Summary

The work I present in this document has been divided into two main parts, the first one related to the IOTA project and the second one related to the study on the lunar interferometer, and an introduction section. Each section can be read independently from the other, however they are presented following the logical order in which the research work has been developed. As a guide for the reader here I describe the content of each chapter, which represents the original contribution (except when it is specifically declared) to the research accomplished.

1. Introduction section

This section consists in the *Introduction* itself, with a presentation of the motivations for this research work, and in the chapter *Interferometry from the Earth and from the Moon.* The first part of this chapter shows the performances which are expected to be reached by ground-based interferometers (*Colavita, 1992*) by using adaptive optics systems (*Beckers, 1993*). The evaluation is made separately for the case of high resolution imaging and for high accuracy astrometric measurements. The most optimistic results expected for ground-based instruments determine the level of the performance that has to be required from a space interferometer (both an orbiting and a lunar instrument). In the second part of the chapter I specifically deal with the case of a lunar interferometer, which allows to put together the advantages offered by a ground-based instrument (very long baseline, a stable platform) and those offered by the space environment (absence of atmospheric turbulence, long integration times, and wavelength range of observation from the ultraviolet to the far infrared). In order to evaluate the limits of the lunar interferometer, I need to consider three subjects with which I did not explicitly dealt for the study on IOTA: the maximum length of the baseline (*Tango and Twiss, 1974*), the maximum integration time, and the performances obtainable at the minimum temperature of operation (*Ridgway, 1990*). The chapter ends with a list of the main reviews which deal with the scientific objectives of space and lunar interferometry.

In *Appendix A* I present an introduction to the principles of optical stellar interferometry. This part is mainly derived by the study and re-elaboration of the contents of the following works: *Armstrong et al. (1995)*, *Shao and Colavita (1992)*, and *Born and Wolf (1980).*

2. Part one: the IOTA project

In this section I present the work I specifically developed within the IOTA project. This work allowed me to, directly or indirectly, acquire the theoretical and technical knowledge I then applied in the study on the lunar interferometer.

After having identified some of the main sources of systematic error for an interferometer, I examined: the problem of the telescope alignment, the beamsplitter behaviour, the effects that thermal variations cause on the optics and their support structures. The results

obtained in these analyses and the evaluations performed on the performances of other subsystems of the instrument, allowed me to proceed in the evaluation of the instrumental visibility loss for IOTA.

In the first chapter (I) I present a general description of the IOTA instrument, avoiding a detailed description of each subsystem. When it is necessary, this is given in its appropriate context.

The second chapter (II) is the result of the largest part of my work done on IOTA: the analisys of the alignment of each telescope of the interferometer. A non-perfect alignment of the telescope optics causes a distortion of the wavefront coming from the observed object. The distortions affecting the wavefront are responsible for the corruption of the interference fringes produced by the instrument, and eventually of the astrophysics information derived from their analysis. In order to study the effect of the optics misalignment on the performances of IOTA, I wrote a program to simulate some misalignment conditions and to evaluate the wavefront aberration they cause. For each case considered, an interferogram is produced by simulating the interference of the distorted wavefront with a plane wavefront. This interferogram is a means to represent the loss in the optical performance of the system. The interferograms produced by the simulation program are collected in an "atlas" (I present an extract of it as an independent appendix of this thesis) which will be used during the telescope alignment operation to help in diagnosing the error in the optics position.

The initial part of the chapter is devoted to a brief introduction to the wavefront aberration theory (from *Schroeder, 1987*) and to the analysis of one of the IOTA telescopes when it is in auto-collimation mode (the configuration used for the alignment operation). On the basis of this analysis I derived the misalignment conditions which I studied by means of the simulation program.

The third chapter (III) is devoted to the beamsplitter, a fundamental element for most Michelson optical stellar interferometers. The beamsplitter is the optical element by means of which the beams are made interfere. In the first part of the chapter I describe the characteristics of a beamsplitter for astronomical applications. This description is mainly based on a study that James D. Phillips accomplished within the project of the space interferometer POINTS. To my knowledge, this is the only study which deals with both the theoretical approach of the problem and the experimental aspect involved in the design of a beamsplitter for astronomical applications. In the second part of the chapter I present the original contribution on this subject, which consists in the analysis of the polarizations effects introduced by the beamsplitter. In particular I obtain the expression of the fringe intensity when considering the polarization effects. This expression contains a term which gives the visibility loss due to the phase difference between the polarization components of the interfering beams. Then I evaluate the loss in visibility for IOTA and

present the results of the simulation of the interfering fringes obtained when polarization effects are present.

In the fourth chapter (IV) I deal with the problem of how thermal effects affect the performance of a stellar interferometer. Knowing these effects and the thermal characteristic of the site when designing the instrument, allows to riduce its sensibility to the thermal changes. In the first part of the chapter I take into consideration the optical components, both reflective and refractive, of the system: I evaluate the wavefront error due to the alteration of the physical characteristics of the optical material as a consequence of a temperature variation. In the second part I consider the effects on the structure which supports and connects the primary and the secondary mirrors: a temperature variation may cause a perturbation in the telescope alignment and hence a wavefront error. The general expressions for the wavefront error obtained in the first and second part are then applied to IOTA. In particular the result of this study are used to perform an *a posteriori* evaluation of the visibility loss for IOTA. Finally, I performed a detailed analysis for the evaluation of the wavefront error introduced by a non-perfect thermal compensation affecting the metric structure which connects the mirrors of the telescope.

The fifth chapter (V) consists in a description of the causes of mechanical instability which can affect the operation of IOTA. In most cases I only present a qualitative description of the phenomena and a rough evaluation of their effects. This is because an accurate evaluation of the wavefront error induced by each of them requires a specialistic study of the vibration propagation throughout the whole structure of the interferometer, a study which is beyond the purposes of this work. I performed a more detailed evaluation only to determine the efficiency of the insolation system applied to the vacuum pumps. One of the problems arisen the first times the instrument was used for observation in the visible was that of the vibrations introduced by the vacuum pump system. Because of them, it was often impossible to detect the interferometric fringes even if the overall conditions for observation were good.

In the sixth chapter (VI) I deal with the evaluation of the wavefront error due to the deformation of the primary mirror caused by the gravitational load.

In the seventh chapter (VII) I eventually use the results presented in the previous chapters to evaluate the instrumental visibility of IOTA. A means to evaluate the efficiency of an interferometer is by determining the fringe visibility loss due to the intrinsic characteristics of the instrument. In fact, the smaller is the fringe visibility the more difficult is to extract from it accurate information about the observed object. In the introduction of this chapter I describe the relation, which has ben derived by *ten Brummelar et al.(1994)*, between the Strehl ratio (a quantity which defines the performance of an optical system by means of the wavefront error due to the system itself) and the factor which relates the

actual visibility of the object and the visibility measured by the instrument. The rest of the chapter is devoted to the calculation, based on the wavefront error estimations made previously, of the visibility loss related to different subsystems of IOTA. In addition, I evaluated the visibility loss due to diffraction effects on the transmitted beams and to the residual error left by the wavefront tilt correction system. The final result is summarized in a table which shows the value of visibility for the four main subsystems in which the interferometer has been divided: the product of these values gives the instrumental visibility for IOTA. Since most of the instrumental parameters are already fixed, the main contribution of this evaluation consists in the estimation of the potential of the instrument in terms of both visibility and magnitude limit (the signal-to-noise ratio of the measurement depends on the visibility) of the IOTA interferometer.

3. Part two: the lunar interferometer

In this section I present the study developed to get a strawman design for the lunar interferometer. In this study I take into consideration some characteristics of the IOTA design (initial system with two collectors, telescope configuration consisting in a siderostat plus a beam-compressor, two delay lines for the compensation of the optical path delay) which are analyzed to be properly make suitable for a lunar interferometer. I also deal with subjects which, even if not explicitly considered in the study for IOTA, constitute a fundamental part of the knowledge derived by working on IOTA, and need to be considered for the definition of the lunar instrument.

The last chapter, devoted to the evaluation of the visibility budget for the lunar interferometer, sums up the relations with the study performed for IOTA. I suggest the analogy between the contribution due to the atmosphere for a ground-based instrument and that due to the environment for a lunar one, and I evaluate the wavefront error related to different subsystems on the basis of the analyses performed for IOTA. However there is a main difference between the two cases. For IOTA I evaluated the instrumental visibility which derives from the already defined instrument characteristics. On the contrary, for the lunar interferometer I first estimated a minimum value for the instrumental visibility so that the instrument has performances better than those of a ground-based instrument, and then I evaluated the maximum contribution to the wavefront error that each subsystem can give. When it was possible, I also derived the subsystem components' specifications which allow to meet the initial budget.

In the first chapter (I) I deal with the orientation of the baseline and with the implication it has both on the variation of the optical path difference (OPD) and on the u-v plane coverage (which is the quantity that determines the imaging performance of the instrument). In the first part of the chapter I obtain the expressions, as a function of the Moon rotation, for the calculation of the OPD variation and of its rate of variation. Then I show the results for two specific cases: baseline orientation N-S and E-W. In the second part of the chapter I present the results for the u-v coverage obtained with different baseline orientations. I considered three orientations (N-S, E-W, NW-SE) and for each of

them I evaluated the *u-v* coverage for three different baseline latitudes and three different object declinations.

In the second chapter (II) I present the analysis on the telescope configuration. In particular, I considered the configuration adopted for IOTA, which consists in a siderostat followed by an afocal beam-compressor. I show that by defining two angular parameters, the maximum and the minimum incidence angle on the siderostat, we can derive a set of conditions and consequent trade-offs related to: the declination and hour angle ranges available for observation, the dimensions of the siderostat and the compactness of the telescope structure. I first explain the reasons for the choice of the siderostat plus beam-compressor solution for the lunar interferometer, then most of the chapter is devoted to the analysis of the telescope parameters. In the last part, I also take into consideration other aspects related to the telescope configuration which are not main drivers for the definition of the telescope parameters.

In the third chapter (III) I deal with the lunar thermal environment and I present the thermal analysis for one of the telescopes of the interferometer. The lunar thermal environment implies extreme conditions for the operation of any instrument and severe constraints may be imposed on the components of an interferometer, as it can be suggested on the basis of the study on the thermal effects performed for IOTA. Without a suited thermal control system, these thermal conditions may cause permanent misalignments of the telescope optics, introduce structural deformations, and damage mechanical and electronic systems. In the first part of the chapter I present an analysis of the thermal characteristics of the lunar surface. It is based on the study, by *Cremers et al. (1972)*, of some samples of lunar soil collected during the *Apollo 12* mission. In the second part I present the results of the thermal analysis, performed by Sherry Walker of the *NASA Marshall Space Center*, of a 16 m lunar telescope. On the basis of several interaction and discussion with engineer Walker I was able to adapt these results to the case of a 1 m telescope, which is the size of the telescope that I would expect for a lunar interferometer. In the third part of the chapter I present the results to be applied for this case, which consist in the temperature and the temperature variations, relative to three periods of the lunar night and two periods of the day, to which the telescope is exposed. Finally, the *Appendix L* is a complementary part of this chapter: in it I specifically deal with some effects induced on the primary mirror by the thermal environment. The data I present are due to a study which *Hughes Danbury Optical System, Inc.* performed for the selection of the material and the geometry of the primary mirror of LUTE (*Lunar Ultraviolet Telescope Experiment*). I used the results obtained by *Hughes* to estimate an appropriate temperature range within which the lunar interferometer should be kept. This result is important because it constitute the starting point for the analysis of the subsystem interactions which I present in the following chapter.

In the forth chapter (IV) I give an example of the relations among the parameters of

different subsystems and show how the choice of the value of these parameters depends on the environmental conditions. In order to define the set of parameters which characterize the strawman design that I propose, I make use of the results from the analyses presented in the chapters *Thermal analysis for a telescope of the lunar interferometer*, *The siderostat plus beam-compressor solution for a lunar interferometer*, and *Baseline orientation: OPD variation and* u-v *coverage*. The instrument consists in two collectors separated by a baseline 1 km long, and it will operate only during the lunar night. From the results of the thermal analysis, in particular those related to the period of radiative equilibrium for the telescope, I derived the hour angle range which is available for observations. On the basis of this result I determine the other parameters which define the telescope configuration. Finally, I consider the baseline orientation and calculate the u-v coverage obtainable with the maximum integration time (defined by the hour angle range previously found) available. I show the results of this exercise for two different baseline latitude, 0° and 30°, and from their comparison I suggest the following preliminary instrument design:

- two-element interferometer with a baseline of 1 km;
- each telescope is provided of a dome-shade which allows to keep within 100 K the maximum temperature variation suffered by the telescope;
- equatorial baseline with E-W orientation;
- the optical path difference compensation is realized by means of two delay lines:
  a coarse delay line with 5 fixed station separated by 100 m;
  a fine delay line, movable, with maximum length of 50 m;
- maximum integration time of 130 h;
- telescope oriented in direction E-W and tilted by 35° with respect to the horizontal;
- primary mirror diameter of 1 m.

This preliminary design should be considered as the starting point for a detailed evaluation of the constraints on the instrument components and the definition, for the parameters involved, of one set of values which allows to reach the mission objectives.

In the fifth chapter (V) I deal with the lunar environment and in particular with those aspects which have main impacts on the operation of a lunar interferometer. The description I present is mainly based on the results of experiments performed during the *Surveyor 3* and several *Apollo* missions. I provide more detailed information about the effects induced by the lunar dust and the lunar seismic activity, since both of them may significantly affect the operation of an interferometer. Other subjects I take into consideration are the flux of meteoroids, the characterization of the lunar atmosphere, and the cosmic and solar radiation that reaches the lunar surface. I do not deal with the thermal environment which is already described in the chapter on the thermal analysis of the telescope.

In the sixth chapter (VI) I perform, as I did for the study on IOTA, a global estimation of the lunar interferometer performances by means of the visibility budget of the instru-

ment. In order to work out a realistic budget it is necessary to know the architecture of the whole design, to single out the main subsystems and to determine their contribution to the degradation of the instrument performance, by taking into account each subsystem architecture, its interaction with the other subsystems and with the environment. Since the visibility budget is one of the means to determine the instrument potential, it is also a means to evaluate the worthwhileness of a lunar interferometer project. In the first part of the chapter I determine the visibility goal for the instrument, by taking into account the best performances expected for a ground-based interferometer of the next generation. In this context, in order to characterize the causes of visibility loss, I introduce the analogy between the role played by the atmosphere for a ground-based interferometer and that played by the environment for a lunar interferometer. In the second part of the chapter I take into consideration various causes of visibility loss which may affect the components of the four main subsystems of the instrument (in particular visibility loss related to the optical components, to a variation of the optical path difference during an integration time, to diffraction effects on the transmitted beam, and to the servo control system for the correction of tilt of the beam). When it is possible, I allocate to each item a value of visibility loss which allows to meet the suggested visibility goal.

The result of this analysis shows the need of improving, both qualitatively and quantitatively, our knowledge about the lunar environment expecially for those aspects which mainly affects the operation an interferometer. This can be achieved by means of one or several missions designed to perform *in situ* experiments to measure the parameters which characterize the thermal environment, the seismic activity, the dust transport mechanism, and the lunar atmosphere. Two missions which, in the near future, could provide part of these information are *LUNAR-A* and *LEDA*, and they are briefly described in the conclusion chapter of this work.

Introduction

My interest in space stellar interferometry and in particular in interferometry from the lunar surface, developed as a consequence of attending the *Summer Session* of the *International Space University* in 1993. In that occasion I was involved, as teaching assistant of the "Space Physical Science" department, in the design project for a scientific laboratory to be located on the far side of the Moon. One of the objectives of this study was to single out the instrument(s) for astrophysical observations whose characteristics and potential are best exploited on the lunar surface (a description of the instruments proposed for being operated on the lunar surface can be found in *Mendell (1985), Burns and Mendell (1988)*, and *Mumma and Smith (1990)*). To perform this evaluation it was taken into account the scientific return of the instruments as well as the possibility of realizing the same project on the Earth or in Space, the uniqueness of the advantages offered by the lunar location), the availability of the required technology, and a cost estimation. Although this kind of multiparameter analysis always implies a certain amount of subjectivity, it allowed to single out the two instruments which seem to be most appropriate for a lunar location. They are a very low frequency array (thanks to the advantages offered by the location on the far side) and an optical stellar interferometer (independently of the location, where optical means the wavelength range from the ultraviolet to the far infrared (for more detailed information on this evaluation, see table 1.1 in *Project Report: ILFOSS, 1993*).

