entire field [3][4]. Throughout literature it is clear that a need exists for a more accurate tracking method and reducing or completely removing the recalibration time. The literature review covers this in detail in Section 2.2.

In this thesis a method is developed whereby the orientations of all heliostats in the field can be simultaneously determined. When the direction of the incident solar irradiation is known, the orientation of the heliostat is used to determine the direction of the reflected beam. This provides near real-time feedback to the control system and completely removes the need for recalibration while maintaining accurate tracking.

1.1 Background

This section provides a brief historical overview of the technologies preceding and leading up to the introduction of the heliostat. The necessary context is created by first looking at the broad field of renewable energy, then narrowing down to solar energy, concentrating solar power, central receiver systems and finally, the introduction of the heliostat. At the end of the section it should be clear to the reader how and where heliostats fit into the bigger picture that is renewable energy.

1.1.1 Renewable Energy

Renewable energies are energy sources that are not depleted when used as they are continually replenished by nature. This is in contrast to conventional energy sources such as fossil fuels of which there are only a finite amount on the earth. Besides the host of benefits it provides to society in the political, social, environmental, economic and technological sectors [5], renewable energy is a crucial step towards ensuring the future of the planet, especially since it is estimated that we would have run out of oil, gas and coal before the end of the century [6].

The main sources of renewable energy are solar energy (energy from the sun), wind energy, biomass energy (energy from plants), geothermal energy (thermal energy from the earth's interior), hydropower energy (energy from moving water) and marine energy (energy from oceans). Whether or not nuclear energy is renewable is a subject of major debate [7]. Although it is considered clean energy, it is generally not included in the list, as the fuel required for nuclear fission is uranium, which is a finite resource.

The Renewable Energy Policy Network for the 21st Century (REN21) stated in their global status report of 2015 that by the end of 2014 renewable energies accounted for an estimated 27.7 % of the world's power generating capacity, enough to supply an estimated 22.8 % of global electricity. This is a considerable increase from the estimated 19.1 % of global energy consumption that renewables provided in 2013. A breakdown of the renewable power capacity by the end of 2014 is shown in Figure 1.1.

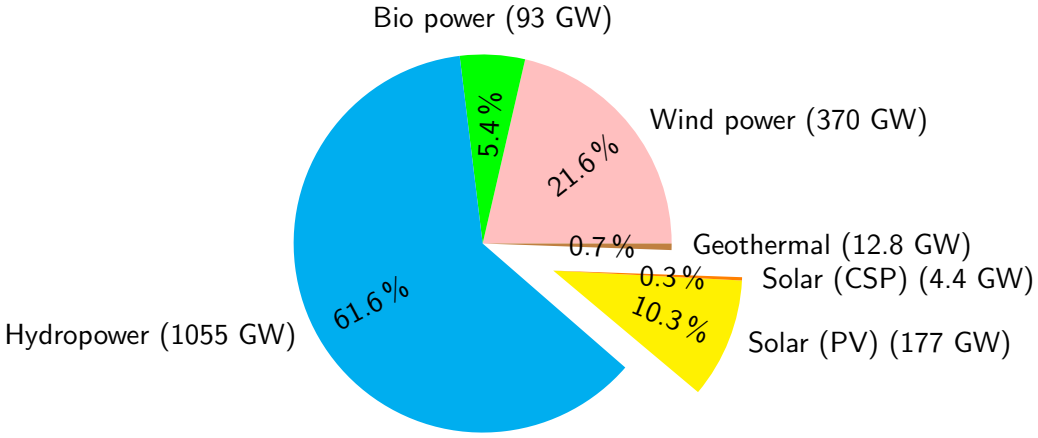


Figure 1.1: Renewable energy breakdown at the end of 2014 (data from [8]).

1.1.2 Solar Energy

Essentially all of the earth's power comes from the sun [9]. Wind energy, biomass energy and fossil fuels, among others, are simply secondary forms of solar energy. Figure 1.2 shows the direct normal irradiation (DNI) world map. DNI, also called direct or beam radiation, is the part of solar radiation that reaches the surface of the earth with essentially no change in direction [10]. The map shows that South Africa, especially the Northern Cape province, is prime territory for the harvesting of solar energy.

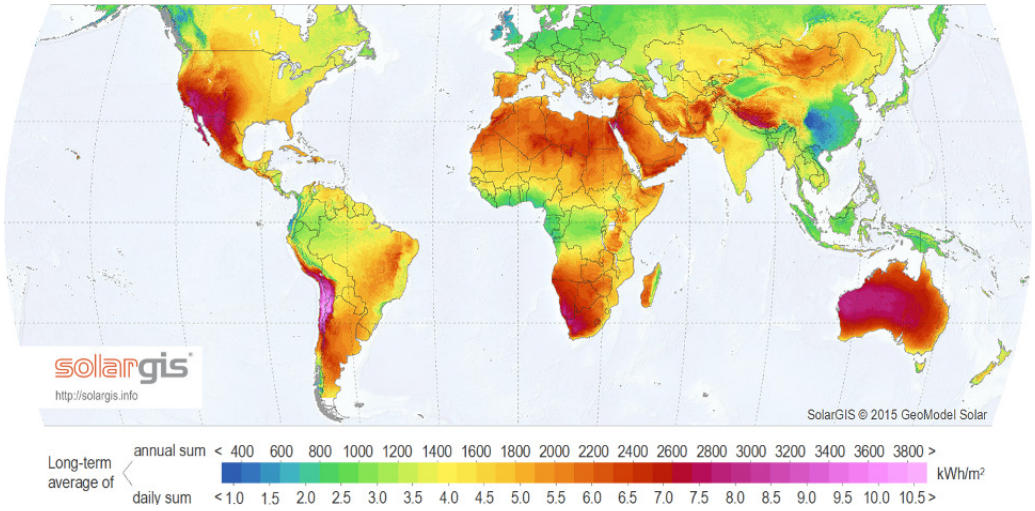


Figure 1.2: Direct normal irradiation (DNI) world map [11].

Solar energy can be used to generate electricity via photovoltaic (PV) and CSP systems. A PV system consists of an array of interconnected PV cells – semiconductor devices that directly convert solar energy into direct current (DC)

electricity. With CSP technologies, the direct-beam solar irradiance is first concentrated onto a medium which absorbs the thermal energy, after which another process uses the heat to generate electricity. The technologies used in solar power plants can be further subdivided as shown in Figure 1.3.

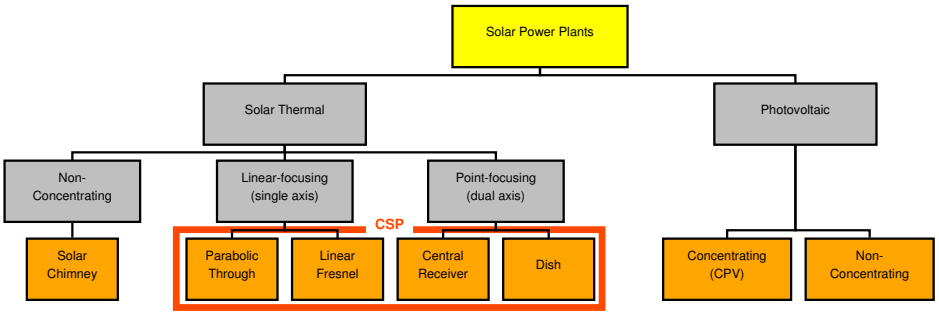


Figure 1.3: Solar power technologies breakdown (adapted from [12],[13]).

In 2014, the global electricity consumption was 20,300 TWh [14]. If the average solar power reaching the surface of the earth is 174.7 W/m^2 [15], and solar power can be harvested at a conservative 10 % efficiency,[1] it means that the global demand for electricity in 2014 could be met in just 132,647 km^2.[2] This is an area roughly the size of Greece (131,957 km^2 [17]), or 0.026 % of the earth's surface. Such figures are often quoted by solar energy enthusiasts, but unfortunately harvesting solar energy is easier said than done [18].

1.1.3 Concentrating Solar Power

Concentrating solar power (CSP) technologies use an optical system to direct the solar radiation incident on a large area onto a smaller area by means of reflection or refraction [10]. The system as a whole is usually referred to as the collector, the optical subsystem as the concentrator, the incident area as the aperture and the target area as the receiver or absorber [19].

Concentrators are generally divided into two categories as shown in Figure 1.3: those that track the sun in one axis (linear-focusing) and those that track the sun in two axes (point-focusing). Tracking of the sun is necessitated by the rotation of the earth around the sun which causes the angle of incident sunlight to change continuously. The different types of collector technologies are illustrated in Figure 1.4.

One distinct advantage that CSP holds over other solar technologies is its ability to continue electricity production even in the absence of sunlight by

[1] Household solar panels generally have a greater than 10 % efficiency, with solar panels for specialised applications reaching efficiencies of up to 46 % [16].

[2] $(20{,}300\,\text{TWh})/[(174.7\,\text{W/m}^2)(365 \cdot 24\,\text{h})(10\,\%)] = 132{,}647\,\text{km}^2$

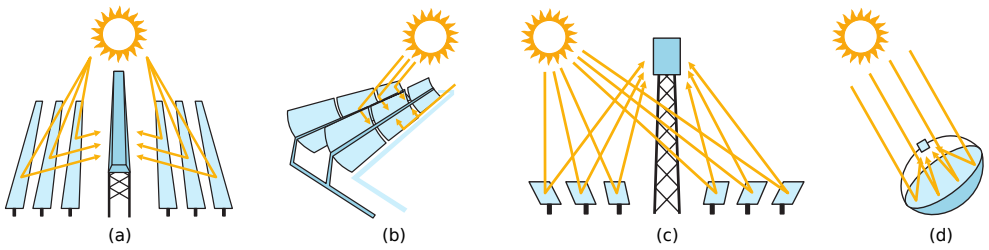


Figure 1.4: Concentrator types with single-axis tracking in (a) linear Fresnel and (b) parabolic trough and dual axis tracking in (c) central receiver and (d) parabolic dish (adapted from [18]).

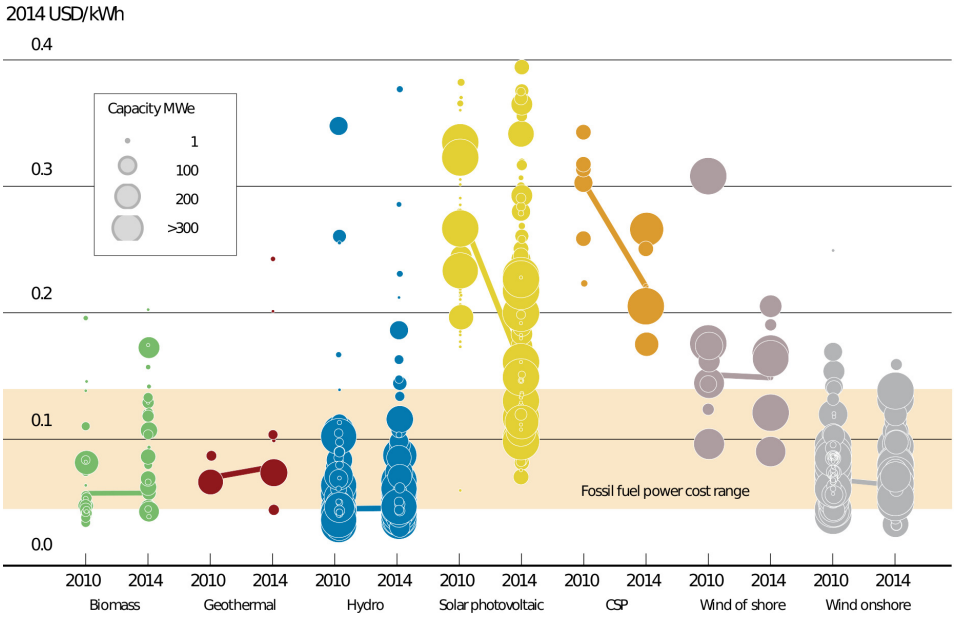


Figure 1.5: Levelised cost of electricity from utility-scale renewable technologies, 2010 and 2014 (sourced from [21]).

using the stored thermal energy accumulated throughout the day. A major drawback, however, is the current high initial cost compared to conventional power plants [20] as discussed next.

CSP is still a relatively expensive form of electricity generation. The levelised cost of electricity (LCoE)[3] of CSP is much higher than that of other renewable energy technologies, as shown in Figure 1.5. In 2014, the LCoE for CSP varied from 17 ¢/kWh to 26 ¢/kWh, compared to 8 ¢/kWh for the most competitive utility scale PV projects, and 4.5 ¢/kWh to 14 ¢/kWh for fossil fuels. What is encouraging is the significant drop in the LCoE for CSP from 2010 to 2014.

[3]LCoE is a measure of the overall competitiveness of different electricity generating technologies. It represents the overall building and maintenance cost of a plant versus electricity output over its expected lifetime [22].

Continued drops in LCoE can be expected since CSP is still in its infancy in terms of deployment, with significant growth only taking place in the last decade, as shown in Figure 1.6.

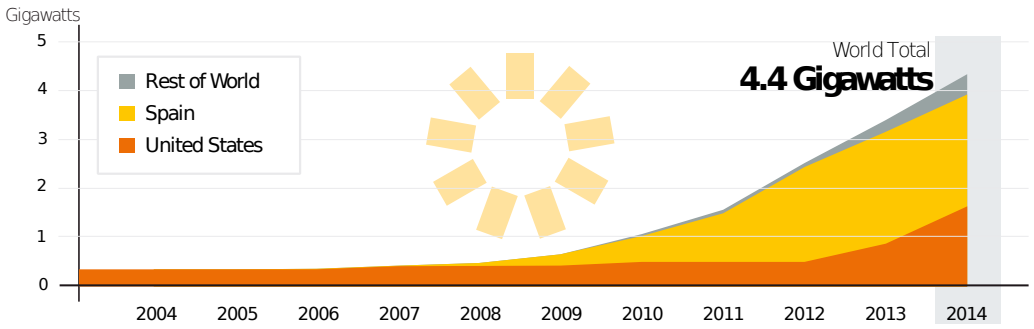


Figure 1.6: CSP growth during 2004 – 2014 (adapted from [8]).

The future of CSP is certainly a bright one. Steady growth is expected as the technology gains continued international support [18]. In South Africa, 400 MW of CSP projects have already been awarded in the first four bidding windows of the renewable energy independent power producer procurement (REIPPP) programme since 2010, led by Abengoa (250 MW), SolarReserve (100 MW) and ACWA Power (50 MW) [23]. Another local advancement is the 100% South African Helio100 technology developed at STERG with construction of a pilot facility at Mariendahl completed in September 2015 [24].

1.1.4 Central Receiver Systems

The central receiver system (CRS), sometimes also referred to as a power tower system, is a type of CSP technology which uses a field of sun-tracking mirrors, called heliostats, as the concentrator to track and direct the incident solar radiation onto a tower-mounted receiver. The reflected sunlight is absorbed by the receiver and converted into thermal energy which can then be used for electricity generation or stored for later use. Figure 1.7 showcases a photograph of the Gemasolar CSP CRS plant near Seville, Spain.

A heliostat field can contain tens of thousands of mirrors. The largest heliostat field is that of the Ivanpah plant in the United States, with 173,500 heliostats amounting to a total aperture area of 2,600,000 m^2 [25]. It is estimated that the heliostats alone contribute 30 % – 50 % of the total cost of a CRS plant [26][27]. This means that a cost reduction for a single heliostat can dramatically reduce the cost for the entire CRS plant.

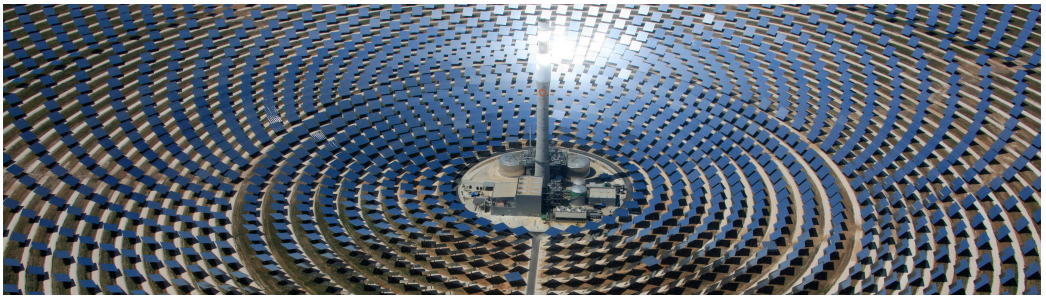


Figure 1.7: Gemasolar Central Receiver CSP plant near Seville, Spain.

1.1.5 Heliostats

A typical heliostat consists of a mirror mounted on a mechanical structure along with the necessary electronics to adjust the orientation of the mirror. Tracking is usually performed by rotating the mirror around two axes: a vertical axis to track the azimuth of the sun and a horizontal axis to track the elevation, where the vertical axis is fixed and the horizontal axis rotates around the vertical axis, as shown in Figure 1.8 [28]. Apart from the standard design, other interesting designs have also been proposed [29].

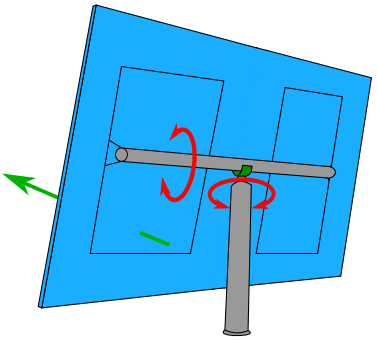


Figure 1.8: Heliostat normal and axes of rotation. Two axes are typically used: a vertical axis to track the azimuth of the sun and a horizontal axis to track the elevation.

The heliostat mirror normal is represented by the green arrow in Figure 1.8. When the orientation of the heliostat is updated, an attempt is made to align the actual mirror normal with the desired theoretical mirror normal. The angle between the actual normal and the desired normal is known as the tracking error. The tracking accuracy of heliostats directly affects the plant efficiency and therefore reducing the tracking error is an important step in the advancement of the technology. It is an area that is receiving much attention today and also the motivation for this project.

1.2 Problem Statement

Heliostat tracking accuracy is an important consideration in every CSP CRS plant. Insufficient tracking accuracy can lead to losses in power output due to the solar irradiation reflected from heliostats missing the central receiver. This also limits the maximum distance a heliostat can be from the central receiver, limiting the field (aperture) size, which in turn limits the maximum power output of the plant. Furthermore, uneven distribution of energy on the central receiver causes thermal gradients which can potentially cause damage [30], as well as improper focusing of heliostats which can potentially lead to other safety or environmental issues such as avian mortality (bird deaths) [31]. To avoid these problems, accurate heliostat tracking is required.

The current industry standard seems to be open-loop tracking using an error model for a tracking accuracy of one milliradian (discussed later in Section 2.2.3). Recalibration of error model parameters can take up to three weeks for the entire field [3][4]. During this time the calibration and tracking accuracy of heliostats may potentially deteriorate, giving rise to all the problems previously mentioned. A need therefore exists for a method that improves on the current tracking accuracy and/or reduces or completely removes the calibration time.

This need is addressed by developing a method that can accurately measure the orientation of all heliostats in a field in near real-time. When the heliostat orientation is known, the normal vector can be calculated and using the position of the sun, the direction of the reflected solar irradiation can be calculated. This information is used to reorientate the heliostat when its reflection is missing the target. Near real-time measurements remove the need for recalibration, which further reduces losses.

The method works by determining the positions of three distinct points on the heliostat mirror surface. These points define a plane which, when parallel to the mirror, is used to calculate the heliostat normal vector. This thesis documents the design and performance analysis of such a method.

1.3 Objectives

The ultimate objective of the project is to develop a method that can be used to accurately measure the orientation of all heliostats in a solar field in near real-time. After the method is fully developed, an initial analysis of its performance is required, both theoretical and experimental, in order to determine the feasibility of the method and whether or not it merits further investigation.

Primary objectives

1. Develop a method to accurately measure the orientation of all heliostats in a solar field.
2. The method must be able to measure the orientations of all the heliostats simultaneously in intervals shorter than the heliostat realignment interval.
3. Validate the method using simulations and experiments.
4. Analyse the performance of individual subsystems as well as the overall system performance to determine the feasibility of the method.

Secondary objectives

1. Maintain a tracking error of less than one milliradian.
2. Demonstrate by constructing a prototype.

1.4 Methodology

A thorough literature study is performed to become acquainted with previous work done in the field of CSP, specifically with regards to CRS and heliostat tracking strategies. This is followed by the identification of an alternative method for the measurement of heliostat mirror orientation and the development of the mathematical models and algorithms for the selected method, drawing from past work and adapting/adding where necessary. The method is validated by simulating the full system as well as individual subsystems to analyse the theoretical performance, and experiments are conducted to determine the practical performance of the method and to serve as an initial feasibility study.

1.5 Scope

The project focuses on the theoretical design and analysis of the specific method. It excludes all work related to the construction and control of the heliostats, and is only concerned with measuring their orientation. For simplicity, certain real-world considerations such as hardware and transmission channel phenomena are ignored and abstracted during the design and analysis of the method, though their influence is considered in the final remarks.

Included in the scope is the conceptual design of a method to measure heliostat mirror orientation as well as the detail design of individual subsystems. A theoretical performance analysis of the system as a whole as well as individual subsystems is also included, along with the design and conducting of experiments to analyse practical performance.

Excluded from the scope are any phenomena specific to the transmission channel, such as multipath propagation and line of sight issues, as well as modulation. All hardware is abstracted for simplicity. Heliostat construction and control is also excluded – only measurement of orientation is considered.

1.6 Contributions

The method developed in this thesis is the first approach of its kind to solving the heliostat tracking problem. It consists of a range of individual logical subsystems, contributing to modularity, all of which are designed and integrated to form a working system. Contributions meriting special mention are:

1. A detailed analysis of the performance of methods used for measuring the phase difference between two sinusoids in the presence of noise.
2. An algorithm to calculate the optimal layout of transmitters with regards to transmitter geometry.
3. A detailed analysis of the effect of transmitter geometry on position errors for errors in both absolute and relative distances.
4. A peer-reviewed research paper presented at the 4th Southern African Solar Energy Conference (SASEC) 2016, published on the conference website at http://www.sasec.org.za and available in Appendix C.

1.7 Document Outline

Chapter 2 contains a literature review consisting of two parts. The first part pertains to solar tracking and the different means by which it is achieved, as well as an overview of heliostat tracking methods, highlighting the current industry norm and the need for improvement. The second part focuses on material relevant to the tracking method developed in this thesis, starting with an overview of wireless positioning technologies, briefly touching on multilateration, and then looking at methods for phase difference calculation as a means of measuring the distance travelled by an electromagnetic wave.

Chapter 3 starts with a conceptual overview of the proposed system. A modular design is proposed and a number of subsystems are identified. The remainder of the chapter is dedicated to the detail design of each of the subsystems.

Chapter 4 evaluates and quantifies the performance of each subsystem. This is done primarily by means of simulations using Monte Carlo analysis and mathematical derivations. The performances of the individual subsystems are then used to predict the overall system performance. Finally, the performance of the full system is simulated and compared with the predicted performance.

Chapter 5 records the experiments performed and results obtained in order to validate the theoretical models and predicted system performance of the previous chapter.

Chapter 6 concludes by reviewing the content and outcome of the thesis in light of the initial problem statement and objectives, discussing the advantages, disadvantages and problem areas of the method, and ending with recommendations for future work.

Chapter 2

Literature Review

The literature review begins with a review of solar tracking methods and solar position algorithms since this forms an integral part of heliostat operation. Once a foundational understanding of these topics is obtained, the focus shifts towards specific applications of solar tracking with regards to heliostat tracking methods, with a comprehensive study of proposed as well as currently implemented methods. This comprises the first part of the literature review.

The second part reviews concepts that form the building blocks for the specific method developed in this project. A review of commercially available as well as other proposed wireless positioning technologies is presented. Then, short sections on position calculation and phase difference calculation are also presented as these topics are used in the conceptual development of Chapter 3.

2.1 Solar Tracking

A fundamental requirement for the operation of a heliostat is the ability to accurately track the sun. When the sun is observed from an arbitrary position on the surface of the earth, its position is defined relative to a coordinate system based at the point of observation. The coordinate system consists of an upwards line which is perpendicular to the surface of the earth, and a horizontal plane which is parallel to the surface of the earth, containing a north-south line and an east-west line. This is illustrated in Figure 2.1.

Using this coordinate system the position of the sun is described by two angles: the *solar zenith* angle (θ_z) and the *solar azimuth* angle (A). Since the sun does not appear as a point in the sky but rather as a disc, these angles are measured to the centre of that disc, that is, the "central ray" from the sun. The solar zenith angle is defined as the angle between the central ray from the sun and the vertical axis, whereas the solar azimuth angle is measured clockwise on the horizontal plane from the northern axis to the projection of the sun's central ray. The *solar altitude* angle (α) is sometimes used as an alternative to the

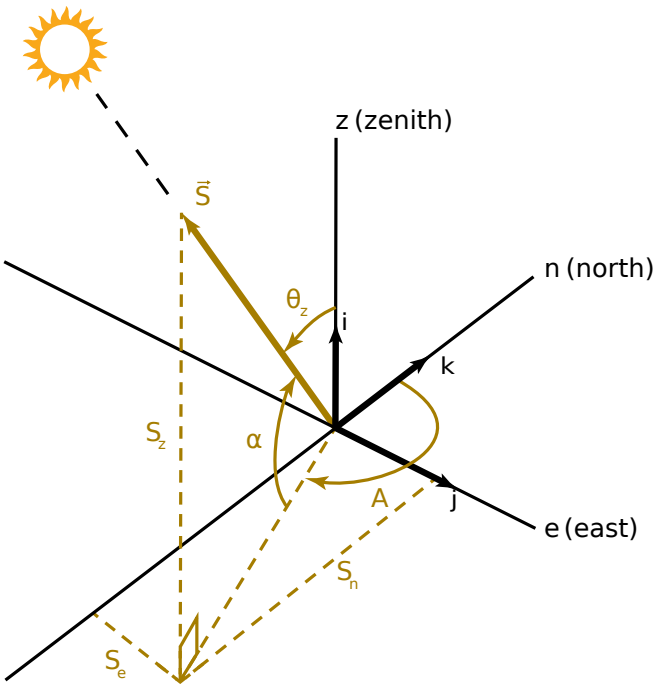


Figure 2.1: Solar vector and geometry as viewed from observer on earth surface (adapted from [32]).

solar zenith angle and is its compliment, measured from the central ray of the sun to the horizontal plane. When the solar zenith and solar azimuth angles are known it is possible to calculate $\vec{S}$, the solar vector as viewed from the position of the observer [32]. A basic understanding of the solar vector and geometry is essential in solar tracking applications, especially those that employ solar position algorithms.

In a 2013 master's thesis [2] Malan notes that solar tracking methods can be classified as being either passive or active [33], as well as being either open-loop or closed-loop [34]. Table 2.1 shows some examples of the technologies in each class. This is discussed in further detail in the following subsections.

Table 2.1: Classifications and examples of solar tracking methods [2].

	Passive Tracking	Active Tracking
Open-loop Tracking	Clockwork mechanisms	Microprocessor-based solar algorithms
Closed-loop Tracking	Trackers using thermal expansion actuators	Electro-optical feedback sensors

2.1.1 Passive Solar Tracking

Passive solar trackers are mechanical in nature and do not require electronic sensors or actuators to operate. An example of an open-loop passive solar tracker is the one designed by Abdulrahim *et al.* in 2011 which operates using a clockwork mechanism [35]. Another example of a closed-loop passive solar tracker is the one discussed by Clifford and Eastwood [33] and shown in Figure 2.2 where two identical cylindrical tubes filled with fluid under partial pressure and partially covered with shades are mounted at either side of a panel. Thermal energy from the sun causes the fluid to evaporate and move from the more exposed cylinder to the less exposed cylinder, with the mass imbalance causing the panel to move back towards facing the sun such that both cylinders always receive an equal amount of solar radiation.

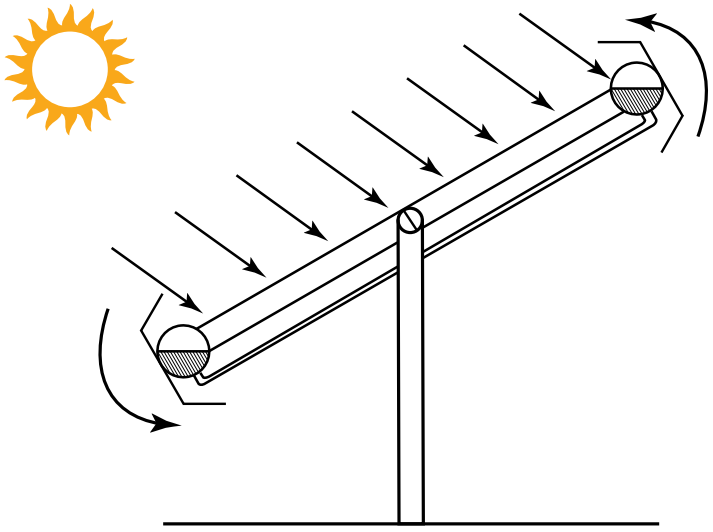


Figure 2.2: A passive closed-loop solar tracker using two identical cylindrical tubes filled with a fluid under partial pressure [33]. Unequal exposure causes vapour from one tube to flow into the other, with the mass imbalance causing the panel to tilt and face the sun.

2.1.2 Active Closed-loop Tracking

A closed-loop control system makes use of feedback obtained from sensors to generate control signals [34]. A commonly used method for active closed-loop solar tracking is electro-optical sensor feedback [36]. In such devices as shown in Figure 2.3, differential illumination on two electro-optical sensors generates a differential control signal which drives the actuators of the apparatus, reorienting the device towards the sun until the differential signal is eliminated. In (a), such a configuration is shown with a shading device separating the two sensors. In (b), the sensors are mounted on tilted planes for increased sensitivity. In

(c), the sensors are placed inside a collimator (container) so as to remove the influence of diffuse radiation.

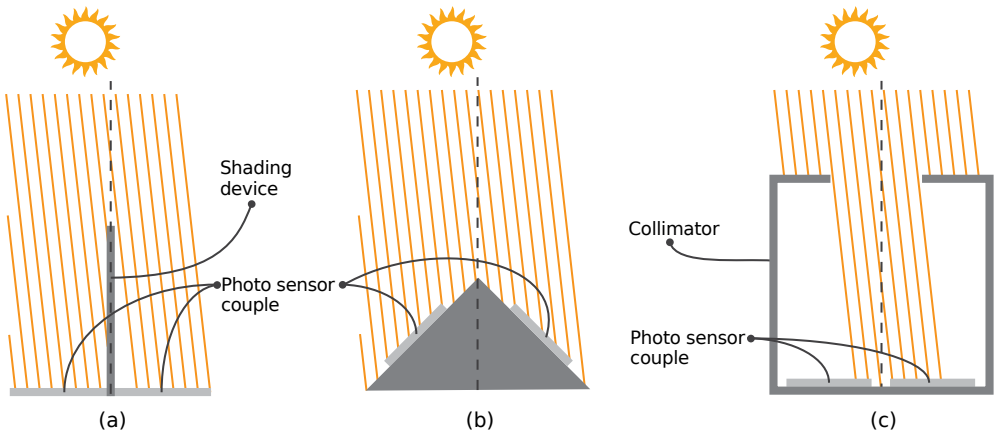


Figure 2.3: Electro-optical sensors illustrating the shade balancing principle using (a) a shading device, (b) tilted mounting of sensors and (c) a collimator to eliminate diffuse irradiation (adapted from [36]).

2.1.3 Active Open-loop Tracking

An open-loop control system uses only its predefined algorithm and current state to determine its actions, without making use of feedback to determine if the actions taken have achieved the desired result [34]. This means that an open-loop system cannot correct for any errors that might occur in the process. Typical examples of such systems are ones using solar position algorithms to periodically account for the movement of the sun. This is a very common method used in solar tracking and as such these algorithms are the subject of the next subsection.

2.1.4 Solar Position Algorithms

In 2001, Blanco-Muriel *et al.* reviews various solar positioning algorithms [37] and identifies four that can be used to estimate the true horizontal coordinates of the sun. These algorithms include Spencer [38], Pitman and Vant-Hull [39], Walraven [40] and Michalsky [41], with the latter being the most accurate having an uncertainty of 0.175 mrad. The Plataforma Solar de Almerýa (PSA) algorithm was later introduced and improved upon the Michalsky algorithm for an uncertainty of 0.15 mrad [37].

The most accurate to date is the Solar Position Algorithm (SPA) [42] published in 2004 by Reda and Andreas with uncertainties of 0.005 mrad valid from the year 2000 BC to 6000 AD. The article provides a step by step version of the algorithm that was originally described in a book by Meeus in 1988 [43].

In 2008, Grena publishes an algorithm with a maximum uncertainty of 0.047 mrad [44]. While this is less accurate than the SPA, it is designed for computational efficiency and much simpler. Four years later he publishes "Five new algorithms for the computation of sun position from 2010 to 2110" [45], however none of these algorithms have an uncertainty less than 0.047 mrad.

2.2 Heliostat Tracking Methods

This section reviews the existing heliostat tracking methods proposed in literature as well as those that are currently in use in large-scale plants. The tracking methods are divided into three groups: those that use closed-loop tracking, those that use open-loop tracking, and those that use open-loop tracking with an error model. The current tracking method used at the Solar Thermal Energy Research Group (STERG) is also discussed.

2.2.1 Closed-loop Tracking

A closed-loop control system relies on feedback obtained from the system. Figure 2.4 shows three typical sources of feedback with regards to heliostat tracking: *local feedback*, referring to feedback from actuators and other sensors on the heliostat itself, *beam feedback*, referring to feedback from sensors measuring the reflected beam angle, and *receiver feedback*, referring to feedback from sensors on the central receiver.

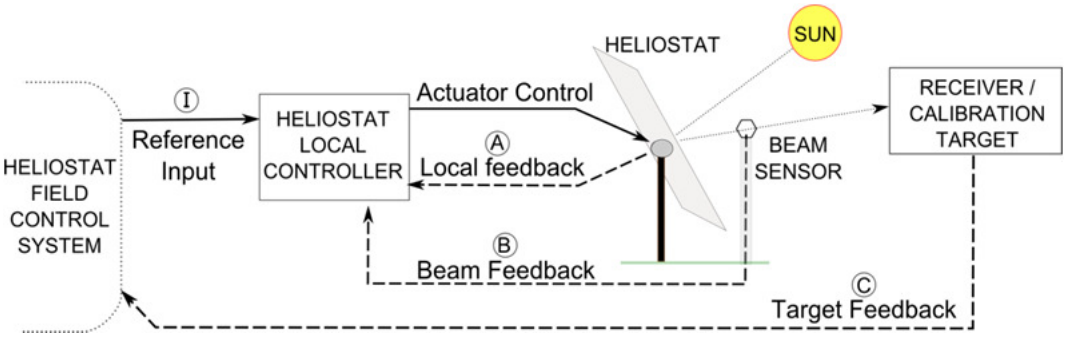


Figure 2.4: Feedback sources for closed-loop heliostat tracking [2].

Local Feedback

In 2007, Roos *et al.* designs and builds a 25 m^2 target aligned heliostat at the Council for Scientific and Industrial Research (CSIR), South Africa [46]. The heliostat design is based on one proposed by Ries and Schubnell in 1990 [47] which was shown to be superior to conventional designs in a study by Chen *et al.* in 2004 [48]. An electro-optical solar tracker consisting of a shading device

and photodiodes is used for closed-loop tracking to achieve a tracking accuracy of 3.3 mrad.

A 2009 patent application by Pfahl *et al.* [49] describes a method for local feedback by making a small hole in the surface of the mirror allowing the images from the sun and receiver target area to be projected onto an imaging device on the plane behind the hole. The imaging device measures the location of the projected images which allows for the calculation of the angles between the heliostat-to-sun vector and the heliostat normal, as well as the heliostat-to-receiver vector and the heliostat normal. When these two angles are equal, the heliostat is correctly aligned.

Beam Feedback

Quero *et al.* presents a technique using "a novel sensor design for generating an alignment error signal" in 2007 [50]. The sensor had already been patented in 1999 [51] and discussed in a 2002 article by Brey *et al.* [52]. A similar sensor is discussed by Aiuchi *et al.* in 2006 [53]. It consists of two adjacent photodiodes covered in a metal shield with a thin aperture that exposes an equal fraction of each photodiode. The difference in photocurrent produced by the photodiodes is proportional to the angle between the incident light and the sensor normal. Very accurate installation of the sensor precisely between the heliostat and the receiver allows any misdirection of the reflected light to be detected by a difference in current between the two photodiodes. Periodic recalibration is performed either mechanically or with an electronic offset adjustment.

Receiver Feedback

A 1999 patent by Yogev and Krupkin presents a system that generates feedback at the receiver on the alignment of individual heliostats [54]. Four cameras are mounted around the receiver at the top, bottom, left and right, all of which are facing the heliostat field. During each feedback interval, the cameras simultaneously photograph the heliostat field. The result is that any misaligned heliostat will result in a brightness imbalance on photographs from opposing cameras. Heliostats can then be commanded to adjust their orientation until all brightness imbalance is eliminated.