This conclusion is in agreement with recent initiatives, above all by the European Space Agency (ESA), for the proposal of new projects for a return to the Moon, where to perform scientific experiments with unique features with respect to those obtainable on Earth and on orbiting satellites. The realization of a lunar interferometer is considered in both the ESA report *Mission to the Moon (1992)* and in the proceedings of the *International Lunar Workshop (1994)*, and this idea is still present in the ESA program *Horizon 2000 plus (1994)*. Finally, the interest showed by ESA in this project led to the *Invitation to Tender*, issued in September 1994, for the study of a long-baseline stellar interferometer either to be located on the lunar surface or to be realized by means of several free-flyers.

Eventually, I like to remind the reader that, although the development of the techniques for stellar interferometry and the interest in space interferometry date quite recently[1], the idea of a lunar stellar interferometer is not new. A list of the projects for a lunar interferometer proposed in the last ten years is presented in tab. 1[2].

[1]The first conference on space interferometry, *Workshop on Optical Interferometry from Space*, was held in Baltimore in 1984. The proceedings of this workshop were published in *Bulletin of AAS, Vol. 16, No. 3*).

[2]For a more complete description of the project presented in tab. 1, see Optical VLA: *Burke (1985)*, LOUISA: *NASA Workshop Proc. (1989)*, LOVLA: *Labeyrie and Mourand (1990)*, OVLA: *Labeyrie et al. (1988)*, LOI: *Colavita et al. (1991)*, MARK III: *Shao et al. (1988).*

tab. 1 Lunar interferometer projects proposed in the last ten years

PROJECT	YEAR	CHARACTERISTICS
Optical VLA	1985	Designed according to the VLA radiointerferometer Maximum baseline: 11 km Configuration: *Y-array* of 27 elements fixed or located along a "Cornwell Circle" with variable separations Telescopes: diameter of 1.6 m; alt-az mounting Angular resolution: from 10 mas to 10 μas Spectral range: 0.1 $\div$ 3 μm Magnitude limit: $m_V \sim 20$
LOUISA	1989	Results from the NASA LOUISA Workshop Maximum baseline: 10 km Configuration: 33 elements on a 10 km diameter circle plus 9 on an inner circle with diameter of 500 m Telescopes: diameter of 1.5 m; alt-az mounting Angular resolution: from 0.2 mas to 10 μas Spectral range: 0.1 $\div$ 1 μm Magnitude limit: $m_V = 26 \div 29$
LOVLA	1990	Design derived from the ground-based interferometer OVLA Maximum baseline: several km Configuration: 28 movable elements located along an ellipse; each element moves independently by means of an articulated structure constituted by 6 "legs" Telescopes: diameter of 1.5 m; Coudé configuration Angular resolution: from the order of mas to the order of μas Spectral range: from UV to far IR Magnitude limit: not specified
LOI	1991	Design based on the experience acquired with the ground-based interferometer MARK III Maximum baseline: about 1 km Configuration: *Y-array* of 12 elements Telescopes: diameter of 1 m; alt-az mounting; configuration: siderostat + beam compressor Angular resolution: from 100 to 10 μas Spectral range: 0.4 $\div$ 10 μm Magnitude limit: $m_V \sim 29$

Notwithstanding the increased interest in lunar interferometry, we are at the proposal stage and the gap between proposals, sometimes clever but without any proof of technical feasibility, and actual possibility of a future realization of a lunar interferometer is still large. In other words, I think it is time to undertake the preliminary design of a lunar interferometer which allow us to determine, both in terms of potential scientific results and in terms of economic and technical resources necessary for the implementation on the lunar surface, the worthwhileness to go on in a detailed system design of the instrument.

To develop such preliminary study requires a team of experts whose work has to be coordinated by a person with technical and scientific knowledge about interferometry. Essentially, since a lunar interferometer resembles closer a ground-based than an orbiting interferometer, we can think to apply the design strategy for a ground-based interferometer and in addition to develop the technologies necessary to allow the system to operate in the lunar environment. To support this statement we can notice that all proposals listed in tab. 1 are derived, or are directly influenced, by the experience acquired with a ground-based interferometer. Obviously, one person can not develop this kind of study within his or her Ph.D. research program. However, it is possible for him or her to build up a cultural background so that he or she can figure out some of the important aspects of the project and have an idea of the complexity and of the interactions involved in the management of the project. This has been the aim I pursued in my research work, during which I tried to look at the problem in terms of interaction among subsystems rather than analyzing in detail the characteristics of one system.

At this regard it was very important for me to work within the IOTA (*Infrared-Optical Telescope Array*) project, which is one of the ground-based interferometers actually in operation (for a list of the ground-based interferometers which have been in operation, or will be, since the the beginning of the 90's, see tab. 2). Observations with IOTA started in 1993, however the instrument has not yet reached its definitive configuration with three apertures instead of the two which have been operative until now. In this context I had the opportunity to follow the upgrade of some components of the instrument, to develop some experience in different technical fields and to get that kind of interdisciplinary mentality, at least as far as interferometry is concerned, which is essential for the study on the lunar interferometer that I aimed to develop. The aspect that I wanted to characterize the originality of this work consists in fact in the development of a multidisciplinary and interdisciplinary approach to the project of a lunar stellar interferometer. This aim appears to be in agreement with the multidisciplinary program of the doctorate course in "Space Science and Technology". Among the objectives of this program is that of *training a researcher with an interdisciplinary background (in the applicable disciplines) able to take scientific responsibilities, at least at a subsystem level but also at a global level, in the definition, design, development, operation, and utilization of an astrophysics mission* (from the "curricula" presentation of the doctorate course in Space Science and Technology).

tab. 2 Ground-based interferometers of the 90's (from *Armstrong et al., 1995*)

Interferometer	Apertures	Maximum baseline (m)	λ (μm)
COAST	4×40 cm	100	$0.4 \div 0.95$
(Cambridge University)			2.2
SUSI	2×14 cm	640	$0.4 \div 0.85$
(Sydney University)			
IOTA	3×45 cm	38	$0.45 \div 0.8$
(Center for Astrophysics)			2.2
NPOI astrometric array	4×12.5 cm	38	$0.45 \div 0.9$
(USNO-NRL)			
NPOI imaging array	6×12.5 cm	437	$0.45 \div 0.9$
(USNO-NRL)			
ASEPS-0 test bed	4×40 cm	100	2.2
(JPL)			
CHARA	7×1 m	354	$0.55 \div 0.9$
(Georgia State University)			$2.1 \div 2.5$
KECK Observatory	2×10 m +	165	$2.2 \div 10$
	6×1.5 m		
VLTI	4×8 m +	200	$0.45 \div 12$
(European Southern Observatory)	8×1.8 m		

Interferometry from the Earth and from the Moon

0. Introduction

Interferometric missions in space have been the subject of numerous proposals during the previous two decades (a meaningful sample of these proposals is presented in *ESA report SP-1135, 1990*). However, any space-borne instrumentation would be expensive and ultimately justified if, and only if, it can provide scientific data which are unaccessible using terrestrial instrumentation. This is particularly true as there has been much activity in design, construction and operation of terrestrial interferometers.

In making a comparison between possible space mission in a given year, and terrestrial instrumentation likely to be available by that date, one should be optimistic concerning future developments in terrestrial technologies. A particular crucial point here is the performance from adaptive optics applied to terrestrial telescopes.

The advantages of the space environment for optical and infrared astronomical interferometry are well known: transparency at all wavelengths bands, no atmospheric turbulence, and the ability to cool optics for observation in the thermal infrared.

However, the Earth certainly offers a few adequate observing sites with the required features for a large optical interferometer: good seeing, large baseline potential, stable subsoil, and seismic quiescence. The essential problem in ground based interferometry is overcoming the wavefront perturbations introduced by the atmosphere, and the vibrations in the instrument provoked by the wind.

As far as disturbance due to the atmosphere is concerned, adaptive optics aided by laser stars shows promise to eliminate some of the effects of turbulence for any ground-based instrument for imaging. Other techniques such as cophasing (phase-referencing) and closure phase can also be used to circumvent some of the effects of atmospheric turbulence, particularly affecting ground-based interferometers.

It is then clear that, even if it is a long step from the theoretical performance predictions for a ground-based instrument to their attainment, a future space interferometer must emphasize those capabilities which are impossible or at least very difficult to have from the ground.

As far as imaging is concerned, among these capabilities there are: UV observations, very high dynamic range measurements, and high resolution (which requires long baselines) measurements of faint, extended objects.

As far as astrometry is concerned, ground interferometry can offer high accuracies over very small fields, but a space interferometer can offer more than two orders of magnitude better performance over wide fields than that achievable with ground-based systems.

Taking into account the previous evaluations, I would finally point out that an interferometer located on the lunar surface could combine the advantages of the space environment,

in terms of wavelength ranges and absence of atmosphere, and the advantage of the similarity with a ground instrument, in terms of very long baselines and the possibility of having massive, rigid foundation that can absorb vibrations with minimal deformation.

In the following sections, I will first present an introduction to adaptive optics (*Beckers, 1993*), in order to show the result of the comparison between performances of large ground-based telescopes and space-based interferometers, and between long-baseline ground-based interferometers and space-based interferometers (*Colavita, 1992*).

Eventually the case for the lunar interferometer is presented with its close relationship with a ground-based interferometer and some of the aspects that have to be considered when evaluating the pros and cons of such instrument. In particular: limits in maximum baseline length due to diffraction effects (*Tango and Twiss, 1974*), advantage of long integration time, possibility to operate at constant low temperature (around 100 K) and to reach lower temperature in order to observe in the far-infrared (*Ridgway, 1990*).

1. Adaptive optics

Adaptive optics removes the wavefront distortions introduced by the Earth's atmosphere by means of an optical component which is introduced in the light beam and which can introduce a controllable counter wavefront distortion which both spatially and temporally follows that of the atmosphere. This optical component is generally a mirror whose surface can be distorted. To control the mirror, the wavefront distortions have to be known. These are measured by means of a wavefront sensor using either the object under study (self-referencing) or a nearby stellar or laser-generated object (also referred to as natural or *laser guide stars*). In the case where the wavefront is measured with the required accuracy and spatial and temporal resolution, and in which the adaptive mirror control is perfect, the atmospheric effects are removed and the optical system will give a diffraction-limited image[3].

The application of adaptive optics to astronomical telescopes requires the development of expensive, complex opto-mechanical devices and their control systems. It is not the scope of this study to explain the details of the different adaptive optics techniques, however according to the current state of the development, we can expect that the goal of diffraction-limited imaging with large telescope will be reached in the years to come, probably sooner or in the same range of time necessary to realize and operate a space-based interferometer.

Finally, for sake of completeness, let us recall the distinction between "Adaptive Optics" and "Active Optics". The latter term is now commonly used to describe ways of controlling the wavefront distortions in a telescope introduced by mechanical, thermal, and

[3]An optical system is said diffraction-limited when a spherical wave divergent from a pointlike source in the direction of the entrance pupil of the system, exit from the exit pupil of the system as a spherical wave converging on one point in the image plane.
A less qualitative definition states that an optical system is diffraction-limited according to the Strehl tolerance system, if the central peak of the intensity of the image is equal to or greater than the 80% of the central peak intensity for a perfect system, i.e. the system Strehl ratio is ≥ 0.8.

optics effects in the telescope itself. Since these effects vary on a rather long time scale, Active Optics is rather slow as compared to Adaptive Optics, whose purpose is to compensate for the rapid varying wavefront distortions (what we generally mean by the term seeing). Astronomical Adaptive Optics uses bandwidths in the vicinity of 10 to 1000 Hz; Active Optics works at less than 1 Hz. While there is no need for Adaptive Optics in space, there is certainly need for Active Optics.

As far as guide stars are concerned, we refer to those objects used for measuring the atmospheric distortions, including, but not restricted to, wavefront tilts. The can be either natural or laser guide stars.

Adaptive optics techniques, using natural guide stars, will have some major limitations. Among these there are: (a) the sky coverage[4] limitation; (b) the limitation to relatively bright objects, (c) the variation of the point spread function with position across the field-of-view; and (d) light losses and increased infrared emissivity that will occur due to the additional optical elements. However, recent developments in the implementation of laser guide stars and associated techniques promise to remove at least part of these limitations (*Beckers, 1993*).

2. Imaging with large telescopes and interferometers

At wavelength transmitted by the Earth's atmosphere, the traditional advantage of a space-based optical system over a ground-based one has been the elimination of atmospheric turbulence. As far as imaging with a ground-based interferometer is concerned, the effect of atmospheric turbulence is two fold: the reduction of the coherence volume (the product of the atmospheric coherence area r_0^2 and the coherence time τ_o) results in a loss of sensitivity[5], and the corruption of the fringe phase that makes it unusable for synthesis imaging. High-performance imaging interferometric systems must address both of these issues. The sensitivity issue can be addressed by using adaptive optics to phase the individual apertures of the array, essentially increasing r_o to the diameter of the aperture, and by cophasing between apertures, to essentially increase τ_o to the available integration time. The corruption of the fringe phase can be addressed by the use of closure phase[6].

[4]The main problem of adaptive optics lies in the need for a sufficiently bright reference source to achieve the correction. The light coming from this object feeds a wavefront sensor which analyses the perturbation induced by the atmosphere (and this implies that the wavefront sensor sets the magnitude limit of the reference star). This correction is only valid in a limited sky angular area, the isoplanatic patch, which is the distance on the sky over which the wavefront distortions are for all practical purposes the same. More precisely, the isoplanatic patch corresponds to the angular separation (isoplanatic angle) between objects for which the atmospheric fluctuations are correlated to ~ 1 rad rms. Like r_o and τ_o it increases as $\lambda^{6/5}$. Together with the limited brightness of the reference source, this defines the "sky coverage" or percentage of the sky where an efficient correction can be achieved (this is equivalent to the probability to find a suitable reference star in the isoplanatic patch). An "efficient correction" is defined as a correction for which the corrected wavefront phase variance is less than or equal to 1 rad^2 (*Rigaut and Gendron, 1992*). Clearly, for the same isoplanatic patch the probability of finding fainter references is higher, but the correction is less accurate. Brighter references assure more accurate correction but the probability of finding them decreases, above all at short wavelengths.

[5]This is a problem that affects a single aperture telescope as well.

[6]See introduction section on *Principles of stellar interferometry* for explanations of phasing, cophasing,

Adaptive optics using laser guide stars currently offers the highest potential for high-resolution, short wavelength imaging on the ground, thanks to the increase in sky coverage that this technique allows.

According to *Colavita (1992)* the potential performance of a large telescope with laser guide adaptive optics is limited only by the sky coverage. An example of sky coverage as a function of telescope diameter and wavelength is given in Tab. SG.1:

tab. SG.1 Potential performance of a large telescope with laser guide adaptive optics: sky coverage as a function of wavelength and aperture diameter, D. From *Colavita (1992)*

wavelength	sky coverage		
	D = 2 m	D = 6 m	D = 10 m
0.6 μm	12 %	25 %	$\sim$ 50 %
1.0 μm	50 %	90 %	$\sim$ 100 %

Although achieving the theoretical performance given above is not straightforward, the implications of these results for a future space-based system are that resolutions less than that equivalent to a 10-m filled aperture are not competitive from space at visible wavelengths.

Laser guide stars can also be used to phase the individual apertures of an interferometer. For an interferometer with large apertures the sky coverage for phasing of the apertures is the same as that given in tab. SG.1.

However, with an interferometer, aperture phasing with adaptive optics is only one-half of the solution to the problem of atmospheric turbulence. The sensitivity of an interferometer is also limited by the coherence time τ_o. By cophasing, the temporal analog to adaptive optics, the goal is to increase the coherent integration time τ_o to the total integration time. A ground interferometer with phased aperture and cophased baselines, to first order, has the same imaging performance as a space interferometer.

Unfortunately, laser guide star technology does not seem applicable to cophasing interferometers. The primary problem is due to atmospheric reciprocity, which has the effect of making the laser spot appear stationary to the instrument, thus preventing it from being used as both a tilt and a cophasing reference. The relative phase of two telescopes will still drift with a time constant characteristic of the atmospheric turbulence or the instrument vibrations. The cophasing of the independent telescope still requires a natural reference.

In order to compete with a space interferometer, it must be possible to cophase a ground

and closure phase.

interferometer for arbitrary targets[7]. When the target is not suitable to be used as a reference star, the reference object must be located within the ordinary isoplanatic patch. This implies that the sky coverage for cophasing, even with phased pupils, is less than that achievable with laser guide star adaptive optics. However, as with aperture phasing, there is some cutoff wavelength above which full sky coverage is available.
The sky coverage for cophasing of ground-based interferometer with large apertures is given in tab. SG.2.

tab. SG.2 Sky coverage for cophasing of large apertures as a function of wavelength and telescope diameter, D from *Colavita (1992)*

wavelength	sky coverage		
	D =2 m	D = 6 m	D = 10 m
0.6 μm	0.1 %	1 %	3 %
1.0 μm	0.5 %	5 %	15 %
2.0 μm	7 %	70 %	100 %

The analysis below is not intended to be rigorous, but it should be adequate to show the possibilities for ground-based interferometric imaging, even if to achieve the performance presented will not be straightforward. According to these performances, the implications for a space interferometer can be derived.