Kribus *et al.* successfully develops and operates a system based on the Yogev and Krupkin patent [54] at the Weizmann Institute heliostat field [55]. In 2004 the results are published, claiming an observed tracking accuracy of 0.1 mrad – 0.3 mrad. This is an improvement of almost an entire order on current tracking strategies which can be in the range of 1 mrad. The authors recommend that this method "should therefore be considered for use in high-concentration commercial scale heliostat fields." The fact that this has not been the case for over a decade is evident of the problems associated with operating cameras so close to a thermal receiver.

In a 2010 patent application [56], Convery presents a novel method for closed-loop heliostat control with a system that is able to uniquely identify the reflection of a single heliostat from among potentially thousands. The surface of each heliostat is fitted with a piezoelectric actuator that induces vibrations in the mirror and modulates the reflected sunlight with a unique frequency. When the reflected sunlight passes over a photodiode an algorithm such as the fast Fourier transform (FFT) is used to detect the modulating frequency and identify the corresponding heliostat. Results show that once a reflection has been brought on target the system is able to keep it there indefinitely. Convery presents the method in 2011 [57] and in 2013 the patent is approved [58].

2.2.2 Open-loop Tracking

In 2001, Chen *et al.* proposes a non-imaging, focusing heliostat using purely open-loop control and consisting of 25 40 cm×40 cm individually controlled mirrors arranged in a 5×5 grid for a total reflective area of 4 m^2 [59]. A year later, reports [60] show that the heliostat was able to rise the temperature of a solar furnace to 3400 °C while displaying a tracking accuracy of 2.5 mrad. The ability of this heliostat to increase the reflected intensity might become less prominent at greater distances, such as in a large central receiver system (CRS) plant, to the point where the added complexity of individually controlled mirrors offers very little benefit.

2.2.3 Open-loop Tracking with Error Model

In 1980, Baheti and Scott [61] present a system for open-loop control of a heliostat using a self-calibrating controller that "corrects the control system for unknown, deterministic errors in the heliostat drives and installation." The error model considered the following sources of errors: pedestal axis tilt, azimuth and elevation bias, and drive wheel radius tolerances. Other mechanical error sources that were deliberately left out include fabrication tolerances and sun sensor parts, as the authors argued that "tighter control over these errors can be maintained." This system uses the sun as reference and during each calibration interval the heliostat aligns its normal vector with the heliostat-sun vector using an accurate solar tracking sensor for feedback. The angle between the heliostat normal using the solar tracker as feedback and the heliostat normal that would have resulted from open-loop tracking is used to update the error model accordingly. Integration of the error model resulted in up to 5 times better tracking accuracy.

In a 1986 patent [62], Stone builds on the system by Baheti and Scott [61], adding the effect of non-orthogonality in the heliostat reference system to the error model. The system also moves away from using the sun as a reference. Instead, the heliostat reflection is projected onto a target location and detected

using a digital image radiometer (DIR). This is later termed the beam characterisation system (BCS). The difference between the actual location of the reflection and the intended location is then used to update the error model accordingly. Many variants of this method have been implemented and decades later, even though specific technologies have evolved, this approach is still the trend in solar concentrator tracking systems [63][64].

A modern version of the BCS proposed by Stone [62] is described by Berenguel *et al.* in 2004 [64] and in a 2012 book, *Control of Solar Energy Systems* [65], by Camacho *et al.*. A black and white charge-coupled device (CCD) camera is used to capture the reflection of the heliostats and digital image processing is used to calculate the centre of the reflected image. This method is still widely used today as is evident in the next section.

2.2.4 Tracking Methods Implemented in Existing Power Plants

A 1999 report by Stone and Jones [66] states that azimuth rotational axis tilt was considered the primary error source at Solar One at the time the heliostat field was installed. The error was minimised by measuring the azimuth tilt angle using an inclinometer and then manually adjusting the foundation bolts at the base of the heliostat pedestal. This process proved expensive and was later abandoned in favour of a BCS which is described in detail in King [67], Blackmon [68], Mavis [69] and Strachan [70]. According to a 1995 report by Stone and Lopez [71] this method provided only a small increase in tracking accuracy and may even decrease the tracking accuracy for some heliostats. In 1999, Jones and Stone evaluate three possible strategies to improve the tracking accuracy of the heliostats at Solar Two [72]. Of the three strategies, the one using an error correction model was found to be far superior. The authors go as far as saying that the future of power towers may depend on the implementation of an error correcting strategy into the control system.

Gemasolar was the first concentrating solar power (CSP) plant with a CRS and molten salt storage in commercial operation [73]. A calibration strategy such as the one described by Berenguel *et al.* is used, with a cylindrical calibration target and four cameras [64]. Additionally, each heliostat also features an inclinometer to speed up the initial approach [74]. Other large CSP CRS plants such as PS10 and PS20 in Spain, Delingha in China and Sierra SunTower in the USA all feature a calibration target mounted below the receiver. This strongly suggests that this method is – or was at some stage – used at these plants.

The calibration strategy used by eSolar at the Sierra SunTower is described in a 2009 patent application [75] and also in a 2011 article by Schell [76]. The heliostat field is surrounded by a number of cameras at elevated locations. Heliostats are periodically commanded to direct their beams into one of these

cameras. Once the beam is centred on the camera, the full state of the heliostat is captured and recorded in a database. These recorded states are used to calibrate the heliostats. The method is similar to the one described by Baheti and Scott [61] but differs in that it uses a camera as reference instead of the sun. Using multiple cameras allows much of the calibration to be performed in parallel, resulting in an entire field of 12,000 heliostats to be calibrated in less than 20 good-weather days [3][76].

A 2010 patent application by Schwarzbach and Kroyzer [77] from Brightsource describes a calibration strategy where multiple heliostats simultaneously sweep their reflected beams over a macro array of light sensors. The brightness reported by each sensor continuously changes during the sweep and analysis of this data is used to distinguish between the individual heliostat reflections. This allows simultaneous calibration of multiple heliostats which significantly speeds up the rate of calibration for the over 170,000 heliostats.

2.2.5 Heliostat Tracking Method used at STERG

In his 2013 Master's thesis, Malan [2] successfully designs and tests the first heliostat field control system within STERG. The prototype system is proudly featured on the solar roof laboratory at the Department of Mechanical and Mechatronic Engineering at Stellenbosch University. The control strategy of the system is based on the one by Berenguel *et al.* [64] and relies on open-loop tracking with local feedback and model-based error correction by means of a calibration target and image processing. This is illustrated in Figure 2.5. The normal vector root mean square (RMS) error of this system was confirmed to be equal to 0.97 mrad during three full days' testing.

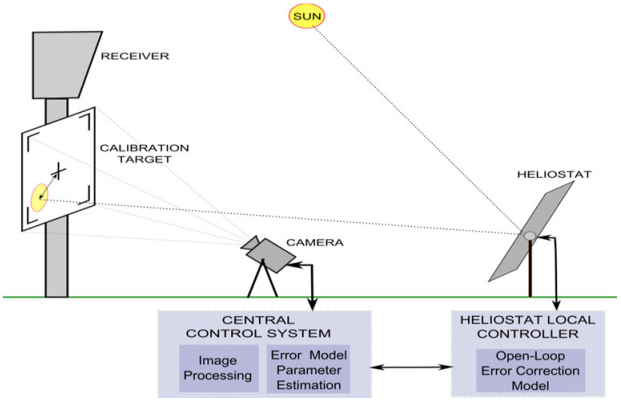


Figure 2.5: Use of a calibration target to estimate error model parameters [2].

2.2.6 The Need for Improvement

It could be argued that the need for improvement in most cases boils down to cost – whether it be directly or indirectly. Since the heliostat field contributes 30 % – 50 % of the total plant cost [26][27], it is a good area to start looking at cost reductions. The amount by which construction costs can be reduced is limited, however, since open-loop tracking requires the hardware to reliably respond to actuating signals as it is unable to perform error correction based on feedback. A method that provides feedback could potentially allow the use of less expensive hardware while maintaining the desired accuracy.

Another important factor to consider is the amount of reflected solar irradiation missing the receiver, which can be substantial [78] and directly impacts overall plant efficiency. This is commonly known as spillage and is reduced by an increased tracking accuracy, reducing or removing the heliostat recalibration time, as well as the use of focusing heliostats [79]. The lifetime of the thermal receiver can also be prolonged by maintaining an even flux distribution, achieved through the use of advanced aiming strategies [30], which require precise tracking accuracy for optimal performance.

A tracking error of 1 mrad causes the reflected beam to be off-target by about 10 cm over a range of 100 m. This may seem insignificant, but for larger fields over a range of 1 km the error grows to 1 m, which may be undesirably large. The method developed in this project attempts to measure heliostat orientation by measuring the positions of three or more distinct points on its mirror surface. The required positional accuracy for each point is loosely approximated at the hand of Figure 2.6 and investigated in detail later in Section 4.5. For a tracking error ϵ_T of 1 mrad, a small heliostat with d_R of 2 m and a large heliostat with d_R of 20 m results in position errors ϵ_R of 1 mm and 1 cm, respectively. The following sections investigate different methods related to determining the positions of these points.

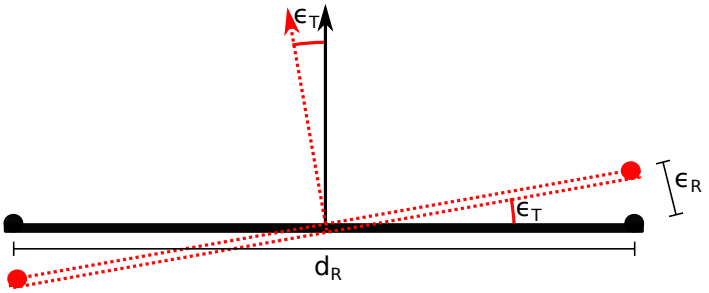


Figure 2.6: Initial estimation of required position accuracy for points on the heliostat mirror surface. When the estimated positions of the points contain errors it causes an error in the calculated heliostat normal.

2.3 Wireless Local Positioning

This section is concerned with determining the position of an object using wireless technologies. Perhaps the most well-known example of this is the Global Positioning System (GPS). The number of publications on the topic of wireless local positioning is vast – more than could be covered in a couple of pages. A wireless positioning system is required that operates with high accuracy, within one order of the estimated requirement ($< 10\,\text{cm}$) in moderately sized outdoor environments ($100\,\text{m} - 1\,\text{km}$). For this reason the focus will only be on literature deemed relevant to the aforementioned requirements or deserving of special mention.

2.3.1 Introduction to Wireless Local Positioning

The content of this introduction is largely based on the 2003 article by Vossiek *et al.* titled *Wireless Local Positioning* [80], which serves as an excellent introduction on the topic. A wireless local positioning system has two foundational components: a signal transmitter and a measuring unit. In the simplest case the transmitter is just a beacon, whereas the measuring unit contains the bulk of the system intelligence and processing load.

One way of classifying the different types of wireless positioning systems is by system topology. A system can be either self-positioning or remote-positioning. In a self-positioning system, the measuring unit is mobile and receives signals from transmitters in known locations from which it calculates its own position. A remote-positioning system works the other way around, where the transmitter is mobile and several stationary measuring units calculate the position of the transmitter.

A second classification can be made based on the measurement principle. Three main types of measurement principles are in common use today: angle-of-arrival (AOA), received-signal strength (RSS) and propagation-time based systems. The latter can further be subdivided into time-of-arrival (TOA), round-trip time-of-flight (RTOF) and time-difference-of-arrival (TDOA). The accuracy of AOA and RSS systems is only moderate when compared to TOA systems. TDOA is used in most solutions available today.

The maximum achievable accuracy of a wireless local positioning system is another consideration which is of particular importance in this work. Proximity-based methods are categorised as *coarse grained* localisation methods, whereas approaches that measure the distance to a given reference point using signal strength or timing measurements are generally much more accurate and categorised as *fine grained* localisation methods [81]. Figure 2.7 presents a rough overview of current wireless positioning systems with regards to their accuracy.

Figure 2.7 indicates a divide between technologies achieving metre accuracy

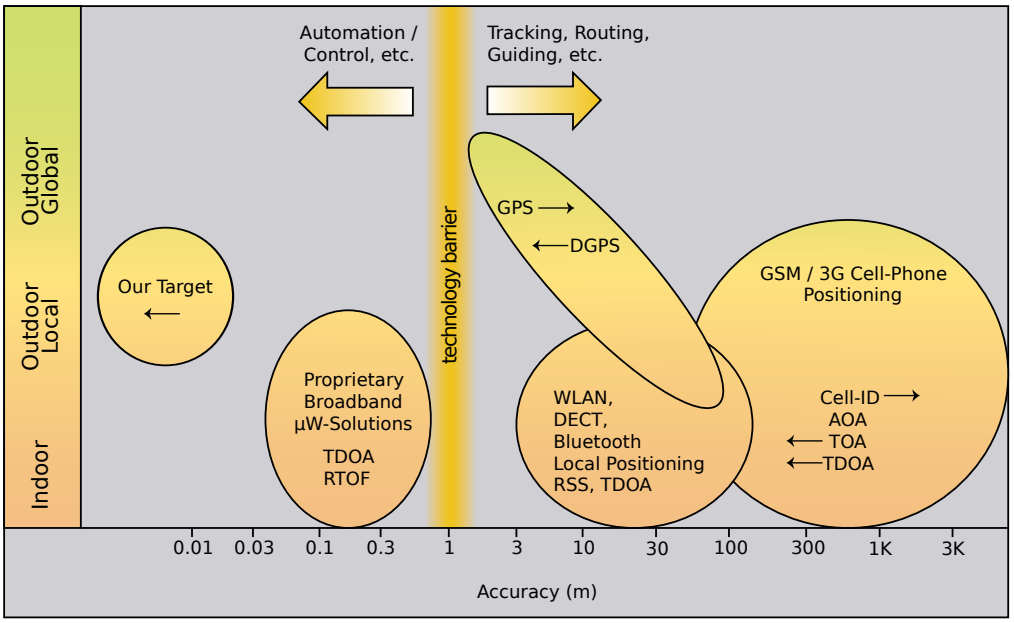


Figure 2.7: Overview of wireless positioning systems (adapted from [80]).

and the more specialised technologies achieving submetre accuracy. This idea is reinforced by a number of publications that survey existing wireless positioning technologies [80][82][83][84][85]. It is evident that much opportunity exists for highly accurate submetre positioning systems, particularly for outdoor applications.

2.3.2 Existing Solutions

Some inexpensive GPS units reach accuracies of 10 metres 95 % of the time, while the more expensive differential units (DGPS) achieve accuracies of 1 – 3 metres 99 % of the time [82]. GPS measures one-way flight time to infer the distance between the unit and the satellite by measuring the time the signal is transmitted with reference to its own internal clock when it is received. This requires accurate synchronisation between clocks: a one microsecond timing error is equivalent to a 300 metre distance error [81].

A 2001 article by Hightower and Borriello [82] lists a number of location sensing technologies and their properties. Most notable on the list is the already mentioned GPS, a technology called PinPoint 3D-iD which uses radio frequency (RF) lateration for an accuracy of 1 – 3 metres in an indoor setting, and a technology called Active Bats, which uses ultrasound time-of-flight lateration for an accuracy of 9 cm 95 % of the time, also indoors. Ultrasound technologies are less effective outdoors because they all use a single transmission frequency (40 kHz) and are thus more susceptible to interference from other ultrasound sources [81].

In 2003 Stelzer *et al.* introduces a novel position measurement technology [86]. The technology requires base stations to be stationed around the measurement field and works in the 5 GHz band using a RTOF principle, similar to conventional radar systems. An accuracy of 5 cm – 15 cm is achieved at a measurement rate of 1 kHz. Siemens' local-positioning radar (LPR) achieves a similar update rate and accuracy [80].

2.3.3 Other Publications

In 2010 Hann *et al.* [87] presents an indoor localisation technique using white light emitting diode (LED) ceiling lights. Each LED is placed at a known location and assigned a unique address wherein its position is encoded. The position of the mobile unit is determined using a correlation based method. Experimental setup is performed in a space with dimensions 30 cm × 30 cm × 50 cm, with maximum and mean position errors of 1.495 cm and 0.651 cm, respectively.

Another indoor localisation technique using white LEDs is proposed by Panta and Armstrong in 2012 [88]. The LEDs are placed at known locations and transmit sinusoids of matching frequency. Pairs of sinusoids are mixed to create an interference pattern. The phase of the sinusoid from one of the LEDs is varied in a predetermined way to cause variations in the peak-to-peak value of the interference pattern from which TDOA is then calculated. An experiment demonstrates proof of concept, though exact position errors are not published.

Jung *et al.* publishes an article in 2011 [89] discussing a method similar to the one used by Panta and Armstrong, except that instead of producing an interference pattern for TDOA, it directly calculates the phase differences between the sinusoids from pairs of LEDs using a Hilbert transform based method. Simulation results for a room with dimensions 5 m × 5 m × 3 m predict maximum and mean errors of 4.5 mm and 1.8 mm, respectively. The same approach is used by Nadeem *et al.* in 2014 [90] with simulations predicting a mean localisation error of 19.8 cm for realistic noise levels. Nadeem *et al.* finds in 2015 that frequency division multiplexing (FDM) is preferable to time division multiplexing (TDM) when the synchronisation error between two LEDs surpasses 10 % [91].

Yang *et al.* identifies the fact that LED based illumination systems have been attracting much research interest. This leads to the publication of a paper in 2009 [92] in which the signal processing application called illumination sensing is studied. FDM is investigated as a means to distinguish signals from different LEDs in order to facilitate simultaneous transmission. A filter bank structure is also proposed to support simultaneous illumination sensing. It is concluded that a large number of LEDs can be accommodated with a simple FDM scheme and a low-cost filter bank based sensor structure.

2.4 Position Calculation

This section is concerned with estimating the position of an unknown point once the positions of a number of reference points, as well as the distances between the unknown point and each reference point, are known. In a previously mentioned work, Nadeem *et al.* uses a linear least squares (LLSQ) approach to estimate the position of the unknown point [90]. A thorough study on the subject is performed by Murphy in his 2007 master's thesis [93]. With the reference points arrange roughly in the xy-plane, it was found that the result of the LLSQ estimation generally contains larger errors in the z-axis which can be improved by using averaging and weighting techniques. Furthermore it was found that the non-linear least squares (NLLSQ) estimation produces better results than LLSQ, even after averaging and weighting is applied. To reduce the number of iterations required for NLLSQ, the algorithm is seeded with the result from the LLSQ estimation.

2.5 Phase Difference Calculation

The first method presented here in calculating the phase difference between two sinusoids is the dot product (sometimes referred to as vector inner product) [94]. The method is simple but lacking as it fails to account for any noise that may be present. Another closely related method is the use of cross-correlation [94]. The main disadvantage of cross-correlation is that the resolution of the phase difference is limited by the sampling rate. Well known Fourier transform based methods such as the discrete Fourier transform (DFT) and fast Fourier transform (FFT) can also be used to calculate the phase of signals, though these methods are often hindered by problems such as spectral leakage [95].

A 2011 paper by Yoshizawa *et al.* [96] recognises some of the problems encountered with using the DFT for phase extraction and proposes a method based on non-harmonic analysis to overcome those problems. The effectiveness of the technique for extracting music from a noisy environment is investigated and it is found to be superior to DFT, providing an approximately 2 dB improvement. A 2011 paper by Dash and Hasan [97] presents an adaptive method for tracking the amplitude, phase and frequency of a time-varying sinusoid in white noise. The algorithm is shown to provide faster convergence and smaller tracking error than other methods, even in the presence of strong Gaussian white noise with low signal-to-noise ratio (SNR).

Peters and Kay investigates the unbiased estimation of the phase of a sinusoid using maximum likelihood estimator (MLE) in a 2004 paper [98]. The approach is to model the frequency and phase as unknown and non-random parameters and then apply MLE. It is proven that no unbiased estimator exists, after which the authors propose several means of obtaining estimators with less bias than

the maximum likelihood estimator. In 2007, Fu and Kam publishes a paper [99] investigating a technique called maximum *a posteriori* probability (MAP), which assumes *a priori* statistical information to be known as opposed to MLE, in which only the distribution of the noise is known. The performance of MAP is shown to be superior to MLE.

Yang *et al.* presents "A Hilbert transform based method for dynamic phase difference measurement" in a 2012 paper [100]. Simulation results show that the proposed method closely tracks the phase difference between two sinusoids, even in the presence of white Gaussian noise. Furthermore it is shown that the Hilbert transform based method is far superior to the sliding Goertzel algorithm (SGA) which was used as benchmark.

2.6 Conclusion

The topics reviewed in this chapter serve as an introduction to solar tracking and heliostat tracking methods. The current industry standard for heliostat tracking seems to be open-loop tracking with an error model for a tracking accuracy of about one milliradian. This does, however, require periodic recalibration of error model parameters which can take up to three weeks. The need for a better tracking method is clear. The remainder of the chapter reviewed topics that are relevant to the concept for an alternative tracking method developed in Chapter 3.

Chapter 3

System Design

This chapter documents and motivates the design of the system being developed. The overall process features a modular design consisting of interdependent yet isolated subprocesses. The chapter starts by giving a conceptual overview of the system, describing the process, identifying the subprocesses, and highlighting their interdependence. Each of the subprocesses is then described in detail in the remainder of the chapter.

3.1 Conceptual Overview

As stated in Chapter 1 and as suggested by the title of the thesis, the primary objective of this project is to develop a method that can be used to accurately measure the orientation of heliostat mirrors. The relevance of heliostat mirror orientation is illustrated in Figure 3.1 which shows how when the solar vector, heliostat position and heliostat mirror orientation are known, the reflected beam vector can be calculated. Additionally, when the position of the target is known, then it can be determined whether or not the beam is correctly reflected onto the target.

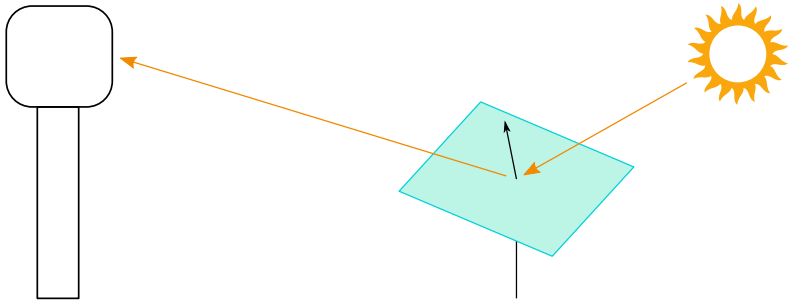


Figure 3.1: Solar irradiation incident on heliostat reflected onto target. When heliostat is improperly oriented, the reflected beam will miss the target.

The selected approach is conceptually simple and the method is based on the ability to measure the coordinates of three or more distinct points on the heliostat mirror surface. These points define a plane which, when parallel to the mirror, is used to calculate the heliostat normal vector. This is illustrated in Figure 3.2.

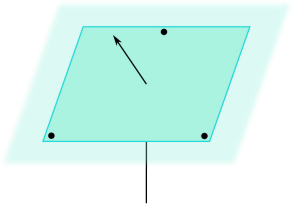


Figure 3.2: Three points mounted on heliostat mirror at distinct positions define a mathematical plane parallel to the mirror surface. The heliostat normal vector is perpendicular to the plane.

Section 2.3 reviews a number of methods used for wireless local positioning, though none are both suitable for outdoor use as well as having subcentimetre accuracy. The closest match is the method proposed by Jung *et al.* [89] on which the method developed here is based. Figure 3.3 illustrates the concept and is discussed in the following paragraphs.

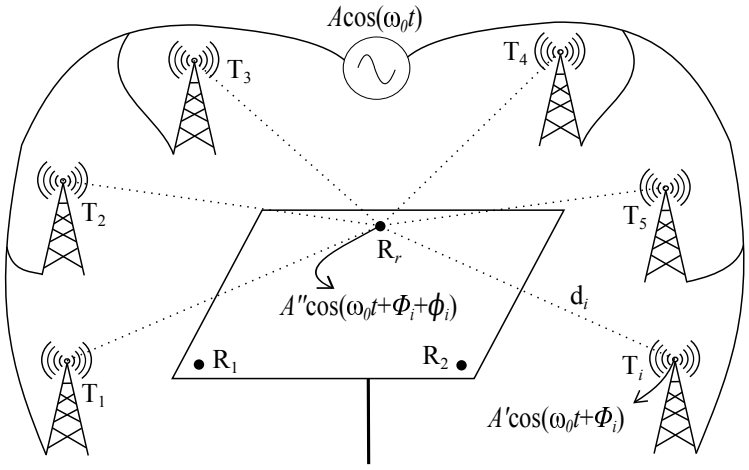


Figure 3.3: Conceptual illustration of transmitter towers surrounding heliostat field and receivers on heliostat mirror surface. A single source ensures that all transmitted signals have the same frequency. The difference between the electronic phase delay Φ_i and the phase delay due to distance ϕ_i is also illustrated.

The position of a point R on the mirror surface is calculated by placing a number of 'beacons' in the vicinity at known positions, and then measuring the distance between the point and each beacon T_i, after which an algorithm is

used to estimate the most likely position of the point. This process is called multilateration, sometimes interchangeably referred to as triangulation.

To facilitate the distance measurement, the beacons are represented by a number of tower mounted electromagnetic transmitters with the points on the mirror being represented by electromagnetic receivers. The phase shift ϕ_i of a signal from the time it is transmitted to the time it is received can be used to calculate the distance travelled d_i when its wavelength λ is known. While measuring the phase shift ϕ directly may prove problematic, it is possible to measure the difference in phase shift $\phi_i - \phi_j$ between two signals from transmitters T_i and T_j. This value is proportional to $d_i - d_j$, the distance between the receiver and transmitter T_i relative to the distance between the receiver and transmitter T_j, plus an unknown number of full wavelengths which must be accounted for. When the wavelength, and thus frequency, of the sinusoid is known, $d_i - d_j$ can be calculated exactly.

To ensure an accurate phase difference measurement requires that all transmitters transmit exactly the same frequency. A single central oscillator is connected to each transmitter to achieve this. Even though the frequency of all transmitted signals is the same, each signal still undergoes a different phase shift Φ_i from the time it is generated at the central oscillator to the time it is transmitted at T_i due to the electronic path. Instead of attempting to synchronise the phases, they are simply measured and accounted for. Their values may, however, slowly change throughout the day and need to be periodically recalculated. This is done by placing a calibration receiver at a known position within the field which uses its own position, the known positions of the transmitters, and the expected phase shift due to the known distance d_i to calculate the value of Φ_i for each T_i.

Once the phase differences have been measured and converted to distance differences, multilateration is used to estimate the position of the receiver. When the positions of all receivers on the mirror surface have been calculated, the corresponding plane and resulting normal vector is calculated. This process is illustrated in Figure 3.4.

In order for receivers to be able to distinguish between the different signals, it is necessary to divide the shared medium (air) into a number of virtual channels. To achieve this, frequency division multiplexing (FDM) is used. This is a well-documented field and an exact study on the optimal way of achieving this is beyond the scope of the project. Briefly, each transmitter T_i modulates the sinusoidal signal onto a unique carrier frequency which allows a receiver to demodulate the received signals and, based on the carrier frequency, determine which T_i each signal originated from. To prevent frequency drift it may be necessary to distribute the carrier to the receivers, for instance by using a pilot signal. Downconversion from the carrier frequency will also introduce an additional phase term which would have to be taken into account.

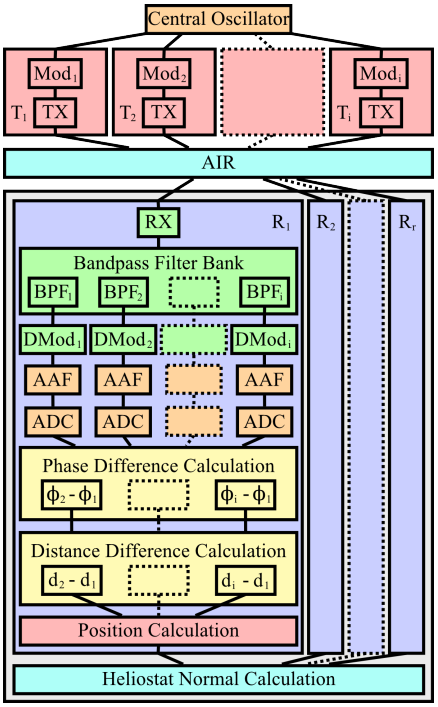


Figure 3.4: Overview of process to calculate heliostat normal vector. RX and TX represent electromagnetic transmitters and receivers, and Mod_i and DMod_i represent modulators and demodulators for the carrier frequency of transmitter T_i. AAF and ADC are the anti-aliasing filters and analog-to-digital converters used in digitisation.

In summary, the process can be logically divided into the following subprocesses: laying out of transmitter towers, modulation and demodulation of the transmitted signals, calculating the electronic phase delay using a calibration receiver, calculating the phase differences at the receiver, converting the phase differences to distance differences, estimating the position of the receiver, and calculating the heliostat normal vector. These subprocesses are addressed in detail in the sections that follow.

3.2 Phase Difference Calculation

The measured phase difference between two signals directly corresponds to the distance difference between the two transmitters. For this reason it is of the utmost importance to calculate phase difference with the highest possible accuracy.

The requirement of the method is to calculate the phase difference between two sinusoids in the presence of noise. The amplitude, frequency and individual phases of the sinusoids are not important. Furthermore only basic information about the noise is known, such as its variance and RMS power, though it is preferable not to rely on these values.

Section 2.5 reviews existing methods for calculating the phase difference between two sinusoids. Some of the methods mentioned, such as cross-correlation and Fourier transform based methods, have a limited resolution and may lack the necessary accuracy. Other methods have the ability to calculate and account for a varying amplitude, frequency and phase, making them overly complex.

The three methods investigated here are the dot product (DP), the dot product with noise compensation (DPNC) and the Hilbert transform (HT). The mathematics for these methods are derived here and their MATLAB implementations can be found in Appendix A.1.

Let $n_1(t)$ and $n_2(t)$ be two stationary,[1] ergodic[2] random processes representing noise that is uncorrelated with the signals, with $n_1(t)$ and $n_2(t)$ denoting the values of the variables at time t, and let $s_1(t)$ and $s_2(t)$ be two sinusoids with frequency f_0, angular frequency $\omega_0 = 2\pi f_0$, phases of $\theta_1(t)$ and $\theta_2(t)$, and amplitudes of a_1 and a_2, respectively:

$$\begin{aligned} s_1(t) &= a_1 \cos[\omega_0 t + \theta_1(t)] + n_1(t), \\ s_2(t) &= a_2 \cos[\omega_0 t + \theta_2(t)] + n_2(t). \end{aligned} \tag{3.1}$$

3.2.1 Dot Product (DP)

One of the simplest ways to calculate the phase difference between two sinusoids is by using the dot product. Let $\vec{s}_1$ be a vector containing S samples from $s_1(t)$ sampled at a frequency of $1/T$, where the ith sample is given by

$$\vec{s}_1[i] = s_1(iT), \qquad 0 \leq i \leq S-1, \tag{3.2}$$

and let $\vec{s}_2$ be given similarly. When $\theta_1(t)$ and $\theta_2(t)$ remain constant, then it follows from the dot product that

$$\vec{s}_1 \cdot \vec{s}_2 = \|\vec{s}_1\| \, \|\vec{s}_2\| \cos(\theta_{21}), \tag{3.3}$$

where $\|\vec{s}_1\|$ and $\|\vec{s}_2\|$ denote the sizes of vectors $\vec{s}_1$ and $\vec{s}_2$, respectively, and θ_{21} is the phase difference between the two sinusoids. The equation is solved for θ_{21} as follows:

$$\theta_{21} = |\theta_2 - \theta_1| = \arccos\left(\frac{\vec{s}_1 \cdot \vec{s}_2}{\|\vec{s}_1\| \, \|\vec{s}_2\|}\right). \tag{3.4}$$

3.2.2 Dot Product with Noise Compensation (DPNC)

One shortcoming of the DP method is that it does not account for noise. When the mean $\mathrm{E}[n]$ and variance $\mathrm{Var}[n]$ of the noise is known and $\theta_1(t)$ and $\theta_2(t)$

[1] A random process is stationary if its statistical properties do not change over time [101].

[2] A random process is ergodic if all its statistical properties can be estimated from a sufficiently large number of samples [102].

remain constant, it is possible to compensate for the effect of noise on the calculated phase difference. A closer look at the mathematics involved when multiplying two signals reveals how this can be done.

Multiplication of $s_1(t)$ and $s_2(t)$ yields

$$\begin{aligned} s(t) &= [s_1(t)][s_2(t)] \\ &= [a_1\cos(\omega_0 t+\theta_1)+n_1(t)][a_2\cos(\omega_0 t+\theta_2)+n_2(t)] \\ &= a_1a_2\cos(\omega_0 t+\theta_1)\cos(\omega_0 t+\theta_2)+ \\ &\qquad a_1\cos(\omega_0 t+\theta_1)n_2(t)+a_2\cos(\omega_0 t+\theta_2)n_1(t)+n_1(t)n_2(t) \\ &= \frac{a_1a_2}{2}\cos(\theta_2-\theta_1)+\frac{a_1a_2}{2}\cos(2\omega_0 t+\theta_2+\theta_1)+ \\ &\qquad a_1\cos(\omega_0 t+\theta_1)n_2(t)+a_2\cos(\omega_0 t+\theta_2)n_1(t)+n_1(t)n_2(t), \end{aligned} \tag{3.5}$$

where the last step makes use of the trigonometric identity

$$\cos A\cos B=\frac{1}{2}\cos(A-B)+\frac{1}{2}\cos(A+B).$$

The mean of the signal $s(t)$ is calculated as

$$\begin{aligned} s_\mu &= \frac{1}{T}\int_0^T s(t)dt \\ &= \frac{1}{T}\int_0^T \frac{a_1a_2}{2}\cos(\theta_2-\theta_1)dt+\frac{1}{T}\int_0^T n_1(t)n_2(t)dt \\ &= \frac{a_1a_2}{2}\cos(\theta_2-\theta_1)+\mathrm{E}[n_1n_2], \end{aligned} \tag{3.6}$$

where the second step follows from the fact that the mean of the periodic terms approach zero as T becomes sufficiently large, and where $\mathrm{E}[n_1n_2]$ is the expected value (mean) of the product of n_1 and n_2. Rearranging the terms gives the phase difference as

$$\theta_{21}=|\theta_2-\theta_1|=\arccos\left(\frac{2(s_\mu-\mathrm{E}[n_1n_2])}{a_1a_2}\right). \tag{3.7}$$

The amplitudes of the original signals, a_1 and a_2, are calculated by multiplying the original signal by itself, taking the average, and then rearranging the terms:

$$\begin{aligned} [s_1(t)]^2 &= [a_1\cos(\omega_0 t+\theta_1)+n_1(t)][a_1\cos(\omega_0 t+\theta_1)+n_1(t)] \\ &= a_1^2\cos^2(\omega_0 t+\theta_1)+2a_1\cos(\omega_0 t+\theta_1)n_1(t)+n_1^2(t) \\ &= \frac{a_1^2}{2}+\frac{a_1^2}{2}\cos 2(\omega_0 t+\theta_1)+2a_1\cos(\omega_0 t+\theta_1)n_1(t)+n_1^2(t) \end{aligned} \tag{3.8}$$

$$[s_1(t)]_\mu^2=\frac{1}{T}\int_0^T[s_1(t)]^2dt=\frac{1}{T}\int_0^T\frac{a_1^2}{2}dt+\frac{1}{T}\int_0^T n_1^2(t)dt=\frac{a_1^2}{2}+\mathrm{E}[n_1^2], \tag{3.9}$$

from which it follows that

$$a_1 = \sqrt{2\left([s_1(t)]_\mu^2 - \mathrm{E}[n_1^2]\right)}, \tag{3.10}$$

with

$$\mathrm{E}[n_1^2] = \mathrm{Var}[n_1] + \mathrm{E}[n_1]^2, \tag{3.11}$$

where $\mathrm{Var}[n_1]$ is the variance of n_1. The value of a_2 is calculated similarly. By letting $\mu_{12} = \mathrm{E}[n_1 n_2]$, $\mu_1 = \mathrm{E}[n_1]$, $\mu_2 = \mathrm{E}[n_2]$, $\sigma_1^2 = \mathrm{Var}[n_1]$ and $\sigma_2^2 = \mathrm{Var}[n_2]$, the final equation for the phase difference becomes

$$\theta_{21} = |\theta_2 - \theta_1| = \arccos\left(\frac{s_\mu - \mu_{12}}{\sqrt{\left([s_1(t)]_\mu^2 - \sigma_1^2 - \mu_1^2\right)\left([s_2(t)]_\mu^2 - \sigma_2^2 - \mu_2^2\right)}}\right). \tag{3.12}$$

It is important to note that neither this nor the previous method includes the sign of the angle θ_{21}. Furthermore, this method requires the mean and variance of the noise to be known, which is undesirable.

3.2.3 Hilbert Transform (HT)

Let $\hat{s}_1(t)$ and $\hat{s}_2(t)$ denote the Hilbert transforms [103] of $s_1(t)$ and $s_2(t)$, respectively, given by

$$\begin{aligned} \hat{s}_1(t) &= a_1 \sin[\omega_0 t + \theta_1(t)] + \hat{n}_1(t), \\ \hat{s}_2(t) &= a_2 \sin[\omega_0 t + \theta_2(t)] + \hat{n}_2(t), \end{aligned} \tag{3.13}$$

and let $h_1(t)$ and $h_2(t)$ be the positive analytical signals given by

$$\begin{aligned} h_1(t) &= s_1(t) + j\hat{s}_1(t), \\ h_2(t) &= s_2(t) + j\hat{s}_2(t). \end{aligned} \tag{3.14}$$

Figure 3.5 shows how an analytical signal can be geometrically represented in the complex plane as a rotating vector with angle

$$\Theta(t) = \omega_0 t + \theta(t) + \gamma(t), \tag{3.15}$$

where $\gamma(t)$ is the variation in phase due to noise. In the figure, $E_n(t)$ is the size of the noise envelope at time t and $\|h(t)\|$ is the size of the analytical signal at time t.

The angle Θ of a complex signal $h(t)$ is

$$\Theta = \angle h = \mathrm{arctan2}\left[\mathrm{imag}(h), \mathrm{real}(h)\right], \tag{3.16}$$

which leads to the difference in angle between two complex signals $h_1(t)$ and $h_2(t)$ being calculated as

$$\Theta_2(t) - \Theta_1(t) = \mathrm{arctan2}\left[\hat{s}_2(t), s_2(t)\right] - \mathrm{arctan2}\left[\hat{s}_1(t), s_1(t)\right] \tag{3.17}$$

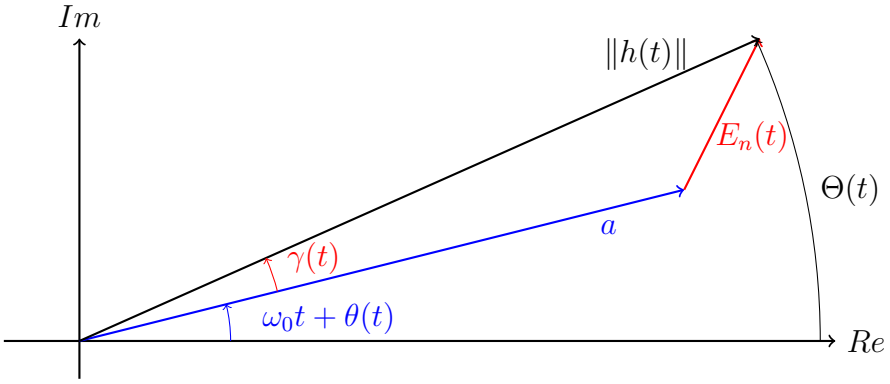


Figure 3.5: Complex geometric representation of analytical signal.

or, from Equation 3.15,

$$\begin{aligned}\Theta_2(t) - \Theta_1(t) &= [\omega_0 t + \theta_2(t) + \gamma_2(t)] - [\omega_0 t + \theta_1(t) + \gamma_1(t)] \\ &= \theta_2(t) - \theta_1(t) + \gamma_2(t) - \gamma_1(t),\end{aligned} \tag{3.18}$$

such that, from Equation 3.17 and Equation 3.18,

$$\theta_2(t) - \theta_1(t) + \gamma_2(t) - \gamma_1(t) = \text{arctan2}\left[\hat{s}_2(t), s_2(t)\right] - \text{arctan2}\left[\hat{s}_1(t), s_1(t)\right]. \tag{3.19}$$

When the SNR is large enough, $\gamma_2(t) - \gamma_1(t) \approx 0$ and Equation 3.19 simply becomes

$$\theta_2(t) - \theta_1(t) = \text{arctan2}\left[\hat{s}_2(t), s_2(t)\right] - \text{arctan2}\left[\hat{s}_1(t), s_1(t)\right]. \tag{3.20}$$

Additionally, if the phase difference is assumed to be constant, it may be desirable to average the results over time for improved accuracy. Using phasor geometry to account for phase wrapping, the averages of the real and imaginary components of the phase difference are calculated as

$$\begin{aligned}x &= \frac{1}{T}\int_0^T \cos[\theta_2(t) - \theta_1(t)]dt, \\ y &= \frac{1}{T}\int_0^T \sin[\theta_2(t) - \theta_1(t)]dt.\end{aligned} \tag{3.21}$$

The resulting phase difference, averaged over time, is then calculated as

$$\theta_{21} = \theta_2 - \theta_1 = \text{arctan2}(y, x). \tag{3.22}$$

3.3 Translating Phase Difference to Distance

When the phase shift ϕ of a sinusoid from the time it is transmitted to the time it is received is known, and the wavelength λ or frequency is also known,

then it is possible to calculate the distance d travelled using the relation

$$d = \frac{\phi}{2\pi}\lambda. \tag{3.23}$$

Recall, however, that $\phi = \theta - \Phi$, the measured phase minus the electronic phase delay. If ϕ is in the range $(-\pi, \pi]$, then an unknown number of wavelengths n must also be accounted for, since for example distances of $\frac{\lambda}{4}$, $\frac{5\lambda}{4}$ and $\frac{9\lambda}{4}$ will all yield a phase shift of $\frac{\pi}{2}$. Equation 3.23 thus becomes

$$d = \frac{\phi}{2\pi}\lambda + n\lambda. \tag{3.24}$$

The same idea applies to phase differences. When the phase difference between two sinusoids is known, then it is possible to calculate the distance difference between the two transmitters using the equation

$$\Delta_{ij} = d_i - d_j = -\left(\frac{\phi_i - \phi_j}{2\pi}\lambda + n\lambda\right), \tag{3.25}$$

where $d_i - d_j$ represents the distance between the receiver and transmitter T_i relative to the distance between the receiver and transmitter T_j. When the signal from T_i leads the one from T_j, the positive phase difference means that T_i is closer, hence the minus prefix in the equation.

When the position of the receiver is approximately known, the correct value of n can be substituted into the equation. The approximate value of n is found by rearranging the terms of Equation 3.25 such that

$$n \approx -\left(\frac{\phi_i - \phi_j}{2\pi} + \frac{d_i - d_j}{\lambda}\right). \tag{3.26}$$

For example, if T_i and T_j are located at $[10\lambda \quad 0 \quad 0]$ and $[-5\lambda \quad 0 \quad 0]$, respectively, and it is known that the receiver is located at approximately $[0 \quad 0 \quad 0]$, then a phase difference of $\frac{\pi}{18}$ would yield, from Equation 3.26,

$$n \approx -\left(\frac{\frac{\pi}{18}}{2\pi} + \frac{10\lambda - 5\lambda}{\lambda}\right) = -5.03. \tag{3.27}$$

Since n is an integer, the value is rounded down to produce $n = -5$. The distance difference is then calculated to be, from Equation 3.25, $d_i - d_j = 4.97\lambda$, which is the expected value, since T_i is approximately 5λ further than T_j.

Another way of determining the value of n is to transmit multiple sinusoids with different wavelengths such that the sinusoid with a longer wavelength is used to approximate the receiver position, while the sinusoid with a shorter wavelength provides the required accuracy. In this thesis, for simplicity, it is assumed that the wavelength is longer than the diameter of the heliostat field such that n is always zero.

3.4 Transmitter Layout

Efficient transmitter layout is an important consideration and there are many factors to consider regarding the optimal layout of transmitter towers. Taking into account all possible factors is a study in its own right and falls beyond the scope of this project. While some of these factors are briefly discussed, the only factor considered here is the sensitivity of the transmitter geometry to noise, which is the focus of this section.

The distances between transmitters and receivers is something to consider during transmitter layout. Signals from transmitters further away may be significantly weaker and thus less reliable. This may require more power at the transmitters, or an approach to intelligently handle weaker signals. One such approach may be to assign a 'reliability factor' to signals, such that stronger and more reliable signals carry more weight when calculating the receiver position.

Line of sight might be another factor to consider, especially if the transmission channel is optical. The rotation of heliostats during each day of the year would have to be taken into account to ensure that each heliostat has line of sight to the minimum number of required transmitter towers at all times. This means that it is facing the direction of the transmitter and nothing (particularly other heliostats) is blocking its vision.

The important factor considered here is the sensitivity of the transmitter geometry to noise. This principle is illustrated in Figure 3.6 for transmitter T_2. The dotted blue line represents the true distance measurement resulting in the receiver position at R_1, whereas the dashed red line represents an erroneous distance measurement resulting in the receiver position at R_2. The same small distance error has a large effect on the position error in (a) and a small effect in (b).

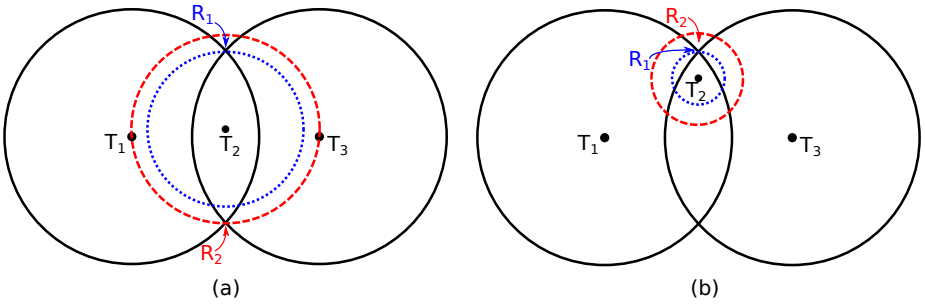


Figure 3.6: Sensitivity of transmitter geometry to noise. The effect of a small error in the measurement from transmitter T_2 on the position error of R is large in (a) and small in (b).

3.5 Position Calculation

Section 2.4 reviews a number of methods to calculate the three-dimensional position of an object when the distances between it and several known points are known. The methods discussed here are based on the ones discussed by Murphy [93] and adapted to work with relative instead of absolute distance measurements.

The calculated position $\hat{R}$ of a receiver at unknown location R is calculated using multilateration. Let there be N transmitters located at known positions T_i, with the measured distance between each T_i and R given by d_i, and the distance between each T_i and $\hat{R}$ given by $\hat{d}_i$. Due to noise and other factors, the measured distances all contain small errors such that an error $f_i = \hat{d}_i - d_i$ exists. These errors must be minimised such that the error in the calculated position $\hat{R}$ is a minimum.

When the measured distances are relative and not absolute, d_i can be expressed as $d_i = d_r + \Delta_i$, where d_r is the reference distance and Δ_i is the difference between d_i and d_r, in other words, how much further T_i is from R than T_r.

A minimum of $N \geq D + 1$ transmitters are required, where D is the number of dimensions. For three dimensions, $N \geq 4$. When relative distances are used instead of absolute distances, an additional transmitter is required, such that $N \geq D + 2$. For three dimensions and relative distances, $N \geq 5$.

3.5.1 Linear Least Squares

Let $R = [x \quad y \quad z]^T$ denote the unknown receiver coordinates and let $T_i = [T_{i,x} \quad T_{i,y} \quad T_{i,z}]^T$ denote the known coordinates of transmitter T_i. The square of the measured distance between R and T_i is given by

$$\begin{aligned} d_i^2 = (d_r + \Delta_i)^2 &= (x - T_{i,x})^2 + (y - T_{i,y})^2 + (z - T_{i,z})^2 \\ d_r^2 + 2d_r\Delta_i + \Delta_i^2 &= x^2 - 2xT_{i,x} + T_{i,x}^2 + y^2 - 2yT_{i,y} + T_{i,y}^2 + z^2 - 2zT_{i,z} + T_{i,z}^2, \end{aligned} \tag{3.28}$$

which can be rearranged as

$$2xT_{i,x} + 2yT_{i,y} + 2zT_{i,z} + 2d_r\Delta_i = T_{i,x}^2 + T_{i,y}^2 + T_{i,z}^2 - \Delta_i^2 + x^2 + y^2 + z^2 - d_r^2. \tag{3.29}$$

This can be written in matrix form as

$$2\begin{bmatrix} T_{1,x} & T_{1,y} & T_{1,z} & \Delta_1 \\ \vdots & \vdots & \vdots & \vdots \\ T_{N-1,x} & T_{N-1,y} & T_{N-1,z} & \Delta_{N-1} \\ T_{N,x} & T_{N,y} & T_{N,z} & \Delta_N \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ d_r \end{bmatrix}$$
$$= \begin{bmatrix} T_{1,x}^2 + T_{1,y}^2 + T_{1,z}^2 - \Delta_1^2 + x^2 + y^2 + z^2 - d_r^2 \\ \vdots \\ T_{N-1,x}^2 + T_{N-1,y}^2 + T_{N-1,z}^2 - \Delta_{N-1}^2 + x^2 + y^2 + z^2 - d_r^2 \\ T_{N,x}^2 + T_{N,y}^2 + T_{N,z}^2 - \Delta_N^2 + x^2 + y^2 + z^2 - d_r^2 \end{bmatrix}. \quad (3.30)$$

By choosing T_N as the reference transmitter and subtracting the last equation from the others, this system of equations becomes

$$\boldsymbol{A}\vec{\beta} = \vec{b}, \quad (3.31)$$

where

$$\boldsymbol{A} = \begin{bmatrix} T_{1,x} - T_{N,x} & T_{1,y} - T_{N,y} & T_{1,z} - T_{N,z} & \Delta_1 - \Delta_N \\ \vdots & \vdots & \vdots & \vdots \\ T_{N-1,x} - T_{N,x} & T_{N-1,y} - T_{N,y} & T_{N-1,z} - T_{N,z} & \Delta_{N-1} - \Delta_N \end{bmatrix},$$
$$\vec{\beta} = \begin{bmatrix} x \\ y \\ z \\ d_r \end{bmatrix},$$
$$\vec{b} = \frac{1}{2}\begin{bmatrix} T_{1,x}^2 + T_{1,y}^2 + T_{1,z}^2 - \Delta_1^2 - T_{N,x}^2 - T_{N,y}^2 - T_{N,z}^2 + \Delta_N^2 \\ \vdots \\ T_{N-1,x}^2 + T_{N-1,y}^2 + T_{N-1,z}^2 - \Delta_{N-1}^2 - T_{N,x}^2 - T_{N,y}^2 - T_{N,z}^2 + \Delta_N^2 \end{bmatrix}. \quad (3.32)$$

In some cases, like when $N > D + 2$ and the system is over-defined, there may be no exact solution for $\vec{\beta}$ and a solution is required that best approximates R such that $\boldsymbol{A}\vec{\beta} \approx \vec{b}$. To achieve this, define the sum of the squares of the residuals as

$$S = (\vec{b} - \boldsymbol{A}\vec{\beta})^T(\vec{b} - \boldsymbol{A}\vec{\beta}) = \vec{b}^T\vec{b} - \vec{b}^T\boldsymbol{A}\vec{\beta} - \vec{\beta}^T\boldsymbol{A}^T\vec{b} + \vec{\beta}^T\boldsymbol{A}^T\boldsymbol{A}\vec{\beta}$$
$$= \vec{b}^T\vec{b} - 2\vec{b}^T\boldsymbol{A}\vec{\beta} + \vec{\beta}^T\boldsymbol{A}^T\boldsymbol{A}\vec{\beta}, \quad (3.33)$$

and find its minimum by setting its derivative with respect to $\vec{\beta}$ equal to zero

$$\frac{\partial S}{\partial \vec{\beta}} = -2\vec{b}^T\boldsymbol{A} + \vec{\beta}^T(\boldsymbol{A}^T\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{A}) = -2\vec{b}^T\boldsymbol{A} + 2\vec{\beta}^T\boldsymbol{A}^T\boldsymbol{A} = 0, \quad (3.34)$$

which yields

$$\boldsymbol{A}^T\boldsymbol{A}\vec{\beta} = \boldsymbol{A}^T\vec{b}. \tag{3.35}$$

If $\boldsymbol{A}^T\boldsymbol{A}$ is non-singular and well-conditioned, then

$$\vec{\beta} = \boldsymbol{A}^{\#}\vec{b}, \tag{3.36}$$

where $\boldsymbol{A}^{\#} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is the pseudoinverse of $\boldsymbol{A}$. The calculated receiver position $\hat{R}$ is then given by the first three values of $\vec{\beta}$:

$$\hat{R} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{3.37}$$

3.5.2 Non-Linear Least Squares

The error function f_i is properly defined as

$$f_i(x,y,z) = \hat{d}_i - d_i = \sqrt{(x-T_{i,x})^2 + (y-T_{i,y})^2 + (z-T_{i,z})^2} - (d_r + \Delta_i), \tag{3.38}$$

where x, y and z are the calculated position of the receiver $\hat{R}$. The goal is to minimise the sum of the squares of the errors

$$F(x,y,z) = \sum_{i=1}^{N} f_i(x,y,z)^2. \tag{3.39}$$

Differentiating F with respect to x yields

$$\frac{\partial F}{\partial x} = 2\sum_{i=1}^{N} f_i \frac{\partial f_i}{\partial x} \tag{3.40}$$

with $\partial F/\partial y$, $\partial F/\partial z$ and $\partial F/\partial d_r$ obtained similarly. The partial derivative of f_i with respect to x is given by

$$\frac{\partial f_i}{\partial x} = \frac{x - T_{i,x}}{\sqrt{(x-T_{i,x})^2 + (y-T_{i,y})^2 + (z-T_{i,z})^2}} = \frac{x - T_{i,x}}{f_i + d_r + \Delta_i}, \tag{3.41}$$

with $\partial f_i/\partial y$ and $\partial f_i/\partial z$ obtained similarly, and $\partial f_i/\partial d_r = -1$. Introducing the vectors $\vec{f}$, $\vec{g}$ and the Jacobian matrix $\boldsymbol{J}$, leads to

$$\vec{g} = 2\boldsymbol{J}^T\vec{f}, \tag{3.42}$$

where

$$\vec{g} = \begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial z} \\ \frac{\partial F}{\partial d_r} \end{bmatrix}, \boldsymbol{J} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial d_r} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} & \frac{\partial f_2}{\partial d_r} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_N}{\partial x} & \frac{\partial f_N}{\partial y} & \frac{\partial f_N}{\partial z} & \frac{\partial f_N}{\partial d_r} \end{bmatrix}, \vec{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}. \tag{3.43}$$

Using the vector

$$\vec{\beta} = \begin{bmatrix} x \\ y \\ z \\ d_r \end{bmatrix}, \tag{3.44}$$

where x, y and z are the calculated position of $\hat{R}$, Newton iteration gives

$$\vec{\beta}_{k+1} = \vec{\beta}_k - (\boldsymbol{J}_k^T \boldsymbol{J}_k)^{-1} \boldsymbol{J}_k^T \vec{f}_k, \tag{3.45}$$

where $\vec{\beta}_k$ denotes the k^{th} approximate solution. The subscript k in $\boldsymbol{J}$ and $\vec{f}$ means that these quantities are evaluated at $\vec{\beta}_k$. Iteration is stopped once the current calculated position differs from the previous calculated position by less than some threshold δ, that is $||\vec{\beta}_{k+1} - \vec{\beta}_k|| < \delta$, or once the maximum allowed number of iterations is exceeded. Also given is

$$\boldsymbol{J}^T\boldsymbol{J} = \begin{bmatrix} \sum_{i=1}^{N} \frac{(x-x_i)^2}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{(x-x_i)(y-y_i)}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{(x-x_i)(z-z_i)}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{-(x-x_i)}{f_i+d_r+\Delta_i} \\ \sum_{i=1}^{N} \frac{(x-x_i)(y-y_i)}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{(y-y_i)^2}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{(y-y_i)(z-z_i)}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{-(y-y_i)}{f_i+d_r+\Delta_i} \\ \sum_{i=1}^{N} \frac{(x-x_i)(z-z_i)}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{(y-y_i)(z-z_i)}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{(z-z_i)^2}{(f_i+d_r+\Delta_i)^2} & \sum_{i=1}^{N} \frac{-(z-z_i)}{f_i+d_r+\Delta_i} \\ \sum_{i=1}^{N} \frac{-(x-x_i)}{f_i+d_r+\Delta_i} & \sum_{i=1}^{N} \frac{-(y-y_i)}{f_i+d_r+\Delta_i} & \sum_{i=1}^{N} \frac{-(z-z_i)}{f_i+d_r+\Delta_i} & \sum_{i=1}^{N} 1 \end{bmatrix}, \tag{3.46}$$

$$\boldsymbol{J}^T\vec{f} = \begin{bmatrix} \sum_{i=1}^{N} \frac{x-x_i}{d_i+d_r+\Delta_i} f_i \\ \sum_{i=1}^{N} \frac{y-y_i}{d_i+d_r+\Delta_i} f_i \\ \sum_{i=1}^{N} \frac{z-z_i}{d_i+d_r+\Delta_i} f_i \\ \sum_{i=1}^{N} -f_i \end{bmatrix}. \tag{3.47}$$

3.6 Receiver Plane Normal Vector

The normal vector $\vec{v}$ of the plane formed by the receivers on the heliostat mirror can be calculated when the positions of at least three points (R_1, R_2, R_3) on the plane are known. The unit vector $\vec{u}$ is obtained by normalising the result of the cross product used to calculate $\vec{v}$

$$\vec{v} = (R_2 - R_1) \times (R_3 - R_1), \qquad \vec{u} = \frac{\vec{v}}{||\vec{v}||}. \tag{3.48}$$

When the receivers are numerically arranged counter-clockwise as viewed from the front, then $\vec{u}$ is in the forward-facing direction of the heliostat, otherwise it is backwards.

For simplicity, it is assumed in this project that the receiver plane and mirror surface are parallel, though it is possible to develop an error model to compensate for any non-parallelism. It might also be desirable to use more than three receivers to reduce the error in the normal vector using redundancy.

3.7 Electronic Phase Delay Calculation

The electronic phase delay Φ_i of the signals before they are transmitted is calculated using a calibration receiver at known position R_c. It is important to realise that the absolute value of each Φ_i is not important as long as their relative values remain the same. The receiver starts by calculating the distance d_i between itself and every transmitter T_i

$$d_i = \sqrt{(T_{i,x} - R_{c,x})^2 + (T_{i,y} - R_{c,y})^2 + (T_{i,z} - R_{c,z})^2}. \tag{3.49}$$

It then calculates the phase differences of the signals it receives in exactly the same way as any other receiver does. If all phase differences are calculated relative to transmitter T_r, then $\Phi_r = 0$ and each $\Phi_i, i \neq r$ is calculated relative to Φ_r. The electronic phase delay Φ_i is calculated using

$$\begin{aligned} \theta_i - \theta_r &= (\phi_i + \Phi_i) - (\phi_r - \Phi_r) \\ \theta_{ir} &= \phi_i - \phi_r + \Phi_i \\ &= \frac{d_i}{\lambda}2\pi - \frac{d_r}{\lambda}2\pi + \Phi_i = \frac{d_i - d_r}{\lambda}2\pi + \Phi_i, \end{aligned} \tag{3.50}$$

such that

$$\Phi_i = \theta_{ir} - \frac{d_i - d_r}{\lambda}2\pi, \tag{3.51}$$

where θ_{ir} is the measured phase difference between the signals from transmitters T_i and T_r, and d_i and d_r are the distances from the receiver R_c to transmitters T_i and T_r, respectively.

3.8 Conclusion

In this chapter an overview of the concept for the proposed system was presented, followed by the theoretical development of each subsystem in detail. These subsystems include the accurate measuring of phase difference between two signals, translating the phase difference to distance, transmitter layout and calculating receiver position based on measured distances, and finally calculating the heliostat orientation. It was also shown how the electronic phase delay from the central oscillator to each transmitter is calculated. In the next chapter the theoretical performance of the system is analysed to determine its preliminary feasibility.

Chapter 4

System Analysis

This chapter performs an in-depth analysis of the performance of all subsystems discussed and developed in the previous chapter. The performance of each individual subsystem is quantified and used to predict the overall system performance. Finally, the performance of the full system is analysed and compared with the predicted performance. An error budget is also provided clearly showing the effect of each subsystem on the overall error.

4.1 Phase Difference Calculation

The different methods for phase difference calculation are evaluated by means of a Monte Carlo analysis. The goal is to determine the influence of a number of factors on the accuracy with which each method is able to calculate the phase difference between two sinusoids. These factors are: varying SNR, varying noise bandwidths, varying number of samples, and varying phase differences (some methods may be more sensitive to noise at certain phase differences).

To emulate real-world conditions, all simulations have additive white Gaussian noise (AWGN) added to the signals for realistic results, that is, zero-mean noise with constant power spectral density (PSD). Figure 4.1 shows the time and frequency domain representations of a sinusoid, to which AWGN is added, after which a band pass filter (BPF) is applied. The exact methodology for the simulations is described next, followed by a detailed analysis of the performance of each method. Finally, the methods are compared in order to identify the most suitable method.

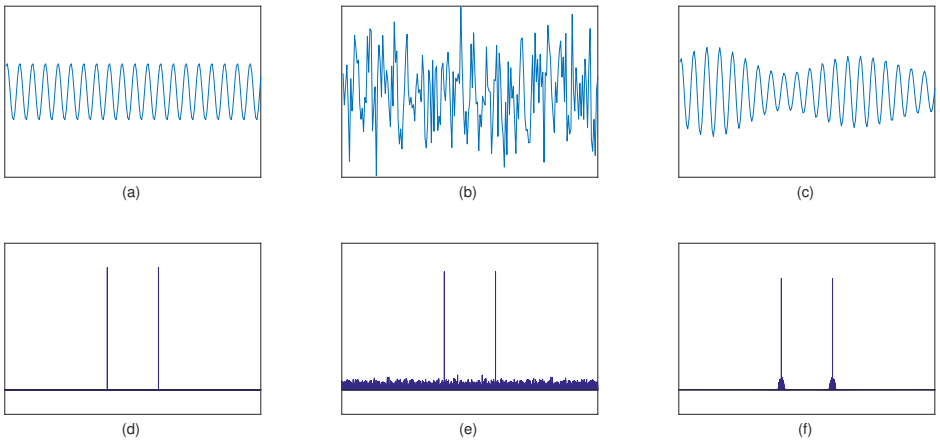


Figure 4.1: Time domain illustration of (a) a pure sinusoid, (b) the sinusoid in the presence of AWGN, (c) the signal after a band pass filter has been applied, and their frequency domain representations in (d), (e) and (f), respectively.

4.1.1 Methodology

The analysis of each method can be divided into two parts. The first part evaluates the performance of the method with varying SNR, varying noise bandwidths, and varying number of samples. The second part evaluates the performance of the method for varying phase differences.

Noise is generated using a random number generator with a normal distribution which is optionally passed through a BPF. The SNR is controlled by measuring the RMS power of the resulting noise and then adjusting the amplitude of the original signal accordingly. Noise is regenerated using a different seed for each signal such that the noise is always uncorrelated. Each simulation is repeated 10,000 times to ensure convergence of results.

Algorithm 1 shows the pseudocode for the core of the simulation, with statements specific to the first and second parts in red and blue, respectively. In the first part, the SNR is varied from 0 dB to 50 dB in 5 dB intervals. Two sinusoids are generated with a common random phase from a uniform distribution in the range $(-\pi, \pi]$. This is done to eliminate any effects caused by the absolute phase of the signals. An additional random phase difference is added to the second signal, generated from a uniform distribution in the range $(-\pi, \pi]$. In the second part, the phase difference is varied from -179 ° to 180 ° in 1 ° intervals, while the SNR and number of samples are kept constant. Sinusoids are again generated with a common random phase from a uniform distribution in the range $(-\pi, \pi]$, with the specified phase difference added to the second signal.

Recall that of the three methods investigated, only the Hilbert transform (HT) method correctly includes the sign of the angle θ_{21}. It therefore only makes sense to include the absolute value of the error for the other two methods,

Algorithm 1 Performance analysis of phase difference calculation method

$S \leftarrow$ number of samples
$SPC \leftarrow$ samples per cycle
$intervals \leftarrow (1 : S)/SPC \times 2\pi$
$SNR \leftarrow$ desired signal-to-noise ratio
$filter \leftarrow$ **false**
if filtering is required **then**
    $filter \leftarrow$ filter object with specified parameters
$phase \leftarrow$ random value uniformly distributed in $(-\pi, \pi]$
$phasediff \leftarrow$ random value uniformly distributed in $(-\pi, \pi]$
$phasediff \leftarrow$ desired phase difference in radians
$noise1 \leftarrow S$ random values from the standard normal distribution
$noise2 \leftarrow S$ random values from the standard normal distribution
if $filter$ is not **false then**
    $noise1 \leftarrow noise1$ after filtering with $filter$
    $noise2 \leftarrow noise2$ after filtering with $filter$
$a1 \leftarrow \text{rms}(noise1) \times \text{sqrt}(2 \times SNR)$
$a2 \leftarrow \text{rms}(noise2) \times \text{sqrt}(2 \times SNR)$
$signal1 \leftarrow noise1 + a1 \times \cos(intervals + phase)$
$signal2 \leftarrow noise2 + a2 \times \cos(intervals + phase + phasediff)$
$phasediffcalc \leftarrow$ calculated phase difference between $signal1$ and $signal2$
$error \leftarrow phasediffcalc - phasediff$

which would inevitably introduce a positive non-zero bias in the error mean. This makes comparison difficult and is avoided by assuming the correct sign for θ_{21} in the simulation for these methods.

Table 4.1: Clarification of the legends for figures later in the section.

Key	Meaning
S	Number of samples.
SPC	Samples per cycle, which is a measure of the normalised sampling frequency.
B	Equivalent normalised noise bandwidth. No filtering is indicated by 'none'.
SNR2	Indicates that the SNR of the second signal is SNR2-times more than the SNR of the first signal.

The RMS phase difference error is denoted by ϵ_ϕ in radians, and in the figures it is shown in degrees, that is, $(180°/\pi)\epsilon_\phi$. The legend used in some of the figures in this section is clarified in Table 4.1. When f_o and f_s denote the signal frequency and sampling frequency, respectively, then samples per cycle (SPC) is defined as $SPC = f_s/f_o$. The normalised sampling frequency is given by

$2f_o/f_s$ [104], or $2/SPC$. The normalised noise bandwidth is obtained similarly by multiplying the bandwidth by $2/f_s$.

Generally, the -3 dB cut-off points are considered as the endpoints of a filter's bandwidth, however, filters with the same -3 dB bandwidth but different filter orders may not always have the same magnitude. For this reason it is simpler to use an ideal filter in the calculations, such that filters with the same equivalent bandwidth always have the same magnitude. Figure 4.2 illustrates the difference between the practical -3 dB bandwidth and ideal equivalent bandwidth, and Table 4.2 shows how the equivalent normalised bandwidths used later in the section are calculated.

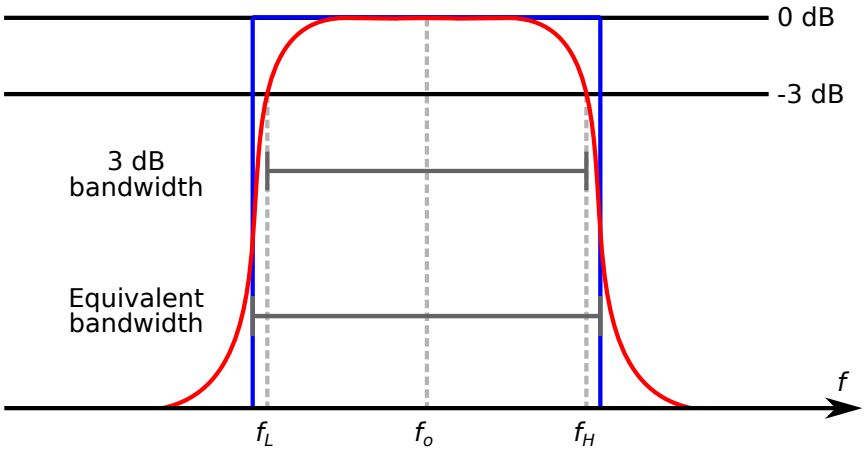


Figure 4.2: Difference between -3 dB bandwidth and equivalent bandwidth. The magnitude response of a practical band pass filter is shown in red and its equivalent ideal filter in blue, with equal area under each curve.

Table 4.2: Obtaining equivalent normalised bandwidth (BW) from bandwidth expressed in terms of signal frequency f_o and samples per cycle (SPC). Bandwidths are chosen to better illustrate simulation results and not for practical considerations.

BW	SPC	Normalised BW	Equivalent Normalised BW
$0.8f_o$	10	0.16	0.163
$0.2f_o$	10	0.04	0.043
$0.75f_o$	3	0.5	0.51

4.1.2 Dot Product (DP)

The performance of the dot product (DP) method for varying SNR, varying bandwidth and varying number of samples is shown in Figure 4.3. As expected, the error decreases as the SNR increases, though it appears that filter bandwidth and number of samples have no effect on the error. The last graph (-·-·-) shows

that a lower sampling frequency negatively affects the results leading to a larger error.

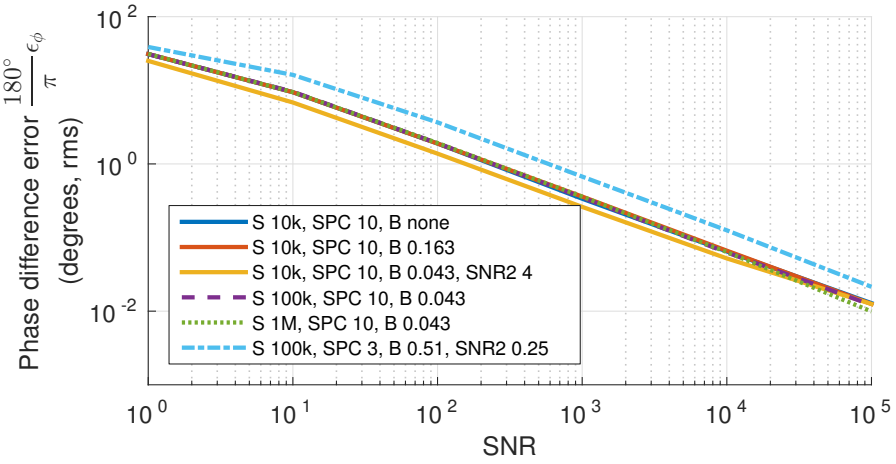


Figure 4.3: Performance of DP method for varying SNR, varying bandwidth B, varying number of samples S, and varying phase difference.

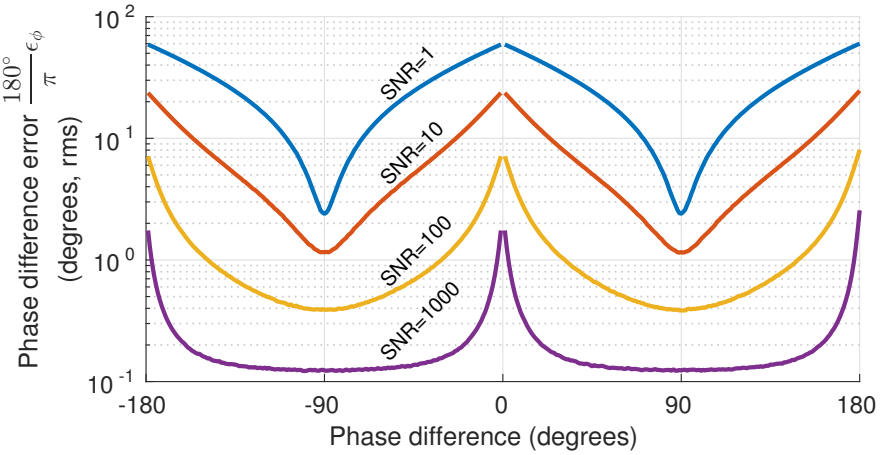


Figure 4.4: Performance of DP method for varying phase differences and varying SNR, with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=0.043.

Figure 4.4 shows how the error varies at various phase differences. The DP method uses the inverse cosine function which is least and most sensitive to amplitude noise where a change in y results in the smallest and largest possible change in x, respectively. To illustrate this, let

$$y = \cos x \quad \Longleftrightarrow \quad x = \cos^{-1} y, \qquad x \in [-180\,^\circ, 180\,^\circ], \tag{4.1}$$

such that the absolute of the derivative of x is

$$\left|\frac{\mathrm{d}\,x}{\mathrm{d}\,y}\right| = \left|\frac{\mathrm{d}}{\mathrm{d}\,y}\left(\cos^{-1} y\right)\right| = \left|\frac{-1}{\sqrt{1-y^2}}\right|, \qquad y \in [-1, 1], \tag{4.2}$$

which has a minimum at $y = 0$ and a maximum as y approaches $y = \pm 1$. It can therefore be expected that the DP method would be least sensitive to noise at $\pm 90\,^\circ$ ($\arccos 0$) and most sensitive to noise at $0\,^\circ$ and $\pm 180\,^\circ$ ($\arccos \pm 1$). This reasoning is confirmed in Figure 4.4 which clearly shows an increased sensitivity to noise at these phase differences.

The error distribution of the DP method is shown in Figure 4.5. While the input noise is normally distributed, the resulting error distribution clearly is not. Notice the correlation with Figure 4.4: at low SNR, errors of all sizes are almost equally likely, but as the SNR is increased, the majority of errors are small, with larger errors only occurring at around 0° and 180°. This is confirmed in Figure 4.5 at higher SNR, with the majority of errors being small and close to 0, though the presence of larger errors is clearly visible despite being less frequent.