As limited coherence volume is the main problem with ground-based systems, it is clear that any space interferometer should be both phased and cophased, in one word phase-stable. However, a small phase-stable space interferometer will find it hard to compete with a filled aperture 8-10 meter telescope (provided with adaptive optics) on the ground, for observations at red/near-infrared wavelengths. The small space interferometer would have advantages at shorter wavelengths where full sky coverage is problematic, and obvious advantages in the UV which is unobservable from the ground.

As far as the comparison with a ground-based, long baseline interferometer is concerned, at near- to mid-infrared wavelengths, where full phasing and cophasing is possible on the ground, the space interferometer is not competitive unless the apertures are cooled, so that it can observe in the far infrared (very faint limits, of the order of 10^{-3} to 10^{-4} *Jansky*, for simple sources, are reachable when the optics are cooled down to 10-20 K *(Ridgway, 1990)*).

For visible to near-infrared wavelengths, where general cophasing is not possible on the ground, a long baseline space interferometer can observe faint, extended objects which

[7] In some cases the target itself can be used as the cophasing reference, but this is possible only for compact objects and if the interferometer configuration is highly redundant. Besides, the source itself has to be bright enough to satisfy the magnitude limit for cophasing imposed by the fringe-tracking system.

are not candidates for self-cophasing using redundant arrays on the ground.

In any case, a space-based interferometer offers potential advantages in the following areas:
– UV observations.
– High dynamic range observations. The contrast between faint point sources and the sky background for an adaptive optics system improves by $S \cdot D^2/r_o^2$ (where S is the Strehl ratio[8]), and the image of a star appears as a diffraction-limited peak surrounded by a halo, whose size is equal to the seeing disk. In space, there is no halo due to the seeing and the quality of the images are limited by the much smaller irregularities in the wavefront due to errors in the figure of the optics.
– Cooling optics. Cooled optics greatly reduces the thermal background for long-wavelength observations compared to an uncooled ground telescope.

3. Astrometry

Astrometry is the part of astronomy that, by means of different techniques, provides the position, the kinematic and dynamic properties of celestial bodies.

For the purpose of defining the limits from the ground and the expectation for space-based astrometry, it is useful to consider three general categories of measurements:

- *narrow-angle astrometry*: This is the determination of the position of celestial bodies relative to neighbouring stars within a stellar field which can vary from 20 *arcsec*, up to 5-8 *degrees* (usually it is given by the field of view of the telescope). Among the applications there are the determination of parallaxes relative to the mean parallax of the background stars, of relative proper motions for the study of local kinematics in clusters and stellar associations, and positioning of small bodies of the solar system. This has traditionally been realized by means of long-focus telescopes.

- *very-narrow angle astrometry*: This is the geometrical description of objects which occupy a field of about 10-20 *arcsec* or less. Its main application is to obtain information on the structure of a light source within its immediate environment. The astrometric information that can be obtained are the dimensions (diameter) of a star or the respective position of the components of a double or multiple star. This is generally performed by means of speckle imaging or long-baseline interferometry.

- *wide angle astrometry*: This is the determination of relative positions of celestial bodies simultaneously visible from a given observatory, whatever is their angular separation. Eventually, by using one (in space) or several (on ground) instruments, it should be possible to determine relative positions over the whole sky and to establish a celestial reference frame which allows a transformation from a relative measurement using small field astrometry into an absolute determination. When the entire sky is scanned by one

[8]The amount of aberration that can be tolerated in an optical system is defined according to the image quality. The image quality is expressed in terms of the Strehl ratio which is defined as the ratio of the intensity at the center of a reference spheric wavefront (with respect to which the aberration is defined) with and without aberration.

instrument, wide-angle measurements are usually referred to as global astrometry.

Wide-angle astrometry has traditionally been accomplished with transit instruments and astrographs. However, since the development of the interferometric techniques, wide angle astrometry has become part of the domain of phase[9] interferometry.

Among the main applications of wide angle, global, astrometry is to obtain absolute parallaxes and to build a consistent system of proper motions without regional biases.

3.1 Limits for ground-based astrometry

According to *Colavita (1992)* the limits of ground-based astrometric measurements, for which the effects of atmospheric turbulence decrease (nonlinearly) with the size of the field, can be summarized as follows.

Wide angle astrometry

For wide angle, or absolute, astrometry, the error due to atmospheric turbulence is only weakly dependent on integration time[10], so that it establishes the fundamental accuracy on a measurement. The limit achievable from the ground is of the order of 50-100 *mas* for a 1-min integration measurement (see also Colavita, Shao, Staelin, App.Opt., 1987). To improve the accuracy, repeated measurements can be performed on time separations of the order of one hour, so that the measurements are not correlated. Assuming ten observations per night for a period of ten nights, it is possible to reach an accuracy of the order of 5-10 *mas*.

At visible wavelengths somewhat higher accuracies are achievable by means of a *two-color* technique (*Shao et al., 1990*)[11]. In this case, the achievable gain in astrometric performance is typically a factor of 3-5, limited by the effects due to the presence of the water vapor.

In conclusion, it seems that an accuracy of the order of 1 *mas* is the limit to be assumed for ground-based wide-angle astrometry, with visible limiting magnitude (due to the limited integration time of the order of 1 minute) of about 10 in the visible and 15 in the infrared.

Narrow angle

The atmospheric effects are less severe with a differential measurement which measures

[9] We make a distinction between amplitude interferometry, which is exploited for narrow angle astrometry, by measuring the variation of the fringe visibility, and phase interferometry, where the phase of the fringe carries the information about the position of the object.

[10] According to *Lindegren, 1980* the atmosphere induced error is given by

$$\sigma_\alpha \sim 0.25 \cdot t^{-1/6}\ arcsec$$

where t is the integration time, and this result is equally applicable to telescopes and interferometers.

[11] The essence of the technique is to observe the fringe position at two-widely separated colors. Because of atmospheric dispersion, the two fringe positions will be different, and the difference will be proportional to the instantaneous atmospheric error, allowing a correction to be made.

the difference in the positions of two stars.

The analysis by *Lindegren (1980)* shows that the turbulence-limited accuracy decreases as the squared root of the integration time. In particular, *Lindegren* shows that for a simultaneous measurement of two stars separated by θ, this error is given by

$$\sigma_\alpha \propto \theta^{1/3} t^{-1/2} \; arcsec \; when \; \theta h \gg D$$

and

$$\sigma_\alpha \propto \theta D^{-2/3} t^{-1/2} \; arcsec \; when \; \theta h \ll D$$

where D is either the telescope diameter or the interferometer baseline, and h is a mean atmosphere height, $\simeq$ 6 km, and the quantity θh is given in radians.

From these expressions it is clear that there is a significant advantage in using long-baseline interferometers for narrow angle astrometry. Also, as far as the stellar separation is concerned, to narrow the field (i.e. to increase the baseline) involves a stronger dependence on the stellar separation, and also improves the accuracy.

As far as narrow angle astrometry is concerned, the limit for ground-based observations is set by the best performance of a long-focus telescope.

For a typical field of 10 arcmin and a 1.5-meter telescope, the achievable accuracy is of the order of 1-2 $mas/\sqrt{hr}$. The limiting magnitude is set by the photon noise of the detector, that with current CCDs is of the order of 18 mag.

As far as very narrow angle astrometry is concerned, the ground limit is definitively set by the performance of long-baseline interferometers. According to *Shao and Colavita (1992)*, a 100-m baseline interferometer can make differential measurements of objects separated by 1 $arcmin$ with an accuracy of the order of 70 μas, in one hour of integration. Extrapolating to longer baselines (and making use of the baseline length dependence of the accuracy, which is proportional to $B^{-2/3}$, where B is the baseline length), a 500-m baseline could reach an accuracy of about 25 μas and a for 1000-m baseline it would reach an accuracy of the order of 15 μas, where all of these values are reached in one hour of integration. If the observations are made in the infrared, by using the phase information from fringe tracking to cophase the interferometer, a 200-m baseline interferometer could reach accuracy of the order of 10 $\mu as/\sqrt{hr}$, for stars separated by 15 $arcsec$ (*Shao and Colavita, 1992*).

In conclusion the approximate capabilities of ground-based visible/near infrared astrometry, presented above, are summarized in tab. SG.3

tab. SG.3 Approximated capabilities of ground based astrometry, in terms of accuracy limit imposed by the atmospheric effects and magnitude limit *(from Colavita, 1992)*

Measurement	Instrument	Field	Accuracy	Sensitivity
wide angle	IR interferometry	20°	15 $mas/\sqrt{night}$	15 mag
wide angle	VIS 2-color interf.	20°	5 $mas/\sqrt{night}$	10 mag
narrow angle	1.5 m telescope	10'	1-2 $mas/\sqrt{hr}$	18 mag
very narrow angle	IR interferometry	15"	10 $\mu as/\sqrt{hr}$	16 mag

The comparison with space interferometers is simpler for astrometry than for imaging. Interferometers outside of the atmosphere are now metrology and photon-noise limited. Proposed space interferometers as POINTS (*Reasenberg et al, 1995*) and OSI (*Colavita and Wolff, 1995*) could provide accuracies better than 10 μas over wide fields[12], which is at least a factor 100 better than what obtainable with ground-based instruments. Ground interferometers can approach the accuracy of space interferometers over very narrow fields, however this performance is limited by finding suitable reference stars within the narrow field under investigation. In addition, the great advantage of space, that have to be exploited as much as possible, is the possibility to reach much fainter magnitudes and to have sample of objects much larger than those available from ground. These two last aspects are fundamental in order to make many of the relevant studies addressed by high accuracy astrometry.

Resolution and maximum baseline

As far as maximum resolution is concerned, this is only dependent on the ratio of wavelength to on the baseline length of the interferometer. It is very likely that the baseline of next generation (actually already this generation, see for example SUSI) ground interferometers will reach lengths of several hundred meters. This is a length longer than those expected to be reached for next generation orbiting space interferometers for which the main issue to overcome is to keep the whole structure stable within the level required by an optical path stability of the order of the nanometer (standard OPD stability values imply that $\delta OPD \simeq \lambda/50$, *(O'Brien et al., 1995)*)[13]. I will not address this problem in this context, however I recall that the rates at which interferometer fringes move at the detector depend on both the internal structural motions and the overall attitude motions

[12]OSI is expected to reach an accuracy of 10 μas on a star of magnitude 20 in about 250 seconds of integration, while POINTS reaches its nominal accuracy of 5 μas in about 150 seconds of integration.

[13]However, the *ESA Invitation to Tender (1994)* suggested to investigate the possibility of reaching kilometric baseline by means of a system of free-flying satellites. This was also based on the assumption that, development in the technology of high precision control and actuator systems, can allow to apply them to a 2-3 free-flying elements, separated by about 1 km, interferometer to control the position of one element with respect to each other within the accuracy required.

of the spacecraft. To overcome most of the stability problem and to be able to reach baselines either of the order of or longer than those feasible for ground-based interferometer, the solution can be offered by a Moon-based interferometer.

4. Interferometry from the lunar surface

The Moon has a number of features making it well suited as a location for an optical interferometric array. Certainly the presence of only a very tenuous atmosphere removes the coherence constraints of ground-based interferometers, allowing for greatly enhanced sensitivity and full spectral coverage, properties shared by an orbiting instrument too. However, in addition, the Moon provides a large, stable, slowly rotating platform on which to assemble kilometric (at least in principle) baselines, allowing for high resolution imaging and high precision astrometry. With its well known orbit it is possible to correct for stellar aberration[14] at the microarcsec level, a requirement for precision astrometry, and with a 100 K nighttime temperature, it is possible to cool the interferometer optics to obtain high infrared sensitivity.

The Moon also has some undesiderable features: the day-time to night-time temperature swing, which requires temperature controls or compensation systems; the dusty nature of the upper regolith requires attention to contamination control; the significant micrometeoroid flux and high energy radiation flux, require attention to shielding of optics and electronics.

There are many similarities between a lunar interferometer and ground-based interferometer. The two-dimensional lunar surface makes many of the technologies which work on the ground applicable. The lunar environment also simplifies certain aspects of the design. For example, the almost complete lack of atmosphere simplifies beam propagation and global metrology (see *Hutter, 1990*), and allows the use of bright reference stars in order to calibrate the array. The low rotation simplifies drive systems and delay line design. On the other hand the problems caused by the lunar environment pose several engineering challenges that are to be taken into serious account when "dreaming" about a lunar interferometer.

The objective of the study that will be presented in the second part of this work, is to present a strawman lunar interferometer with the description of some of the required subsystems of the instrument. The idea is to go one step beyond the "pure conception" stage, in order to give an idea of the order of complexity of the instrument, to illustrate some of the choices which need to be made, and to discuss the kinds of trade-offs which

[14]Because of the finite speed of light, an observer in motion sees an objected shifted in the direction of his motion. This change of apparent direction is called *aberration*. The amount of the aberration, a, is given by

$$a = \frac{v}{c} sin\theta$$

where v is the velocity of the observer, c is the speed of light and θ is the angle between the true direction of the object and the velocity vector of the observer.

are available. In addition, the strawman design allow us to individuate some of the areas requiring technology development. Since a lunar interferometer can very closely resemble a ground-based interferometer, allowing for the exploitation of the experience gained from those instruments, I will mainly concentrate on those aspects for which I have some experience from working on the ground-based interferometer IOTA. In particular, I will consider the aspects related to the choice of a siderostat plus beam-compressor configuration for the telescope (as it is the case for IOTA), the thermal environment effects on the performance of the instrument, and the instrument visibility budget. However, some other fundamental aspects need to be addressed to put together the main features of a strawman design for the lunar interferometer. As said before, the main reason to design a lunar interferometer is that of combining the advantages offered by the space environment with those due to the large stable platform offered by the Moon surface. While full spectral coverage and potential high sensitivity are not put into question, some words should be addressed to the maximum baseline attainable, the maximum integration time, and the minimum temperature of operation.

4.1 Baseline length

Ground-based interferometers could in principle have baselines as long as lunar-based ones, as far as a suitable site can be found on Earth. However there are two aspects which will eventually limit the maximum extension of the baseline: dispersion along the air path travelled by the beams, and diffraction effects on the transmitted beam, both of which have deleterious effects on the fringe visibility. While the first problem concerns only ground-based instruments, the second one affects also space-based instruments, even if in this case, since the diffracted beam has not been already distorted by the atmospheric turbulence, the overall wavefront distortion will be less.

Let us consider a two-aperture interferometer situated in a horizontal plane with a baseline vector $\mathbf{B}$. The optical path difference (OPD) between the two interfering beams can be written as the sum of three terms:

$$OPD(t) = \mathbf{B} \cdot \hat{\mathbf{s}}(t) + l_0 + l_{atm}(t).$$

The first term represents the astrometric OPD and can be of the order of the baseline length $|\mathbf{B}|$. The second term is the internal OPD within the instrument while the third term represents the randomly fluctuating contribution to OPD made by the turbulent atmosphere. In this context our interest will focus on the term l_0, which is the parameter that, in most designs, is made variable in order to compensate for the astrometric term and get the condition of zero-delay. In long baseline instruments the amount of variable path that is required is considerable. Because of the wavelength dependence of the index of refraction of air, $n(\lambda)$, different wavelengths will travel different paths, if the compensation path is not in vacuum. For example, the actual ground-based longest baseline, which is the maximum baseline of SUSI (Sydney University Stellar Interferometer) and is of 640 m, implies an internal path length variation up to 420 m, and for a wavelength bandwidth from 400 to 800 nm the largest path length difference could be as much as 3 mm.

Tango (1990) examined the effects of uncompensated dispersion on the fringe visibility of a two-beam long baseline interferometer with a path compensator in air rather than in vacuum. He showed that by limiting the optical bandwidth to $\sim 100\ nm$ and using a compensating system with two glasses along the beam path, it is possible to achieve high fringe visibility, for beam path up to about 500 m. This in turn would imply that, without making use of vacuum delay line systems, the maximum baseline conceivable for a ground-based interferometer is also of the order of 500 m. The reason for Tango not to consider a vacuum system for the internal path travelled by the beam, is the complexity and the very high cost of the vacuum compensator.

As a matter of fact there are several discordant opinions about this statement, and when considering the cost and complexity of a mighty lunar interferometric project, we could reasonably wonder which project has the most demanding constraints.