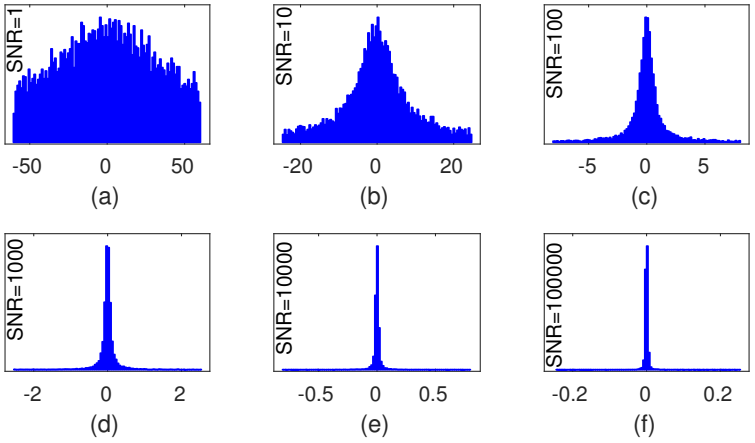


Figure 4.5: Error distribution in degrees of DP method for varying SNR in (a) – (f), with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=none (no filtering).

4.1.3 Dot Product with Noise Compensation (DPNC)

Figure 4.6 shows the performance of the dot product with noise compensation (DPNC) method for varying SNR, varying bandwidth, and varying number of samples. Again, as expected, the error decreases as the SNR increases. The fourth (– – –) and fifth (·····) graphs indicate that an increase in the number of samples leads to a smaller error. Additionally, the first (——) and second (——)

graphs indicate that a wider bandwidth also contributes towards a smaller error. At first this may seem counter-intuitive, but for the same SNR in

$$\text{SNR} = 2\frac{\mathcal{N}}{2}B = \mathcal{N}B, \tag{4.3}$$

where $\mathcal{N}/2$ is the PSD and B the bandwidth, a lower PSD and wider bandwidth results in a smaller phase difference error.

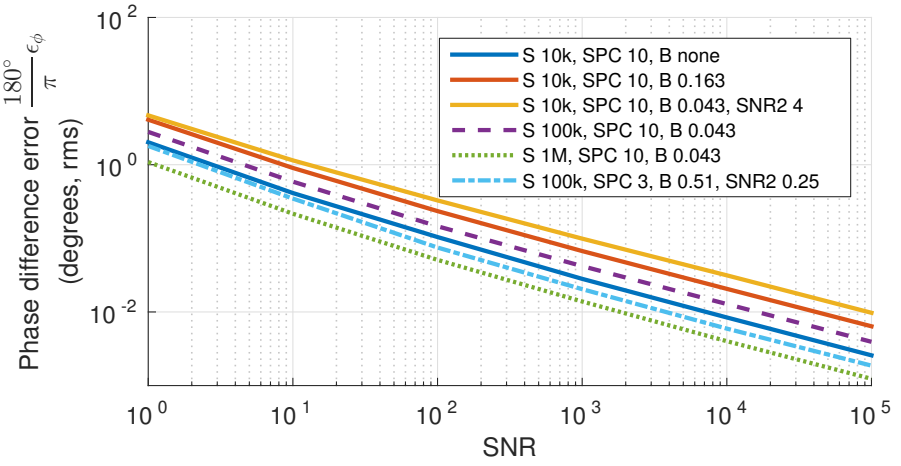


Figure 4.6: Performance of DPNC method for varying SNR, varying bandwidth B, varying number of samples S, and varying phase difference.

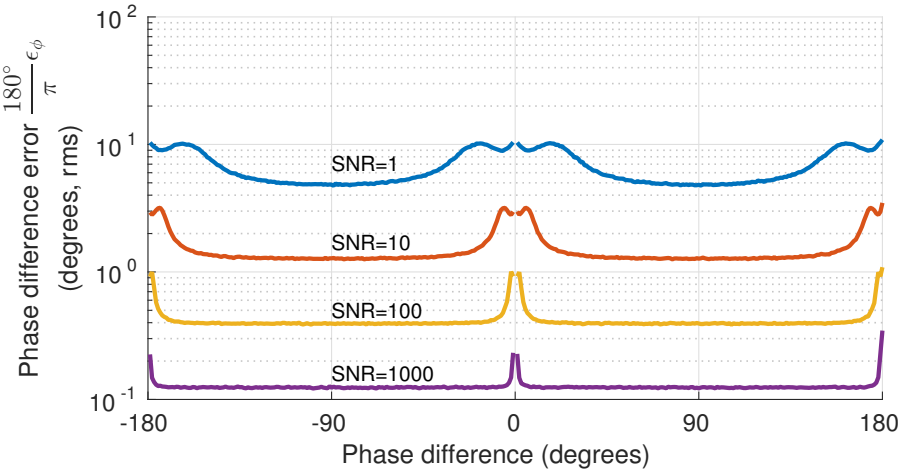


Figure 4.7: Performance of DPNC method for varying phase differences and varying SNR, with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=0.043.

The performance of the DPNC method at various phase differences is shown in Figure 4.7. Like the DP method, the DPNC method shows a similar increased

sensitivity to noise at certain phase differences. This is to be expected as both methods use the inverse cosine in the final step of calculating the phase difference. Notice, however, that there is much less deviation between the maximum and minimum in Figure 4.7 than in Figure 4.4. The shape of the graph is also more favourable as only small ranges are very sensitive to amplitude noise, whereas in Figure 4.4 only small ranges are *not* very sensitive to amplitude noise.

Figure 4.8 shows the error distribution of the DPNC method. This time the distribution is closer to normal. As with the DP method, a correlation exists with Figure 4.7, such that the majority of the errors are small and close to 0, though larger errors are also present despite being less frequent. The distribution in (f) suggests that as the SNR increases, the distribution would tend to become normal. Figure 4.7 supports this, showing an ever-flattening graph as the SNR increases.

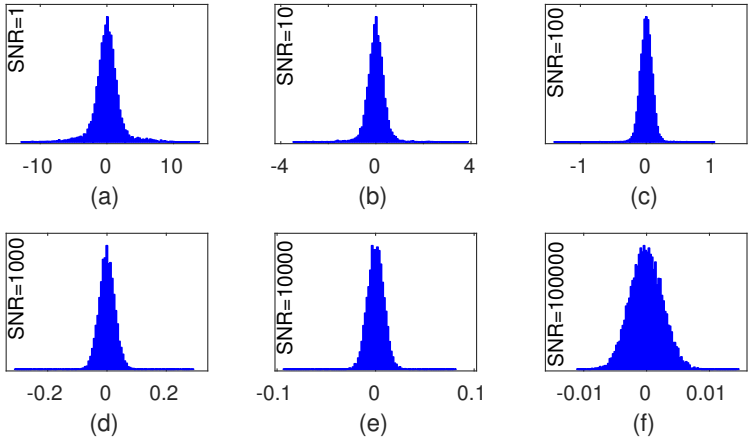


Figure 4.8: Error distribution in degrees of DPNC method for varying SNR in (a) – (f), with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=none (no filtering).

Recall that the DPNC method requires the mean and variance of the noise to be known beforehand, as the knowledge and integration of these numbers is the very thing that separates the DPNC method from the DP method. Figure 4.9 shows how an error in the estimation of the noise standard deviation affects the result. Five graphs are shown: one where the variance is estimated without error, and four with an increasing error. The figure shows that the performance of the method rapidly decays when even a small error is made in the noise estimation. The sensitivity of the DPNC method to the accuracy of the noise estimation is an undesirable property making the method less robust.

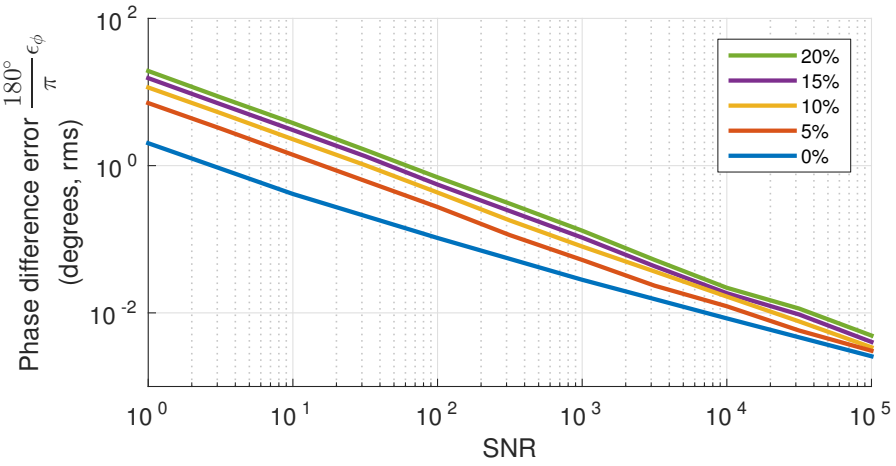


Figure 4.9: Influence of noise estimate errors on performance of DPNC method for varying SNR, with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=none (no filtering).

4.1.4 Hilbert Transform (HT)

The performance of the HT method for varying SNR, varying bandwidth, and varying number of samples is shown in Figure 4.10. Once again, as expected, the error decreases as the SNR increases. As with the DPNC method, it is clear that an increase in the number of samples (– – –, ·····) as well as a wider bandwidth (while maintaining the same SNR, thus lower PSD) (——, ——) both lead to a smaller error.

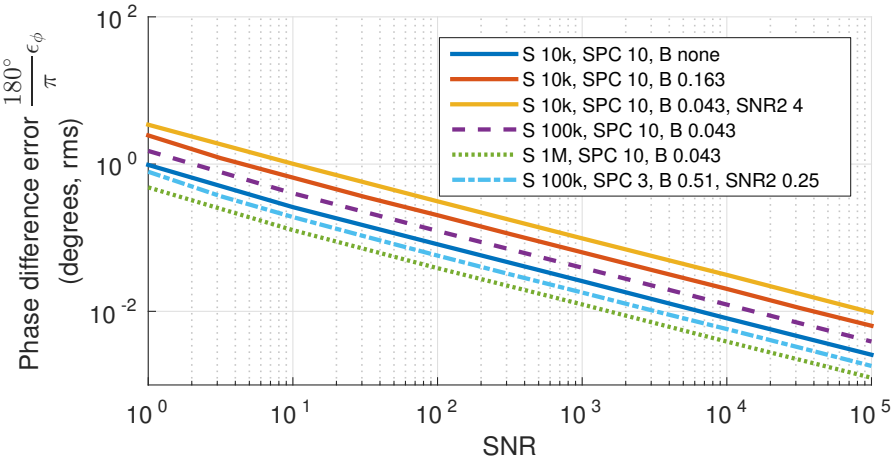


Figure 4.10: Performance of HT method for varying SNR, varying bandwidth B, varying number of samples S, and varying phase difference.

The performance of the HT method is approximated by [105]

$$\epsilon_\phi = \sqrt{\frac{1}{S}\left(\frac{1}{B_1(\mathrm{SNR}_1)} + \frac{1}{B_2(\mathrm{SNR}_2)}\right)}, \tag{4.4}$$

where ϵ_ϕ is the RMS error in radians, S is the number of samples, B_1 and B_2 is the equivalent normalised bandwidth, and SNR_1 and SNR_2 is the SNR of the first and second signal, respectively. This equation is used later when the overall performance of the system is estimated using the performance of individual subsystems.

The accuracy with which the approximation in Equation 4.4 predicts the results in Figure 4.10 is shown in Figure 4.11. The figure graphs the equation

$$\mathrm{Error}\% = \frac{\epsilon_\phi - \epsilon'_\phi}{\epsilon_\phi} \times 100\%,$$

where ϵ_ϕ and ϵ'_ϕ denote the simulated and predicted phase difference errors, respectively. When the SNR is above 10^2, the approximation error is below 2%, proving that the approximation is a very accurate representation of the actual results. Keep in mind that the approximation is simply used when predicting the performance of the system and does not in any way influence the actual performance of the system.

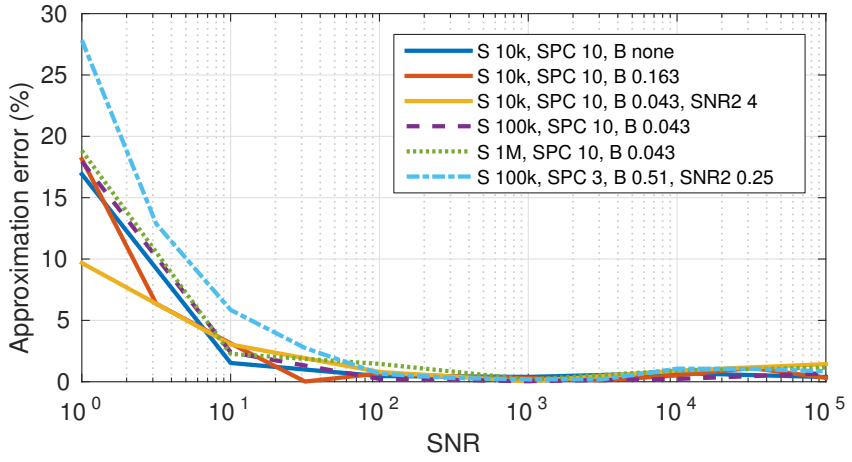


Figure 4.11: Error in HT method analytically predicted performance for varying SNR, varying bandwidth B, and varying number of samples S.

Figure 4.12 shows the performance of the HT method at various phase differences. Unlike the other methods, the accuracy of the HT method is not dependent on the phase difference. This is a desirable property and means that the RMS errors depicted in Figure 4.10 are a very true representation of the expected performance of the method.

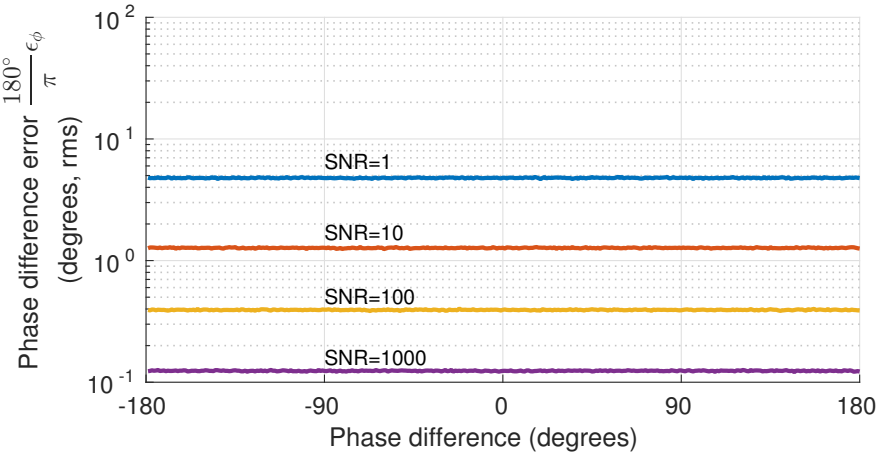


Figure 4.12: Performance of HT method for varying phase differences and varying SNR, with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=0.043.

The error distribution of the HT method is shown in Figure 4.13. The figure clearly shows that a normal distribution for input noise also results in a normal distribution for the resulting error. This correlates with Figure 4.12 which shows that the method does not tend to amplify errors under certain conditions. The HT method also does not make use of the inverse cosine function in calculating the phase difference, which preserves the shape of the noise distribution.

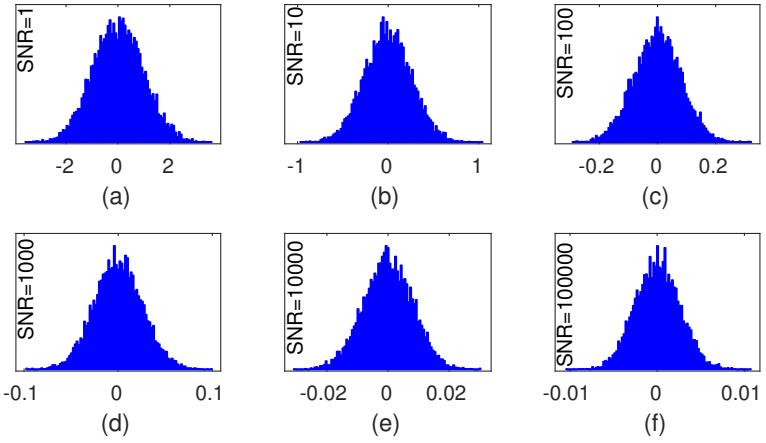


Figure 4.13: Error distribution in degrees of HT method for varying SNR in (a) – (f), with number of samples S=10 k, samples per cycle SPC=10, bandwidth B=none (no filtering).

4.1.5 Comparison of Methods

The relative performance of the DP, DPNC and HT methods is shown in Figure 4.14 by plotting corresponding graphs from Figure 4.3, Figure 4.6 and Figure 4.10 on the same axes. The increased performance of the DPNC method over the DP method is clear. The figure also shows little difference between the performance of the DPNC and HT methods, though other important differences are discussed later.

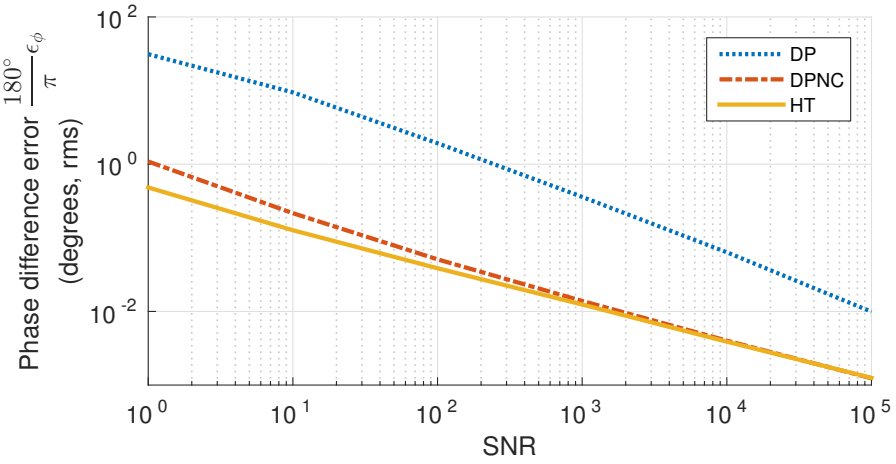


Figure 4.14: Comparison of methods for varying SNR, with number of samples S=1 M, samples per cycle SPC=10, bandwidth B=0.043, and varying phase difference.

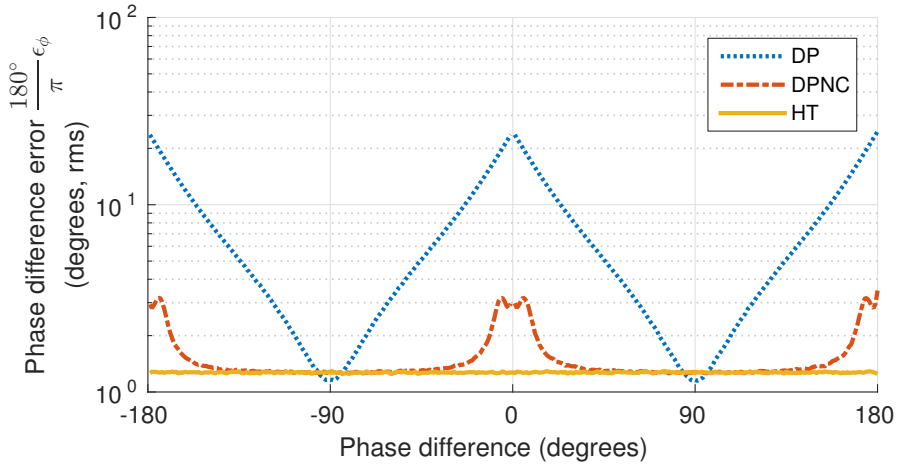


Figure 4.15: Comparison of methods for varying phase differences, with SNR=10, number of samples S=10 k, samples per cycle SPC=10, bandwidth B=0.043.

Figure 4.15 compares the sensitivities of the three methods to varying phase differences. This reaffirms the fact that the performance of the DP method

varies greatly as the phase difference varies. It also shows how the performance of the DPNC method for the most part approaches that of the HT method, except at phase differences of $0\,^\circ$ and $\pm 180\,^\circ$, where it is more sensitive to noise. This confirms the small performance difference between these two methods shown in Figure 4.14.

From the error distributions in Figure 4.5, Figure 4.8 and Figure 4.13 it can be seen that all methods produce a zero-mean error for the zero-mean input noise. This means that the standard deviation σ_ϕ of the error is equal to the RMS ϵ_ϕ of the error.

While the DPNC and HT methods may seem similar in performance, they are separated by two important differences. Firstly, the performance of the DPNC method rapidly decays when the noise estimation contains errors, whereas the HT method does not require any noise characteristics to be known. Secondly, the DPNC method is more sensitive to noise at certain phase differences, whereas performance of the HT method is not influenced by phase difference. For these reasons the HT method is deemed superior to the other methods.

4.2 Translating Phase Difference to Distance

The distance error ϵ_Δ is easily calculated once the phase difference error ϵ_ϕ is known. Recall from Equation 3.25 that the measured phase difference $\phi_j - \phi_i$ is translated to a distance difference $d_j - d_i$ by essentially multiplying $\phi_j - \phi_i$ by $\lambda/2\pi$. This means that the distance difference error is obtained by simply scaling the phase difference error appropriately, so that

$$\epsilon_\Delta = \frac{\lambda}{2\pi}\epsilon_\phi. \tag{4.5}$$

Notice how the distance difference error is directly proportional to the wavelength. This means that a smaller error can be obtained by transmitting a shorter wavelength, but this, however, results in more possibilities for the value of n in Equation 3.26, such that the approximate position of the heliostat must be more accurately known. One way to work around this issue is to transmit multiple sinusoids with different wavelengths, as mentioned earlier.

4.3 Transmitter Layout

The placement of transmitter towers plays an important part in the overall accuracy of the system. Poor transmitter placement can result in large receiver position errors even when distance measurement errors are small. To understand the effect of transmitter placement on the overall accuracy, the problem is first investigated in two dimensions before moving to three dimensions. Three

factors influencing the position errors are considered: transmitter geometry, number of transmitters, and the distance measurement errors.

4.3.1 Methodology

Transmitter layouts are evaluated by means of a Monte Carlo analysis. All simulations assume a circular heliostat field within which all transmitters and receivers are placed. The dimensions of the field are normalised such that it always has a radius of unity. Using normalised units allows the results of the simulations to be applicable to any arbitrarily sized field by scaling appropriately. The purpose of each simulation is to find the expected RMS error as a function of the distance from the centre of the field.

The heliostat field in the simulation consists of two virtual areas, namely a transmitter area and a receiver area, which designate where transmitters and receivers may occur. Both areas are constrained within the unit circle on the xy-plane, with $z = 0$ in two dimensions. In three dimensions, minimum and maximum height constraints are also introduced to both areas. This ensures that both areas are above ground level and can only extend to heights that are practically achievable.

Each simulation analyses the performance of a particular transmitter layout. Transmitters are placed within the transmitter area according to the layout to analyse after which a receiver is placed at a random location uniformly distributed within the receiver area. The distance from the receiver to each transmitter is measured and noise from a normal distribution with the desired magnitude is added to each measurement, after which the noisy measurements are passed to the positioning algorithm. The position error is the distance between the actual and calculated positions of the receiver. To analyse the expected error at all locations in the field and to ensure convergence of results, this process is repeated for 100,000 different receiver positions. Algorithm 2 shows the pseudocode for the core of the simulation.

Algorithm 2 Performance analysis of transmitter layout

```
txs ← list of transmitter coordinates according to layout
rx ← random receiver coordinates uniformly distributed within receiver area
absdist ← empty list
reldist ← empty list
for all t, tx in txs do
    absdist(t) ← distance from rx to tx
    absdist(t) ← absdist(t) + random value from normal distribution
    reldist(t) ← absdist(t) − absdist(0)
rxcalc ← calculated position using txs and reldist
error ← distance between rxcalc and rx
```

4.3.2 Two Dimensions

The purpose of the two-dimensional analysis is to gain an understanding of the nature of the problem of transmitter layout more so than producing quantitative results. The problem is first investigated in its simplest form which is the one with the least number of transmitters, from where the complexity is incrementally increased. Recall from Section 3.5 that the number of transmitters $N \geq D + 1$ when using absolute measurements and $N \geq D + 2$ when using relative measurements, where D is the number of dimensions.

Figure 4.16 shows three different transmitter layouts using absolute distance measurements for $N = 3$ transmitters laid out in (a) equilateral, (b) acute isosceles and (c) obtuse isosceles triangular configurations, with the bottom images representing zoomed-in views of the top images. A normalised distance measurement RMS error of $\epsilon_\Delta = 0.01$ is used which is deliberately large to illustrate the effect of distance errors on position errors. Transmitters are represented by the black dots and the colour map represents the RMS error at a specific field location, with blue through yellow representing an increase in error.

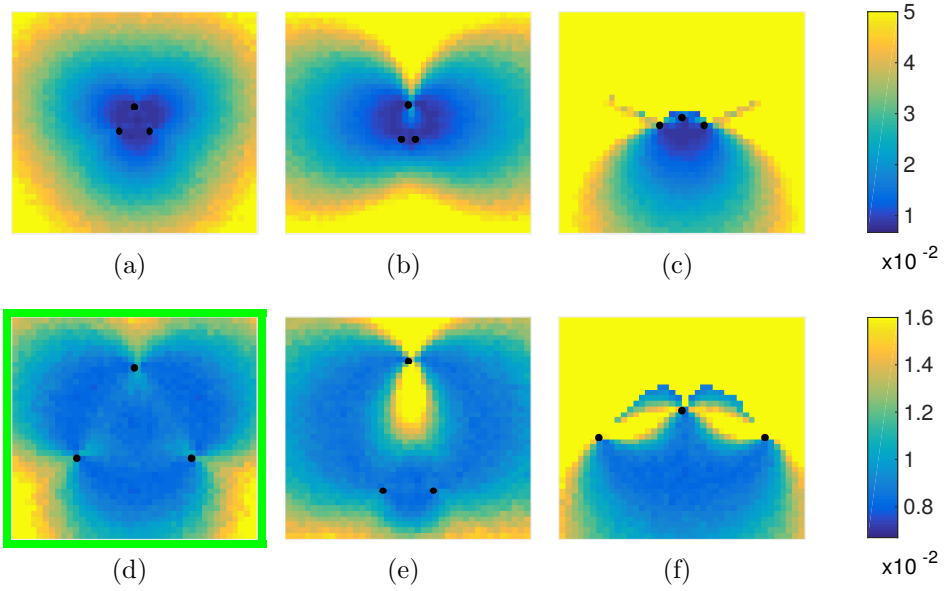


Figure 4.16: Error sensitivity of transmitter layout in 2D with 3 TX and absolute distances laid out in (a) equilateral, (b) acute isosceles and (c) obtuse isosceles triangular configurations, with (d) – (f) representing zoomed-in views of (a) – (c), respectively. The largest even distribution of low error areas is shown in (d).

All layouts show a small error in the proximity of the transmitters and an ever increasing error outwards. This suggests that for an optimal layout the transmitters must surround the field as opposed to being located at the centre.

In (e) and (f), high error areas are seen amidst the low error areas (yellow inside blue). This is due to an ambiguity existing in those areas when noise is present, as demonstrated previously in Figure 3.6. It is clear that the largest even distribution of low error areas is achieved using an equilateral triangle configuration in (d).

In Figure 4.17 the number of transmitters is increased to $N = 4$. Transmitters are laid out in (a) square, (b) rectangular and (c) centred triangular configurations. Again the lowest error is in the proximity of the transmitters with errors increasing outwards, which reinforces the observation of Figure 4.16. All layouts show a relatively even distribution of low error areas, however the largest even distribution of low error areas is clearly seen in the square configuration in (d).

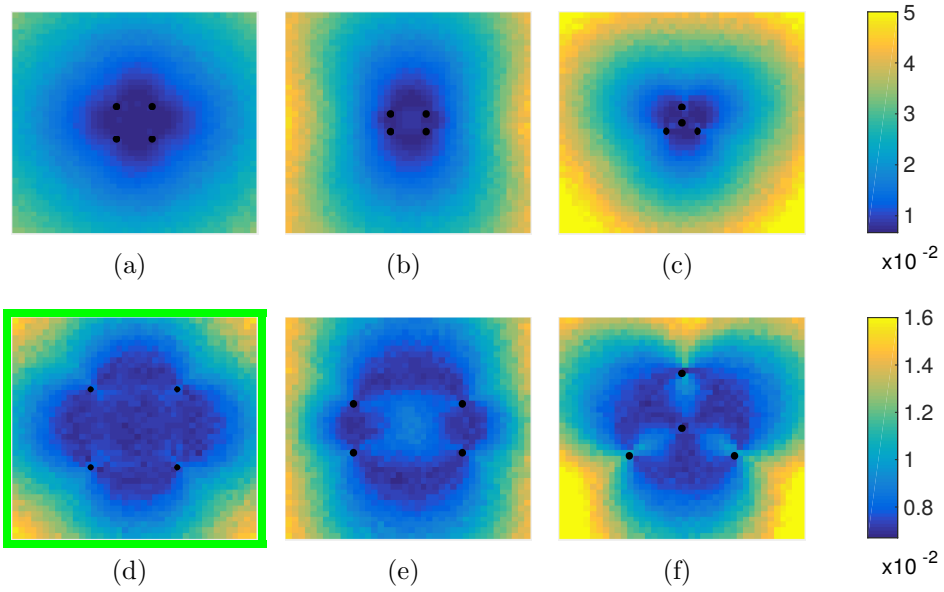


Figure 4.17: Error sensitivity of transmitter layout in 2D with 4 TX and absolute distances laid out in (a) square, (b) rectangular and (c) centred triangular configurations, with (d)–(f) representing zoomed-in views of (a)–(c), respectively. The largest even distribution of low error areas is shown in (d).

Figure 4.18 shows the same configurations as Figure 4.17, only with relative distance measurements. With absolute distances, the position error slowly increases as the position is moved away from the proximity of the transmitters, however with relative distances the error drastically increases the moment the position is outside the area enclosed by the transmitters. This observation necessitates the transmitters to be laid out such that they surround the field, since positions outside the area enclosed by the transmitters are extremely sensitive to distance measurement errors. As with absolute distances, the largest distribution of low error areas is achieved using a square layout in (d).

The equilateral triangle and square layouts found to be optimal in the previous figures can be generalised and viewed as transmitters equally distributed on the edge of the unit circle of the field. Figure 4.19 shows this configuration for $5 \leq N \leq 8$ using absolute distances in the top images and relative distances in the bottom images, with all images using the same colour scale. The figure illustrates how an increase in the number of transmitters reduces the overall error as the blue areas become darker from (a) – (d) and (e) – (h).

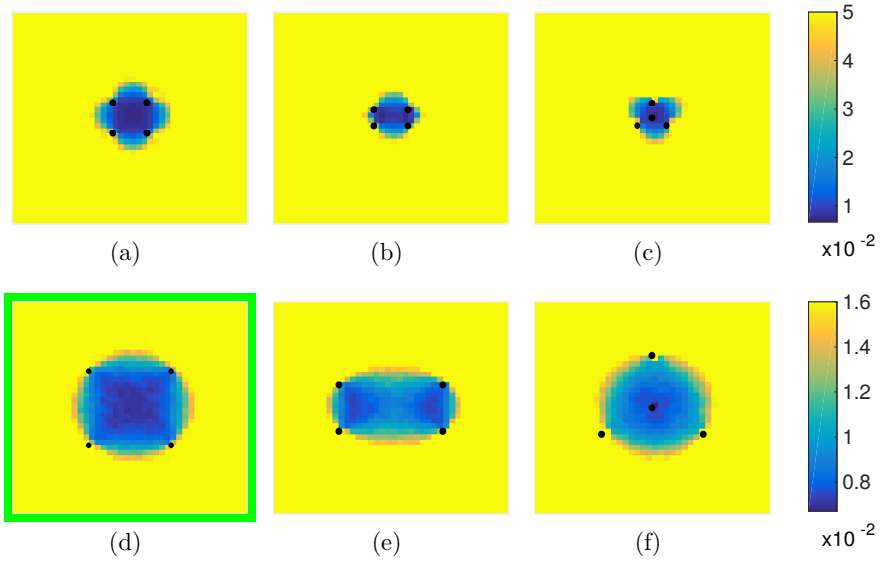


Figure 4.18: Error sensitivity of transmitter layout in 2D with 4 TX and relative distances laid out in (a) square, (b) rectangular and (c) centred triangular configurations, with (d) – (f) representing zoomed-in views of (a) – (c), respectively. The largest even distribution of low error areas is shown in (d).

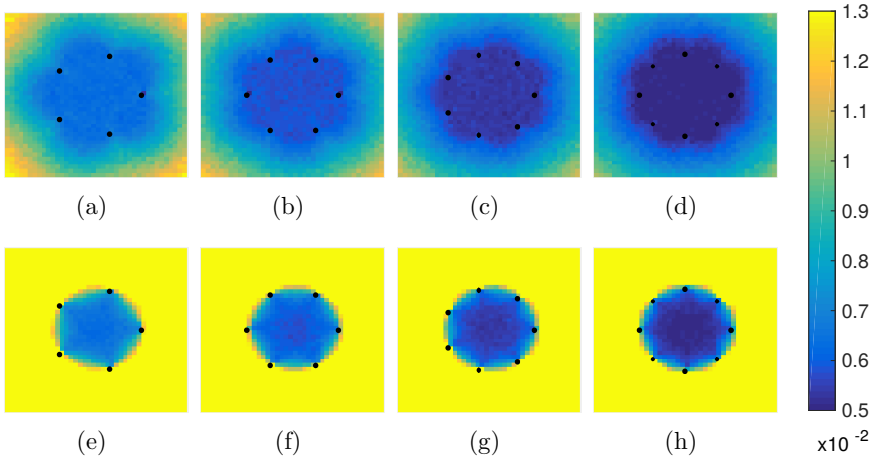


Figure 4.19: Error sensitivity of transmitter layout in 2D with 5 – 8 TX comparing absolute and relative distances in the top and bottom images, respectively.

Figure 4.20 shows the position RMS error as a function of the distance from the centre of the field for varying distance errors and varying number of transmitters using absolute distances. Figure 4.21 is similar, only using relative distances. The top group of graphs, where $\epsilon_\Delta = 10^{-2}$, can be seen to correlate with the results in Figure 4.19. In both cases, it can be seen that the lowest error occurs at the centre of the field and that the error increases as the receiver is moved towards the edge. Figure 4.21 quantitatively shows how the error rapidly increases once the receiver is outside the field perimeter, confirming the earlier observation.

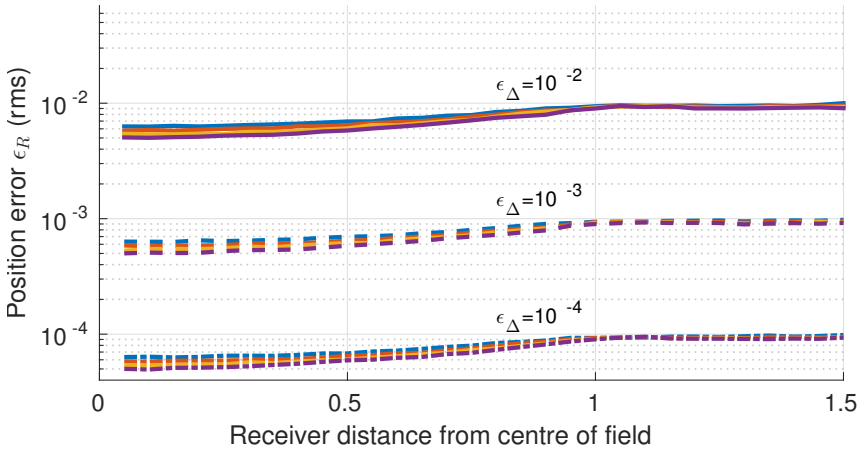


Figure 4.20: Receiver position error in 2D as affected by number of transmitters and error on absolute distances. The number of transmitters is 5 (blue), 6 (red), 7 (orange) and 8 (purple).

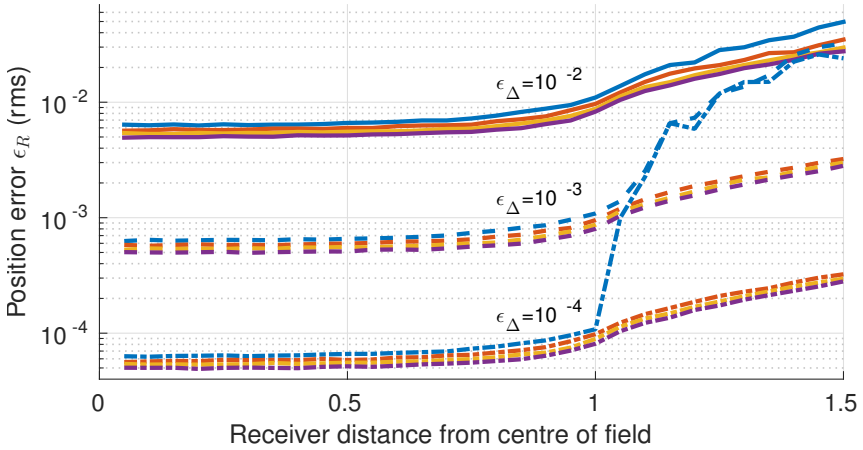
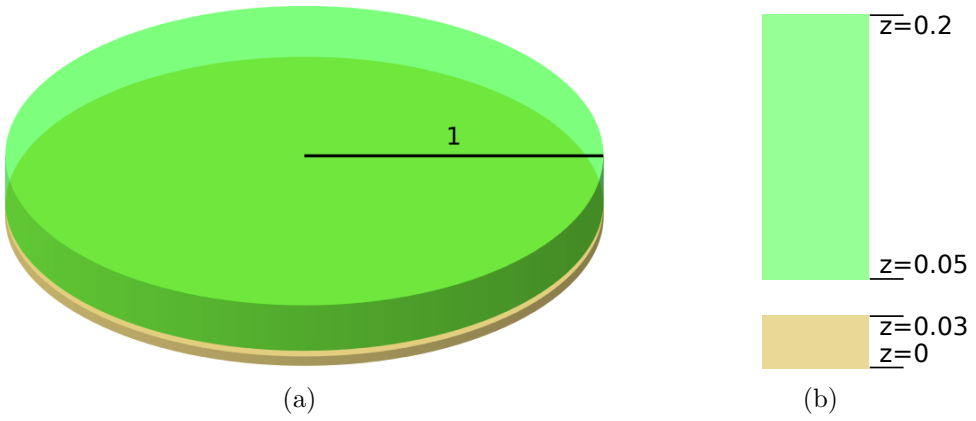


Figure 4.21: Receiver position error in 2D as affected by number of transmitters and error on relative distances. The number of transmitters is 5 (blue), 6 (red), 7 (orange) and 8 (purple).

The 2D analysis reveals that both a decrease in distance errors ϵ_Δ and an increase in the number of transmitter N lead to a lower position error ϵ_R. It also shows the importance of transmitter layout and how it affects the overall error. Furthermore, it shows how the smallest errors occur inside the area enclosed by the transmitters, especially when using relative distances. This necessitates the transmitters to be laid out such that they surround the field.

4.3.3 Three Dimensions

In two dimensions, transmitters and receivers occupy the same area constrained within the unit circle on the xy-plane with $z = 0$. When considering a practical field in three dimensions, heliostats (and therefore receivers) are typically located near ground level, extending only a few metres upward. Additionally, it may be beneficial to ensure that transmitters are slightly elevated above the receivers to prevent blocking line of sight (LoS) from neighbouring heliostats. The receiver and transmitter areas are shown in Figure 4.22 along with the normalised constraints of each area. The vertical constraints are chosen such that they are realistic for a field with a radius of 100 m, that is, receivers are located between 0 m and 3 m, and transmitters are slightly elevated and located between 5 m and 20 m. Note that even though the two areas are both circular discs in this case, they can take on any shape required.



(a) (b)

Figure 4.22: Normalised field dimensions in 3D showing transmitter area in green and receiver area in brown, along with height constraints of each area.

Due to the added complexity, an algorithm is developed to aid with the layout of transmitters in three dimensions. It starts by placing all N transmitters arbitrarily inside the transmitter area. Then, it greedily[1] moves the transmitters around inside the transmitter area until the lowest RMS error is achieved for

[1]Greedy algorithms focus on making the locally optimal choice at each decision point with the hope of finding the global optimum [106].

receivers uniformly distributed throughout the receiver area. The distance transmitters are moved, here called the *step* size, is initially large and then decreased until the desired resolution is achieved. In this case, the step sizes are chosen as $step = \{0.2, 0.05, 0.01\}$, for a maximum resolution of 0.01. The greedy nature of the algorithm may cause it to encounter suboptimal local minima. For this reason the algorithm is run multiple times with different initial transmitter positions and when the same result is achieved multiple times the validity of the result is confirmed. Algorithm 3 shows the pseudocode for this process.

Algorithm 3 Optimal transmitter layout

```
txs ← list of arbitrary transmitter tower coordinates
steps ← list of step sizes decreasing in magnitude
error ← rms error for current txs layout
for all s, step in steps do
    change ← true
    while change do
        change ← false
        for all t, tx in txs do
            newtxs ← list of neighbouring coordinates step distance from tx
            for all n, newtx in newtxs do
                curerror ← rms error as if current tx was located at newtx
                if curerror < error then
                    error ← curerror
                    txs(t) ← newtx
                    change ← true
```

The results obtained by running the transmitter layout optimisation algorithm are shown in Figure 4.23 for $5 \leq N \leq 10$. The exact results of the algorithm are shown – no manual adjustments were made. Each image represents a top view of the circular field, with the dots representing transmitter locations. Interestingly enough, transmitters are always located at either the minimum or maximum allowed height, as indicated by the blue and red dots, respectively. Furthermore, one might expect the results to show an arbitrary distribution of transmitters, however a predictable pattern emerges. The exact coordinates for each layout is shown in Appendix B.

The position RMS error map of each layout in Figure 4.23 is shown in Figure 4.24. As with observations in 2D, it is clear that the best distribution of low error areas once again occurs within the perimeter of the area enclosed by the transmitters. Even though relative distances are used, the error does not increase as rapidly once the receiver is outside the field perimeter. At first it may seem as if this contradicts the 2D observations, but keep in mind that in 3D the transmitters form a sphere, or more specifically a flattened dome, over

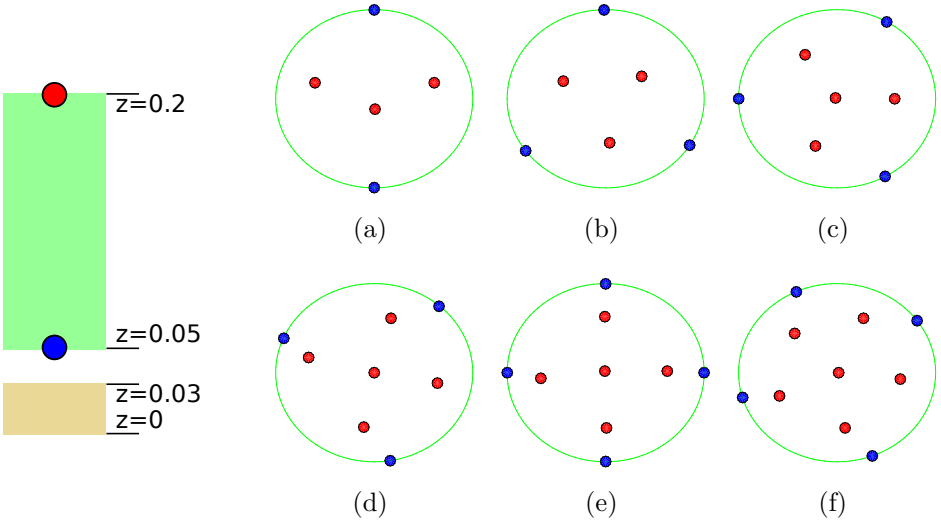


Figure 4.23: Optimal transmitter layouts in 3D with 5 – 10 TX in (a) – (f) as viewed from top with the field perimeter indicated by the green circle. Transmitters are located either at the top (red) or bottom (blue) of the transmitter area.

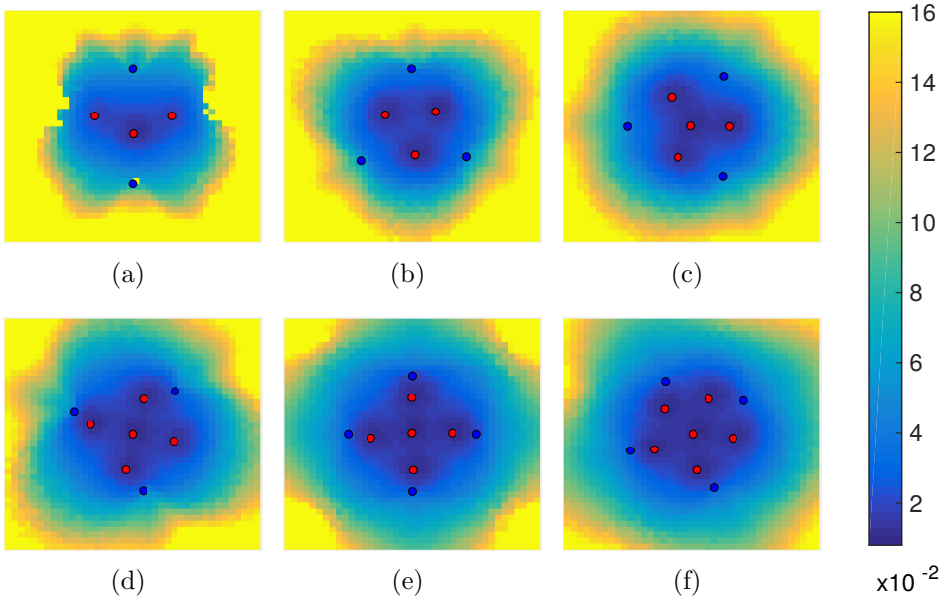


Figure 4.24: Error sensitivity of transmitter layout in 3D with 5 – 10 TX in (a) – (f) and relative distances, corresponding to the optimal layouts in Figure 4.23.

the field, so the area enclosed by the transmitters is much larger than the field itself.

Figure 4.25 shows the position RMS error as a function of the distance from the centre of the field for the transmitter layouts of (c) – (f) in Figure 4.23 and varying distance errors. To prevent the figure from being cluttered, (a) and (b)

are omitted. Just as before, the top group of graphs, where $\epsilon_\Delta = 10^{-2}$, can be seen to correlate with the results of (c) – (f) in Figure 4.24. The figure also shows how the lowest errors occur near the centre of the field, and how the error increases in magnitude and tempo as the distance from the centre of the field increases.

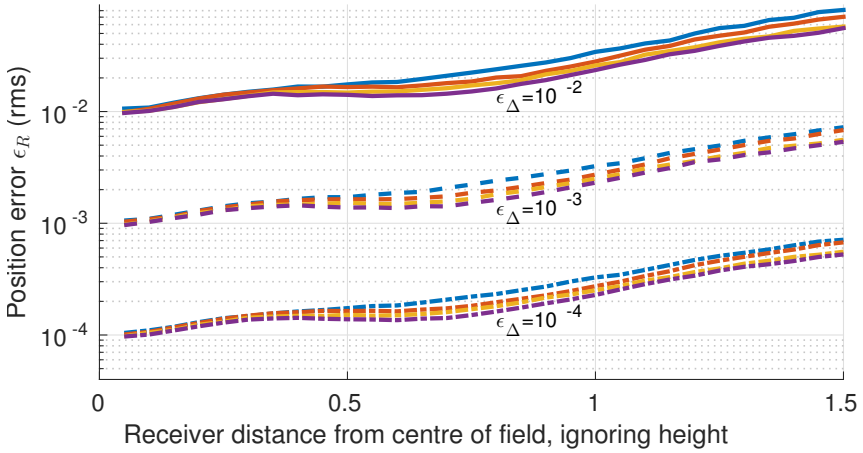


Figure 4.25: Receiver position error in 3D as affected by number of transmitters and error on relative distances. The number of transmitters is 7 (blue), 8 (red), 9 (orange) and 10 (purple), corresponding to the optimal layouts in Figure 4.23.

4.4 Position Calculation

Recall from Section 2.4 that Murphy suggests seeding the non-linear least squares (NLLSQ) algorithm with the results of the linear least squares (LLSQ) estimation as opposed to the centre (origin) of the field for faster computation time. The first part of this section investigates the difference between a LLSQ seed and an origin seed and it is found that in some cases the LLSQ seed causes the algorithm to get stuck in a local minimum yielding an erroneous position. The second part of the section presents a normalised version of the results from the previous section which can be used in estimating the overall system performance.

4.4.1 Seeding with LLSQ vs. Seeding with Origin

When the LLSQ seed results in an error greater than that of the origin seed by more than one percent, then the result of the LLSQ seed is deemed erroneous. Table 4.3 shows the results of the comparison for the optimal layouts in Figure 4.23 with varying number of transmitters and varying distance errors, using both absolute and relative distances. The results are representative of 1,000,000 simulation runs for receivers uniformly distributed within the field,

comparing the runtime of the two seeds, as well as the number of erroneous results yielded by the LLSQ seed.

Table 4.3: Comparison of LLSQ seed vs. origin seed for position calculation with varying number of transmitters corresponding to the optimal layouts in Figure 4.23 and varying distance errors, using both absolute and relative distances. The runtime of the LLSQ seed vs. the origin seed is shown as a percentage, and the number of errors out of 1 M simulation runs is also shown.

# TX	Distance errors ϵ_Δ	Absolute or Relative distances	LLSQ vs. origin seed runtime %	Errors	Errors where LLSQ seed is further than d from field centre			
					$d = 0.5$	$d = 0.6$	$d = 0.9$	$d = 1$
7	10^{-2}	R	77	3,107	3,079	2,937	1,464	814
		A	65	2,362	2,362	2,358	1,630	140
	10^{-4}	R	60	0	0	0	0	0
		A	49	0	0	0	0	0
8	10^{-2}	R	76	855	853	851	515	323
		A	65	520	520	520	483	68
	10^{-4}	R	59	0	0	0	0	0
		A	49	0	0	0	0	0
9	10^{-2}	R	76	504	504	498	277	208
		A	64	333	333	333	331	53
	10^{-4}	R	59	0	0	0	0	0
		A	48	0	0	0	0	0
10	10^{-2}	R	76	340	340	333	145	78
		A	64	563	563	563	557	65
	10^{-4}	R	59	0	0	0	0	0
		A	49	0	0	0	0	0

The results show that the LLSQ seed causes the algorithm to complete in 48 % – 77 % of the time required by an origin seed, with less time required as the distance error decreases. The trade-off, however, is that the LLSQ seed sometimes causes erroneous results, though only with larger distance errors. The results also show that the seed does not need to be at the edge or outside the field to cause an error, which removes the option of falling back to the origin seed when the LLSQ seed is too close to the edge. Even though the percentage of errors is small ($\sim 0.3\,\%$), any error whatsoever is undesirable and the 23 % – 52 % gain in computation time does not merit the use of the LLSQ seed over seeding with the origin.

4.4.2 Normalised Position Calculation Results

To quantify the performance of the position calculation a more generic version of Figure 4.25 is required. This is done in Figure 4.26 by normalising the graphs in Figure 4.25 with regards to the distance error ϵ_Δ. The dotted lines in Figure 4.26 represent the normalised graphs of Figure 4.25 which, once normalised, are seen to be very similar. The solid line represents the RMS of the corresponding dotted lines. This allows the position error ϵ_R to be calculated by simply multiplying the distance error ϵ_Δ with the graph ϵ_{Rr}, or

$$\epsilon_R = \epsilon_{Rr}\epsilon_\Delta. \tag{4.6}$$

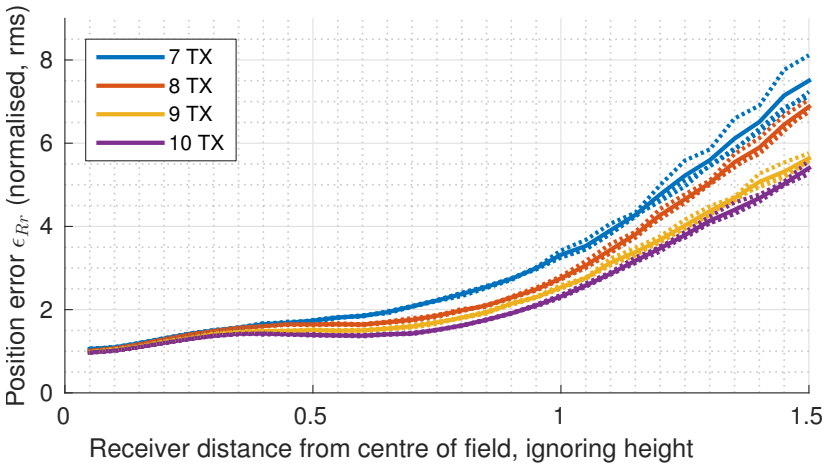


Figure 4.26: Normalised receiver position error in 3D for relative distances as affected by number of transmitters optimally arranged as in Figure 4.23.

Further inspection reveals that the z component of the position error varies in magnitude in comparison to the magnitude of the x and y components, which are mostly similar. This is in agreement with the findings of Murphy, who also observed that the z component generally contains larger errors than the x and y components [93]. The factor by which the z component is larger than the x and y components is shown in Figure 4.27. The significance of this becomes apparent later when predicting the overall system performance.

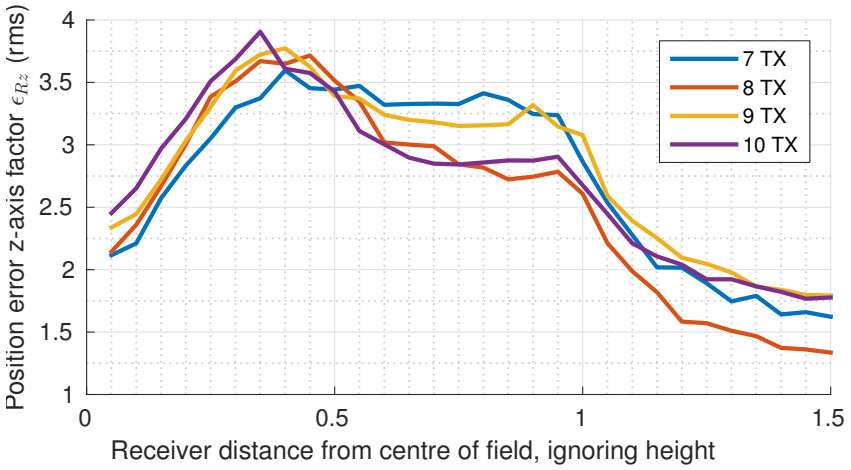


Figure 4.27: Factor by which position error z component is greater than x and y components for varying number of transmitters optimally arranged as in Figure 4.23.

4.5 Receiver Plane Normal Vector

This section investigates how the accuracy of the receiver plane normal vector is influenced by errors in the receiver positions ϵ_R. Recall that when the receiver plane is parallel to the mirror surface, its normal vector is equal to the heliostat normal vector, hence the receiver plane normal error is equal to the heliostat tracking error. Analysis is done for three receivers, spaced out evenly over the mirror surface so as to form an equilateral triangle. The distance between the receivers d_R is referred to as the inter-receiver distance, shown in Figure 4.28.

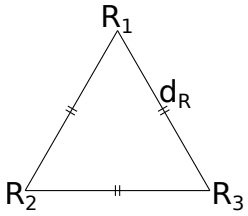


Figure 4.28: Inter-receiver distance.

Figure 4.29 shows how the tracking error is influenced by errors on the receiver positions. It is clear that the same receiver position error ϵ_R can have different effects on the tracking error ϵ_T depending on the inter-receiver distance d_R. For this reason, ϵ_T is analysed as a function of both ϵ_R and d_R.

Figure 4.30 shows the tracking error as a function of the receiver error and inter-receiver distance. The relationship d_R/ϵ_R is shown on the x-axis with the tracking error on the y-axis. The figure shows clearly that the tracking error decreases as d_R/ϵ_R increases, that is, as either the position error decreases or

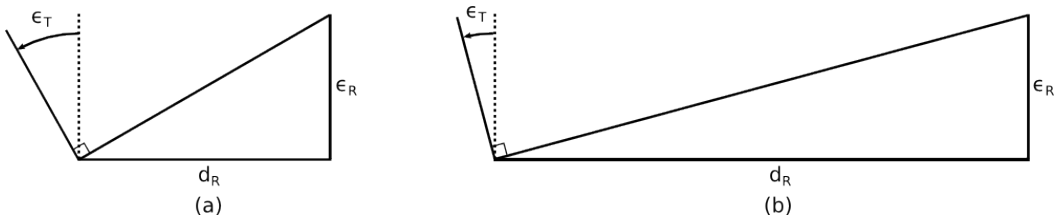


Figure 4.29: Tracking error as affected by receiver position errors and inter-receiver distance

the receivers are placed a greater distance apart. From Figure 4.30, the RMS of the tracking error ϵ_T is approximated by

$$\epsilon_T = \sqrt{4/3}\frac{\epsilon_R}{d_R} \tag{4.7}$$

in orange (——), where ϵ_T represents the overall tracking accuracy of the system, in radians. For a tracking error of less than one milliradian, a d_R/ϵ_R relationship of above $\sim$ 1,150 is required.

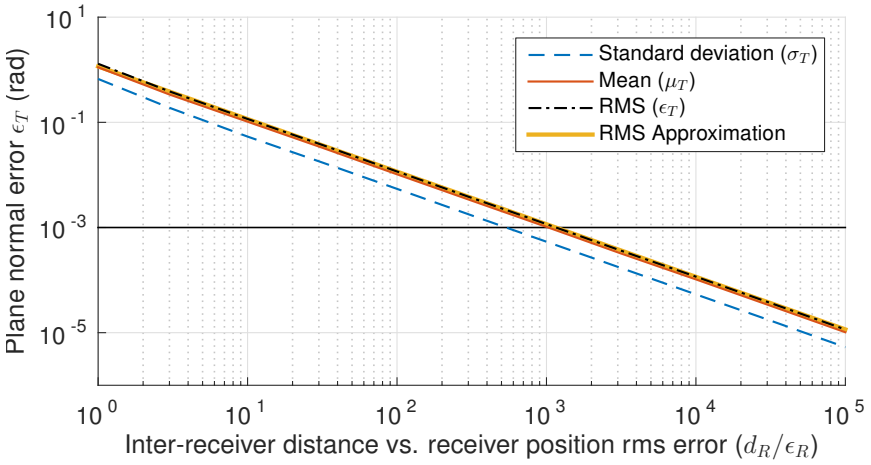


Figure 4.30: Receiver plane normal error as a function of receiver position error and inter-receiver distance.

The results assume that receiver position errors are uncorrelated, though it can be shown that a positive correlation leads to a smaller RMS tracking error ϵ_T for the same d_R/ϵ_R. This is easily argued: if all receiver positions contain an error equal in direction and magnitude, then the resulting plane orientation remains unaffected. Additionally, if the error on all receiver positions is such that they remain on the original plane, then the plane orientation also remains unaffected.

The last point represents a significant opportunity with regards to the hardware implementation of the receiver. Stated otherwise, it means that only the position

error component perpendicular to the mirror surface affects the tracking error. This allows the shape of the receiver to extend much further along the plane parallel to the mirror surface as long as it is constrained in the perpendicular dimension, as shown in Figure 4.31. Possible implementations include antennas etched on printed circuit board (PCB) or flat optical sensors.

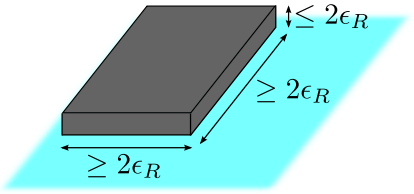


Figure 4.31: Position errors only constrain the height of the receiver since errors parallel to the mirror surface have practically no effect on the tracking error.

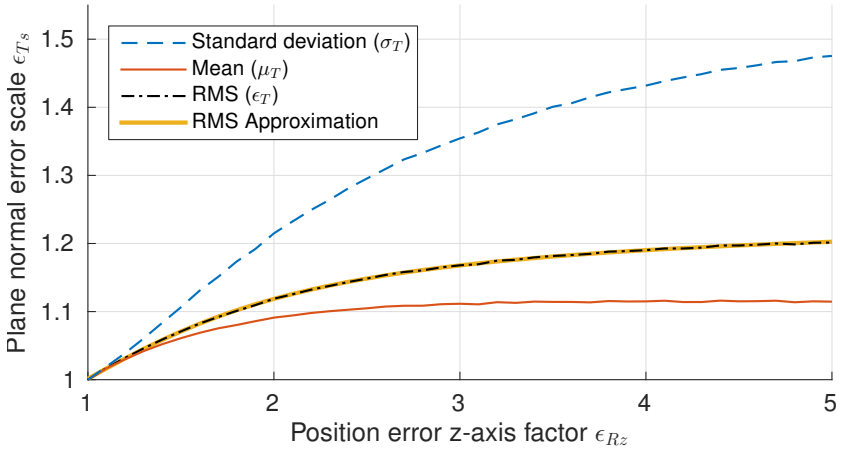


Figure 4.32: Plane normal error scale factor for varying position error z component factors.

When the z component of the position error varies in magnitude with respect to the x and y components – even while the combined magnitude remains constant – the plane normal error also changes. Figure 4.32 shows how the plane normal error scales when the magnitude of the z component factor is varied. The RMS value is approximated in the orange graph (——) by

$$\epsilon_{Ts} = a \arctan(b\epsilon_{Rz} - c) + d, \tag{4.8}$$

where the parameters a – b are obtained through curve fitting as

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0.16014 \\ 0.95672 \\ 0.89325 \\ 0.99126 \end{bmatrix}, \tag{4.9}$$

and thus the new expression for ϵ_T becomes

$$\epsilon_T = \sqrt{4/3}\frac{\epsilon_R}{d_R}\epsilon_{Ts}. \tag{4.10}$$

4.6 Overall Performance

In this section all the individual subsystems are integrated and the system is simulated as a whole. The accuracy with which the performance of the full system can be predicted when only the performance of individual subsystems is known is also investigated. Algorithm 4 shows the pseudocode for the core of the simulation. Some detail is omitted where steps have already been described in previous algorithms.

Algorithm 4 Performance analysis of full system

```
txs ← list of transmitter coordinates according to layout
elecpshift ← list of electronic phase shifts from source to each txs
d_R ← required inter-receiver distance
hstat ← random heliostat coordinates uniformly distributed in receiver area
zenith ← random value uniformly distributed in [0, π/2]
azimuth ← random value uniformly distributed in [0, 2π)
normal ← heliostat normal vector according to zenith and azimuth
plane ← mathematical plane perpendicular to normal
rxs ← coordinates of 3 receivers d_R apart on plane centred around hstat
rxscalc ← empty list
for all r, rx in rxs do
    signals ← empty list
    phasediffs ← empty list
    for all t, tx in txs do
        dist ← distance from rx to tx
        distpshift ← phase shift corresponding with dist
        signal ← sinusoid with phase distpshift + elecpshift(t)
        noise ← normally distributed noise for desired SNR
        signals(t) ← signal + noise
        phasediffs(t) ← phase difference between signals(t) and signals(0)
    distdiffs ← relative distances from phase differences in phasediffs
    rxcalc(r) ← calculated position using txs and distdiffs
normalcalc ← calculated normal vector using rxscalc
error ← angle between normalcalc and normal
```

The normalised system parameters are listed in Table 4.4. Parameters are chosen to realistically scale to a field with a radius of 100 m as demonstrated later in the section. Recall from Section 4.1.4 that an increased sampling

frequency has no effect on the HT method, so the SPC value of 10 is arbitrarily chosen. A SNR of 10^5 is high but achievable. A wavelength of 4 is chosen such that n is always 0 in Equation 3.24. The number of transmitters is chosen to be double that of the minimum required amount, which is 5 for three dimensions and relative distances. Figure 4.33 shows how the inter-receiver distance is derived from an equilateral triangle with all corners 0.01 from the centre.

Table 4.4: Normalised system parameters.

Parameter	Value	Description
SPC	10	Samples per cycle
SNR_i	10^5	Signal-to-noise ratio
B_i	0.043	Equivalent normalised noise bandwidth
S	varied	Number of samples
N	10	Number of transmitters (layout in Figure 4.23)
r	1	Field radius
λ	4	Wavelength
d_R	0.0173	Inter-receiver distance

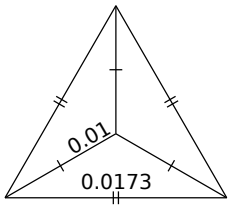


Figure 4.33: Normalised inter-receiver distance.

The simulated tracking error of the full system is shown by the solid lines in Figure 4.34. Keep in mind that the results may differ when the system parameters are adjusted, so they serve primarily as an indication of the order of the error that can be achieved. It can be seen that for the system parameters in Table 4.4 it is theoretically possible to obtain a tracking error in the order of one milliradian.

The dashed lines in Figure 4.34 represent the predicted system performance using the performance of the individual subsystems. The figure shows that the performance of individual subsystems can be used to predict the overall system performance with high accuracy. This is a desirable property and favourable towards a modular design. Figure 4.35 shows the error budget for the integrated system as well as the equations modelling the performance of each subsystem. These equations are combined to find the final equation for the tracking error as

$$\epsilon_T = \epsilon_{Ts}\sqrt{4/3}\frac{\epsilon_r}{d_R}\frac{\lambda}{2\pi}\sqrt{\frac{1}{S}\left(\frac{1}{B_1(\text{SNR}_1)}+\frac{1}{B_2(\text{SNR}_2)}\right)}, \tag{4.11}$$

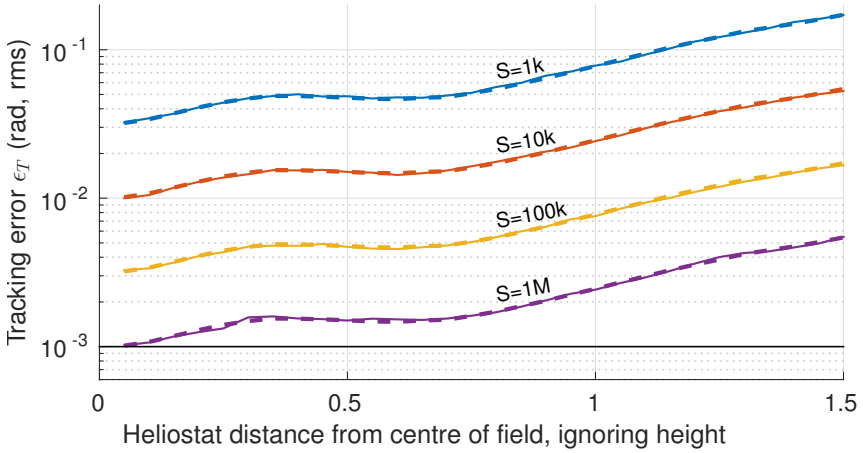


Figure 4.34: Overall tracking error with normalised system parameters as in Table 4.4. Simulated graphs are solid and predictions are dashed.

from which it is clear what the influence of each parameter is on the overall tracking error.

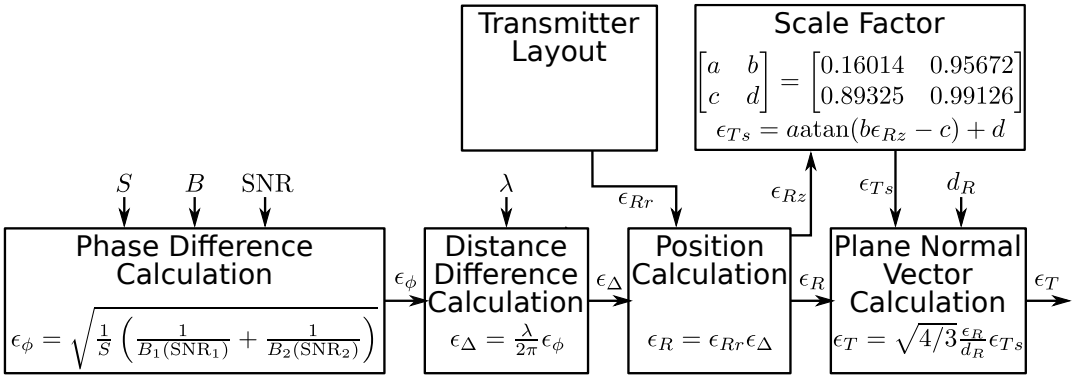


Figure 4.35: Error budget for integrated system.

All results thus far have been normalised where possible and can be applied to a realistically-sized heliostat field by appropriately scaling the necessary parameters. Table 4.5 shows the scaled values for the parameters in Table 4.4. Keep in mind that the transmitter coordinates must also be scaled. Since the normalised field radius is one, a field radius of 100 m is obtained by a scale factor of 100.

Table 4.5: Scaled system parameters.

Parameter	Value	Scaled Value	Description
SPC	10		Samples per cycle
SNR_i	10^5		Signal-to-noise ratio
B_i	0.043		Equivalent normalised noise bandwidth
S	varied		Number of samples
N	10		Number of transmitters (layout in Figure 4.23)
r	1	100 m	Field radius
λ	4	400 m	Wavelength
d_R	0.017	1.73 m	Inter-receiver distance

4.7 Conclusion

This section provided an in-depth analysis of the performance of each subsystem as well as the system as a whole. The factors influencing the overall error were identified and their effects investigated and quantified. It was found that the performance of the full system can be accurately predicted when the performance of the individual subsystems are known. The simulations show that it is theoretically possible to achieve a tracking error in the order of one milliradian. It was also demonstrated how the normalised results can be applied to any field size by scaling appropriately.

While the accuracy required for the position of receivers is very high, there are two properties specific to this application that can be exploited. Firstly, the receivers are mostly stationary and only move when the heliostat is reoriented. This allows averaging over long periods for increased accuracy. Secondly, it was shown how only position errors perpendicular to the mirror surface affect the tracking error, allowing the use of larger receivers, as long as they are flat.

Since many of the subsystems are digital in nature they can be accurately analysed using only simulations. In fact, once the signals are sampled at the receiver, all the subsystems that follow are digital and therefore the simulated system performance is a very true representation of the expected real-world performance, provided that the assumptions about the sampled signals are accurate. The validity of these assumptions is investigated in the next chapter.

Chapter 5

Experimental Results

This chapter describes in detail the experiments conducted and results obtained in order to verify the theoretical results of the previous chapter. Due to the complexity associated with constructing a working prototype,[1] and the limited time in which to do so, the decision was made to only verify individual subsystems and not the system as a whole. Recall that most of the subsystems are digital in nature and therefore their performance can be accurately determined using only simulations. The critical subsystem that requires experimental verification is the phase difference measurement. The accuracy with which phase difference can be measured in practice is investigated and any unforeseen factors not accounted for in the theoretical models are also identified and investigated.

For simplicity, experiments are conducted using cables instead of wirelessly. This not only guards against interference, thus removing the need for modulation, but also allows the use of more commonly available equipment such as ordinary signal generators and oscilloscopes. The distance between the wireless transmitter and receiver is emulated by using cables of corresponding length. To emulate the measurement of signals received from different transmitters, the output of the signal generator is split into two cables which are connected to two separate channels on the oscilloscope, where the relative delay between the signals is measured.

5.1 Equipment and Components

This section lists and motivates the equipment and components used in the experiments, with all the necessary specifications. Conceptually, the experiments require only a signal generator, oscilloscope, two cables and a way to split the output of the signal generator. In practice, some adapters are required for

[1] A working prototype would require at least 5 transmitters, 3 receivers, all positioned with submillimetre precision, as well as the necessary electronics for modulation and demodulation.

interconnection, as well as impedance matching components and attenuators to minimize the effect of reflections [107].

The list of equipment and components used is shown in Table 5.1. It is important to properly clean all components beforehand to ensure good electrical contact. A network analyser is used to aid in impedance matching and minimising reflections. A frequency of 1 MHz is used for measurements, corresponding to a wavelength of 300 m, which is close enough to the 400 m wavelength in Table 4.5 for the purpose of demonstration.

Table 5.1: List of equipment and components, and their connector types.

#	Description	Connector
1	Hewlett-Packard 8647A Signal Generator	N-F
2	Tektronix TDS 380 Digital Real-Time Oscilloscope	BNC-F
3	Rohde&Schwarz ZVB8 Vector Network Analyzer	N-F
4	LMR400 low loss 50 Ω coaxial cable, $VF = 0.85$	N-F/F
5	50 Ω feed-through terminator	BNC-M/F
6	10 dB attenuator	SMA-M/F
7	Adapter	N-M/SMA-M
8	Adapter	N-M/M
9	Adapter	N-F/BNC-M
10	T-piece	SMA-F/F/F
11	Custom matched T-piece	SMA-F/F/F

The maximum phase difference that can be measured depends on the length between the two cables. A roll of high-quality cable is used to minimise interference and attenuation. Individual cables are manually assembled such that their lengths allow for a phase difference corresponding to at least 10 % of the wavelength, which is easily measurable on an oscilloscope. This length is calculated next.

The speed v of an electromagnetic wave in a medium is given by

$$v = c \times VF, \tag{5.1}$$

where c is the speed of light and VF is called the velocity factor. The wavelength in the medium λ' then becomes

$$\lambda' = v/f. \tag{5.2}$$

A 10 % phase difference requires that the cable lengths differ by at least l, where

$$\frac{l}{\lambda'} = \frac{lf}{c \times VF} = 10\,\%. \tag{5.3}$$

For the cable's specified velocity factor of 0.85, this means that the cable lengths must differ by at least $l = 25.5$ m.

A roll of approximately 50 m is purchased from which three cables are created as listed in Table 5.2. After assembly a 5 m tape measure is used to measure the lengths of the cables, with cable C measured piece-wise in 10 intervals. At each measurement interval an uncertainty of ±1 mm is assumed for tape measure alignment, with an additional 2 mm per metre uncertainty in extra length due to the cable not being perfectly straight. The measured cable lengths are listed in the table as the gross length, which includes the connector. The next column shows the minimum and maximum lengths when the uncertainties are included. The connector head measures 18 mm in length, with an additional roughly 15 mm crimped onto the cable, both of which are regarded as uncertainty in shorter cable length since the cable ends somewhere in this region. The mechanical cable lengths, excluding the connectors, are listed in the table as the net length.

Table 5.2: Mechanical cable lengths as per tape measure.

Cable	Gross Length (m) measured	Gross Length (m) min–max	Net Length (m) min–max
A	1.031	1.030–1034	0.964–1.034
B	1.034	1.033–1037	0.967–1.037
C	45.055	45.045–45.155	44.979–45.155

5.2 Impedance Matching

This section describes how impedance matching is done in the experimental setup. It starts by motivating the need for impedance matching by discussing the effect of reflections on the measurements. Then it documents the process of systematically identifying and removing the impedance mismatches until the setup produces the minimum amount of reflections for best possible results. The initial minimal setup is shown in Figure 5.1, with component numbers according to Table 5.1.