However, even assuming that we are able to extend the baseline to lengths of the order of 1 km, another aspect needs to be considered, that related to diffraction effects on the beam.

Diffraction caused by finite apertures in long baseline interferometers leads to a significant loss of signal, decreased fringe visibility and a phase shift in the interference pattern. *Tango and Twiss (1974)* developed formulas for evaluation of the diffraction effects for the general case of an interferometer with unequal path lengths travelled by the beams in the instrument, and for incident plane waves propagating along these paths. On the basis of their results consequences for the performance of long baseline stellar interferometers can be derived.

First off all diffraction causes a loss of signal which, as for any instrument designed for astronomical use, is detrimental because it reduces the limiting magnitude reachable with the interferometer. However, this is not such a serious problem as it is the loss of fringe visibility resulting from the finite aperture of the instrument. This effect is baseline length dependent and, if neglected, would lead one to conclude, for example, that the star had an angular diameter larger than its actual one. This loss of visibility would not arise if the path length in each arm of the interferometer were exactly equal, but this could only be the case if the interferometer were steerable and oriented so that the baseline was always normal to the direction of the star. This does not seem practical with a Michelson interferometer when the baseline length is of the order of hundreds of meters in length. With an instrument fixed on the ground (it does not matter if on the Earth or on the Moon) one must correct for the external path difference, which arises continuously as the star moves across the sky, by using some sort of compensation, i.e. delay lines. Apart from the necessity, in the case of visible operation and ground-based instruments, of evacuating the delay lines, just discussed, another requirement for both ground- and lunar-based instrument is to keep the beam diameter in this part of the instrument enough smaller than the size of the collected beam, above all when large apertures are used, to keep small the dimensions of the optics and of the whole structure of the delay lines. However this is not a straightforward task to accomplish, when considering the deleterious effects due to

diffraction. *Tango and Twiss* show that the fringe visibility loss increases by increasing the length of the compensation path and the wavelength of observation, and by reducing the beam size. As an example, for a compensating path of 300 m and $\lambda = 0.5$ μm, to keep the visibility loss less than 5 % requires a beam size of 5.75 cm. However, if the visibility loss has to be less than 1 %, the beam size has to be about 5 times larger.

For a kilometric baseline the beam size needs to be even larger: for a maximum loss visibility of 2%, at visible wavelengths, it has to be of the order of 20 cm.

The effects of visibility degradation due to diffraction are even worse when the incoming wave is already distorted due to the turbulence of the atmosphere (see for example *Bagnuolo, 1988*). This will be an additional cause for the limitation in the maximum baselines which can actually be operative on Earth (even if extensive use of adaptive optics will certainly allow to get close to the condition of diffraction of an almost plane wave). However, even on the Moon, independently of the absence of distortion in the incident wave, the decrease of visibility due to diffraction together with the necessity of keeping the relay beam reasonably small (an even harder task if we want our instrument to be able to operate on large range of wavelengths, up to the infrared) will imply a maximum baseline length limit. For a baseline longer than 1 km designs with several stages for beam compression should be considered.

4.2 Integration time

To address the topic of the maximum integration time available to a lunar interferometer implies that we must take into account several aspects. In particular I will deal with how this depends on the combination of the parameters related to the baseline orientation, the telescope configuration and on the thermal environment conditions. These topics are addressed in three chapters of the second part of this study. The combination of these aspects is typical for a lunar instrument, and I did not deal with it when working on the IOTA project, for which the maximum integration time is set by the coherence time of the atmosphere, and the orientation of the baseline is that allowed by the natural configuration of the site (the two IOTA baselines are arranged on a L-shaped track with arms of the "L" aligned approximately NE and SE from a central station). In addition, an evaluation of the u-v coverage obtainable with IOTA, was performed by *Schloerb* (*1990, 1993*).

4.3 Minimum temperature of operation

According to *Ridgway (1990)*, a comparison of opportunities for optical aperture synthesis from the ground and from the Moon suggests that the lunar surface (at night) will be an excellent site for interferometry to faints limits in the intermediate wavelengths, 4-20 μm. At shorter wavelengths, ground-based observations are competitive. In the 5-20 μm region, ground-based will be competitive if the background photon limit can be reached. In the far infrared ($\lambda > 40$ μm) the faintest IRAS sources could be measured, and many other brighter sources imaged. However the temperature of the lunar environment limits the likely far infrared performance of a lunar system.

In particular *Ridgway (1990)* calculated the sensitivity limits for four telescope temperatures: 10 and 20 K as well as 80 and 100 K. The range 80-100 K is likely to be the lower limit for a lunar interferometer. In spite of the relatively high temperature, the predicted sensitivity is reasonably matched to the IRAS survey limit, owing to the long integration time available. The lower temperatures, 10 and 20 K, are in the range which might be chosen for a telescope cooled below ambient. Of course, much greater sensitivity would be possible in the far-infrared with telescope temperatures reduced to this range. However, the Moon is a poor location for a cooled interferometer. With the large horizontal optical paths (up to one km or more), multiple reflections, moving optical delay lines, and the problem of diffraction, it would be extremely difficult to design a thermally clean system. For these reasons, a very low temperature interferometer might be best realized in space. Also, with respect to a space-based instrument, the gain in performance with the slightly reduced temperature of a permanently shaded polar region does not, in itself, justify the presumably greater cost. On the other hand, the improved stability of a constant temperature environment would be a great advantage for any interferometer operating at any wavelength range.

Due to the importance of the thermal environment for the operation of a lunar instrument, a whole chapter of the second part of this study will be devoted to a detailed analysis of the thermal conditions for a lunar interferometer.

4.4 Scientific objectives for space- and lunar-based interferometry

Since the first conference on space interferometry, held in Baltimore in 1984, many papers describing both astrometric and imaging objectives have been published. I do not intend to propose again the list of the scientific objectives reachable with a medium- or long-baseline space interferometer. Very good reviews are presented in the following works and in the references wherein.

Workshop on Optical Interferometry from Space, Bulletin of the American Astronomical Society, Vol. 16, No.3 (Part 2), 1984. This review includes several papers on space astrometry science (*Russell, Black*) and of the possibility of imaging science of stars (*Dupree et al.*) and of galaxies (*Lauer*).

Science objectives for ground- and spaced-based optical/IR interferometry, S.T. Ridgway, in *A Lunar Optical-Ultraviolet-Infrared Synthesis Array (LOUISA)* Workshop Proceedings, NASA Conf. Publ. 3066, 1992. This paper presents some example of objectives for space interferometry in the following fields: primeval galaxies and galaxy formation, quasar and mass distribution of the Universe, the structure of active galactic nuclei, the scale of the Universe, galactic structure, planet formation around T Tauri stars.

The scientific support for a space interferometry mission, S.T. Ridgway, in *SPIE Proceedings* **1947**, *Spaceborne interferometry*, 1993. This work presents a brief history of

space interferometry and summarizes the status and the scientific case for both astrometry and imaging of a space interferometry mission.

Scientific support for space interferometry, D.M. Peterson in *SPIE Proceedings* **2477**, *Space Born Interferometry II*, 1995. This work represents the update continuation of that by *Ridgway (1993)*. In particular Peterson presents seven areas and scientific problems of currents interest which were identified by the NASA Space Interferometry Science Working Group. The subjects considered are: evolution of interacting binary systems, stellar luminosities, trigonometric parallaxes and the cluster-universe age problem, dynamics of small stellar systems, stellar dynamics of the Galaxy, astrometry of AGNs, rotational parallaxes: distance to nearby galaxies. The entire document produced by the Working Group is available by *"anonymous ftp"* at *sbast3.ess.sunysb.edu* in directory *"pub/siswg"*.

Infrared-Optical Telescope Array (IOTA): description of the instrument

In this chapter I present a description of the IOTA interferometer and its subsystems, and a review of scientific objectives (from the *Science Observing Plan for IOTA*). A more detailed description of the instrument can be found in *Traub et al. (1995)*.

0. Introduzione

Il progetto IOTA é una collaborazione fra la *Harvard University*, la *University of Massachusetts*, lo *Smithsonian Astrophysical Observatory* e la *University of Wyoming*.

IOTA consiste di un *array* di 3 collettori (per ora due sono operativi) di 0.45 m di diametro separati da una *baseline* che puó raggiungere la massima lunghezza di 38 metri. Lo strumento é stato studiato per realizzare osservazioni tanto nella banda visibile dello spettro (da 0.45 μm a 0.8 μm) che nell'infrarosso (da 1 a 2.4 μm). La massima risoluzione angolare dello strumento é di circa 1.4 *mas* nel visibile e di circa 6 *mas* nell'infrarosso.

Il sito dell'interferometro é sul monte Hopkins, in Arizona, poco distante dall'installazione del Multi Mirror Telescope. Le dimensioni e la forma dall'area geografica hanno determinato la geometria dell'*array*. I collettori sono disposti lungo i bracci di una *"L"* lungo i quali essi possono essere spostati in 17 stazioni diverse (Si veda fig. I.1).

fig. I.1 IOTA: i tre collettori del sistema possono essere collocati in ognuna delle 17 stazioni che si trovano lungo i bracci della dell'*array* a forma di L

Da ogni stazione i fasci ottici raccolti dai collettori sono trasferiti alla stazione posta

all'angolo della *"L"*, e da qui sono rimandati indietro lungo il braccio piú lungo dell'*array*, dove si trova la *delay line* utilizata per rendere uguali i cammini ottici percorsi dai fasci. Dopo aver percorso la *delay line* i fasci sono diretti in un'area definita il "laboratorio" dove si trovano due banchi ottici, uno usato per le osservazioni in infrarosso e uno usato per le osservazioni in banda ottica. Su questi banchi ottici sono collocati i sistemi per la combinazione dei fasci e i rivelatori del segnale interferometrico.

Un diagramma del percorso seguito dai fasci prima di arrivare nella zona di combinazione dei fasci é presentato in *Appendice B*.

1. Sottosistemi

Si descrivono brevemente le caratteristiche dei sottosistemi dello strumento. Informazioni dettagliate su ogni sottosistema si possono trovare in *Traub et al., 1995*, e nel *Manuale per l'utente di IOTA* (in fase di preparazione).

1.1 Collettori

Le principali componenti di ogni unitá che rappresenta un collettore sono un siderostato e un telescopio Cassegrain afocale. Queste componenti sono montate su di un sopporto di acciaio e posizionate in modo tale che l'asse del telescopio formi un anglo di 30° con il piano orizzontale parallelo al suolo.

La luce proveniente dalla sorgente viene riflessa dal siderostato, uno specchio piano di 46×69 *cm*, verso il telescopio che agisce da compressore del fascio (*beam-compressor*). Il telescopio é costituito da uno specchio primario parabolico (f/2.5) di 45 cm di diametro e da un secondario, anch'esso parabolico, di 6 cm di diametro. Insieme producono una riduzione della sezione del fascio incidente pari ad un fattore 10. Sia il primario che il secondario sono costruiti in zerodur.

La specifica sulla superficie delle ottiche, tanto per lo specchio piano che per la combinazione primario-secondario, richiede che l'errore sul fronte d'onda introdotto sia minore di $\lambda/10$ (variazione picco-picco nella distorsione del fronte d'onda). Questa specifica é verificata, per lunghezze d'onda infrarosse (2 μm), sul 90% dell'area delle ottiche maggiori. Per lunghezze d'onda nel visibile (0.5 μm) la specifica vale entro qualsiasi apertura di 15 cm localizzata intorno al centro del raggio del primario. Un'ulteriore specifica che la coppia di compressori deve soddisfare é che i loro fattori di riduzione siano uguali al meglio del 2% di accuratezza.

Le copertura delle ottiche é cosituita da un sottile strato d'argento che assicura il 98-99% di riflettanza fino a lunghezze d'onda di 0.5 μm (nel blu le prestazioni sono inferiori).

Ogni insieme siderostato piú *beam-compressor* é collocato all'interno di un "capanno" che ho lo scopo di isolare e proteggere le ottiche dall'ambiente esterno. Il tetto é costituito da diverse unitá disposte una adiacente all'altra che, mediante azione manuale, possono essere fatte scorrerre una sull'altra, in modo da aprire il tetto ed esporre il siderostato al cielo. La costruzione che ospita il collettore si puó muovere lungo il braccio dell'interferometro percorrendo un binario di acciaio di 5 metri di larghezza. Durante la fase di trasporto del

capanno, gli elementi del collettore sono rimossi dalla posizione di lavoro per mezzo di un sistema di sollevamento idraulico.

1.2 Sistema per il trasporto del fascio

Il fascio che emerge dal *beam-compressor* viene indirizzato verso il basso da uno specchio piano attivo ($60 \times 70\ mm$) mosso da una coppia di attuattori piezoelettrici. Lo specchio attivo viene riposizionato in base all'informazione che arriva da un rivelatore che é posto nella zona dove avviene la ricombinazione dei fasci. In questo modo si assicura la stabilitá della posizione dell'immagine al momento della ricombinazione.

Il fascio ottico raggiunge quindi il sistema per il trasferimento (*relay*) del fascio vero e proprio, che consiste di una serie di specchi opportunamente disposti per la guida del fascio, collocati all'interno di un tubo a vuoto.

Per gli specchi del sistema di *relay* si richiede un'accuratezza nella superficie pari a $\lambda/20$ (picco-a-picco). Gli specchi sono provvisti di motori che permettono il loro riposizionamento a distanza.
La disposizione degli specchi del sistema di *relay* assicura che ogni fascio subisca esattamente la stessa sequenza di riflessioni (cioé dopo lo stesso numero di riflessioni i fasci viaggiano nella stessa direzione) in modo da assicurare che arrivino al punto di ricombinazione nello stesso stato di polarizzazione. Questo é necessario per evitare una diminuizione nella visibilitá della frangia interferometrica (*Traub, 1988*).

1.3 Linee di ritardo

Lo strumento deve disporre di un sistema per la variazione del cammino ottico dei fasci, sia per compensare la differenza nella distanza dei telescopi dal laboratorio, sia per compensare la differenza di cammino dall'oggetto ai telescopi, quando si osserva in direzioni diverse dallo zenith.

Il sistema di linee di ritardo (*delay lines*) di IOTA é suddiviso in due parti. Una *delay line* é collocata lungo il braccio piú lungo dell'interferometro e copre una distanza di circa 30 metri ed é usata per una compensazione grossolana della differenza nel cammino ottico dei fasci. Il meccanismo che permette la variazione di questa *delay line* é costituito da un carrello che si muove su una rotaia di acciaio la cui posizione é controllata per mezzo di un laser. Il carrello trasporta uno specchio diedro il cui angolo rispetto alla direzione del fascio incidente puó essere aggiustato, dopo che la posizione di compensazione é stata acquisita, per mezzo di un sistema meccanico attivato a distanza.

La *delay line* piú corta (2.4 metri in estensione) é quella responsabile della compensazione fine della differenza di cammino ottico. Per variare la *delay line* si usa uno specchio diedro montato su un carrello che si sposta, su di un cuscino d'aria, lungo una guida di granito perfettamente levigato. La posizione del carrello é controllata per mezzo di un fascio laser. L'assenza di contatto meccanico fra il carrello e la guida assicura una linearitá nello spostamento tale che l'errore introdotto nella fase del fronte d'onda é minore di quello dovuto alla turbolenza atmosferica.

La lunghezza di questa *delay line* permette di seguire l'oggetto in cielo per un periodo di circa mezz'ora prima che si necessario riposizionare la *delay line* piú lunga.

1.4 Sistema a vuoto

La necessitá di far viaggiare i fasci ottici all'interno di un sistema a vuoto, origina fondamentalmente da due motivi. Il primo é che la differenza di cammino ottico dalla sorgente a due collettori diversi é quasi del tutto percorsa nello spazio al di sopra dell'atmosfera. Una corretta compensazione dovrebbe avvenire nelle stesse condizioni di assenza di atmosfera. Il secondo motivo, piú vincolante, é che a causa della dipendenza dell'indice di rifrazione dell'atmosfera dalla lunghezza d'onda, lunghezze d'onda diverse percorrono cammini ottici diversi. Al crescere delle distanze percorse dal fascio, questa differenza aumenta, con gravi effetti sulla visibilitá della frangia interferometrica.
Per evitare gli effetti di dispersione introdotti dall'atmosfera, tanto gli elementi per il *relay* del fascio che le *delay lines* sono collocati all'interno di condotti dove si mantiene il vuoto.

Per IOTA, la parte di dimensioni maggiori del sistema a vuoto é quella che contiene la *delay line* per l'aggiustamento fine e tutte le componenti associate con il sistema di trasferimento del fascio da e verso la *delay line*. Questa parte ha un diametro di 0.8 m e una lunghezza di 9 metri. Ad essa sono connessi i due tubi entro i quali sono trasferiti i fasci provenienti dai telescopi: un tubo di 10 cm di diametro per il braccio piú corto dell'interferometro e un tubo di 40 cm di diametro per quello piú lungo (che contiene anche la *delay line* lunga).