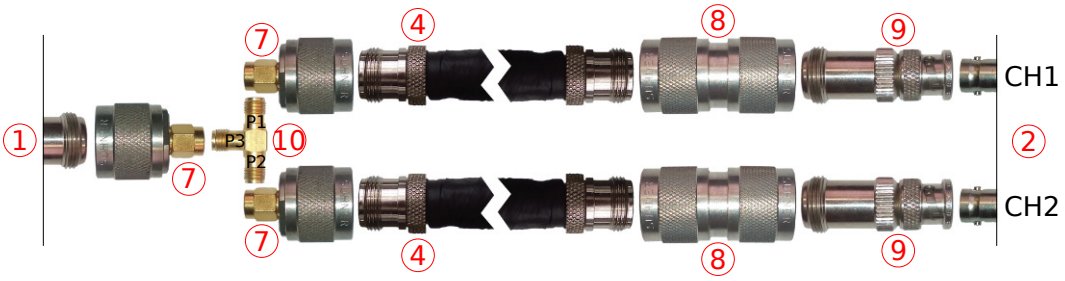


Figure 5.1: Initial minimal unmatched experimental setup. Component numbers are according to Table 5.1.

5.2.1 Motivation

The effect that a reflected wave has on the measured wave is shown in Figure 5.2. Each image shows the original wave in blue, the reflected wave in red, and the sum of the waves or measured wave in yellow. In (a), the reflected wave is shifted 90 °, causing a phase shift between the original and measured waves. In (b), the reflected wave is shifted 180 °, causing a difference in amplitude between the original and measured waves.

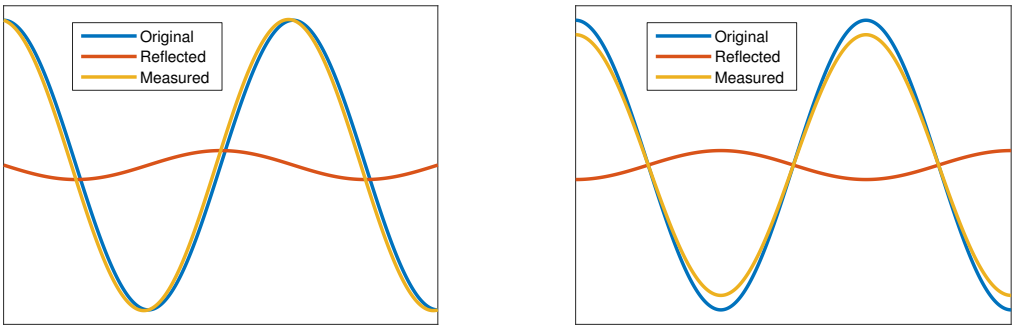


Figure 5.2: Effect of reflections on measured signal, causing (a) a phase shift and (b) a change in amplitude.

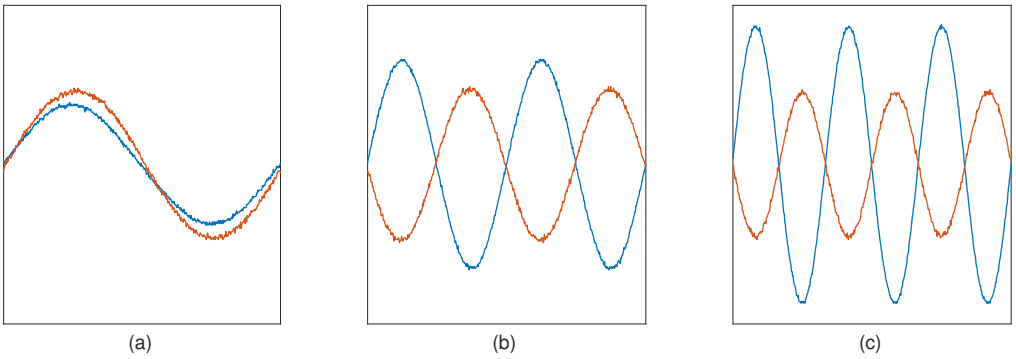


Figure 5.3: Reflections in measurements causing phase and amplitude changes, indicating that the setup in Figure 5.1 is not correctly matched. Cables A and C are used at oscilloscope channels 1 and 2, respectively. The frequency is 1 MHz, 2 MHz and 3 MHz in (a), (b) and (c), respectively, with channel 1 in blue and channel 2 in red.

Figure 5.3 shows the signals measured by the oscilloscope when it is connected to the signal generator according to Figure 5.1 and the frequency of the signal generator is set to 1 MHz, 2 MHz and 3 MHz, respectively. Cables A and C are used at oscilloscope channels 1 and 2, respectively. The same phenomena illustrated in Figure 5.2 is clearly visible in the measurements, which show how

the phase difference and amplitude between the two signals varies as only the frequency is varied. This leads to the conclusion that reflections are almost certainly present in the setup. Since the difference in length of the cables is calculated by measuring the phase difference between the two signals, any reflections could potentially alter the measured phase difference and must be kept to a minimum.

5.2.2 Matching

When a wave travels from one medium to another, the portion of the wave reflected back is given by [107]

$$\Gamma = \frac{Z_L - Z_S}{Z_L + Z_S}, \tag{5.4}$$

where Γ is known as the reflection coefficient, Z_S is the impedance of the current medium (source) and Z_L is the impedance of the next medium (load). In this case, the potential sources of impedance mismatch are at the signal generator output, both oscilloscope channel inputs, and each port on the T-piece. Each of these sources are investigated and matched as effectively as possible, and the magnitude of the reflection of each is quantified to determine the remaining reflections that can be expected, if any.

Impedance mismatches in the network are measured using the network analyser to measure the scattering parameters [107] of the network. Up to three ports (P1, P2, P3) of the network analyser are used, with s_{ij} denoting the scattering parameter of ports i and j. Briefly, s_{ij} measures the portion of the wave incident at port j and measured at port i. The network analyser has a built-in calibration routine performed with the aid of a calibration kit. The calibration is menu-guided and straight-forward and is described in detail in the device manual.

5.2.2.1 Signal Generator Output

The output impedance of the signal generator, according to specification, is 50 Ω. To test this, it is connected directly to P1 of the network analyser, which also has a 50 Ω impedance. Figure 5.4 shows s_{11}, with around -30 dB of the incident wave reflected back. This indicates a good match, hence no additional impedance matching is required.

5.2.2.2 Oscilloscope Channel Inputs

The input impedance of the oscilloscope channels is 1 MΩ according to specification, hence a large mismatch with the 50 Ω cables can be expected. To test this, P1 and P2 of the network analyser are connected to cables A and B, which are connected to channels 1 and 2 of the oscilloscope as in Figure 5.1. The

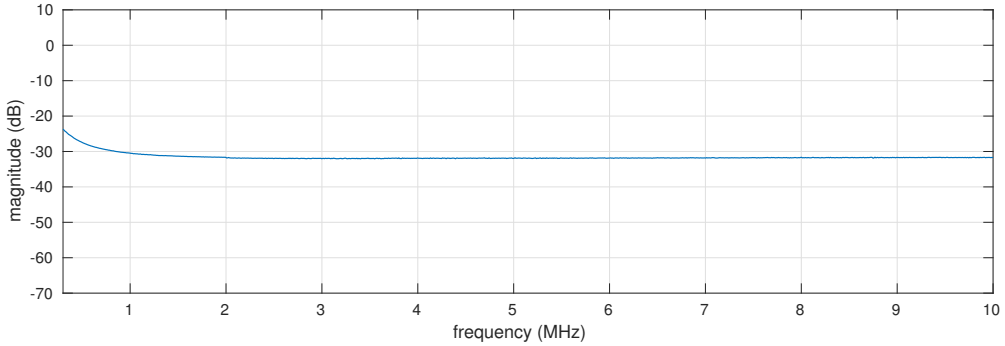


Figure 5.4: Reflections at signal generator indicating that no additional matching is required.

mismatch is confirmed by the black line in Figure 5.5, showing a 0 dB (100 %) reflection for both channels. Matching is done by placing a 50 Ω feed-through terminator between the cable and oscilloscope channel input – a common method of matching 50 Ω cables with high-impedance oscilloscope channels. The result is shown by the blue and red lines in Figure 5.5, with -50 dB or less reflected at 1 MHz, which indicates an excellent match.

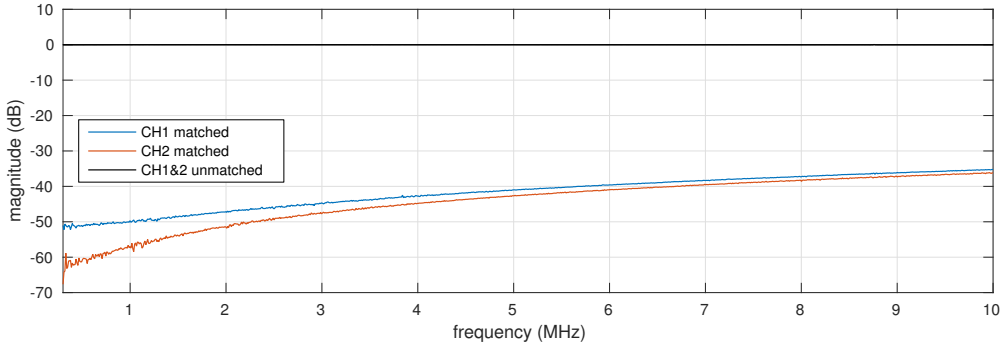


Figure 5.5: Reflections at oscilloscope channels before (black) and after (blue, red) matching using a 50 Ω feed-through terminator.

5.2.2.3 T-piece

Reflections at the T-piece are measured by connecting all three ports to the network analyser and measuring the scattering parameters. P3 of the network analyser is connected directly to the T-piece port facing the signal generator, while P1 and P2 are connected with cables A and B to the T-piece ports facing oscilloscope channels 1 and 2, respectively. Figure 5.6 shows how about -9.6 dB of the wave incident at each port is reflected back, which is about 33 %. This indicates an impedance mismatch at the T-piece as shown in Figure 5.7 (a),

since each 50 Ω cable is connected to two 50 Ω cables in parallel, such that each Z_{in} is equal to 25 Ω, resulting in a reflection coefficient of $\Gamma = 0.33$.

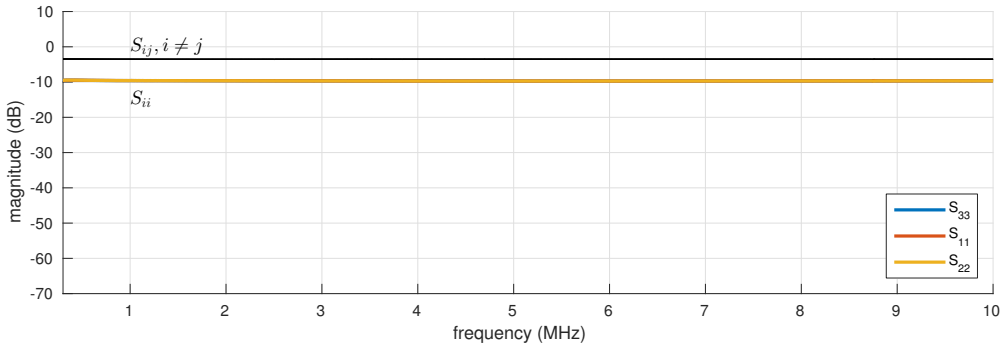


Figure 5.6: Reflections at T-piece before matching.

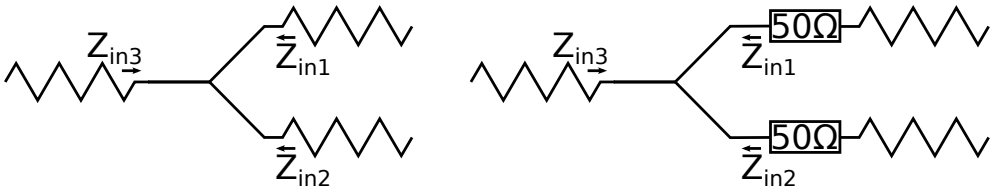


Figure 5.7: Schematic of (a) unmatched T-piece and (b) T-piece matched at port 3.

The most important port on the T-piece to match is the one connected to the signal generator (port 3) for maximum power transfer to the oscilloscope. Since the oscilloscope channels are very well matched, the reflections expected into ports 1 and 2 of the T-piece are minimal making matching less critical. Figure 5.7 (b) shows how port 3 is matched. Due to availability, 51 Ω resistors are used. The quality of the matching is shown in Figure 5.8, with port 3 reflecting about -45 dB at 1 MHz and the other two ports reflecting -11.8 dB, or about 26 %, which is better than the 33 % before matching.

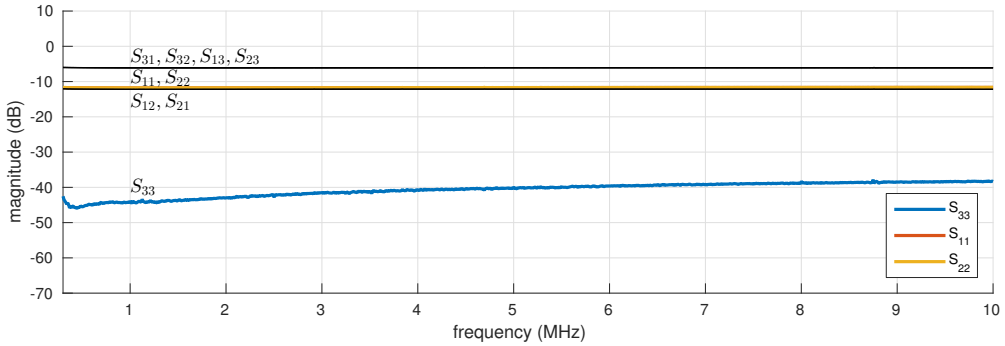


Figure 5.8: Reflections at T-piece after matching as in Figure 5.7 (b).

5.2.3 Isolation by Attenuation

The final step to reduce reflections is by adding attenuation. When a reflection occurs at one of the oscilloscope channels, it heads back to the T-piece, from where it can reflect again towards the channel from whence it came, or pass to the other oscilloscope channel. To limit the effect of these reflections, attenuation is added to ports 1 and 2 of the T-piece, causing these reflections to undergo twice the attenuation of the incoming signal, thus improving the SNR. Due to availability, 10 dB attenuators are used. Figure 5.9 shows the scattering parameters after the oscilloscope channels have been isolated in this manner. As expected, the reflections at ports 1 and 2 (s_{11}, s_{22}), as well as their influence on one another (s_{12}, s_{21}), are reduced by roughly 20 dB, while the signal reaching ports 1 and 2 from port 3 (s_{13}, s_{23}) are reduced by only 10 dB.

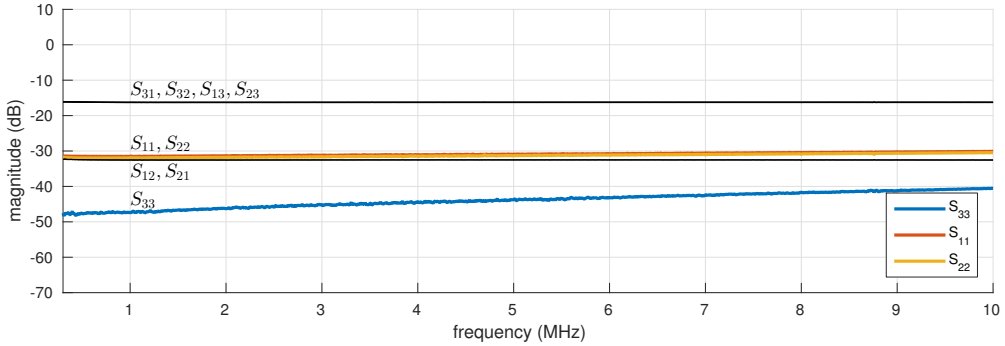


Figure 5.9: Reflections at T-piece after matching as in Figure 5.7 (b) and attenuation at ports 1 and 2.

The final experimental setup and interconnectivity of all components is shown in Figure 5.10. The feed-through terminators (5), custom matched T-piece (11) and attenuators (6) all contribute to reducing the reflections in the network.

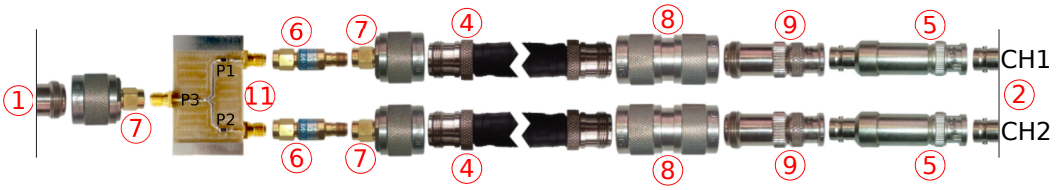


Figure 5.10: Final matched experimental setup with attenuation. Component numbers are according to Table 5.1.

5.2.4 Evaluation

The final matched experimental setup in Figure 5.10 represents a considerable improvement over the initial unmatched setup in Figure 5.1 with regards to impedance matching. While reflections are greatly reduced, a small amount is still present and it may be desirable to know their influence on the measurements.

Consider the following equation which gives the resulting sinusoid when two sinusoids of the same frequency but different phases are added together [108]

$$A\cos(\omega t) + B\cos(\omega t + \alpha) = \sqrt{A^2 + B^2 + 2AB\cos(\alpha)} \cdot \cos\left(\omega t + \arctan\left[\frac{B\sin(\alpha)}{A + B\cos(\alpha)}\right]\right), \tag{5.5}$$

where the dot $(\cdot)$ indicates multiplication at the line break. Let A be the amplitude of the original wave and let B be the amplitude of the net reflections, with a phase difference of α between the waves. The phase of the resulting sinusoid is given by the arctan term, which has a maximum when the argument is a maximum, in this case at $\alpha = 90\,^\circ$. This means that the maximum amount by which the reflections can shift the phase of the original signal is given by

$$\beta = \arctan\left(\frac{B}{A}\right). \tag{5.6}$$

The reflections at the oscilloscope ports after matching are -50 dB according to Figure 5.5. These reflections can head back to the T-piece and return to the oscilloscope channels with magnitude -50 dB + -32.5 dB = -82.5 dB according to Figure 5.9. Another -50 dB reduction is encountered at the oscilloscope port and further reflections are regarded as negligible. Reflections between the signal generator and T-piece are ignored as they affect each oscilloscope channel equally. The total ratio of the reflections vs. the original wave is

$$\frac{B}{A} = 10^{-50/20} + 10^{-82.5/20} = 0.003237, \tag{5.7}$$

such that the maximum phase shift caused by the reflections at each channel is

$$\beta = 0.1855\,^\circ, \tag{5.8}$$

with the net effect on the measured phase difference being 2β, or $0.371\,^\circ$.

5.3 Measurements

The primary purpose of the measurements is to verify that the Hilbert transform (HT) method successfully calculates the phase difference between measured signals. This is done by comparing the output of the phase difference calculation algorithm with manual measurements from the oscilloscope. Furthermore, the theoretical model is verified by measuring the practical SNR and comparing the practical method performance with the predicted performance at the same SNR. The SNR is determined by the ambient noise floor and the power output of the signal generator.

Measurements are taken using the experimental setup in Figure 5.10. As stated previously, a frequency of 1 MHz is used, with an output power of 10 dBm at the signal generator. The oscilloscope is setup such that each pixel represents 1 ns on the horizontal axis, and the scale of the vertical axis is left unchanged throughout all measurements. First, a set of calibration routines are described and performed. Then, measurements are taken by manually reading from the oscilloscope. Each manual measurement is repeated ten times in order to establish the measurement uncertainty. Additionally, ± 1 ns uncertainty is added due to the maximum oscilloscope resolution. Finally, the oscilloscope data is digitally exported and put through the HT based phase difference calculation algorithm. The section is concluded by comparing and commenting on the results.

5.3.1 Calibration

It is important to ensure proper calibration of all equipment before use. For the signal generator, the accuracy of the output frequency is important, while amplitude and initial phase are not. At the oscilloscope, frequency is again important, while amplitude is not. The output of the signal generator is connected directly with a single cable to the oscilloscope – first to channel one (CH1) and then to channel two (CH2). In both cases it is found that the output frequency of the signal generator matches the frequency measured by the oscilloscope. For this reason both machines are deemed sufficiently calibrated with regards to frequency.

Next, the calibration routines for the relative delay between the oscilloscope channels as well as the relative delay between the two paths of the T-piece are described. This calibration ensures that the relative delay between the two signals measured at the oscilloscope is solely due to the difference in cable lengths. This is done at the hand of Figure 5.11, which shows how the signal generator output is split and connected to the two oscilloscope channels. The grey areas in the figure indicate which components are regarded as a single unit for simplicity. Cables A and B are used as p1 and p2, respectively.

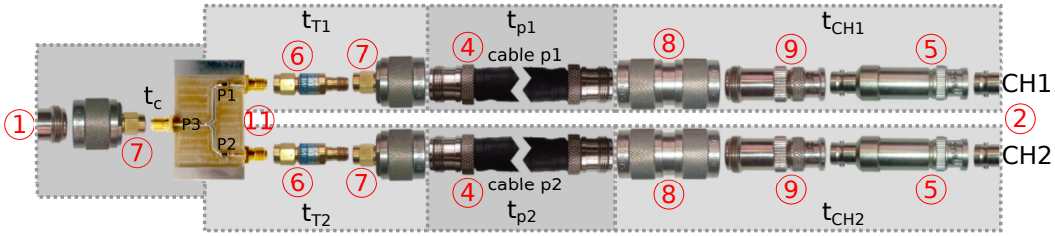


Figure 5.11: Calibration setup for oscilloscope channels and T-piece.

5.3.1.1 Oscilloscope Channels

The calibration steps to determine the relative delay between the oscilloscope channels are as follows:

1. Connect the components as in Figure 5.11 such that CH1 is $t_c + t_{T1} + t_{p1} + t_{CH1}$ and CH2 is $t_c + t_{T2} + t_{p2} + t_{CH2}$, and measure the relative delay on the oscilloscope, $t_{\Delta CH1}$, where
$$t_{\Delta CH1} = (t_c + t_{T2} + t_{p2} + t_{CH2}) - (t_c + t_{T1} + t_{p1} + t_{CH1}) . \tag{5.9}$$
2. Now disconnect both cables at the oscilloscope channels and swap them around such that CH1 is $t_c + t_{T2} + t_{p2} + t_{CH1}$ and CH2 is $t_c + t_{T1} + t_{p1} + t_{CH2}$, and measure the relative delay on the oscilloscope, $t_{\Delta CH2}$, where
$$t_{\Delta CH2} = (t_c + t_{T1} + t_{p1} + t_{CH2}) - (t_c + t_{T2} + t_{p2} + t_{CH1}) . \tag{5.10}$$
3. Add the previous two equations together to find the relative channel delay $t_{\Delta CH}$ as
$$t_{\Delta CH} = t_{CH2} - t_{CH1} = \frac{t_{\Delta CH1} + t_{\Delta CH2}}{2}. \tag{5.11}$$

Table 5.3 shows the calibration measurements, with the final calibration value for the oscilloscope relative channel delay $t_{\Delta CH}$ as

$$-2\,\text{ns} \leq t_{\Delta CH} \leq 2\,\text{ns}. \tag{5.12}$$

5.3.1.2 T-piece

The calibration steps to determine the relative delay between the two paths of the T-piece are as follows:

1. This step is identical to the previous first step. Connect the components as in Figure 5.11 such that CH1 is $t_c + t_{T1} + t_{p1} + t_{CH1}$ and CH2 is $t_c + t_{T2} + t_{p2} + t_{CH2}$, and measure the relative delay on the oscilloscope, $t_{\Delta T1}$, where
$$t_{\Delta T1} = (t_c + t_{T2} + t_{p2} + t_{CH2}) - (t_c + t_{T1} + t_{p1} + t_{CH1}) . \tag{5.13}$$

Table 5.3: Calibration measurements for oscilloscope relative channel delay in ns.

Measurement #	$t_{\Delta CH1}$	$t_{\Delta CH2}$
1	0	0
2	1	-1
3	-1	1
4	0	-1
5	0	0
6	-1	-1
7	0	1
8	-1	1
9	0	-1
10	-1	0
Value range	-1 – 1	-1 – 1
Add 1ns uncertainty	-2 – 2	-2 – 2
Value range for $t_{\Delta CH}$	-2 – 2	

2. Now disconnect both cables at the T-piece and swap them around such that CH1 is $t_c + t_{T2} + t_{p1} + t_{CH1}$ and CH2 is $t_c + t_{T1} + t_{p2} + t_{CH2}$, and measure the relative delay on the oscilloscope, $t_{\Delta T2}$, where

$$t_{\Delta T2} = (t_c + t_{T1} + t_{p2} + t_{CH2}) - (t_c + t_{T2} + t_{p1} + t_{CH1})\,. \tag{5.14}$$

3. Subtract the last equation from first to find the relative T-piece delay $t_{\Delta T}$ as

$$t_{\Delta T} = t_{T2} - t_{T1} = \frac{t_{\Delta T1} - t_{\Delta T2}}{2}. \tag{5.15}$$

Table 5.4 shows the calibration measurements, with the final calibration value for the T-piece relative path delay $t_{\Delta T}$ as

$$-2\,\text{ns} \leq t_{\Delta T} \leq 2\,\text{ns}. \tag{5.16}$$

Table 5.4: Calibration measurements for T-piece relative path delay in ns.

Measurement #	$t_{\Delta T1}$	$t_{\Delta T2}$
1	0	-1
2	1	1
3	-1	0
4	0	0
5	0	1
6	-1	-1
7	0	-1
8	-1	0
9	0	-1
10	-1	1
Value range	-1 – 1	-1 – 1
Add 1ns uncertainty	-2 – 2	-2 – 2
Value range for $t_{\Delta T}$	-2 – 2	

5.3.2 Time Delay by Manual Measurements

The steps for taking measurements from the oscilloscope manually are:

1. Connect components as in Figure 5.10 such that CH1 is $t_c+t_{T1}+t_{p1}+t_{CH1}$ and CH2 is $t_c + t_{T2} + t_{p2} + t_{CH2}$, and measure the relative delay on the oscilloscope, $t_{\Delta p1}$:

$$t_{\Delta p1} = (t_c + t_{T2} + t_{p2} + t_{CH2}) - (t_c + t_{T1} + t_{p1} + t_{CH1})\,. \tag{5.17}$$

2. Simplify and add calibration:

$$\begin{aligned} t_{\Delta p} = t_{p2} - t_{p1} &= t_{\Delta p1} - t_{T2} + t_{T1} - t_{CH2} + t_{CH1} \\ &= t_{\Delta p1} - t_{\Delta T} - t_{\Delta CH}. \end{aligned} \tag{5.18}$$

Figure 5.12 shows some examples of the waveforms captured on the oscilloscope for cables A–B, A–C and B–C, and Table 5.5 shows the measurements and the uncertainty range. It can be seen that the resulting uncertainty is quite large due to the limitations of the hardware. The table also shows the resulting range of values when the time values are converted to phase difference by the equation

$$\text{phase difference} = -\text{time delay} \times \text{frequency} \times 360°. \tag{5.19}$$

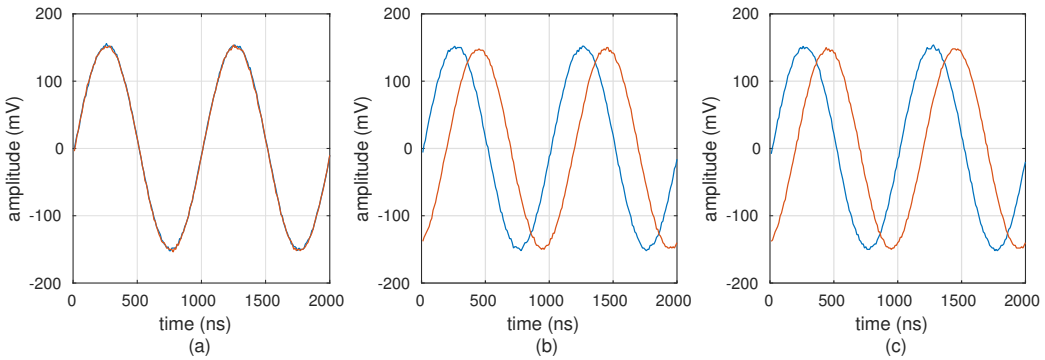


Figure 5.12: Examples of the waveforms captured on the oscilloscope for cables (a) A–B, (b) A–C and (c) B–C. The first cable is connected to CH1 (blue) and the second is connected to CH2 (red).

Table 5.5: Time delay between cables A–B, A–C and B–C, manually measured from oscilloscope in ns. The resulting phase difference is also shown in degrees.

Measurement #	p1: A | p2: B	p1: A | p2: C	p1: B | p2: C
1	0	181	181
2	1	182	182
3	-1	180	181
4	0	182	179
5	0	181	180
6	-1	180	182
7	0	181	179
8	-1	182	181
9	0	179	182
10	-1	182	182
Value range	-1 – 1	179 – 182	179 – 182
Add 1ns uncertainty	-2 – 2	178 – 183	178 – 183
Add calibration	-6 – 6	174 – 187	174 – 187
Phase difference (°)	-2.16 – 2.16	-67.32 – -62.64	-67.32 – -62.64

5.3.3 Phase Difference by Hilbert Transform (HT)

To eliminate the large uncertainty introduced by the hardware limitations and calibration values, the following differential measurement procedure is followed when calculating the phase difference using the HT method:

1. Connect components as in Figure 5.10 such that CH1 is $t_c+t_{T1}+t_{p1}+t_{CH1}$ and CH2 is $t_c+t_{T2}+t_{p2}+t_{CH2}$, and capture data from oscilloscope, then calculate the phase difference $\phi_{\Delta p1}$, where $\phi_{\Delta p1} = 2\pi f t_{\Delta p1}$ and

$$t_{\Delta p1} = (t_c + t_{T2} + t_{p2} + t_{CH2}) - (t_c + t_{T1} + t_{p1} + t_{CH1})\,. \tag{5.20}$$

2. Completely disconnect the cables and swap them around such that CH1 is $t_c+t_{T1}+t_{p2}+t_{CH1}$ and CH2 is $t_c+t_{T2}+t_{p1}+t_{CH2}$, and capture data from oscilloscope, then calculate phase difference $\phi_{\Delta p2}$, where $\phi_{\Delta p2} = 2\pi f t_{\Delta p2}$ and

$$t_{\Delta p2} = (t_c + t_{T2} + t_{p1} + t_{CH2}) - (t_c + t_{T1} + t_{p2} + t_{CH1})\,. \tag{5.21}$$

3. Subtract the last equation from the first

$$t_{\Delta p} = t_{p2} - t_{p1} = \frac{t_{\Delta p1} - t_{\Delta p2}}{2}, \tag{5.22}$$

which shows how the relative time delay between the cables can be obtained without having to calibrate the oscilloscope channels or T-piece. The phase difference is then calculated as

$$\phi_{\Delta p} = 2\pi f t_{\Delta p} = \frac{\phi_{\Delta p1} - \phi_{\Delta p2}}{2}. \tag{5.23}$$

Each measurement is repeated 1,000 times in order to properly characterise the mean and distribution of the result. A MATLAB script, shown in Appendix A.2, is used to automate the process of exporting data from the oscilloscope. The calculated phase differences are shown in Table 5.6, and the corresponding distributions are shown in Figure 5.13.

Table 5.6: Phase difference mean μ and standard deviation σ in degrees between cables A–B, A–C and B–C, calculated using HT method from exported oscilloscope data.

	$\phi_{\Delta p1}$		$\phi_{\Delta p2}$		$\phi_{\Delta p}$		
	μ	σ	μ	σ	μ	σ	min–max
A–B	0.0412	0.0475	0.0420	0.0479	-0.0004	0.0477	-0.0986 – 0.0883
A–C	-65.0940	0.0473	65.1658	0.0475	-65.1299	0.0474	-65.2390 – -65.0325
B–C	-65.1015	0.0479	65.1692	0.0438	-65.1354	0.0459	-65.2403 – -65.0411

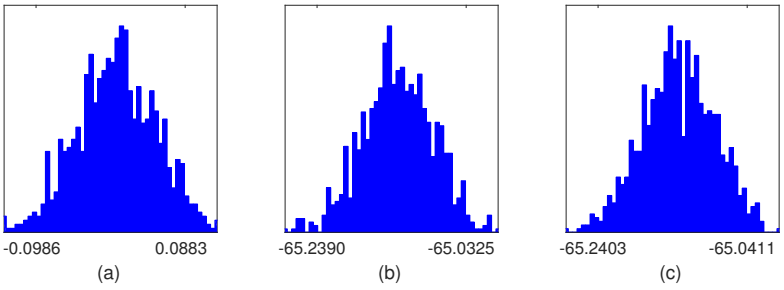


Figure 5.13: Phase difference distribution in degrees for cables (a) A–B, (b) A–C, and (c) B–C, corresponding to Table 5.6.

5.3.4 Evaluation and Validation

The first step in validation is to determine whether or not the HT method produces correct results when presented with actual measured data. This is done by comparing the manual measurements with the results from the HT method. Figure 5.14 shows the distribution of the HT method results in comparison with the manual measurements. In each case, the distribution of the results obtained using the HT method is in the centre of – and much smaller than – the manual measurements uncertainty range, leaving no doubt about the correctness and accuracy of the method.

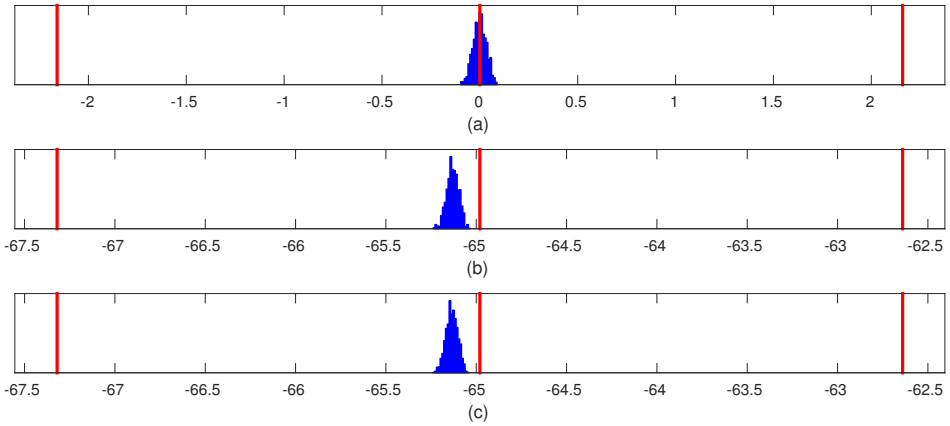


Figure 5.14: Phase differences calculated using HT method compared to the uncertainty range of the manual measurements between cables (a) A–B, (b) A–C and (c) B–C. The distribution of the HT results is shown in blue and the lower, middle and upper values of the uncertainty range for the manual measurements is indicated by the red lines.

The next step is to compare the predicted and actual practical performance of the HT method. The noise on each oscilloscope channel is measured by turning off the signal generator output and then exporting the oscilloscope measurements. The distribution of the noise on each channel is shown in Figure 5.15. Since there is no way to separate the noise N from the signal S, the RMS of $S + N$ is measured such that the SNR is given by

$$SNR = \frac{S}{N} \approx \frac{S+N}{N}, \quad N \ll S. \tag{5.24}$$

The RMS values of the signal for each cable and the noise on each channel are given in Table 5.7. Since cable C is longer, the signal is attenuated more and the RMS of the measured signal is less.

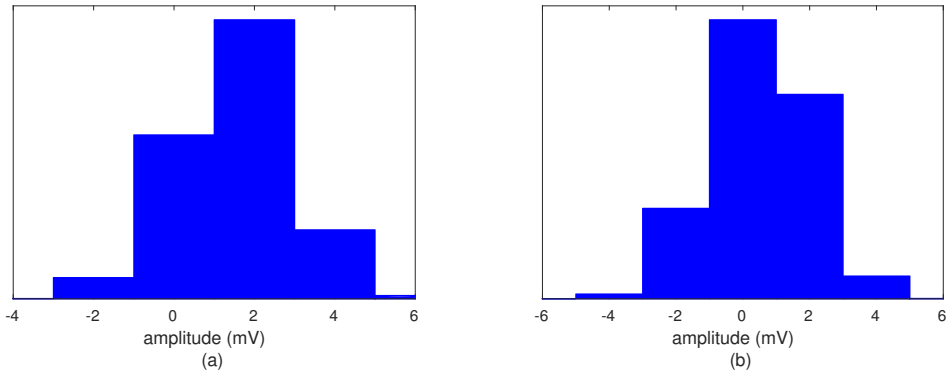


Figure 5.15: Measured noise distribution, mean and standard deviation on (a) channel 1 and (b) channel 2.

Table 5.7: Measured RMS for the signal in each cable (S_A, S_B, S_C) and the noise on each oscilloscope channel (N_1, N_2).

Parameter	Value (mV, RMS)
S_A	108.00
S_B	108.00
S_C	105.30
N_1	2.125
N_2	1.646

Table 5.8: Predicted ϵ'_ϕ and actual ϵ_ϕ performance of the HT method as well as the error %.