Il vuoto é ottenuto per mezzo di un sistema di due pompe che insieme sono in grado di mantenere la pressione al di sotto di 1 $torr$, anche in presenza del gas emesso dal sistema a cuscino d'aria che trasporta il carrello della linea di ritardo fine. Questo sistema immette nel tubo 15 litri di gas al minuto. Le pompe sono collocate in una costruzione separata, posta ad alcuni metri di distanza dal laboratorio. Al fine di isolare il laboratorio dalle vibrazioni indotte dalle pompe, queste sono montate, mediante un sistema di molle la cui frequenza di risonanza é di 1 Hz, su di un supporto di cemento di 30 cm di spessore.

1.5 Rivelatori

All'uscita dal sistema a vuoto i fasci si trovano nel laboratorio. Per le osservazioni in infrarosso il fascio passa attraverso uno specchio dicroico che invia verso il *beamsplitter* infrarosso solo la radiazione con lungheza d'onda maggiore di 1 μm. I due fasci che escono dal *beamsplitter* sono raccolti da due rivelatori ad antimoniuro di indio (InSb) a elemento singolo, che sono mantenuti ad una temperatura di circa 60 K per mezzo di un sistema di raffreddamento ad azoto liquido.
La parte di radiazione visibile che arriva al dicroico é invece diretta verso il rivelatore che invia i segnali di guida allo specchio attivo posto dopo il *beam-compressor*.

Per osservazioni nel visibile solo una parte del fascio va al sistema di guida, il resto viene convogliato verso il *beamsplitter* e i fasci in uscita da esso sono focalizzati, per mezzo di un sistema di 4 lenti, su fibre ottiche multimodali. Ogni lente raccoglie la luce proveniente

da un'area del fronte d'onda pari all'area di coerenza r_o. Per mezzo delle fibre ottiche questa porzione del fronte d'onda raggiunge la fenditura di uno *spettrometro-grism*[1] che é stato progettato per ottenere una dispersione costante della radiazione rispetto alle frequenza (*Traub,1990*). La radiazione cosí dispersa é focalizzata sul catodo di un rivelatore a conteggio di fotoni, che in questo caso é una *PAPA* [2] *camera* (*Papaliolios et al., 1985*) che opera nella banda spettrale compresa fra 0.45 e 0.80 μm.

Infine una parte della radiazione visibile viene inviata al rivelatore per il sistema di guida dello specchio attivo. Il rivelatore é cosituito da un piccolo CCD di 32×64 pixels, di cui la parte attiva (cioé che riceve i fotoni) é data solo da 32×32 pixels. Dopo una breve integrazione (meno di 0.01 s) la carica raccolta nella metá attiva viene rapidamente spostata nella metá schermata, dove avviene la lettura. In questo modo acquisizione e lettura della quantitá di carica immagazzinata possono avvenire contemporaneamente, il che riduce i tempi morti nel sistema di controllo della posizione dei fasci. I centroidi delle due immagini provenienti dai due collettori, e le correzioni da apportare per farli sovrapporre, sono calcolate da un computer, che invia i segnali di guida agli attuattori piezoelettrici degli specchi attivi all'uscita dei *beam-compressors*.

[1] Un grism é una componente ottica costruita combinando un reticolo di diffrazione (grating) e un prisma (prism). In questo modo la radiazione dispersa dal reticolo emerge dal prisma in direzioni parallele per tutte le lunghezze d'onda.

[2] Precision Analog Phothon Address.

2. Programma scientifico per IOTA

Al momento IOTA dispone di due telescopi di 45 cm di diametro e puó sfruttare una *baseline* massima di circa 38 m. Con i nuovi rivelatori che sono in fase di costruzione (un CCD 512×512 da sostituire alla PAPA camera, e una camera NICMOS-3 256×256 HgCdTe, con cui sostiture i rivelatori infrarossi attuali) IOTA sará in grado di osservare oggetti di magnitudine limite $V = 9$ con una risoluzione angolare di 1.5 mas nel visibile, e di magnitudine limite $K \simeq 8$ con una risoluzione angolare di 6 mas nell'infrarosso.

In quanto segue vengono presentati le classi di oggetti e il tipo di indagine scientifica che sono alla portata della potenzialitá osservativa dello strumento.

2.1 Oggetti Becklin-Neugebauer: molteplicitá, asimmetria e distribuzione della polvere

Gli oggetti Becklin-Neugebauer sono sorgenti infrarosse compatte, fortemente oscurate nel visibile, probabilmente protostelle di grande massa in fase di collasso. In questa classe ci sono almeno 7 oggetti che possono essere osservati con IOTA. Il prototipo di questi oggetti, che si trova in Orione, é sufficientemente vicino e brillante perché IOTA possa misurare il suo diametro angolare e di consguenza si possa risalire alla sua temperatura efficace superficiale. Di questo oggetto si vogliono studiare possibili deviazioni della superficie dalla simmetria sferica, e asimmetrie nella distribuzione della polvere attorno all'oggetto, nella forma di emissioni bipolari. Per altri oggetti di questa classe si vuole anche indagare sulla possibilitá che essi siano in realtá oggetti multipli.

2.2 Molteplicitá in stelle T-Tauri

Alle distanze delle regioni di formazione stellare Taurus e Ofiuco ($\sim 160 - 200$ pc), IOTA é in grado di risolvere sistemi binari fino ad un limite di separazione dell'ordine di 1 - 1.5 AU, alla lunghezza d'onda di 2.2 μm. Questo rappresenta un miglioramento di un ordine di grandezza rispetto al limite di risoluzione (14 AU) per sistemi binari T Tauri raggiunto per mezzo dell'interferometria *speckle* in infrarosso. IOTA permetterá di verificare se l'elevata incidenza di sistemi binari riscontrata su separazioni maggiori di 14 AU si mantiene anche per separazioni molto minori. Inoltre, la maggior parte delle binarie strette che IOTA sará in grado di individuare, avranno periodi sufficientemente brevi da permettere di ricostruire l'orbita degli oggetti e di stimare (facendo uso di dati sulla velocitá radiale ottenuti con metodi spettroscopici) la loro massa. IOTA sará anche in grado di verificare la presenza o meno di un disco attorno alle binarie T Tauri. Nel caso la ricerca dia risultati positivi per tutte le binarie, questo costituirá un ulteriore incentivo per lo sviluppo del programma di ricerca di pianeti extrasolari associati a sistemi binari (si veda *2.5*).

2.3 Brillamenti e molteplicitá in oggetti Herbig Ae/Be

Ci sono circa 100 oggetti Herbig Ae/Be che sono alla portata di IOTA, e tutti questi risultano interessanti da esaminare perché ci si aspetta che ognuno di questi oggetti presenti caretteristiche nella struttura assai diverse. Si ritiene che questi oggetti abbiano caratteristiche simili a quelle delle T-Tauri ma abbiano masse maggiori e dischi in fase

di collasso verso la stella. I *flares* (brillamenti) che sono stati osservati potrebbero essere dovuti a materiale che si é spostato dal disco alla superficie stellare: un tale evento potrebbe essere direttamente osservato con IOTA. Inoltre, se IOTA rivelasse che questi sistemi sono multipli, ne deriverebbero conseguenze per quanto riguarda l'interpretazione del tipo spettrale e la stima della massa.

2.4 Sistemi binari e massa di giganti, supergiganti e stelle di tipo spettrale M

Lo studio di oggetti binari fornisce gli elementi chiave per una accurata (1%) determinazione della massa di stelle giganti, supergiganti, e di stelle nane di tipo M. Le osservazioni con IOTA si concentreranno su questi tre gruppi di oggetti perché la massa di stelle di tipo solare appartenenti alla sequenza principale sono giá state determinate con accuratezze dell'ordine dell' 1%, o meglio, con altre tecniche osservative. La conoscenza della massa di una stella, insieme con la sua luminositá e il suo raggio, é fondamentale per determinare le proprietá astrofisiche dell'oggetto e il suo processo evolutivo. La determinazione di stime della massa molto piú accurate di quelle finora disponibili, permetterá di predire le proprietá di un singolo oggetto appartenente a questi gruppi con maggiore accuratezza di quanto possibile oggi, e di fornire elementi per la verifica della validitá dei modelli stellari teorici.

2.5 Pianeti in sistemi binari

Con IOTA sará possibile investigare la possibilitá che pianeti possano formarsi e rimanere in orbite stabili presso sistemi stellari doppi. Questa indagine sará condotta facendo accurate osservazioni dei sistemi binari accessibili allo strumento, alla ricerca di una piccola ma significativa variazione nella separazione vettoriale fra le due stelle, che dovrebbe essere il segnale della presenza di un pianeta orbitante attorno ad uno dei due oggetti. L'interferometro verrá equipaggiato con una nuova componente ottica (si veda *Traub, Carleton, and Porro, 1995*) per mezzo della quale sará possibile individuare la variazione nell'orbita di una componente di un sistema binario ad una distanza dell'ordine di 40 *pc*, indotta da un pianeta delle dimensioni di Giove, Saturno o Nettuno.

2.6 Cefeidi

Accurate misure della dimensione angolare delle stelle Cefeidi piú vicine saranno un elemento utile per verificare il valore, ottenuto per via teorica, del diametro lineare degli oggetti di questa classe e per migliorare la misura della distanza scala che é utilizzata per la determinazione delle dimensioni dell'universo. La misura della distanza delle Cefeidi puó essere ottenuta confrontando le misure dei diametri angolari e delle velocitá di espansione delle atmosfere stellari, per mezzo di una teoria di trasporto radiativo in un'atmosfera in espansione che tenga conto di entrambi questi parametri. Poiché peró il proptotipo delle stelle Cefeidi ha dimensioni angolari dell'ordine della massima risoluzione di IOTA, questo tipo di misure saranno piú accuratamente realizzate con strumenti che possono sfruttare *baselines* maggiori di quella di IOTA.

2.7 Distribuzione di polvere attorno a stelle di tipo spettrale avanzato

Alcune stelle in fase evolutiva avanzata producono copiose quantitá di polvere, che si distribuisce attorno alla stella sottoforma di disco o *shell* (guscio). Al momento non esiste ancora un buon modello che descriva questo processo. Per poter sviluppare modelli piú accurati, occorrerebbe conoscere, per esempio, se la produzione di polvere é continua o episodica, se si forma isotropicamente oppure si forma di preferenza ai poli o all'equatore, se é riscaldata direttamente dalla stella o se bisogna tenere conto dell'effettto della profonditá ottica dovuta a meteriali intermedi, e conoscere le proprietá dei materiali. Tutte queste questioni possono in qualche modo essere investigate per mezzo dell'interferometria infrarossa e osservazioni sono state fatte recentemente su 13 oggetti di questo tipo, alla lunghezza d'onda di 11 *mum* e con *baselines* dell'ordine della decina di metri. IOTA sará in grado di ampliare la gamma di oggetti osservati, con una risoluzione angolare 15 volte migliore.

2.8 Regioni attive sulla superficie delle stelle

I risultati di alcune indagini fotometriche hanno dimostrato che é probabile che alcune stelle siano decisamente piú attive del sole, il che suggerisce che sulle loro superfici siano presenti ampie zone sede di fenomeni quali macchie stellari, protuberanze, brillamenti, facole, etc. Questi fenomeni essere individuati e messi in evidenza sfruttando la massima risoluzione di un inteferometro a lunga base. Fra l'altro, la scoperta di regioni attive sulle superfici stellari potrebbe fornire utili informazioni per la comprensione del meccanismo di produzione di polvere attorno a stelle di tipo spettrale evoluto. A questo riguardo una delle ipotesi proposte é che la polvere si formi in piccole regioni fredde della superficie e venga poi immessa nello spazio circostante direttamente da questi siti. Comunque, affinché possano essere osservate, occorre che queste zone di emissione di polvere siano sufficientemente estese e abbiano un buon contrasto rispetto alla fotosfera sottostante. Piú in generale, la ricerca di zone di asimmetria nella distribuzione della polvere in prossimitá della stella, potrá fornire informazioni utili nella definizione delle caratteristiche della sorgente che la produce. Se la polvere invece che essere prodotta in piccole zone della superficie, é prodotta per mezzo di onde di shock (un'altra delle ipotesi in proposito), la distribuzione spaziale della polvere in prossimitá della stella dovrebbe apparire alquanto diversa.

2.8 Andamento della temperatura negli strati superiori dell'atmosfera di stelle giganti e supergiganti

Il profilo della distribuzione della temperatura atmosferica in stelle giganti e supergiganti non é in genere ben determinato, a causa della scarsa conoscenza che si ha nell'ambito della spettroscopia molecolare dei loro cositutenti, TiO e H_2O. Un significativo contributo in questo ambito puó essere offerto da misure accurate di visibilitá, per mezzo delle quali si possono determinare i diametri angolari degli oggetti e l'oscuramento al bordo (*limb-darkening*). Queste due misure insieme permettono di stimare l'andamento verticale della temperatura negli strati superiori dell'atmosfera della stella. Questi effetti infatti

cambiano con la lunghezza d'onda e i diametri sono in genere piú estesi in regioni di maggiore opacitá.

2.9 Temperatura effettiva di stelle fredde

Il diagramma H-R presenta una evidente lacuna nella regione estrema delle stelle di tipo M di bassa temperatura. Osservazioni fatte con IOTA permetteranno di incrementare un ordine di magnitudine il numero di punti in questa zona del diagramma. La misura della visibilitá delle frange in funzione della lunghezza della *baseline*, permette di risalire al diametro angolare dell'oggetto e all'entitá di *limb-darkening* che, insieme a misure di flusso assoluto emesso dalla stella, permettono di ricavare la sua temperatura efficace (misure di questo tipo saranno fatte in un ampio intervallo di lunghezze d'onda, dal visibile, 0.5 μm, all'infrarosso, 2.2 μm).

2.10 Stelle variabili di tipo Mira

Le stelle di tipo Mira sono variabili, giganti o supergiganti, di classe spettrale M. Si trovano nella fase evolutiva della combustione dell'idrogeno e dell'elio, il loro raggio cambia periodicamente di un fattore due e la temperature effettiva cambia di alcune centinaia di gradi Kelvin. IOTA sará in grado di ottenere misure del diametro angolare durante l'intero ciclo di variazione, fornendo dati particolarmente utili per la verifica dei modelli teorici di questi oggetti. Inoltre le osservazioni saranno fatte a lunghezze d'onda diverse, sia nelle regioni del continuo, sia nelle bande di assorbimento che sono state riscontrate tanto nel visibile che nell'infrarosso.

Telescope alignment

0. Introduction

One of the causes of distortion in the wavefronts that are going to interfere in a stellar interferometer, is the misalignment of the optics of the telescopes. Wavefront distortions are responsible for deterioration in the interfering fringes, above all by causing a reduction in the fringe visibility and a variation of the optical path difference between the two interfering beams, which alters the astrometric information that is deduced by the interferometric measurement.

In order to study the effects of telescope misalignments on the performance of IOTA, I performed a simulation of some misalignment conditions and evaluated the distortion suffered by the wavefront. The simulation was in particular intended to produce the interferograms resulting from the interference of the distorted wavefront with a perfect one, which is a measure of the degrading of the optical performance of the system. Also the simulation produces the diagram of the intensity distribution of the beam when it reaches the laboratory.

The results of this simulation are collected in an atlas (see Appendix: *Interferograms Atlas*) that is intended to be used, as a reference for characteristic misalignment conditions, during the operations for the alignment of the telescope optics.

In the following sections, I present first a brief introduction to the theory of wavefront aberration. Then I present an analysis of the optical system of IOTA, in the limit of the paraxial approximation, in order to study the effects of the optics misalignments on the beam position in the laboratory.

Finally, the rest of the chapter is devoted to the description of the programs I used to produce the interferograms and the beam intensity diagrams, and to present the results of the simulation.

1. Wavefront as a function of aberration coefficients

The theory of aberrations, generally called the Seidel theory, has been treated by many authors, including Born and Wolf who derived it in a detailed analytical form. Here I follow the approach by Schroeder (*Schroeder, 1987*) which leads to results that can easily be applied to the optical systems of interest for our instrument.

This approach consists in applying Fermat's principle to a generic surface of revolution and then in deriving the relations which describe the aberrations suffered from a wavefront interacting with the surface. The wavefront eventually is given as a function of defined coefficients, the *aberration coefficients*. Only the main steps and the final results of this derivation are presented in this section.