Cables	SNR_1	SNR_2	Predicted error ϵ'_ϕ (°)	Actual error ϵ_ϕ (°)	Error %
A – B	2583.03	4305.15	0.04509	0.04765	5.37
A – C	2583.03	4092.58	0.04553	0.04739	3.92
B – C	2583.03	4092.58	0.04553	0.04588	0.76

The predicted ϵ'_ϕ and actual ϵ_ϕ performance of the HT method is given in Table 5.8. In each case the number of samples N is 1,000 and no filtering is applied to the noise. Equation 4.4 is used to predict the value of ϵ_ϕ given SNR_1, SNR_2 and N. The actual error is assumed to be only the standard deviation of the value of $\phi_{\Delta p}$ in Table 5.6. The percentage error between the predicted and actual performances is calculated as

$$\text{Error}\% = \frac{\epsilon_\phi - \epsilon'_\phi}{\epsilon_\phi} \times 100\,\% \tag{5.25}$$

and seen to be in the order of 5 % which is very small and confirms the validity of the theoretical model. The average SNR is roughly 3,340, or about 35 dB. Using a custom wireless implementation it should be possible to achieve an even better SNR.

Finally, the difference in length between the cables is calculated using the calculated phase difference and compared to the measured lengths. Table 5.9 shows the phase difference uncertainty from Table 5.6 after the effect of reflections are added in, the resulting calculated difference in cable length, and measured difference in cable length from Table 5.2. The two ranges of values for cables A–C and B–C are seen not to overlap. This is most probably due to non-idealities in the cables such as non-uniformity and parameter tolerances. An attempt was made to reconcile the values and explain the differences, but was abandoned due to time limitations and the fact that it would not contribute to the original purpose of the experiments. That which was set out to be achieved has been achieved.

Table 5.9: Comparison of calculated and measured difference in cable lengths.

Cables	Phase difference after reflections	Calculated difference in length (m)	Measured difference in length (m)
A – B	-0.4696 – 0.4593	-0.325 – 0.332	-0.067 – 0.073
A – C	-65.6100 – -64.6615	45.770 – 46.442	43.945 – 44.191
B – C	-65.6113 – -64.6701	45.776 – 46.443	43.942 – 44.188

5.4 Conclusion

This chapter investigated the real-world performance of the Hilbert transform (HT) method. It was found that the phase differences calculated by the method match those that were manually read from the oscilloscope. Additionally it was found that the performance of the method could be accurately predicted using the measured value of the SNR in the experiments. These two findings successfully validate the theoretical model of the method.

Chapter 6

Conclusion

The work in this thesis addresses the problems associated with current heliostat tracking methods. These problems include insufficient tracking accuracy, long recalibration times and lack of real-time feedback. A method is developed which is able to determine the orientations of all heliostats in the field simultaneously, providing near real-time feedback to the control system and eliminating the need for recalibration while maintaining a high tracking accuracy. The performance of the method is analysed using theoretical models and simulations which are validated by experiments. This chapter reports on the findings of the previous chapters.

The proposed tracking method consists of a number of subsystems contributing towards a modular design. The performance of each subsystem is quantified using simulations, the results of which are reliable due to the fact that most of the subsystems are digital in nature. The subsystem that calculates the phase difference between two sampled signals is validated by means of experiments, and it is found that the theoretical and experimental results agree closely, differing only by about 5%.

The findings of the previous chapters show that the most important factors for an increased tracking accuracy are an increased number of samples S, an increased signal-to-noise ratio SNR, a smaller wavelength λ, and a larger inter-receiver distance d_R. The importance of transmitter placement is also shown, with a properly optimised transmitter layout resulting in a smaller position error, thus smaller tracking error. A simulation of the full system shows that for the given parameters it is theoretically possible to obtain a tracking accuracy in the order of one milliradian or less.

A larger d_R favours larger heliostats over smaller ones. Since heliostats need to update their orientations multiple times per minute, a higher sampling rate is favourable as it will yield a larger number of samples within the allotted time. Using multiple wavelengths to approximate and pinpoint receiver positions introduces additional complexity but offers considerable increases in accuracy.

Increasing the number of transmitters also increases accuracy, albeit only marginally. It was also shown how the receiver can be implemented by an antenna etched on a PCB, or a flat optical sensor, as only the position error component perpendicular to the mirror surface affects the tracking accuracy.

While this thesis focuses mainly on one of the unavoidable error sources, namely phase difference error due to noise, other possible error sources also exist, such as errors in the electronic phase delay Φ, errors in the transmitter positions, and other hardware-specific errors, the effects of which will most certainly lead to an increased tracking error.

With regards to the objectives stated in Section 1.3, all primary objectives have been achieved. This includes the development of a method to accurately measure the orientation of all heliostats in a solar field simultaneously in intervals shorter than the heliostat realignment interval. In the system developed by Malan this interval is 13 seconds [2], which is more than enough time to obtain a million samples like in the simulated system. Furthermore, the method is validated using simulations and experiments, and the performance of individual subsystems are quantified and used to predict the overall system performance with high accuracy. The secondary objective of maintaining a tracking error of less than one milliradian is also achieved. Due to time constraints a prototype system has not been constructed.

From a financial viewpoint the proposed method offers significant advantages. While the required infrastructure may be expensive, it is also minimal: only the central oscillator and transmitter towers. The important consideration is the added cost per heliostat. This amounts to only the three electromagnetic receivers, some electronics, a power source and a processing unit, with the last two already being present in most cases.

In the case where the time between heliostat orientation updates limits the number of samples to such an extent that a tracking error of less than one milliradian is unachievable, the system may become unsuitable for real-time use. It can, however, still be used to aid in the calibration process. The orientation of a heliostat is made stationary for as long as it takes to obtain the required amount of samples, after which its normal vector is calculated. It can also be used to detect which heliostats require urgent recalibration. This will still improve the required time for a full field calibration from three weeks to a few minutes. Another application of the system is the localisation of any object within the heliostat field by simply fixing one of the electromagnetic receivers to the object, provided that it can remain stationary long enough to obtain the required number of samples.

The main advantages of the method are that it determines the orientation of all heliostats simultaneously with high accuracy which completely removes the need for recalibration, allowing all heliostats to be operational 100% of the

time. The main disadvantage is that the method is very sensitive to noise and other inaccuracies as it requires receiver positions to be determined with an error of 10mm or less, depending on heliostat size. This is especially true with phenomena such as reflections and non-idealities in the medium, which can significantly alter the measured phase differences and introduce errors when converting the phase differences to distances. Reflections and non-idealities in a cable can be compared to multipath propagation and mirages[1] in air.

Future work should start with the construction of a small scale prototype which would allow moving from a wired to a wireless environment. This would also require the implementation of a modulation scheme to separate transmission channels. The effects of multipath propagation and mirages, as well as other phenomena introduced by the hardware implementation should be investigated. The initial analysis of the proposed method shows very promising results, but ultimately its feasibility will be determined by an implementation of the full system with all subsystems integrated.

[1]Referring to the bending of light which causes it to travel longer paths, typically observed by the light above a hot object appearing distorted.

Bibliography

[1] United Nations Sustainable Development. Goal 7: Ensure access to affordable, reliable, sustainable and modern energy for all. United Nations. [Online]. Available: http://www.un.org/sustainabledevelopment/energy/

[2] K. Malan, "A heliostat field control system," Master's thesis, Department of Mechanical and Mechatronic Engineering, Stellenbosch University, 2013.

[3] R. Sherif, "Concentrating solar power technologies," in *iNEMI Alternative Energy Workshop*, San Jose, California, October 2010.

[4] M. Zavodny, M. Slack, R. Huibregste, and A. Sonn, "Tower-based CSP artificial light calibration system," *Energy Procedia*, vol. 69, pp. 1488–1497, May 2015.

[5] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy sources: Current status, future prospects and their enabling technology," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 748–764, 2014.

[6] Ecotricity. The end of fossil fuels. [Online]. Available: https://www.ecotricity.co.uk/our-green-energy/energy-independence/the-end-of-fossil-fuels

[7] N. Chowdhury. (2012, March) Is nuclear energy renewable energy? Stanford University. [Online]. Available: http://large.stanford.edu/courses/2012/ph241/chowdhury2/

[8] REN21, "Renewables 2015 global status report," Renewable Energy Policy Network for the 21st Century, Tech. Rep., 2015.

[9] M. Hall, D. McKemy, and A. Lee. (2013, August) Earth's energy balance. North Carolina State University. [Online]. Available: https://climate.ncsu.edu/edu/k12/.eeb

[10] D. Y. Goswami, F. Kreith, and J. F. Kreider, *Principles of Solar Engineering*, 2nd ed. Taylor & Francis Group, LLC, 2000.

[11] Solargis. (2016) Solargis world DNI solar resource map. [Online]. Available: http://solargis.com/assets/graphic/free-map/DNI/Solargis-World-DNI-solar-resource-map-en.png

[12] P. Konstantin and J. Kretschmann. (2010, December) Assessment of

technology options for development of concentrating solar power in South Africa for the World Bank. Fichtner GmbH & Co. KG. [Online]. Available: https://www-cif.climateinvestmentfunds.org/sites/default/files/Presentation-WB(Eskom)Project-2010_12_07.pdf

[13] Sustainable Solar Services. Solar power - harnessing the sun. Sustainable Solar Services. [Online]. Available: http://www.solarsponsoring.com.au/going-solar/how-solar-power-works/

[14] Enerdata. (2015) Global energy statistical yearbook. [Online]. Available: https://yearbook.enerdata.net/electricity-domestic-consumption-data-by-region.html

[15] J. Tsao, N. Lewis, and G. Crabtree, "Solar FAQs," *US Department of Energy*, p. 10, 2006.

[16] Soitec. (2014, December) New world record for solar cell efficiency at 46%. [Online]. Available: http://www.soitec.com/pdf/pr_smart_cell_final-en.pdf

[17] United Nations Statistics Division. (2007) Total surface area as of 19 January 2007. [Online]. Available: http://unstats.un.org/unsd/environment/totalarea.htm

[18] IEA, "Technology roadmaps – concentrating solar power," International Energy Agency, Tech. Rep., 2010.

[19] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*, 4th ed. John Wiley & Sons, Inc., 2013.

[20] IRENA, "Renewable energy technologies: cost analysis series – concentrating solar power," International Renewable Energy Agency, Tech. Rep. 1, 2012.

[21] IRENA, "Renewable power generation costs in 2014," International Renewable Energy Agency, Tech. Rep., 2015.

[22] U.S. Energy Information Administration. (2000) International energy statistics. U.S. Energy Information Administration. [Online]. Available: http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=6&pid=29&aid=12

[23] Eskom. (2016) Renewable energy independent power producer procurement programme. Eskom. [Online]. Available: http://www.eskom.co.za/Whatweredoing/Pages/RE_IPP_Procurement_Programme.aspx

[24] Corporative Marketing. (2015, August) Helio100 promises cost-effective solar power breakthrough. Stellenbosch University. [Online]. Available: http://www.sun.ac.za/english/Lists/news/DispForm.aspx?ID=2863

[25] NREL. (2014, November) Ivanpah solar electric generating system. National Renewable Energy Laboratory. [Online]. Available: http://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=62

[26] J. I. Ortega, J. I. Burgaleta, and F. M. Téllez, "Central receiver system solar power plant using molten salt as heat transfer fluid," *Journal of Solar Energy Engineering*, vol. 130, no. 2, February 2008.

[27] G. J. Kolb, S. A. Jones, M. W. Donnelly, D. Gorman, R. Thomas, R. Davenport, and R. Lumia, "Heliostat cost reduction study," Sandia National Laboratories, Tech. Rep. SAND2007-3293, June 2007.

[28] P. Schramek and D. R. Mills, "Heliostats for maximum ground coverage," *Energy*, vol. 29, no. 29, pp. 701–7113, 2004.

[29] A. Ghosal and R. A. Shyam, "A three-degree-of-freedom parallel manipulator for concentrated solar power towers: Modeling, simulation and design," *SolarPaces 2015*, 2015.

[30] S. M. Besarati, D. Y. Goswami, and E. K. Stefanakos, "Optimal heliostat aiming strategy for uniform distribution of heat flux on the receiver of a solar power tower plant," *Energy Conversion and Management*, vol. 84, pp. 234–243, 2014.

[31] C. K. Ho, "Review of avian mortality studies at concentrating solar power plants." in *International Conference on Concentrating Solar Power and Chemical Energy Systems (SolarPaces) 2015*, vol. 1734, no. 1. AIP Publishing, 2016.

[32] W. B. Stine and M. Geyer. (2001) Power from the sun. [Online]. Available: http://www.powerfromthesun.net/book.html

[33] M. J. Clifford and D. Eastwood, "Design of a novel passive solar tracker," *Solar Energy*, vol. 77, pp. 269–280, 2004.

[34] C.-Y. Lee, P.-C. Chou, C.-M. Chiang, and C.-F. Lin, "Sun tracking systems: A review," *Sensors*, vol. 9, pp. 3875–3890, 2009.

[35] A. T. Abdulrahim, I. S. Diso, and M. B. Oumarou, "Development of a solar tracking bi-focal collectors," *Continental Journal of Renewable Energy*, vol. 2, pp. 19–33, 2011.

[36] H. Mousazadeh, A. Keyhani, A. Javadi, H. Mobli, K. Abrinia, and A. Sharifi, "A review of principle and sun-tracking methods for maximizing solar systems output," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 1800–1818, 2009.

[37] M. Blanco-Muriel, D. C. Alarcón-Padilla, T. López-Moratalla, and M. Lara-Coira, "Computing the solar vector," *Solar Energy*, vol. 70, no. 5, pp. 431–441, 2001.

[38] J. W. Spencer, "Fourier series representation of the position of the sun," *Search*, vol. 2, pp. 172–173, 1971.

[39] C. L. Pitman and L. L. Vant-Hull, "Errors in locating the sun and their effect on solar intensity predictions," in *Meeting of the American Section of the International Solar Energy Society*, Denver, August 1978,

pp. 701–706.

[40] R. Walraven, "Calculating the position of the sun," *Solar Energy*, vol. 20, pp. 393–397, 1978.

[41] J. J. Michalsky, "The astronomical almanac's algorithm for approximate solar position," *Solar Energy*, vol. 40, no. 3, pp. 227–235, 1988.

[42] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *Solar Energy*, vol. 76, pp. 577–589, 2004.

[43] J. Meeus, *Astronomical Algorithms*, 2nd ed. Willmann-Bell, Inc., 1988.

[44] R. Grena, "An algorithm for the computation of the solar position," *Solar Energy*, vol. 82, pp. 462–470, 2008.

[45] R. Grena, "Five new algorithms for the computation of sun position from 2010 to 2110," *Solar Energy*, vol. 86, pp. 1323–1337, 2012.

[46] T. Roos, N. Zwane, E. Kruger, S. Perumal, and R. Cathro, "A 25m^2 target-aligned heliostat with closed-loop control," 2007.

[47] H. Ries and M. Schubnell, "The optics of a two-stage solar furnace," *Solar Energy Materials*, vol. 21, pp. 213–217, 1990.

[48] Y. T. Chen, A. Kribus, B. H. Lim, C. S. Lim, K. K. Chong, J. Karni, R. Buck, A. Pfahl, and T. P. Bligh, "Comparison of two sun tracking methods in the application of a heliostat field," *Journal of Solar Energy Engineering*, vol. 126, pp. 638–644, February 2004.

[49] A. Pfahl, R. Buck, and K. Rehschuh, "Method for controlling the alignment of a heliostat with respect to a receiver, heliostat device and solar power plant," U.S. Patent Application US 2009/0 249 787 A1, October 8, 2009.

[50] J. M. Quero, C. Aracil, L. G. Franquelo, J. Ricart, P. R. Ortega, M. Domínguez, L. M. Castañer, and R. Osuna, "Tracking control system using an incident radiation angle microsensor," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 2, pp. 1207–1216, April 2007.

[51] J. M. Quero Reboul, J. García Ortega, and L. G. Franquelo, "Sensor electrónico para la medida de la posición angular de un objeto luminiscente," Spanish Patent P9 901 385, June 21, 1999.

[52] J. J. Brey, J. M. Quero, L. G. Franquelo, M. Dominquez, and L. Castañer, "Light source tracking microsensor for solar power plants," *MST News*, vol. 3, pp. 45–46, June 2002.

[53] K. Aiuchi, K. Yoshida, M. Onozaki, Y. Katayama, M. Nakamura, and K. Nakamura, "Sensor-controlled heliostat with an equatorial mount," *Solar Energy*, vol. 80, pp. 1089–1097, 2006.

[54] A. Yogev and V. Krupkin, "Control of a heliostat field in a solar energy plant," U.S. Patent 5 862 799, January 26, 1999.

[55] A. Kribus, I. Vishnevetsky, A. Yogev, and T. Rubinov, "Closed loop control of heliostats," *Energy*, vol. 29, pp. 905–913, 2004.

[56] M. R. Convery, "System and method for aligning heliostats of a solar power tower," U.S. Patent Application US 2010/0 252 024 A1, October 7, 2010.

[57] M. R. Convery, "Closed-loop control for power tower heliostats," in *Proceedings of SPIE*, ser. High and Low Concentrator Systems for Solar Electric Applications VI, vol. 8108, 2011.

[58] M. R. Convery, "System and method for aligning heliostats of a solar power tower," U.S. Patent US 8 344 305 B2, January 1, 2013.

[59] Y. T. Chen, K. K. Chong, T. P. Bligh, L. C. Chen, Jasmy Yunus, K. S. Kannan, B. H. Lim, C. S. Lim, M. A. Alias, Noriah Bidin, Omar Aliman, Sahar Salehan, Shk. Abd. Rezan S. A. H., C. M. Tam, and K. K. Tan, "Non-imaging, focusing heliostat," *Solar Energy*, vol. 71, pp. 155–163, 2001.

[60] Y. T. Chen, K. K. Chong, C. S. Lim, B. H. Lim, K. K. Tan, Omar Aliman, T. P. Bligh, B. K. Tan, and Ghazally Ismail, "Report of the first prototype of non-imaging focusing heliostat and its application in high temperature solar furnace," *Solar Energy*, vol. 72, no. 6, pp. 531–544, 2002.

[61] R. S. Baheti and P. F. Scott, "Design of self-calibrating controllers for heliostats in a solar power plant," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 6, pp. 1091–1097, December 1980.

[62] K. W. Stone, "Automatic heliostat track alignment method," U.S. Patent 4 564 275, January 14, 1986.

[63] E. F. Camacho, M. Berenguel, I. Alvarado, and D. Limon, "Control of solar power systems: a survey," in *Proceedings of the 9th International Symposium on Dynamics and Control of Process Systems (DYCOPS)*, Leuven, Belgium, July 2010.

[64] M. Berenguel, F. Rubio, A. Valverde, P. Lara, M. Arahal, E. Camacho, and M. López, "An artificial vision-based control system for automatic heliostat positioning offset correction in a central receiver solar power plant," *Science Direct*, vol. 76, pp. 563–575, December 2004.

[65] E. F. Camacho, M. Berenguel, F. R. Rubio, and D. Martínez, *Advances in Industrial Control: Control of Solar Energy Systems*. Springer, 2012.

[66] K. W. Stone and S. A. Jones, "Analysis of Solar Two heliostat tracking error sources," Sandia National Laboratories, Tech. Rep. SAND99-0239, 1999.

[67] D. L. King, "Beam quality and tracking accuracy evaluation of second-generation and Barstow production heliostats," Sandia National Labora-

tories, Tech. Rep. SAND82-0181, 1982.

[68] J. B. Blackmon, "Solar One beam characterization system design description and requirements document," McDonnell Douglas Astronautics Co., Tech. Rep. SAND-86-8179, 1986.

[69] C. L. Mavis, "10MW$_e$ solar thermal central receiver pilot plant heliostat and beam characterization system evaluation," Sandia National Laboratories, Tech. Rep. SAND874003, 1988.

[70] J. W. Strachan, "Revisiting the BCS, a measurement system for characterizing the optics of solar collectors," in *International Instrumentation Symposium*, no. 39, Albuquerque, May 1993.

[71] K. W. Stone and C. W. Lopez, "Evaluation of the Solar One track alignment methodology," *Solar Engineering*, vol. 1, pp. 521–526, 1995.

[72] S. Jones and K. Stone, "Analysis of strategies to improve heliostat tracking at Solar Two," Sandia National Laboratories, Tech. Rep. SAND99-0092, 1999.

[73] J. I. Burgaleta, A. Ternero, D. Vindel, I. Salbidegoitia, and G. Azcarrraga, "Gemasolar, key points for the operation of the plant," in *SolarPACES*, Marrakech, Morocco, September 2012.

[74] J. M. Lata, S. Alcalde, D. Fernández, X. Lekube, I. Agarraberes, E. Martínez, J. Olano, and I. Equidazu, "Commissioning and operation of commercial surrounding heliostat at fields for maximum but safe production," in *SolarPACES*, Granada, Spain, September 2011.

[75] D. S. Reznik, A. D. Azarchs, A. Csaszar, and M. B. Hartshorn, "Calibration and tracking control of heliostats in a central tower receiver solar power plant," U.S. Patent Application US 2009/0 107 485 A1, April 30, 2009.

[76] S. Schell, "Design and evaluation of eSolar's heliostat fields," *Solar Energy*, vol. 85, pp. 614–619, 2011.

[77] J. Schwarzbach and G. Kroyzer, "Heliostat calibration," U.S. Patent Application US 2010/0 139 644 A1, June 10, 2010.

[78] S. Jones, R. Edgar, and R. Houser, "Recent results on the optical performance of Solar Two heliostats," Sandia National Laboratories, Tech. Rep. SAND94-2776C, December 1994.

[79] K. Chong, "Optical analysis for simplified astigmatic correction of non-imaging focusing heliostat," *Solar Energy*, vol. 84, pp. 1356–1365, 2010.

[80] M. Vossiek, L. Wiebking, P. Gulden, J. Wieghardt, C. Hoffmann, and P. Heide, "Wireless local positioning," *IEEE Microwave Magazine*, vol. 4, no. 4, pp. 77–86, December 2003.

[81] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor

localization for very small devices," *IEEE personal communications*, vol. 7, no. 5, pp. 28–34, 2000.

[82] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *IEEE Computer*, vol. 34, no. 8, pp. 57–66, August 2001.

[83] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 13–32, 2009.

[84] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, vol. 37, no. 6, pp. 1067–1080, November 2007.

[85] P. Vorst, J. Sommer, C. Hoene, P. Schneider, C. Weiss, T. Schairer, W. Rosenstiel, A. Zell, and G. Carle, "Indoor positioning via three different RF technologies," in *4th European Workshop on RFID Systems and Technologies (RFID SysTech)*. VDE, 2008, pp. 1–10.

[86] A. Stelzer, A. Fischer, F. Weinberger, and M. Vossiek, "RF-sensor for a local position measurement system," in *NDE for Health Monitoring and Diagnostics*. International Society for Optics and Photonics, 2003, pp. 136–144.

[87] S. Hann, J.-H. Kim, S.-Y. Jung, and C.-S. Park, "White LED ceiling lights positioning system for optical wireless indoor applications," in *European Conference and Exhibition on Optical Communication (ECOC)*, September 2010.

[88] K. Panta and J. Armstrong, "Indoor localisation using white LEDs," *Electronics Letters*, vol. 48, no. 4, February 2012.

[89] S.-Y. Jung, S. Hann, and C.-S. Park, "TDOA-based optical wireless indoor localization using LED ceiling lamps," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1592–1597, November 2011.

[90] U. Nadeem, N. U. Hassan, M. A. Pasha, and C. Yuen, "Highly accurate 3D wireless indoor positioning system using white LED lights," *Electronics Letters*, vol. 50, no. 11, pp. 828–830, May 2014.

[91] U. Nadeem, N. U. Hassan, M. A. Pasha, and C. Yuen, "Indoor positioning system using visible LED lights: performance comparison of TDM and FDM protocols," *Electronics Letters*, vol. 51, no. 1, pp. 72–74, January 2015.

[92] H. Yang, J. W. M. Bergmans, and T. C. W. Schenk, "Illumination sensing in LED lighting systems based on frequency-division multiplexing," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4269–4281, November 2009.

[93] W. S. Murphy, "Determination of a position using approximate distances

and trilateration," Master's thesis, Colorado School of Mines, July 2007.

[94] W. A. Sethares, *Rhythm and Transforms*. Springer-Verlag London Limited, 2007, ch. 5, pp. 106–121.

[95] D. Lyon, "The discrete Fourier transform, part 4: Spectral leakage," *Journal of Object Technology*, vol. 8, no. 7, pp. 23–34, December 2009.

[96] T. Yoshizawa, S. Hirobayashi, and T. Misawa, "Noise reduction for periodic signals using high-resolution frequency analysis," *EURASIP Journal on Audio, Speech, and Music Processing*, 2011.

[97] P. K. Dash and S. Hasan, "A fast recursive algorithm for the estimation of frequency, amplitude, and phase of noisy sinusoid," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4847–4856, October 2011.

[98] K. Peters and S. Kay, "Unbiased estimation of the phase of a sinusoid," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2004, pp. 493–496.

[99] H. Fu and P. Y. Kam, "MAP/ML estimation of the frequency and phase of a single sinusoid in noise," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 834–845, March 2007.

[100] H. Yang, Y. Tu, H. Zhang, and K. Yang, "A Hilbert transform based method for dynamic phase difference measurement," in *24th Chinese Control and Decision Conference (CCDC)*, 2012.

[101] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, second edition ed. Springer, 2002, ch. Introduction, p. 15.

[102] S. K. Mitra, *Digital Signal Processing*. McGraw-Hill, 2011.

[103] A. D. Poularikas, *The Handbook of Formulas and Tables for Signal Processing*. CRC Press, 1999, ch. 15.

[104] MathWorks. (2016) Butterworth filter design. [Online]. Available: https://www.mathworks.com/help/signal/ref/butter.html

[105] B. Lathi, *Modern Digital and Analog Communication Systems*, third edition ed. Oxford University Press, Inc., 1998, ch. 12, pp. 541–545.

[106] S. S. Skiena and M. A. Revilla, *Programming Challenges*. Springer-Verlag New York, Inc., 2003, ch. 11, p. 246.

[107] S. J. Orfanidis, *Electromagnetic Waves and Antennas*. Rutgers University, 2016, ch. 5&14, pp. 153–185&663–708.

[108] R. G. Lyons. (2011) Sum of two sinusoids. [Online]. Available: http://dspguru.com/sites/dspguru/files/Sum_of_Two_Sinusoids.pdf

Appendix A

MATLAB Code

This appendix showcases some code snippets that are self-contained and simple enough to understand without additional context.

A.1 Methods for phase difference calculation

A.1.1 Dot Product

```
function pd = pddp(s1, s2)
    % PDDP Returns the phase difference between the two signals s1 and s2
    % using the Dot Product method.
    %   The returned phase difference is in the range [0, pi], regardless
    %   of which signal is leading or lagging.

    arg = dot(s1, s2) / (norm(s1) * norm(s2));
    pd = acos(max(-1, min(1, arg)));
end
```

A.1.2 Dot Product with Noise Compensation

```
function pd = pddpnc(s1, s2, o1, o2, covn)
    % PDDPNC Returns the phase difference between the two signals s1 and s2
    % using the Dot Product with Noise Compensation method. o1 and o2 are
    % the standard deviations of the noise on each signal, and covn is the
    % covariance of the noise on the two signals.
    %   The returned phase difference is in the range [0, pi], regardless
    %   of which signal is leading or lagging.

    if ~exist('covn', 'var')
        covn = 0;
    end
    u1 = mean(s1);
    u2 = mean(s2);
    A1 = real(sqrt(max(0, (mean(s1.^2) - u1^2 - o1^2)*2)));
    A2 = real(sqrt(max(0, (mean(s2.^2) - u2^2 - o2^2)*2)));
    arg = (mean(s1.*s2) - u1*u2 - covn) / (A1*A2/2);
    pd = acos(max(-1, min(1, arg)));
end
```

A.1.3 Hilbert Transform

```
function pd = pdht(s1, s2)
    % PDHT Returns the phase difference between the two signals s1 and s2
    % using the Hilbert Transform method.
    %   The returned phase difference is in the range (-pi, pi], and
    %   positive when s2 is leading.

    s1h = hilbert(s1 - mean(s1));
    s2h = hilbert(s2 - mean(s2));
    pdvec = atan2(imag(s2h), real(s2h)) - atan2(imag(s1h), real(s1h));
    pd = atan2(sum(sin(pdvec)), sum(cos(pdvec)));
end
```

A.2 Oscilloscope Export Automation

The oscilloscope is connected to the computer using a General Purpose Interface Bus (GPIB) bus which allows the export of measurements from the oscilloscope to be automated.

```
function rs = readscope(S)
    % READSCOPE Reads data from the scope.
    %   readscope(S) reads S times into resultset rs.

    rs = struct('data', {}, 'settings', '', 'preamble', '');

    spec = gpib('ni', 0, 6, 'EOSMode', 'read');
    set(spec, 'InputBufferSize', 10000);
    fopen(spec);
    disp('Configure and read scope settings...');
    fprintf(spec, 'DATA:SOURCE CH1,CH2');
    fprintf(spec, 'SET?');
    rs.settings = fscanf(spec);
    fprintf(spec, 'WFMPR?');
    rs.preamble = fscanf(spec);
    disp('Done');

    while length(rs.data) < S
        fprintf(spec, 'CURVE?');
        out = fscanf(spec);
        if out(1:6) ~= ':CURVE'
            disp('ERROR: Invalid response for CURVE');
            break;
        end
        data = str2num(out(7:end));
        ch1 = data(1:1000);
        ch2 = data(1001:end);
        if length(rs.data) > 1 && all(all(rs.data{end} == [ch1; ch2]))
            disp('WARNING: No change in waveform. Retrying.');
            pause(1);
            continue;
        end
        rs.data{end+1} = [ch1; ch2];
        pause(.5);
        disp(sprintf('Progress: %d/%d', length(rs.data), S));
    end
    fclose(spec);
end
```

Appendix B

Optimal Transmitter Layouts

Table B.1: Transmitter coordinates for optimal layouts in 3D, with (a) - (f) corresponding to Figure 4.23

TX #	x	y	z
1	0.010	-0.120	0.20
2	0.607	0.180	0.20
3	-0.597	0.180	0.20
4	0	1.000	0.05
5	0	-1.000	0.05

(a)

TX #	x	y	z
1	0.367	0.250	0.20
2	-0.432	0.199	0.20
3	0.040	-0.490	0.20
4	-0.014	1.000	0.05
5	-0.809	-0.588	0.05
6	0.853	-0.521	0.05

(b)

TX #	x	y	z
1	-0.020	0.010	0.20
2	-0.321	0.497	0.20
3	-0.223	-0.530	0.20
4	0.585	-0.002	0.20
5	0.500	0.866	0.05
6	-1.000	0	0.05
7	0.488	-0.873	0.05

(c)

TX #	x	y	z
1	0	0	0.20
2	0.170	0.607	0.20
3	-0.664	0.167	0.20
4	-0.109	-0.614	0.20
5	0.640	-0.120	0.20
6	0.661	0.744	0.05
7	-0.921	0.390	0.05
8	0.166	-0.986	0.05

(d)

TX #	x	y	z
1	-0.010	0.020	0.20
2	0.628	0.018	0.20
3	-0.010	0.630	0.20
4	-0.655	-0.067	0.20
5	0.0121	-0.622	0.20
6	1.000	0	0.05
7	0.001	1.000	0.05
8	-1.000	0	0.05
9	0	-1.000	0.05

(e)

TX #	x	y	z
1	0.020	0	0.20
2	0.267	0.609	0.20
3	-0.428	0.439	0.20
4	-0.583	-0.261	0.20
5	0.078	-0.619	0.20
6	0.643	-0.070	0.20
7	0.811	0.580	0.05
8	-0.416	0.910	0.05
9	-0.959	-0.283	0.05
10	0.355	-0.935	0.05

(f)

Appendix C

SASEC2016 Paper

A METHOD FOR ACCURATE MEASUREMENT OF HELIOSTAT MIRROR ORIENTATION

Benjamin D. Swart[1], Herman A. Engelbrecht[2], and Johann Treurnicht[3]

[1] Solar Thermal Energy Research Group, Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa; Phone: +27 21 808 4016; Fax: +27 21 808 4933; E-Mail: bdswart@sun.ac.za

[2] MIH Media Lab, Stellenbosch University; E-Mail: hebrecht@sun.ac.za

[3] Stellenbosch University; E-mail: jtreurn@sun.ac.za

Abstract

A method is developed for the measurement of the mirror orientations of all heliostats in a field simultaneously. Tower mounted electromagnetic transmitters are placed at known positions around the perimeter of the heliostat field, transmitting identical sinusoidal signals. An electromagnetic receiver in the field uses a Hilbert transform-based method to calculate the phase difference between a pair of signals, and using the wavelength, calculates the difference in distance between itself and the corresponding pair of transmitters. The receiver position is then approximated using multilateration with a non-linear least squares algorithm. By attaching three receivers to the heliostat mirror surface at distinct points, a plane parallel to the mirror surface is formed which is used to determine the mirror orientation. A simulation of the system implementing the proposed method is constructed to verify that the method works, and results show that for the error sources included in the model it is theoretically possible to achieve a tracking error of one milliradian or less.

Keywords: Heliostat; tracking; closed-loop; orientation.

1. Introduction

The central receiver system (CRS) is a relatively new technology within the field of concentrating solar power (CSP) and predictions indicate significant opportunities to reduce the levelised cost of electricity (LCOE) by as much as 40% from 2010 to 2020 [1] by increasing performance and lowering costs. The performance of CRS plants is highly dependent on the efficiency of the heliostat field, which in itself contributes an estimated 30% - 50% of the total plant cost [2] [3].

Heliostat field efficiency is increased by reflecting a greater portion of the solar radiation incident on the field onto the receiver. This typically requires an increased accuracy in heliostat tracking. Open-loop tracking is hindered by problems such as mechanical tolerances, while current closed-loop tracking methods are either impractical, not accurate enough, or too expensive [4] [5] [6] [7]. The current industry norm seems to be open-loop tracking using an error model [8] for a tracking error of less than one milliradian. This, however, requires periodic recalibration of error parameters which can take up to three weeks for the entire field [9] [10].

This paper investigates heliostat mirror orientation as a possible source of feedback for closed-loop tracking systems. A method is developed to accurately determine the mirror orientations of all heliostats simultaneously with an accuracy that aims to match or exceed current standards. The concept for such a system is developed in the following section, after which some of the core aspects are developed in detail. The results demonstrate a working simulation of the system and investigate the effects of some error sources on its performance. The paper is concluded with some final remarks.

2. Conceptual Overview of Proposed System

In the past, sites such as Solar One have used inclinometers to measure the heliostat azimuth axis tilt angle [11]. While such methods can certainly be used to further improve results, they are not the focus of this paper. Instead, this paper focuses on an approach where the heliostat orientation is obtained by accurately measuring the coordinates of three or more distinct points on its mirror surface. These points define a plane which, when parallel to the mirror, is used to calculate the heliostat normal vector. For simplicity, it is assumed in this paper that the plane and mirror surface are parallel, though it is possible to develop an error model to compensate for any non-parallelism.

The approach for determining the coordinates of a single point within a predetermined three-dimensional space is based on a 2007 master's thesis by Murphy [12] which addresses a similar problem. Electromagnetic receivers are placed at three or more points on the mirror surface for which the coordinates are required. Tower mounted transmitters are placed at known positions around the perimeter of the heliostat field as shown in Fig. 1. A single central oscillator is connected to each transmitter to ensure that all transmitters transmit a sinusoidal signal of exactly the same frequency, though the phases may differ. The distance d_i between a receiver and transmitter T_i results in a phase shift of ϕ_i from the time the signal is transmitted to the time it is received. While measuring ϕ_i may prove problematic, it is possible to measure the difference in phase shift $\phi_i - \phi_j$ between two signals (§3.1) from transmitters T_i and T_j. This value is proportional to $d_i - d_j$, the distance between the receiver and transmitter T_i relative to the distance between the receiver and transmitter T_j. When the wavelength, and thus frequency, of the sinusoid is known, $d_i - d_j$ can be calculated exactly (§3.2). These measured distance differences, along with the known positions of the transmitters (§3.3), are then used in an algorithm using multilateration and non-linear least squares to calculate the position of the receiver (§3.4). When the positions of all receivers on the mirror surface have been calculated, the corresponding plane and resulting normal vector is calculated (§3.5). This process is illustrated in Fig. 2.

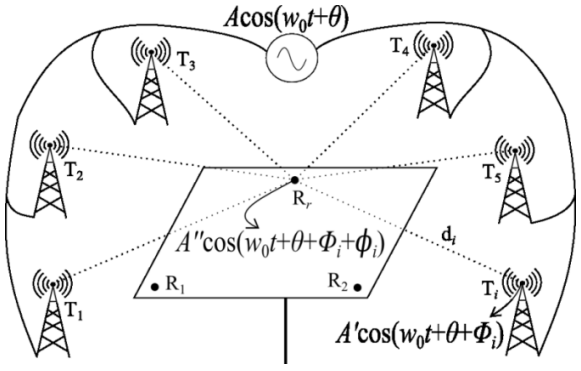


Fig. 1 Conceptual illustration of transmitter towers surrounding heliostat field.

Even though the frequency of all transmitted signals is the same, each signal undergoes a different phase shift Φ_i from the time it is generated at the central oscillator to the time it is transmitted at T_i. These values may change slowly throughout the day and need to be periodically recalculated. This is done by placing a calibration receiver at a known position within the field which uses its own position, the known positions of the transmitters, and the expected phase shift due to the known distance d_i to calculate the value of Φ_i for each T_i (§3.6).

Transmitters use frequency division multiplexing (FDM) for concurrent transmission of all signals over the shared medium (air). That is, each transmitter T_i modulates the sinusoidal signal onto a unique carrier frequency which allows a receiver to demodulate the received signals and, based on the carrier frequency, determine which T_i each signal originated from.

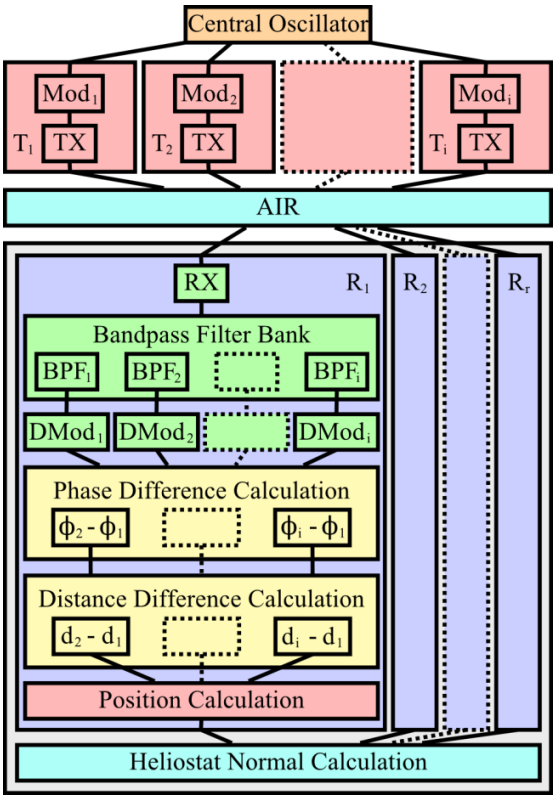


Fig. 2 Overview of process leading to calculation of heliostat mirror normal vector. RX and TX represent electromagnetic transmitters and receivers, and $\mathbf{Mod}_i$ and $\mathbf{DMod}_i$ represent modulators and demodulators for the carrier frequency of transmitter $\mathbf{T}_i$.

3. Development of Core Aspects

3.1. Phase Difference Calculation

Three methods are considered for the calculation of phase difference between two sinusoids of matching frequency: dot product, dot product with noise compensation, and the Hilbert transform. Omitted methods include the discrete Fourier transform due to the problem of spectral leakage, and cross correlation due to the maximum achievable accuracy being limited by sampling frequency. The two signals are defined as

$$s_1(t) = a_1 \cos\big(\omega_0 t + \phi_1(t)\big) + n_1(t)$$
$$s_2(t) = a_2 \cos\big(\omega_0 t + \phi_2(t)\big) + n_2(t),$$

where $n_1(t)$ and $n_2(t)$ are ergodic random variables representing noise that is uncorrelated with the sinusoids.

3.1.1. Dot Product

One of the simplest ways to calculate the phase difference between two sinusoids is by using the dot product. Let $\vec{s}_1$ be a vector containing S samples from $s_1(t)$ sampled at a frequency of $1/T$, where the ith sample is given by

$$\vec{s}_1[i] = s_1(iT), \qquad 0 \le i \le S-1,$$

and let $\vec{s}_2$ be given similarly. When $\phi_1(t)$ and $\phi_2(t)$ remain constant, then it follows from the dot product that

$$\vec{s}_1 \cdot \vec{s}_2 = \|\vec{s}_1\| \, \|\vec{s}_2\| \cos(\phi_{21}),$$

where $\|\vec{s}_1\|$ and $\|\vec{s}_2\|$ denote the sizes of vectors $\vec{s}_1$ and $\vec{s}_2$, respectively, and ϕ_{21} is the phase difference between the two sinusoids. The equation is solved for ϕ_{21} as follows:

$$\phi_{21} = |\phi_2 - \phi_1| = \operatorname{acos}\left(\frac{\vec{s}_1 \cdot \vec{s}_2}{\|\vec{s}_1\| \, \|\vec{s}_2\|}\right).$$

3.1.2. Dot Product with Noise Compensation

One shortcoming of the dot product method is that it does not account for noise. When the mean $\mathrm{E}[n]$ and variance $\mathrm{Var}[n]$ of the noise is known, it is possible to compensate for the effect of noise on the calculated phase difference.

Multiplication of $s_1(t)$ and $s_2(t)$ yields

$$\begin{aligned} s(t) &= [s_1(t)][s_2(t)] \\ &= \frac{a_1 a_2}{2}\cos(\phi_2 - \phi_1) + \frac{a_1 a_2}{2}\cos(2\omega_0 t + \phi_2 + \phi_1) \\ &\quad + a_1 \cos(\omega_0 t + \phi_1)\, n_2(t) \\ &\quad + a_2 \cos(\omega_0 t + \phi_2)\, n_1(t) + n_1(t) n_2(t). \end{aligned}$$

The mean of the signal $s(t)$ is calculated as

$$s_u = \frac{1}{T}\int_0^T s(t)\mathrm{d}t = \frac{a_1 a_2}{2}\cos(\phi_2 - \phi_1) + \mathrm{E}[n_1 n_2],$$

which follows from the fact that the mean of the periodic terms approach zero as T becomes sufficiently large, and where $\mathrm{E}[n_1 n_2]$ is the expected value (mean) of the product of n_1 and n_2. Rearranging the terms gives the phase difference as

$$\phi_{21} = |\phi_2 - \phi_1| = \operatorname{acos}\left(\frac{2(s_u - \mathrm{E}[n_1 n_2])}{a_1 a_2}\right).$$

The amplitudes of the original signals, a_1 and a_2, are calculated by multiplying the original signal by itself, taking the average, and then rearranging the terms:

$$[s_1(t)]_u^2 = \frac{1}{T}\int_0^T [s_1(t)]^2 \mathrm{d}t = \frac{a_1^2}{2} + \mathrm{E}[n_1^2],$$

from which it follows that

$$a_1 = \sqrt{2([s_1(t)]_u^2 - \mathrm{E}[n_1^2])}, \qquad \mathrm{E}[n_1^2] = \mathrm{Var}[n_1] + \mathrm{E}[n_1]^2,$$

where $\mathrm{Var}[n_1]$ is the variance of n_1. The value of a_2 is calculated similarly. By letting $u_{12} = \mathrm{E}[n_1 n_2]$, $u_1 = \mathrm{E}[n_1]$, $u_2 = \mathrm{E}[n_2]$, $\sigma_1^2 = \mathrm{Var}[n_1]$, and $\sigma_2^2 = \mathrm{Var}[n_2]$, the final equation for the phase difference becomes

$$\phi_{21} = \operatorname{acos}\left(\frac{s_u - u_{12}}{\sqrt{([s_1(t)]_u^2 - \sigma_1^2 - u_1^2)([s_2(t)]_u^2 - \sigma_2^2 - u_2^2)}}\right).$$

It is important to note that neither this nor the previous method includes the sign of the angle ϕ_{21}.

3.1.3. Hilbert Transform

Let $\hat{s}_1(t)$ and $\hat{s}_2(t)$ denote the Hilbert transforms of $s_1(t)$ and $s_2(t)$, respectively, given by

$$\begin{aligned} \hat{s}_1(t) &= a_1 \sin\big(\omega_0 \mathrm{t} + \phi_1(t)\big) + \hat{n}_1(t) \\ \hat{s}_2(t) &= a_2 \sin\big(\omega_0 t + \phi_2(t)\big) + \hat{n}_2(t), \end{aligned}$$

and let $h_1(t)$ and $h_2(t)$ be the positive analytical signals

$$h_1(t) = s_1(t) + j\hat{s}_1(t), \qquad h_2(t) = s_2(t) + j\hat{s}_2(t).$$

The analytical signals can be viewed as rotating vectors with angle $\Theta(t) = \omega_0 t + \phi(t) + \gamma(t)$, where $\gamma(t)$ is the variation in phase due to noise. The angle Θ of a complex signal h is

$$\Theta = \arg(h) = \operatorname{atan2}\big(\operatorname{imag}(h), \operatorname{real}(h)\big),$$

which leads to the difference in angle between two complex signals $h_1(t)$ and $h_2(t)$ being calculated as

$$\begin{aligned} \Theta_2(t) - \Theta_1(t) &= \phi_2(t) - \phi_1(t) + \gamma_2(t) - \gamma_1(t) \\ &= \operatorname{atan2}\big(\hat{s}_2(t), s_2(t)\big) - \operatorname{atan2}\big(\hat{s}_1(t), s_1(t)\big). \end{aligned}$$

When the signal-to-noise ratio is large enough,

$$\begin{aligned} \gamma_2(t) - \gamma_1(t) &\approx 0, \\ \phi_2(t) - \phi_1(t) &= \operatorname{atan2}\big(\hat{s}_2(t), s_2(t)\big) - \operatorname{atan2}\big(\hat{s}_1(t), s_1(t)\big). \end{aligned}$$

Additionally, if the phase difference is assumed to be constant, the accuracy of the results can be improved by averaging over time. Using phasor geometry to account for phase wrapping, the averages of the real and imaginary components of the phase difference are calculated as

$$\begin{aligned} x &= \frac{1}{T}\int_0^T \cos\big(\phi_2(t) - \phi_1(t)\big)\,\mathrm{d}t \\ y &= \frac{1}{T}\int_0^T \sin\big(\phi_2(t) - \phi_1(t)\big)\,\mathrm{d}t. \end{aligned}$$

The resulting phase difference, averaged over time, is then

$$\phi_{21} = \phi_2 - \phi_1 = \operatorname{atan2}(y, x).$$

3.2. Translating Phase Difference to Distance

A phase difference $\phi_i - \phi_j \in (-\pi, \pi]$ between two sinusoids with matching frequency and wavelength λ is translated into distance by the equation

$$d_i - d_j = -\left(\frac{\phi_i - \phi_j - (\Phi_i - \Phi_j)}{2\pi}\lambda + n\lambda\right),$$

where $d_i - d_j$ represents the distance between the receiver and transmitter T_i relative to the distance between the receiver and transmitter T_j. A negative value of $\phi_i - \phi_j$ indicates that T_j is further away than T_i, hence the minus prefix.

The $n\lambda$ term indicates that any number of full wavelengths may be added to the result due to the periodic nature of a sinusoidal wave. For example, a distance of $\lambda/4$ and $5\lambda/4$ will both yield a phase difference of $\pi/2$. When the position of the receiver is approximately known, the correct value of n can be substituted into the equation. Another solution is to transmit more than one sinusoid such that the sinusoid with a longer wavelength is used to approximate while the sinusoid with a shorter wavelength is used to accurately determine the receiver position. In this paper it is assumed that the wavelength λ is longer than the diameter of the field such that n is always zero.

3.3. Transmitter Layout

There are many factors to consider for the optimal layout of transmitter towers. Examples of such factors include the distances between transmitters and receivers (greater distances require higher signal power) and line of sight (the changing orientation of a heliostat limits the set of transmitters it has line of sight to). Taking into account all possible factors is a study in its own right. The only factor considered here is the sensitivity of the transmitter geometry to noise.

3.4. Position Calculation

The calculated position $\hat{R}$ of a receiver at R is determined using multilateration. Let there be N transmitters located at known positions T_i, with the distance between each T_i and R given by d_i, and the distance between each T_i and $\hat{R}$ given by $\hat{d}_i$. Due to noise and other factors, a small error $\epsilon_i = \hat{d}_i - d_i$ exists, which is minimised using a non-linear least squares algorithm.

When d_i is measured relative to a reference distance d_r instead of being an absolute measurement, then $d_i = d_r + \Delta_i$, where d_r is the reference distance and Δ_i is the difference between d_i and d_r. When relative distances are used, $N \geq D + 2$, where D is the number of dimensions. For three dimensions, $N \geq 5$.

The error function is properly defined as

$$\begin{aligned}\epsilon_i(x,y,z) &= \hat{d}_i - d_i \\ &= \sqrt{(x-T_{i,x})^2 + (y-T_{i,y})^2 + (z-T_{i,z})^2} - (d_r + \Delta_i),\end{aligned}$$

where x, y and z are the calculated position of the receiver $\hat{R}$. The goal is to minimise the sum of the squares of the errors

$$E(x,y,z) = \sum_{i=1}^{N} \epsilon_i(x,y,z)^2.$$

Differentiating E with respect to x yields

$$\frac{\partial E}{\partial x} = 2\sum_{i=1}^{N} \epsilon_i \frac{\partial \epsilon_i}{\partial x}$$

with $\partial E/\partial y$, $\partial E/\partial z$ and $\partial E/\partial d_r$ obtained similarly. The derivative of ϵ_i with respect to x is given by

$$\frac{\partial \epsilon_i}{\partial x} = \frac{x - T_{i,x}}{\sqrt{(x-T_{i,x})^2 + (y-T_{i,y})^2 + (z-T_{i,z})^2}} = \frac{x - T_{i,x}}{\epsilon_i + d_r + \Delta_i}$$

with $\partial \epsilon_i/\partial y$, $\partial \epsilon_i/\partial z$ obtained similarly, and $\partial \epsilon_i/\partial d_r = -1$.

Introducing the vectors $\vec{\epsilon}$, $\vec{g}$ and the Jacobian matrix $\boldsymbol{J}$, leads to

$$\vec{g} = 2\boldsymbol{J}^T\vec{\epsilon},$$

$$\vec{g} = \begin{bmatrix} \frac{\partial E}{\partial x} \\ \frac{\partial E}{\partial y} \\ \frac{\partial E}{\partial z} \\ \frac{\partial E}{\partial d_r} \end{bmatrix}, \quad \boldsymbol{J} = \begin{bmatrix} \frac{\partial \epsilon_1}{\partial x} & \frac{\partial \epsilon_1}{\partial y} & \frac{\partial \epsilon_1}{\partial z} & \frac{\partial \epsilon_1}{\partial d_r} \\ \frac{\partial \epsilon_2}{\partial x} & \frac{\partial \epsilon_2}{\partial y} & \frac{\partial \epsilon_2}{\partial z} & \frac{\partial \epsilon_2}{\partial d_r} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \epsilon_N}{\partial x} & \frac{\partial \epsilon_N}{\partial y} & \frac{\partial \epsilon_N}{\partial z} & \frac{\partial \epsilon_N}{\partial d_r} \end{bmatrix}, \quad \vec{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}.$$

Using the vector

$$\vec{\beta} = [x \quad y \quad z \quad d_r]^T,$$

where x, y and z are the calculated position of $\hat{R}$, Newton iteration gives

$$\vec{\beta}_{k+1} = \vec{\beta}_k - (\boldsymbol{J}_k^T\boldsymbol{J}_k)^{-1}\boldsymbol{J}_k^T\vec{\epsilon}_k,$$

where $\vec{\beta}_k$ denotes the k^{th} approximate solution. The subscript k in $\boldsymbol{J}$ and $\vec{\epsilon}$ means that theses quantities are evaluated at $\vec{\beta}_k$. Also given is

$$\boldsymbol{J}^T\boldsymbol{J} = \begin{bmatrix} \sum_{i=1}^{N} \frac{(x-T_{i,x})^2}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{(x-T_{i,x})(y-T_{i,y})}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{(x-T_{i,x})(z-T_{i,z})}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{-(x-T_{i,x})}{\epsilon_i + d_r + \Delta_i} \\ \sum_{i=1}^{N} \frac{(x-T_{i,x})(y-T_{i,y})}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{(y-T_{i,y})^2}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{(y-T_{i,y})(z-T_{i,z})}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{-(y-T_{i,y})}{\epsilon_i + d_r + \Delta_i} \\ \sum_{i=1}^{N} \frac{(x-T_{i,x})(z-T_{i,z})}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{(y-T_{i,y})(z-T_{i,z})}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{(z-T_{i,z})^2}{(\epsilon_i + d_r + \Delta_i)^2} & \sum_{i=1}^{N} \frac{-(z-T_{i,z})}{\epsilon_i + d_r + \Delta_i} \\ \sum_{i=1}^{N} \frac{-(x-T_{i,x})}{\epsilon_i + d_r + \Delta_i} & \sum_{i=1}^{N} \frac{-(y-T_{i,y})}{\epsilon_i + d_r + \Delta_i} & \sum_{i=1}^{N} \frac{-(z-T_{i,z})}{\epsilon_i + d_r + \Delta_i} & \sum_{i=1}^{N} 1 \end{bmatrix},$$

$$\boldsymbol{J}^T\vec{\epsilon} = \begin{bmatrix} \sum_{i=1}^{N} \frac{x-T_{i,x}}{\epsilon_i + d_r + \Delta_i}\epsilon_i & \sum_{i=1}^{N} \frac{y-T_{i,y}}{\epsilon_i + d_r + \Delta_i}\epsilon_i & \sum_{i=1}^{N} \frac{z-T_{i,z}}{\epsilon_i + d_r + \Delta_i}\epsilon_i & \sum_{i=1}^{N} -\epsilon_i \end{bmatrix}^T.$$

3.5. Receiver Plane Normal Vector

When the positions of three points (R_1, R_2, R_3) on the mirror plane are known, the unit vector $\vec{u}$ normal to the plane is calculated using the cross product as

$$\vec{v} = (R_2 - R_1) \times (R_3 - R_1), \qquad \vec{u} = \frac{\vec{v}}{\|\vec{v}\|}.$$

When the receivers are numerically arranged counter-clockwise as viewed from the front, then $\vec{u}$ is in the forward facing direction of the heliostat, otherwise it is backwards.

3.6. Electronic Phase Delay Calculation

The calibration receiver at R_c calculates the phase differences of the signals it receives in exactly the same way as any other receiver does. The value of the phase shift Φ_i from the central oscillator to transmitter T_i is then equal to the sum of the expected phase shift due to the receiver's distance from T_i and the measured phase difference ϕ_i:

$$d_i = \sqrt{(T_{i,x} - R_{c,x})^2 + (T_{i,y} - R_{c,y})^2 + (T_{i,z} - R_{c,z})^2},$$

$$\Phi_i = \frac{2\pi d_i}{\lambda} + \phi_i.$$

The value of Φ_i is wrapped to fit in the range $(-\pi, \pi]$.

4. Results

4.1. Phase Difference Calculation

Fig. 3, Fig. 4 and Fig. 5 show the accuracy with which the phase difference between two sinusoids of matching frequency can be calculated using the methods described in section 3.1. All simulations assume zero-mean noise with constant power spectral density (PSD). The signal-to-noise ratio (SNR) is indicated on the x-axis and the standard deviation of the error is indicated on the y-axis in degrees. The SNR is the same for both signals, except where indicated otherwise. Table 1 clarifies the legend used in the figures.

Key	Meaning
S	Number of samples.
SPC	Samples per cycle, which is a measure of the normalised sampling frequency.
B	Equivalent normalised noise bandwidth. No filtering is indicated by 'none'.
SNR2	Indicates that the SNR of the second signal is SNR2-times more than the SNR of the first signal.

Table 1 Clarification of the legends for Figures 3, 4 and 5.

Recall that of the three methods investigated, only the Hilbert Transform method correctly includes the sign of the angle ϕ_{21}. It therefore only makes sense to record the absolute value of the error for the other two methods, which would inevitably introduce a positive non-zero bias in the error mean. This makes comparison difficult and is avoided by assuming the correct sign for ϕ_{21} in the simulation for these methods.

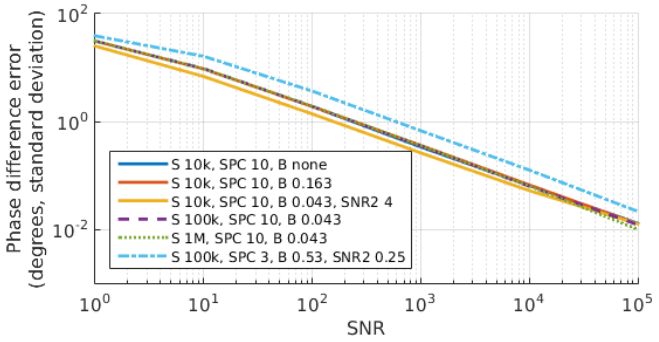


Fig. 3 Performance of Dot Product method.

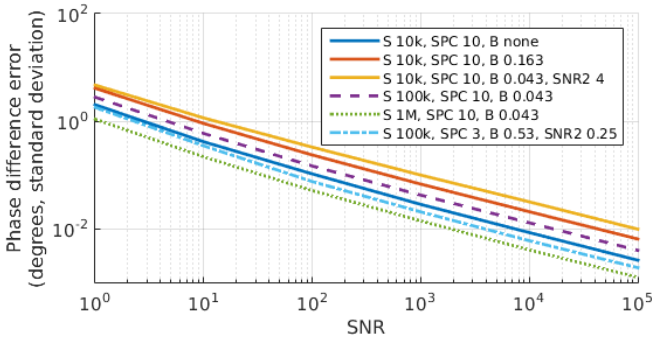


Fig. 4 Performance of Dot Product with Noise Compensation method.

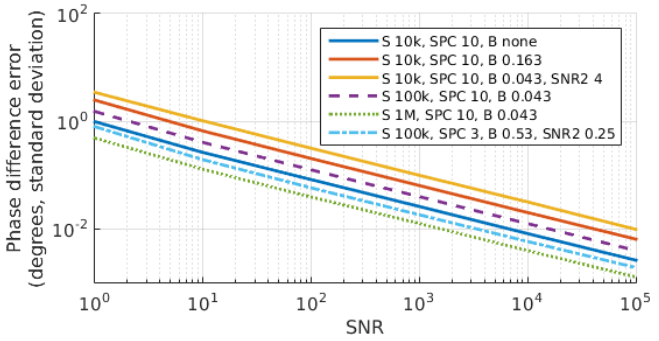


Fig. 5 Performance of Hilbert Transform method.

The resulting error mean for all graphs is approximately zero (mean $\ll$ standard deviation) so the standard deviation of the error is used as the metric to compare results. All figures clearly show a decrease in the resulting phase difference error as the SNR is increased. Fig. 3 shows that the sampling frequency, number of samples and filter bandwidth have little to no effect on the results of the first method. For the latter two methods, however, it is clear from Fig. 4 and Fig. 5 that an increase in the number of samples as well as a wider bandwidth (while maintaining the same SNR, thus lower PSD) both contribute towards a smaller error. As expected, the accuracy of the dot product method is significantly increased when noise compensation is added, to the point where it is almost identical

to the accuracy of the Hilbert transform method. Keep in mind, however, that the former requires the noise characteristics to be exactly known, whereas the latter has no such requirement. The results in Fig. 4 quickly approach those of Fig. 3 (or worse) when the noise characterisation contains errors. Fig. 6 shows how the accuracy of the Hilbert transform method is independent of the phase difference, whereas the other methods are more sensitive to noise at certain phase differences. The Hilbert transform is clearly superior to the other methods.

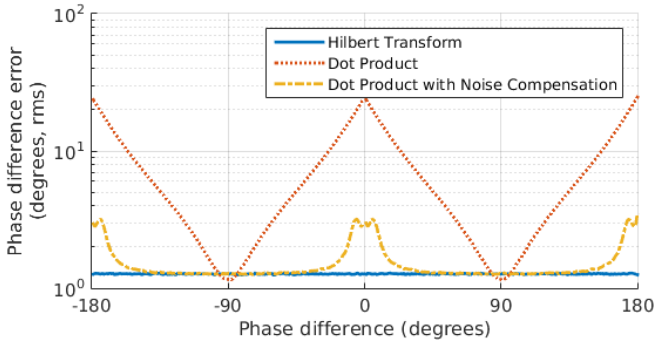


Fig. 6 Sensitivity of methods at various phase differences with SNR = 10.

4.2. Transmitter Layout

The placement of transmitter towers plays an important part in the overall accuracy of the system. Poor transmitter placement can result in large receiver position errors even when phase difference errors are small. From experimentation with the layout of transmitter towers the following guidelines for their optimal placement emerge:

- Any two transmitters should not be placed close to or on the same location.
- Any three transmitters should not be placed close to or on the same line.
- Any four transmitters should not be placed close to or on the same plane.
- Transmitters should surround and be placed outside the field of receivers, not within.

The last guideline is especially important when working with relative distances as the error rapidly increases once a receiver is outside the enclosure of transmitters as shown in Fig. 7. The optimal layout of transmitters in two dimensions is equally distributed on a circle surrounding the field of receivers. For three dimensions, the optimal layout is equally distributed on a sphere surrounding the field of receivers. Fig. 7 shows how the error on the receiver position is affected by varying the number of transmitters and the error on the measured distances. An optimal layout for two dimensions is used for simplicity and transmitters are equally distributed on a circle with normalised radius of unity. It is evident that an increased number of transmitters as well as an increased accuracy in the measurement of the distances both contribute to a smaller error in receiver position.

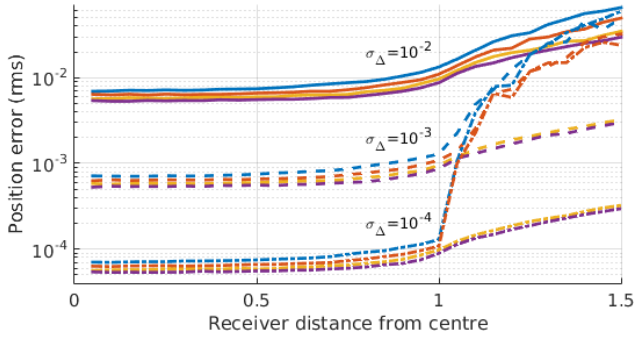


Fig. 7 Receiver position error as affected by number of transmitters and error on measured distances. The number of transmitters is 4 (blue), 5 (red), 6 (orange) and 7 (purple).

For practical reasons an ideal layout may not always be possible. One such constraint is the maximum possible height of a transmitter tower. The transmitter layout used in the remainder of this paper is shown in Fig. 8 and described in Table 2. Note that this does not necessarily represent an optimal layout as transmitter positions are constrained by both a minimum and a maximum height. The layout represents a flattened, elevated dome above the heliostat field. Coordinates are normalised to a field with radius of unity to enable the prediction of results for a larger field by simply scaling the normalised results. The performance of this layout is analysed in the following section.

TX #	X	Y	Z
1	1	0	0.2
2	0.5	0.866	0.3
3	-0.5	0.866	0.2
4	-1	0	0.3
5	-0.5	-0.866	0.2
6	0.5	-0.866	0.3
7	0.433	0.25	0.4
8	-0.433	0.25	0.4
9	0	-0.5	0.4
10	0	0	0.5

Table 2 Normalised coordinates for transmitter layout.

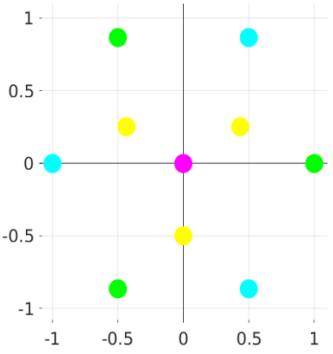


Fig. 8 Top view of transmitter layout. Colours correlate to height.

4.3. Position Calculation

The simulated receiver position error for the transmitter layout proposed in Table 2 is shown in Fig. 9. The results show that for the proposed layout, an error on the distance measurement (σ_Δ) will result in a position RMS error ϵ_R of $0.99\sigma_\Delta$ at the

centre of the field and $1.85\sigma_\Delta$ at the edge. The error at the edge of the field can be reduced by increasing the transmitter radius, while the overall error can be reduced by adding more transmitters and optimising the transmitter layout.

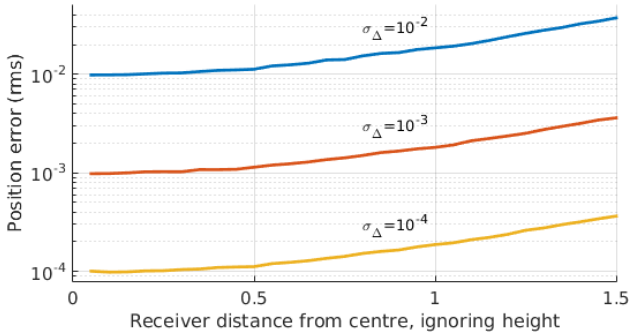


Fig. 9 Receiver position error for proposed transmitter layout in Table 2 for varying error on measured distances.

4.4. Receiver Plane Normal Vector

Fig. 10 shows how the accuracy with which the receiver plane normal can be determined is affected by errors in the estimated receiver positions. Three receivers, all a distance d_R apart, form the receiver plane and the RMS error in each receiver position is ϵ_R. The relationship d_R/ϵ_R is shown on the x-axis, with the resulting error of the plane normal on the y-axis. Fig. 10 shows clearly that the plane normal error decreases as d_R/ϵ_R increases, that is, as either the position error decreases or the receivers are placed a greater distance apart. From Fig. 10, the RMS of the tracking error T_{RMS} is approximated by

$$T_{RMS} = \left(1000\sqrt{4/3}\right)\frac{\epsilon_R}{d_R},$$

where T_{RMS} represents the overall tracking accuracy of the system, in milliradians. For a tracking error of less than one milliradian, a d_R/ϵ_R relationship of above ~ 1150 is required. The results assume that the receiver position errors are uncorrelated, though it can be shown that a positive correlation leads to a smaller RMS tracking error T_{RMS} for the same d_R/ϵ_R.

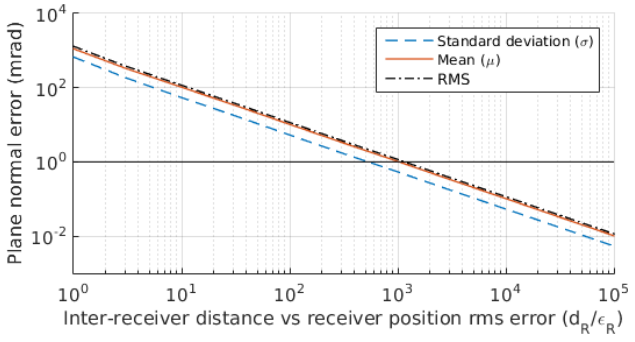


Fig. 10 Receiver plane normal error as a function of receiver position error and inter-receiver distance.

4.5. Overall

The combined performance of all the previously discussed components is shown in Fig. 11 with the system parameters in the caption. The figure shows that an RMS tracking error of less than 10 milliradians is achievable with 100k samples and that the error approaches the 1 milliradian mark as the number of samples is increased. The maximum value shown for S is 10^6 samples, though this number can be increased until the desired accuracy is achieved.

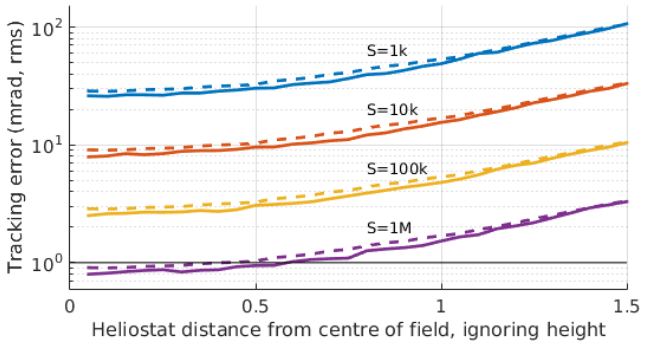


Fig. 11 Overall tracking error with transmitters as in Table 2. $SPC = 10$, $SNR = 10^5$, $B = 0.04$, $\lambda = 4$, $d_R = 0.0173$. Simulated graphs are solid and predictions are dashed.

The above figure also shows the accuracy with which the final error can be predicted, given the error of the individual components and the system parameters. The equation that accurately describes the variance of the error of the Hilbert transform method shown in Fig. 5 is given by

$$\sigma_\phi^2 = \frac{1}{S}\left(\frac{1}{B_1(\mathrm{SNR}_1)} + \frac{1}{B_2(\mathrm{SNR}_2)}\right),$$

where S is the number of samples, B_i is the equivalent normalised noise bandwidth ($\times \sim 2/f_s$) of signal i with sampling frequency f_s, and SNR_i is the signal to noise ratio of signal i. With the given parameters and $S = 10^6$ samples, the tracking error at the edge of the field is predicted by calculating the phase difference error variance $\sigma_\phi^2 = 4.65 \times 10^{-10}$, giving a distance difference error variance $\sigma_\Delta^2 = (\lambda/2\pi)^2\sigma_\phi^2 = 1.89 \times 10^{-10}$, giving a position RMS error of $\epsilon_R = 1.85\sigma_\Delta = 2.54 \times 10^{-5}$, so that $d_R/\epsilon_R = 682$ which results in a tracking RMS error of $T_{RMS} = 1000\sqrt{4/3}\epsilon_R/d_R = 1.69$ milliradians. The difference between the predicted and simulated results is due to a small positive correlation in the simulated receiver position errors which reduces the tracking error. The same results can be applied to a larger field by scaling the transmitter coordinates in Table 2, signal wavelength λ and inter-receiver distance d_R by a factor x. For example if $x = 100$, then each transmitter coordinate is multiplied by x such that the field radius is 100, $\lambda = 400$, and $d_R = 1.73$.

5. Conclusion

The results in the previous section show that the most important factors for an increased tracking accuracy are an increased number of samples S, an increased signal-to-noise ratio SNR, a smaller wavelength λ, and a larger inter-receiver distance d_R. The importance of transmitter placement is also shown. A properly optimised transmitter layout will result in an even smaller tracking error.

A larger d_R favours larger heliostats over smaller ones. Since heliostats need to update their orientations multiple times per minute, a higher sampling rate is favourable as it will yield a larger number of samples within the allotted time. Using multiple wavelengths to approximate and pinpoint receiver positions introduces additional complexity but offers considerable increases in accuracy. Increasing the number of transmitters also increases accuracy, albeit only marginally.

While this paper focuses mainly on one of the unavoidable error sources, namely phase difference error due to noise, other possible error sources also exist, such as errors in the electronic phase delay Φ, errors in the transmitter positions, and other hardware-specific errors, the effects of which will most certainly lead to an increased tracking error.

From a financial viewpoint the proposed method offers significant advantages. While the required infrastructure may be expensive, it is also minimal: only the central oscillator and transmitter towers. The important consideration is the added cost per heliostat. This amounts to only the three electromagnetic receivers, some electronics, and a processing unit, which in most cases is already present.

In the case where the time between heliostat orientation updates limits the number of samples to such an extent that a tracking error of less than one milliradian is unachievable, the system may become unsuitable for real-time use. It can, however, still be used to aid in the calibration process. The orientation of a heliostat is made stationary for as long as it takes to obtain the required amount of samples, after which its normal vector is calculated. This will still improve the required time for a full field calibration from three weeks to a few minutes. Another application of the system is the localisation of any object within the heliostat field by simply fixing one of the electromagnetic receivers to the object, provided that it can remain stationary long enough to obtain the required number of samples.

Although the results in this paper are only simulated, they indicate that a tracking error of one milliradian or less is theoretically achievable which is promising enough for the method to merit further investigation. Future work includes verifying the simulated results with laboratory experiments leading up to field measurements.

Acknowledgements

Johann Treurnicht, for sharing his invaluable insights and experience, and for his gracious assistance with administration, particularly with regards to finances.
Herman Engelbrecht, for his invaluable insights in all things signal processing, and his willingness to take over the role of lead supervisor halfway through the work.
Everyone else, particularly the people at STERG, who contributed in some way or another.
My Lord and Saviour Jesus Christ for his faithfulness and provision. The fear of the Lord is the beginning of wisdom, and the knowledge of the holy is understanding.

References

[1] Gregory J. Kolb et al., "Power Tower Technology Roadmap and Cost Reduction Plan," Sandia National Laboratories, Tech. rep. April 2011.

[2] J. Ignacio Ortega, J. Ignacio Burgaleta, and Félix M. Téllez, "Central Receiver System Solar Power Plant Using Molten Salt as Heat Transfer Fluid," *Journal of Solar Energy Engineering*, vol. 130, no. 2, February 2008.

[3] Gregory J. Kolb et al., "Heliostat Cost Reduction Study," Sandia National Laboritories, Tech. rep. June 2007.

[4] Abraham Kribus et al., "Closed loop control of heliostats," *Energy*, vol. 29, pp. 905-913, 2004.

[5] Andreas Pfahl, Reiner Buck, and Karsten Rehschuh, "Method for controlling the alignment of a heliostat with respect to a receiver, heliostat device and solar power plant," October 2009.

[6] Thomas Roos et al., "A 25m2 target-aligned heliostat with closed-loop control," , 2007.

[7] Mark R. Convery, "Closed-loop control for power tower heliostats," in *Proceedings of SPIE*, vol. 8108, 2011.

[8] M. Berenguel et al., "An artificial vision-based control system for automatic heliostat positioning offset correction in a central receiver solar power plant," *Science Direct*, vol. 76, pp. 563-575, December 2004.

[9] Raed Sherif, "Concentrating Solar Power Technologies," in *iNEMI Alternative Energy Workshop*, San Jose, California, October 2010.

[10] M. Zavodny et al., "Tower-based CSP artificial light calibration system," *Energy Procedia*, vol. 69, pp. 1488-1497, May 2015.

[11] Kenneth W. Stone and Scott A. Jones, "Analysis of solar two heliostat tracking error sources," Sandia National Laboritories, Tech. rep. 1999.

[12] William S. Murphy, "Determination of a Position using Approximate Distances and Trilateration," Colorado School of Mines, Master's thesis July 2007.