In its generalized form Fermat's principle states that the actual path a ray follows is such that the time of travel between two fixed points has a stationary value with respect

to small changes of that path[3]. An equivalent, here adopted, statement is obtained by replacing the words "time of travel" with "optical path length" (OPL).[4] Application of Fermat's principle to a pair of conjugate points allows to find a surface which acts as a perfect imaginery system for that pair. However, for any other pair of points outside the range defined by the paraxial approximation[5] Fermat's principle is not strictly satisfied and the geometrical image of a point is no longer a point, but becomes a blur. When an optical system produces a blurred image, where the blur is in addition to that produced by diffraction, the system has aberrations.

In terms of Fermat's principle this means that different rays have different optical path lengths, while in terms of wavefront this means that a spherical wavefront emerging from the image point is converted by the optical system into a wavefront which departs from a spherical one.

[3]Hence, in first approximation, the time of travel when the ray follows two strictly adjacent paths is the same. If the time of travel from P_0 to P_1 is denoted by τ, then the condition that τ has a stationary value for the actual path is

$$\frac{\partial \tau}{\partial x} = \frac{\partial \tau}{\partial y} = 0$$

where x and y are the generalized coordinates of the point where the ray intersects the surface.

[4]The optical path length (OPL) is expressed in terms of the geometric path length and of the index of refraction, n; if dt is an infinitesimal time of travel, we can write

$$d(OPL) = n \cdot ds = \frac{c}{v} \cdot vdt = c \cdot dt$$

where v is the speed of the light in the medium of refractive index n and c is the velocity of light in vacuum. Considering the variations $\delta(OPL)$ and $\delta\tau$, we can write

$$\begin{aligned} \delta(OPL) &= \frac{\partial(OPL)}{\partial x}\delta x + \frac{\partial(OPL)}{\partial y}\delta y = \\ &= c\left(\frac{\partial \tau}{\partial x}\delta x + \frac{\partial \tau}{\partial y}\delta y\right) = c \cdot \delta \end{aligned}$$

Hence the general statement for the Fermat's principle is either $\delta\tau = 0$ or $\delta(OPL) = 0$.

[5]The paraxial approximation is used to describe the properties of an optical system when rays close to the axis and nearly parallel to it are considered, i.e. when sines and tangents of angles can be replaced with the angles themselves. In this approximation there is always a one-to-one correspondence between object and image point.

fig. TA.1 Path of an arbitrary ray through a refracting surface. Points Q and Q' lie in the xz plane; point B is on the surface. The chief ray passes through the origin of the coordinate system.

In what follows Fermat's principle is applied to a surface of revolution and the general expression for the transverse and the angular aberration is found.

With reference to the sketch in fig. TA.1, and for a general surface of revolution, the optical path length (OPL) between point Q and $Q^{'}$ is given by

$$OPL = n[QB] + n^{'}[BQ^{'}]$$

where $[QB]$ and $[BQ^{'}]$ are the segments of the ray to the left and right of the point B on the surface, n and $n^{'}$ are the indices of refraction of the homogeneous medium to the left and to the right of the surface. The dashed line from Q to $Q^{'}$ passes through the vertex of the surface. It represents the chief ray[6] through the system with angles θ and $\theta^{'}$ with respect to the z-axis.

The OPL measured along the chief ray is simply $n^{'}s^{'} - ns$ (it is $s' > 0$ and $s < 0$, because of the sign convention that distances are positive when measured rightward and upwards with respect to the center of the coordinates system). By writing the appropriate relations for $[QB]$ and $[BQ^{'}]$ and applying the binomial expansion, the OPL is eventually expressed in terms of the quantities which define the position of Q and $Q^{'}$ (s , θ , $s^{'}$, $\theta^{'}$), the surface parameters (radius of curvature, R, conic constant, K) and the aperture radius ($\rho = \sqrt{x^2 + y^2}$). The whole equation is given in *Appendix C*. As far as the application of Fermat's principle is concerned, it is useful to note that in eq.(12) in *Appendix C* the

[6]The chief ray is a ray from an object point which passes through the center of the aperture stop of the system, in this case the vertex of the surface

first set of parentheses denotes the OPL of the chief ray. Because Fermat's principle is concerned with optical path differences and stationary values it is appropriate to remove this term by defining a quantity G as the optical path difference (OPD) between the general ray and the chief ray. Given this definition G is expressed by

$$G = OPL - OPL(chief\ ray) = A_0 x + A_1 x^2 + A_1^{'} y^2 + A_2 x^3 + A_2^{'} x y^2 + A_3 \rho^4 \qquad (1)$$

where the A_i's are the multiplying coefficients given in *Appendix C*. Applying Fermat's principle, in the form $\delta(\mathrm{OPL}) = 0$, to equation (1) gives

$$\frac{\partial}{\partial x}(OPL) = \frac{\partial G}{\partial x} = 0,\ \frac{\partial}{\partial y}(OPL) = \frac{\partial G}{\partial y} = 0 \qquad (2)$$

For an optical system that satisfies eq.(2) for any (x, y) within the aperture, each of the coefficients in eq.(1) must be zero, and the system is perfect. If one or more of these coefficients is nonzero, then aberrations are present. The size of the given aberration is directly proportional to the corresponding coefficient in eq.(1), and each of them can be evaluated for a given surface of revolution. In particular, for $x = y = 0$, eq.(2) is satisfied only if $A_0 = 0$, i.e. $n^{'} sin\theta^{'} - nsin\theta = 0$, which is the Snell's law for the chief ray.

If we limit the analysis to third-order angular and transverse aberration[7], it is shown that $A_1 = A_1^{'}$ and $A_2 = A_2^{'}$ so that we can write

$$G = A_0 x + A_1 (x^2 + y^2) + A_2 (x^3 + x y^2) + A_3 \rho^4 \qquad (3)$$

It eventually results that A_1 is a measure of the astigmatism of the system, A_2 is a measure of coma, and A_3 of spherical aberration.

The last step is to consider the relationship between ray and wavefront aberrations. An optical system free of aberrations takes light from an object point Q, for which the wavefront is a sphere centered on Q, and images it at the point Q'. The wavefront of the light converging toward $Q^{'}$ is a sphere centered at $Q^{'}$, and the OPL along any ray through the system, i.e. the OPL for any pair of conjugate points imaged by the system, is constant. Thus G in eq.(1) is zero for any ray. This spherical wavefront is taken as reference and designated Σ_r.

[7] This implies retaining in eq.(12) in *Appendix C* only those terms for which the sum of power of θ and ρ, or x, or y is not greater than four, because the angular and transverse aberration coefficients are evaluated by differentiating G (see eqq.(6) and (7)).

fig. TA.2 Cross sections of reference and aberrated wavefronts, Σ_r and Σ_a. The radius of curvature of the reference wavefront is $s^{'}$.

For a system with aberrations the wavefront converging toward $Q^{'}$ is no longer spherical and, depending on the sign of G, is either advanced or retarded with respect to the corresponding point on the reference wavefront. This aberrated wavefront is designated Σ_a. The two wavefronts, sketched in fig. TA.2, are in contact at their centers, where $G = 0$, because the ray passing through the center is the chief ray.

At any other point on the actual wavefront, G is the OPD between Σ_r and Σ_a. The geometric distance along any ray between the wavefronts is $G/n^{'}$ and by designating this distance as Δ we have

$$\Delta = \frac{G(x,y)}{n^{'}} = \frac{1}{n^{'}}\left[\Sigma_r(x,y) - \Sigma_a(x,y)\right] \tag{4}$$

Differentiation of eq.(4) gives

$$\frac{\partial \Delta}{\partial x} = \frac{1}{n^{'}}\frac{\partial G}{\partial x} = \frac{1}{n^{'}}\left(\frac{\partial \Sigma_r}{\partial x} - \frac{\partial \Sigma_a}{\partial x}\right), \tag{5}$$

with a similar relation in which y replaces x. The quantity in parentheses in eq.(5) is the difference in slopes between the reference and the aberrated wavefront in a plane parallel to the xz plane. Because rays are perpendicular to wavefronts, this is also the difference between the slopes of the ray for a perfect system and the actual ray, each at point (x, y) on the respective wavefronts. Given this difference in slopes there is a consequent transverse aberration in the x direction of the image given by

$$TA_x = s^{'}\frac{\partial \Delta}{\partial x} = \frac{s^{'}}{n^{'}}\frac{\partial G}{\partial x} \tag{6}$$

where TA_x denotes Transverse Aberration in the x-direction. The corresponding Angular Aberration, denoted by AA, is given by

$$AA_x = \frac{TA_x}{s'} \tag{7}$$

and similar relations are valid for the y-direction.

As far as ray aberrations are concerned, the previous equations show that they depend on aperture radius ρ and field angle θ according to the relation

$$aberration \propto \rho^n \theta^m, \tag{8}$$

where n + m gives the order of approximation of the ray aberration (the wavefront aberration is given by eq.(4), hence the wavefront aberration order is n + m + 1). The Seidel aberrations, referred to as the third order aberrations, are derived considering the third-order ray approximation for both angular and transverse aberrations. In this case five principal aberrations and their dependence on ρ and θ can be identified: spherical aberration ($\propto \rho^3$), coma ($\propto \rho^2\theta$), astigmatism ($\propto \rho\theta^2$), curvature of field ($\propto \rho\theta$), and distortion ($\propto \theta^3$). The first three aberrations affect image quality, while the last two affect only image position.

For an optical system with many optical elements the same analysis can be performed to derive the aberration coefficients for the wavefront at the end of the system: Fermat's principle is applied to the OPD between an arbitrary ray and the chief ray, following the ray from the original object to the final image through all the elements of the system. The OPD is then given by

$$G_s = G_1 + G_2 + \ldots + G_f = \sum G_i, \tag{9}$$

where f denotes the last surface. Each term in eq.(9) can be replaced by eq.(1) with the appropriate (x, y) at each surface, so that eventually the geometric distance, along an arbitrary ray, between the actual and reference wavefronts is given by

$$\Delta = \frac{G_s(x, y)}{n'} \tag{10}$$

At this point of the analysis it is possible to calculate the aberration to the third-order for a general centered optical system. It is useful to remember that even if results in this approximation are not exact, they are usually sufficient for the analysis of most systems used in optical astronomy. Nevertheless, the raytrace program used in the simulation for the study of the telescope misalignments evaluates the aberration coefficients to higher order, including more than the classical terms.

2. Relation with the alignment operations for IOTA

From the previous section we know that a perfect optical system is one in which the wavefront emerging from the final surface is spherical and that there is a close relation between deviations from a spherical wavefront and the appearance of aberrations. Along any ray

the actual wavefront may be behind or in front of the ideal wavefront, and the retardation or advance of the ray is expressed by the quantity G. When the emerging wavefront is made interfere with an ideally perfect wavefront the interferogram that results is a function of G, the actual wavefront error. On the other hand G is also a consequence of any error in the position of one or more of the optical elements of the system. This position error can be a decenter or tilt of one of the mirrors or a displacement of one of the mirrors towards another one.

In the specific case of IOTA, these two aspects are taken together to analyze the misalignment and defocus of the telescope when it is operated in *auto-collimation* mode. In short the process can be described with reference to fig. TA.3.

fig. TA.3 Schematic of an IOTA telescope operated in auto-collimation mode

Light from a laser source in the laboratory[8] is directed towards the primary mirror. After entering the system through the inner hole of the primary mirror the light beam is reflected by the secondary towards the primary and reflected back to the siderostat. The beam bounces between the primary and the secondary one more time and eventually exits the telescope. Between the primary and secondary mirrors, the beam diameter changes by a factor of 10, the compressing factor of the IOTA telescope. The emerging wavefront is then made interfere with the reference wavefront from the laser source and the resulting interferogram analyzed in the laboratory.

What we expect to see is fringes which differ in shape and frequency depending on the different misalignment affecting the system. In order to eliminate this disturbance it is

[8]Laboratory is the place, the optical table, where the beams, which are going to interfere, arrive after travelling along the delay lines, and are eventually combined.

necessary either to correct the position of the element which is misaligned or compensate for the misalignment effects. To do this, we need first to relate the observed fringes to specific misalignments.

The scope of the performed simulation is to produce an atlas of interferograms and beam intensity plots obtained from well defined, both in direction and in magnitude, misalignments. By a comparison of the observed and simulated fringes we should then be able to find the appropriate adjustment of the optical elements which aligns (or re-aligns) the system.

3. Analysis of the IOTA optical system in the limit of the paraxial approximation

In this section I present a description of the effects, due to typical misalignments of the optics, expected when the IOTA telescope is operated in autocollimation mode.

First I briefly present the characteristics of the beam compressor and how it images an object. Then I consider some typical misalignments and analyse how they affect the beam position (by predicting the beam shift in the laboratory) and which wavefront distortion they cause.

I remind the reader that all the calculations hereafter are performed in the limit of the paraxial approximation and that this description is not intended to be an accurate analysis of the optical behaviour of the IOTA system. Here I present what is useful to know in order to test the simulation program, to interpret the simulation results and to utilize them during the alignment operations.

3.1 The IOTA beam compressor

The schematic layout, reported in *Appendix B*, of the IOTA optics consists of the siderostat, the beam compressor, a piezo mirror, a coarse delay line, corner mirrors, a fine delay line, and a reference surface in the laboratory. For the analysis of the telescope misalignment the elements to be considered in the simulation are the siderostat and the beam compressor.

The siderostat consists in a flat mirror that, when the system is operated in autocollimation, is oriented to be parallel to the primary mirror. The beam compressor is what usually we would refer to as the telescope, but in our case when we say telescope we refer to the beam compressor and siderostat together. As far as the beam compressor is concerned, it is an afocal system whose specifications have to provide a reduction in the size of the incoming beam by a factor which meets the requirements due to the interferometer architecture, and optical performance.

An afocal beam compressor consists of a pair of paraboloids, a concave primary and a convex secondary, whose focal points coincide. This combination converts an input beam of diameter D into a collimated output beam of diameter $k\,D$, where $k = h_S/h_P$ which is the ratio of the marginal ray's height at the margin of the secondary and of the primary

respectively. The pair of paraboloids is free of spherical aberration and also has zero third-order coma and astigmatism (*Schroeder, 1987*).

The physical and optical specifications for the IOTA beam compressor are:

Primary mirror

Material:	Zerodur	
Diameter:	18.0 ± 0.1 in	45.72 ± 0.25 cm
Central hole diameter:	3.0 ± 0.05 in	7.62 ± 0.13 cm
Focal length:	45.0 ± 1.0 in	114.3 ± 2.5 cm
F/#:	$45/18 = 2.5$	

Secondary mirror

Material:	Zerodur	
Diameter:	2.50 ± 0.05 in	6.35 ± 0.13 cm
Focal length:	4.5 ± 0.01 in	11.43 ± 0.025 cm
F/# effective:	2.5	
Distance pri-sec:	40.5 ± 10^{-4} in	$102.87\ cm \pm 2.54 \cdot 10^{-6}\ \mu m$

3.2 Geometrical calculation of the demagnification factor

The specifications of the beam-compressor are determined by the reduction in the beam and image size required to meet the requirements on different subsystems of the instrument. In what follows I present first a straightforward way to calculate the beam compression factor, and then I present in some detail the calculation of the image position and size at each surfaces of the telescope. This last part is useful when we look into the telescope system in autocollimation and need to know where the images of the various surfaces lie.

3.2.1 Beam compression factor

The beam compression factor, m, can be calculated by applying the relation which holds for similar triangles. With reference to fig. TA.4.a we derive:

$$m_{SP} = \frac{h_S}{h_P} = \frac{f_S}{f_P}.$$

For IOTA, $f_S = 4.5$ inch, $f_P = 45$ inch hence the beam-compressor reduces the size of the beam to 1 tenth of the original size.

fig. TA.4.a Calculation of the beam compression factor

3.2.2 Images of mirrors within the telescope system in autocollimation mode

We follow a ray from an object at the primary mirror surface in its travel within the reducer system. By means of geometrical constructions we calculate the position and height of the image after reflection at each surface. For a schematic representation of the 3 main step involved, see fig. TA.4.b.

fig. TA.4.b Steps 1, 2, 3 for the geometrical calculation of the images of mirrors.

1. The image of the primary mirror edge formed by the secondary mirror, results to be virtual, behind the secondary and inside its focus.

$$\mid z_{image,1} \mid = \; 10.287\, cm$$

$$h_{image,1} \; = \; 4.572\, cm.$$

2. To see what apparent image is sent toward the secondary by the primary, we calculate now the image of the secondary mirror edge formed by the primary. It results to be a large virtual image located behind the primary (towards the observer).

$$\mid z_{image,2} \mid = \; 1028.7 cm$$

$$h_{image,2} \; = \; 63.5\, cm.$$

3. To find the eventual image, we now calculate the image of the secondary mirror edge formed by the secondary itself. It results to be a virtual image, erect, located inside the focal point behind the secondary.

$$\mid z_{image,3} \mid = \; 11.315\, cm$$

$$h_{image,3} \; = \; 0.635\, cm$$

$$\alpha_{image} \; = \; \frac{h_{image,3}}{Z_{lab}} \; \simeq \; 155\, \mu rad$$

(where Z_{lab} is the distance from the secondary mirror to the telescope $\sim$ 41 m)

As a check, the ratio of the image height at the secondary mirror (step 3.) and at the primary mirror (step 2.) gives the factor by which the secondary mirror edge is demagnified by reflection in the primary-secondary pair:

$$m_{SP} = \frac{0.635}{6.35} = 0.1 \;\; \rightarrow \; m_{SP} = \frac{1}{10}$$

as we also shown in the section above.

3.3 Deflection of the reflected beam

I now consider some of the displacements and rotations which can affect the optical elements of the telescope and analyze the consequences for a beam which travels through the system.

3.3.1 Tilt of the flat

If a ray goes from the laboratory to Secondary, Primary, Flat, Primary, Secondary and back to laboratory, and the flat is tilted by θ_{FLAT}, then the beam shift in the laboratory is given by

$$\begin{aligned} r_{lab} \; &\simeq \; m_{PS} \cdot 2\theta_{FLAT} \cdot Z_{lab} \\ r_{lab}\,(cm) \; &\simeq \; 1.43 \cdot \theta^{\star}_{FLAT}\,(\mu rad) \end{aligned} \qquad (11)$$

where m_{PS} is the magnification factor ($m_{PS} = 1/m_{SP}$), and $\theta^{\star}_{FLAT} = 0.001^{\circ} \simeq 17.45\, \mu rad$ is the physical tilt of the flat. The corresponding angle on the sky is given by $\theta_{sky} = 2\ \theta_{FLAT}$.

By tilting the flat by about 55.8 μrad we would produce a beam shift in the laboratory of about 4.572 cm which is the size of the beam in the laboratory and hence this is assumed to be the maximum displacement in the focal plane of the instrument.

3.3.2 Displacement of secondary mirror

With reference to fig. TA.5 and fig. TA.6 we show that if the secondary mirror moves

laterally by Δx, then the ray in the laboratory is deflected opposite to the direction of the displacement of the secondary mirror by a quantity $\propto \Delta x$.

fig. TA.5 Secondary mirror laterally displaced by Δx: a ray entering parallel to the axis is deflected by θ after reflection at the flat

Let us consider a bundle of rays parallel to the focal axis of the secondary mirror. Since parallel rays are focused in the same point of the focal plane, we choose then to follow that directed towards the vertex of the secondary mirror (the vertex is shifted upward by Δx), named v.

v is reflected back by the secondary in the same direction and then, since it is parallel to the focal axis of the primary, it is reflected by the primary along the line through the focal point of the primary, and eventually hits the flat by which is reflected by an angle 2θ, where

$$\theta = \frac{\Delta x}{f_P}$$

Among the parallel rays reflected by the flat in direction θ (whose images, after reflection at the primary, will focus in the same point of the focal plane) two convenient rays are chosen:
a, the ray that after reflection at the primary is directed toward the vertex of the secondary. It is reflected by an angle θ_{out} with respect to the focal axis of the secondary mirror. θ_{out} results to be the total deflection angle for a ray coming in the system along a line parallel to the optical axis of the primary.

b, the ray that hits the primary at its vertex: it is reflected by an angle θ with respect to the focal axis of the primary.

fig. TA.6 Secondary mirror laterally displaced: a ray entering parallel to the axis eventually exits the system deflected by θ_{out}

Let us consider rays a and b after reflection at the primary. The image point is formed at the intersection of a and b, behind the secondary. The intersection point is located below the primary focal axis by the distance $\theta \cdot f_P = \Delta x$.
From fig. TA.6 we get

$$\begin{aligned} \theta_{out} &\simeq \frac{2 \cdot \Delta x}{f_S} \qquad (12) \\ &= 2 \cdot \theta \cdot \frac{f_P}{f_S} \\ &= 2 \cdot m_{PS} \cdot \theta \\ &= 2 \cdot m_{PS} \cdot \frac{\Delta x}{f_P} \qquad (13) \end{aligned}$$

where $m_{PS} = \frac{f_P}{f_S}$ is the magnification factor.

A ray entering the primary-secondary pair along the focal axis of the primary falls in the laboratory displaced downward (when the secondary mirror is shifted upward) by

$$r_{lab} = Z_{lab} \cdot \theta_{out} = 2 \cdot m_{PS} \cdot Z_{lab} \cdot \frac{\Delta x}{f_P} \qquad (14)$$

For a 1 μm lateral displacement of the secondary mirror we have

$$r_{lab} = 2 \cdot 10 \cdot 4100\ cm \cdot \frac{0.0001\,cm}{114.3\,cm}$$

$$r_{lab}\ (cm) \quad \simeq \quad 0.072 \cdot \Delta x\ (\mu m) \tag{15}$$

To get $\simeq 4.572\ cm$ displacement in the laboratory, the secondary mirror has to be displaced by about $63.5\ \mu m$.

3.3.3 Relation between tilt and shift of a paraboloidal mirror

For any paraboloid surface, any *(x, z)* point on the parabola resulting from the intersection with the xz plane is simply related to the slope of the parabola in that point by (see fig. TA.7)

$$\tan\theta = \frac{dz}{dx} = 2kx$$

where $z = kx^2$ is the parabola shape.

fig. TA.7 Cross section of a paraboloidal surface.

With reference to fig. TA.7 a short calculation shows that $k = 1/(4f)$ so that, in the paraxial approximation, we have

$$\theta \simeq \frac{x}{2f} \tag{16}$$

which says that for a ray parallel to the focal axis, a rotation of the parabolical mirror around its vertex by an angle θ equals to a lateral shift of the mirror by $\mathrm{x} = 2 \cdot f \cdot \theta$.

This also means that if the mirror is shifted by Δx, then to get the reflected beam back in the original position, the mirror has to be tilted by $-\theta$. In particular, if we consider the secondary mirror of the telescope, a lateral displacement of 1 μm equals to a rotation

by

$$\theta \simeq \frac{1\mu m}{2f_S} = 4.37\ \mu rad \tag{17}$$

around the axis perpendicular to the line of the displacement.

From fig. TA.7 we see that a rotation by θ also implies a displacement toward -z by $(f - fcos\theta) \simeq f \cdot \frac{\theta^2}{2}$ so that

$$\Delta z \simeq \frac{x^2}{8f_S} \tag{18}$$

and for a lateral displacement of 1 μm the sagittal shift is $\simeq 1.1 \cdot 10^{-6}\ \mu m$.

Let us now consider the relative position of the mirror with respect to the focal axis of the system.

The starting position is perfect alignment, which means that the vertex of the mirror is on the focal axis and the mirror surface is perpendicular to the axis at the vertex.

Then the mirror is displaced by Δx, i.e. the vertex of the mirror is now apart from the focal axis by Δx and the axis intersects the mirror in a new point $(x_I,\ z_I)$. The tangent to the surface in the point $(x_I,\ z_I)$ is not perpendicular to the focal axis (we can deduce from fig. TA.7 that it is tilted by the angle θ), and a ray parallel to the axis that hits the mirror in this point, will not be reflected back in the same direction.

When the mirror is rotated by $\theta = \frac{\Delta x}{2f}$, the tangent to the mirror at $(x_I,\ z_I)$ is normal to the focal axis, and a ray parallel to the axis that hits the mirror in this point will be reflected back on itself.

This is why we can say that, when the mirror is laterally displaced by Δx, the rotation $\theta = \frac{\Delta x}{2f}$ is the rotation that restores the beam position, or equivalently that it is the rotation for which the mirror and the focal axis are normal at the point of their intersection. Actually this is always true for a spherical surface for any lateral displacement and tilt angle. For a paraboloidal section this is only an approximation: it exactly holds only for the small area around the vertex where the parabola can be approximated by a sphere.

3.3.4 Secondary mirror shift-and-tilt compensation and sagittal displacement

Taking into account the previous relations, let us examine some cases of interest for the IOTA alignment. In practice, in tests performed in '93 and '94, we explored a range of angles corresponding to about ± 10 times the angle needed to shift the beam by ≈ 1 FoV in the laboratory ($\theta_{1FoV} = 4.5/4100 \simeq 1143\ \mu rad$), which is about 11430 μrad. As we saw before, this angle in the laboratory is related to a shift of the secondary mirror by $\theta_{lab} = 2 \cdot m_{PS}\, \Delta x / f_P$ so that for $\theta_{lab} = 11430\ \mu rad$, we get a corresponding $\Delta x = 653\ \mu m$. Considering also the travelling range of the secondary mirror translation stage[9] ($\pm 1\ mm$)

[9]The secondary translation stage is the device by which the secondary mirror is translated along the three axis and by which it is aligned with the primary mirror.

this can be considered a good upper limit on the amount of the linear displacement of the secondary mirror that we should explore.

In tab. TA.1 I present some values of a possible lateral displacement, the related compensating tilt, the consequent defocus, and the expected shift of the beam in the laboratory, calculated according to the previous formulas.

tab. TA.1 Secondary mirror lateral displacement, compensating tilt, associated defocus, and expected beam shift in the laboratory.

displacement $\Delta x(\mu m)$	tilt $\theta(\mu rad)$	defocus $\Delta z(\mu m)$	beam shift $r_{lab}(cm)$
0.0	0.0	0.0	0.0
1.0	4.37	/	0.07
3.0	13.12	/	0.22
10.0	43.75	/	0.72
30.0	131.2	/	2.16
50.0	218.7	0.002	3.6
100.0	437.5	0.022	7.2
650.0	2843.4	0.462	46.8

3.3.5 Shift of the secondary mirror: relation with the tilt of the flat

From fig. TA.5 we find that a transverse displacement of the secondary mirror by Δx causes the outgoing beam to strike the flat at an angle θ where $\theta = \frac{\Delta x}{f_P}$. To send the beam back into the telescope parallel to its original direction, the flat must be tilted by $\theta_{FLAT} = \theta$. If the secondary is displaced upwards, then the flat must tilt up also. In general the flat must tilt to "follow" the motion of the secondary. In tab. TA.2 I show some upward displacements of the secondary and the compensating positive rotation of the flat.

tab. TA.2 Upward secondary displacement and compensating positive tilt of the flat.

displacement $\Delta x(\mu m)$	tilt of the flat $\theta_{FLAT}(\mu rad)$
0.0	0.0
1.0	0.875
3.0	2.625
10.0	8.75
30.0	26.25
50.0	43.75
100.0	87.5
650.0	568.75

3.3.6 Primary mirror: shift-and-tilt compensation

This is the same case as that analysed for the secondary. However, here the error due to shift and tilt may be relatively large and not as adjustable as for the secondary mirror.

In particular, the primary will be geometrically centered with respect to its outside diameter, to about $\pm$ 0.1 mm. The center of curvature can be found by placing an object near this point, and adjusting it in (x, y, z) until its image falls exactly back on itself. When the beam axis is found, the secondary (removed during the previous operation) will be placed on this line so that the beam returns on itself. In short, the whole operation aims to put both optical surfaces on a line at the point (vertex) where the surface is perpendicular to the line itself (it will be shown that this is the ideal situation, but it may not be the actual final situation). For detailed explanation of the operation see *section 4.3.6*

3.4 Combination of primary and secondary mirror misalignments

Due to the small values of the uncertainties in the positions of both the primary and the secondary mirrors, we may expect that, to first order, the optical system behaves approximately like a perfectly aligned one (return beam on itself in the laboratory) even if it is not. If we fix any three parameters among displacement and tilt of the primary mirror (Δx_P, $\Delta\theta_P$) and displacement and tilt of the secondary mirror (Δx_S, $\Delta\theta_S$), there is always a value for the fourth, free, parameter that compensates for the other three, as far as the beam displacement in the laboratory is considered. As a general case, let us consider any combination of primary and secondary mirror misalignment, and calculate either the rotation or the translation of the secondary mirror that compensates for the beam displacement in the laboratory.

3.4.1 Compensation by translation of the secondary mirror

Let us consider the beam displacement in the laboratory Δx_{BEAM}, and let us assume that

it is due only to tilt of the secondary mirror. The expression for the lateral translation of the secondary which compensates for the effects of the tilt is:

$$\Delta x_{S,COMP} \;=\; -2f_S \cdot \Delta\theta_{S,SYS} \tag{19}$$

where $\Delta\theta_{S,SYS}$ is the the tilt of the secondary mirror which is equivalent to the sum of the actual tilt due to both the secondary and primary mirrors. It is given by

$$\Delta\theta_{S,SYS} \;=\; \Delta\theta_{S,TOT} \;-\; m_{PS}\Delta\theta_{P,TOT}$$

The factor $-m_{PS}$ takes into account the relation between equivalent tilt angles[10]

$$\Delta\theta_S \;=\; -m_{PS}\Delta\theta_P. \tag{20}$$

Taking into account that a negative (hereafter a negative rotation is assumed to be clockwise, CW, and a positive rotation is counterclockwise, CCW)) rotation is equivalent to a positive (upward) displacement for both the primary and the secondary (even if with opposite effects on the beam position), we calculate now the total displacement Δx_{TOT} (i.e. due both to a real shift of the mirror and to the shift equivalent to a given rotation) for both the secondary and the primary mirror. For the secondary mirror we have:

$$\Delta x_{S,TOT} = \Delta x_S - \Delta x_S(\theta_S) \tag{21}$$

where

$$\begin{aligned} \theta_S < 0 \rightarrow CW, &\quad beam\ down \rightarrow \Delta x_S(\theta_S) < 0 \\ \theta_S > 0 \rightarrow CCW, &\quad beam\ up \rightarrow \Delta x_S(\theta_S) > 0 \end{aligned}$$

For the primary mirror we have:

$$\Delta x_{P,TOT} = \Delta x_P - \Delta x_P(\theta_P) \tag{22}$$

where

$$\begin{aligned} \theta_P < 0 \rightarrow CW, &\quad beam\ up \rightarrow \Delta x_P(\theta_P) < 0 \\ \theta_P > 0 \rightarrow CCW, &\quad beam\ down \rightarrow \Delta x_P(\theta_P) > 0 \end{aligned}$$

Finally, since $\Delta x_i(\Delta\theta_i) = 2f_i\Delta\theta_i$, where i = P,S, we get the expression to calculate the displacement of the secondary mirror that, in principle, compensates for beam displacement due to the sum of all the misalignment effects. This is given by:

[10] As far as the effects on the beam position is concerned, a rotation by θ of the primary mirror is equivalent to a rotation by 10 θ of the secondary mirror. Also, the effect of a CCW rotation of the primary is equivalent to a rotation CW of the secondary, and viceversa (see *Appendix D*)

$$\begin{aligned}\Delta x_{S,COMP} &= -2f_S\Delta\theta_{S,TOT} + 2f_P\Delta\theta_{P,TOT} \\ &= -\Delta x_{S,TOT} + \Delta x_{P,TOT} \\ &= \Delta x_S(\theta_S) - \Delta x_S + \Delta x_P - \Delta x_P(\theta_P) \end{aligned} \quad (23)$$

3.4.2 Compensation by rotation of the secondary mirror

This time let us assume that the beam displacement is expressed as due to a shift of the secondary mirror, $\Delta x_{S,SYS}$, equivalent to the sum of all the other misalignments present in the system. We want to find the expression for the tilt of the secondary mirror that compensates for the beam displacement. It will be given by

$$\Delta\theta_{S,COMP} = -\frac{\Delta x_{S,SYS}}{2f_S} \quad (24)$$

where

$$\Delta x_{S,SYS} = \Delta x_{S,TOT} - \Delta x_{P,TOT}$$

so that

$$\begin{aligned}\Delta\theta_{S,COMP} &= -\frac{1}{2f_S}(\Delta x_{S,TOT} - \Delta x_{P,TOT}) \\ &= -\Delta\theta_{S,TOT} + m_{PS}\cdot\Delta\theta_{P,TOT} \end{aligned} \quad (25)$$

For the secondary mirror we have:

$$\Delta\theta_{S,TOT} = \Delta\theta_S - \Delta\theta_S(x_S) \quad (26)$$

where

$$\begin{aligned} x_S < 0 \rightarrow beam\ up &\rightarrow \Delta\theta_S(x_S) < 0 \\ x_S > 0 \rightarrow beam\ down &\rightarrow \Delta\theta_S(x_S) > 0 \end{aligned}$$

For the primary mirror we have:

$$\Delta\theta_{P,TOT} = \Delta\theta_P - \Delta\theta_P(x_P) \quad (27)$$

where

$$\begin{aligned} x_P < 0 \rightarrow beam\ down &\rightarrow \Delta\theta_P(x_P) < 0 \\ x_P > 0 \rightarrow beam\ up &\rightarrow \Delta\theta_P(x_P) > 0 \end{aligned}$$

Finally, the tilt of the secondary mirror that compensates for the misalignment due to the component the whole system is given by:

$$\Delta\theta_{S,COMP} = \Delta\theta_S(x_S) - \Delta\theta_S + m_{PS}(\Delta\theta_P - \Delta\theta_P(x_P)) \tag{28}$$

3.4.3 Conclusions

First of all, according to what I presented in the two previous paragraphs, we see that any time Δx_S, $\Delta\theta_S$, $\Delta x_P, and\ \Delta\theta_P$ combine with either $\Delta x_{S,COMP}$ or $\Delta\theta_{S,COMP}$ the return beam in the laboratory is back on itself and there is no way to detect the misalignment of any of the optical elements. Second, we can evaluate, for the parameters which are easier for us to modify, the variation that allows us to keep the telescope aligned.

In particular, we turn our attention on the tilt of the primary mirror, because its rotation by small angles has usually no disturbance effects on the other elements of the system (remember that for compensate for the same effect the secondary mirror should be tilted by an angle 10 times larger). Also, the tilt adjustment operated with the secondary mirror produces substantial translation of the vertex of the secondary mirror, so that the tilt and translation of the secondary mirror should be adjusted iteratively.

We need the expression for the compensation tilt angle for the primary mirror. It is given by

$$\begin{aligned} \Delta\theta_{P,COMP} &= -\frac{1}{m_{PS}}\Delta\theta_{S,COMP} \\ &= \Delta\theta_P(x_P) - \Delta\theta_P - \frac{1}{m_{PS}}(\Delta\theta_S - \Delta\theta_S(x_S)) \end{aligned} \tag{29}$$

Let assume that the telescope was perfectly aligned. Now, any displacement of the beam that we can detect in the laboratory, may be compensated by tilting the primary mirror by the appropriate angle. For example, if the displacement of the beam is due to an equivalent shift of the secondary mirror, $\Delta x_{S,SYS}$, we get

$$\begin{aligned} \Delta\theta_{P,COMP} &= \frac{1}{m_{PS}}\Delta\theta_{S,SYS} \\ &= \frac{1}{m_{PS}}\frac{\Delta x_{S,SYS}}{2f_S} \\ &= \frac{1}{2f_P}\cdot\frac{100}{0.072}\cdot\Delta x_{BEAM}\ \mu rad\ per\ cm \\ \Delta\theta_{P,COMP}\ (\mu rad) &\simeq 6.1\cdot\Delta x_{BEAM}\ (cm) \end{aligned} \tag{30}$$

where the relation between Δx_S and Δx_{BEAM} is taken from eq.(15).

From eq.(15) we derived also that the beam disappears from the field of view in the laboratory when the secondary mirror is laterally shifted by 63.5 μm. This turns to be the

maximum equivalent lateral shift, either of the secondary or of the primary, that our system can tolerate before loosing the beam in the laboratory. The maximum compensating tilt angle for the primary mirror that we have to considered is then about 27.9 μrad.

As far as the alignment operation is concerned, we may expected that, after having reached the alignment condition, small variations in any of the optical elements may occur without that any significant change in the beam position in the laboratory being detected. However these variations may be detected by observing the interferograms. From eq.(15) and eq.(37) (in *section 3.5*) we see that for a given Δx_{SYS} and the related Δx_{BEAM} the expected number of fringes is given by

$$N_f \equiv \frac{w}{\lambda} \simeq \frac{11 \cdot \Delta x_{BEAM}}{\lambda} \tag{31}$$

where Δx_{BEAM} is in cm and λ in μm. A small displacement that gives rise to an almost undetectable deviation in the beam can be in any case detected by observing the change in the fringe pattern of the interferogram. Some beam deviations and the related change in the fringe number, calculated with $\lambda = 0.6328$ (the wavelength of the He-Ne laser source used for the operation of alignment of the telescope), are shown in tab. TA.3

tab. TA.3 Expected number of fringes, N_f, for a given displacement of the beam in the laboratory, Δx_{BEAM}

Δx_{BEAM} (cm)	Δx_{SYS} (μm)	$\Delta \theta_{P,COMP}$ (μrad)	N_f
0.05	0.7	0.3	1
0.1	1.4	0.6	2
0.3	4.2	1.8	5
0.5	7.0	3.0	8
0.6	8.4	3.6	10
1.0	14.0	6.1	16

3.5 Wavefront changes due to displacements of the secondary mirror

The five wavefront errors, here considered, that contribute to the total wavefront error may be grouped as due to axial displacement (Δz), to lateral displacement (Δx, Δy), and to tilt ($\Delta \theta_x$, $\Delta \theta_y$).

3.5.1 Secondary mirror axially displaced

By mean of simple geometrical optics a light beam is followed through the whole system in order to find the eventual image position when the secondary is axially displaced with respect to the primary.

With reference to fig. TA.8 and the sign convention there adopted, we see that if the

secondary moves a distance Δz toward the primary, the wavefront will diverge.

fig. TA.8 Secondary mirror axially displaced towards the primary mirror: the wavefront diverges

(1) An object at infinity has a virtual image at $-f_S$ after reflecting from the secondary.
(2) For the primary, the object is inside its focus by Δz, so the virtual image is at z_{im} where

$$\frac{1}{z_{obj}} + \frac{1}{z_{im}} = \frac{1}{f_P} \tag{32}$$

so that

$$z_{im} = \frac{f_P(f_P - \Delta z)}{\Delta z} \simeq \frac{f_P^2}{\Delta z} \tag{33}$$

that is, the virtual image is far away behind the primary, toward the observer. If the flat is close to the secondary mirror, and z_{im} is large with respect to the distance between the secondary and the flat, then

$$\frac{1}{z_{obj}} + \frac{1}{z_{im}} = \frac{1}{\infty}$$

and, (3), for the primary, the beam reflected from the flat will appear to come from $-f_P^2/\Delta z$, far beyond the flat.
(4) From fig. TA.8 we see that

$$\begin{aligned}
\frac{1}{-\frac{f_P^2}{\Delta z}} + \frac{1}{z_{im}} &= \frac{1}{-f_P} \\
z_{im} &= -\frac{f_P^2}{f_P - \Delta z} \\
&= -f_P \cdot \frac{1}{1 - \frac{\Delta z}{f_P}} \\
&\simeq -f_P \cdot (1 + \frac{\Delta z}{f_P}) \\
&= -(f_P + \Delta z)
\end{aligned}$$

so that the virtual image formed by the primary is now just beyond f_P by approximately Δz.
(5) The beam hits the secondary before actually forming at this position, and from

$$\frac{1}{-(f_S + 2\Delta z)} + \frac{1}{z_{im}} = \frac{1}{-f_S}$$

$$\frac{1}{z_{im}} = -\frac{2\Delta z}{f_S(f_S + 2\Delta z)}$$

we get

$$z_{im} \simeq -\frac{f_S^2}{2\Delta z} \tag{34}$$

The virtual image at the end of the system will be located far beyond the secondary, opposite from the observer, and will be enlarged.
The real image, in terms of the emerging wavefront, will come toward the observer, erect, which means not reversed with respect to the z-axis, and diverging (If the secondary moves toward the primary, then the wavefront bulges in the same direction, outward toward the primary or into the laboratory).

With reference to fig. TA.9 we calculate now the spherical sagitta of the wavefront, w_z, in terms of the maximum displacement with respect to the wavefront emerging from a perfect system.

fig. TA.9 Secondary mirror axially displaced by Δz and the resulting wavefront sagitta $w(z)$.

Let us assume that the diameter of the beam emerging from a perfect system is equal to d_S, the measure of the secondary diameter, and that the plane reference wavefront is located at the distance z_{im} on the focal axis. From fig. TA.9 we see that, assuming $z_{im} \gg f_S$ so that $z_{im} + f_S \simeq z_{im}$, the spherical sagitta w_z for the diverging wavefront is given by

$$\begin{aligned} (z_{im})^2 &= (z_{im} - w_z)^2 + \left(\frac{d_S}{2}\right)^2 \\ 0 &= -2z_{im} \cdot w_z + \frac{d_S^2}{4} \\ w_z &\simeq \frac{d_S^2}{4} \cdot \frac{1}{2z_{im}} \\ w_z &\simeq \frac{d_S^2 \Delta z}{4 f_S^2} = \frac{\Delta z}{4 f_{\#,S}^2} \end{aligned} \tag{35}$$

For the IOTA telescope, considering the full aperture, $f_{\#,S} = f_{\#,P} = 2.5$, so that $\frac{w_z}{\Delta z} = 0.040$.

3.5.2 Secondary mirror laterally displaced

According to previous calculations, when the secondary is displaced by Δx, or equivalently Δy, the beam in the laboratory is shift by an angle $\theta_{lab} = \frac{2 \cdot m_{PS} \Delta x}{f_P}$.

Since rays are perpendicular to wavefronts, θ_{lab} is also the angle between the reference

wavefront and the distorted one. With reference to fig. TA.10, where d_{lab} is the diameter of the beam in the laboratory and it is assumed $d_{lab} = d_S$, the wavefront error is given by

$$\begin{aligned} w_x &= \theta_{lab} \cdot d_{lab} \qquad (36) \\ w_x &= \frac{2 \cdot m_{PS} \Delta x}{f_P} \cdot d_{lab} \\ &= \frac{2 \cdot d_S}{f_S} \cdot \Delta x \\ &= \frac{2 \cdot \Delta x}{f_\#} \qquad (37) \end{aligned}$$

For the IOTA telescope, $\frac{w_x}{\Delta x} = 0.8$

fig. TA.10 Secondary mirror laterally displaced: the aberrated wavefront, Σ_a is tilted by θ_{lab} with respect to the incoming wavefront, Σ_r. w_x is the wavefront error.

3.5.3. Secondary mirror tilted around its vertex

From previous calculations, we have that a rotation θ_x around y-axis is equivalent to a shift along x given by $\theta_x = \frac{\Delta x_{rot}}{2 \cdot f_S}$. According to eq.(33) we can write

$$\begin{aligned} w_{\theta_x} &= \frac{2 \cdot \Delta x_{rot}}{f_\#} \\ w_{\theta_x} &= \frac{4 \cdot f_S}{f_\#} \cdot \theta_x \qquad (38) \end{aligned}$$

For the IOTA telescope, $\frac{w_{\theta_x}}{\theta_x} = 0.183\ \mu m$, if θ_x is in μrad.

3.5.4 Preliminary error budget

For a full-aperture wavefront, the total wavefront error is w where

$$w^2 = w_x^2 + w_y^2 + w_z^2 + w_{\theta_x}^2 + w_{\theta_y}^2 = \sum_{i=1}^{5} w_i^2. \tag{39}$$

By applying the *Rayleigh's quarter wavelength rule*[11] for the evaluation of the maximum amount of aberration that can be tolerated in an image-forming system, the maximum (total) tolerable wavefront aberration is given by $w_{max} = \lambda/4$. Assuming that each misalignment factor contributes with the same weight to the total wavefront error, we have

$$w_i \simeq \frac{w_{max}}{\sqrt{5}} = \frac{\lambda}{4 \cdot \sqrt{5}}.$$

The error tolerance depends on the wavelength selected for the observation. In tab. TA.4 I summarized the estimated maximum displacements and rotations of the secondary which are tolerable for the IOTA telescope[12]. The first two columns present the results for the infrared (K-band centered at 2.2 μm) and the visible (V-band centered at 0.55 μm). The third column presents the results for the ultraviolet, assuming that a telescope with the same optical specifications could be operated in space.

tab. TA.4 Estimated maximum lateral, Δx, axial, Δz, displacements, and tilt, θ_x, of the secondary mirror tolerable in the IOTA telescope, assuming $w_{max} = \lambda/4$.

	K-band	V-band	UV-band
λ	2.2 μm	0.55 μm	0.15 μm
w_i	0.25 μm	0.06 μm	0.020 μm
Δx	0.31 μm	0.08 μm	0.025 μm
Δz	6.25 μm	1.50 μm	0.50 μm
θ_x	1.36 μrad	0.33 μrad	0.011 μrad

[11] It was first shown by Lord Rayleigh that when a system is affected by aberration of such an amount that the wavefront in the exit pupil departs from the reference sphere by less than a quarter wavelength, the intensity at the focus of the system is diminished less than 20 per cent, a loss of light that can usually be tolerated. This rule, however, is only a rough guide as to the desiderable state of correction of a system, since the loss of light that can be tolerated depends on the particular use to which the instrument is put.

[12] In this case we have evaluated the maximum wavefront error regarding the telescope as an imaging device in itself, not as a component of an interferometric array. For the evaluation of the maximum wavefront error that can be tolerated when considering the interferometric performance of the system, we will make use of a more stringent criterion, based on the Strehl ratio (see *0.* in *Visibility budget for IOTA*).

The same results as for Δx and θ_x apply when considering shift along y-axis and rotation around x-axis.

4. Simulation

4.1 Simulation programs

In order to perform the simulation to evaluate the wavefront distortion due to misalignments of the telescope optics, we have to run two programs. The first one, *RayTrace*, originally created for the analysis of the optical system of POINTS[13], provides as output a matrix of parameters which, when they are properly combined, give the aberration coefficients of the wavefront function. The second program, *IOTA_misalignment*, uses this matrix of parameters as input to compute the aberration coefficients. Then it combines them to calculate the aberrated wavefront and simulate its interference with a perfect plane wavefront to finally calculate the interferogram pattern. The program produces interferogram plots and plots of the intensity distribution of the beam, before interference occurs, for any misalignment considered.

4.1.1 Raytrace

At its core, the program RayTrace (*Murison, 1993*) propagates a grid of "pencil" rays through an optical system. Radial grids were used for the calculations here reported. The optical system is defined in a *specification* input file and the information on the perturbation affecting the optical assembly is specified in a *variation* input file. After propagating an entire grid of rays, called a ray bundle, the program performs certain averaging and analysis tasks, resulting in several *normal points*. They consist of several diagnostics, including the OPL averaged over the output beam, the average tilt and orientation of the wavefront, the curvature(s) of the wavefront, the beam centroid displacement on the detector plane, and the change of the OPL due to the wavefront distortions, caused by 16 different kind of aberrations[14]. They are stored internally as a function of the assembly perturbations. The perturbations used in this simulation are of two types: rotations around an arbitrary axis and translations in any direction. For any given specification of the optical system, all of the perturbations specified in the variation file can be analysed in the same run.

[13]POINTS (Precision Optical INTerferometer in Space) is an Earth-orbiting instrument designed to measure the angle between a pair of stars, separated by about 90°, with a nominal accuracy of 5 μas. It comprises a pair of independent Michelson stellar interferometers and a metrology system based on laser gauges. Because of the high accuracy goal, the mitigation of systematic error has been the central theme of the study of the instrument architecture and of the development of the data-analysis methods. The conditions to meet the high degree of alignment required for the optics have been studied by means of the *RayTrace* program, which properly adapted, I utilize in the raytracing analysis for the IOTA telescope optics.

[14]Each aberration is considered independently and the calculated normal point is the value of the aberration coefficient in the expression of the wavefront. If, for example, we considered the expression of the wavefront limited to the Seidel's approximation, we would have four normal points, each of them being one of the coefficients A_i (see eq.(3))

Let us label each perturbation as the parameter p_i. If there are N different perturbation parameters, we can form a vector of N parameters, $\mathbf{p}$, with a vector of normal values, $\mathbf{P}$, calculated for each p_i vector element. RayTrace then fits least squares third-order polynomials to the calculated normal point parameter grids, according to:

$$P_k(p_i) = A_{ik} + B_{ik} \cdot p_i + C_{ik} \cdot p_i^2 + D_{ik} \cdot p_i^3 \tag{40}$$

where P_k are the normal points introduced before, and A_{ik}, B_{ik}, C_{ik}, D_{ik} are the parameters which minimize the chi-square function and which are calculated for each P_k and p_i. In particular, for the IOTA simulation, we are interested in the variation in OPL due to wavefront distortions, $\Delta(OPL)$. For each aberration considered in the expression of the wavefront surface, $\Delta(OPL)$ is calculated as the maximum variation in OPL with respect to the mean OPL calculated from all the rays of the bundle. Since RayTrace was created to perform an analysis of the wavefront distortion to the fifth-order approximation in the ray aberration[15], this implies that there are 16 independent aberration coefficients to be evaluated. Each of them is a normal point, P_k, and for each of them the program calculates four parameters (A, B, C, D). Hence, the output we are going to use in *IOTA_misalignment*, consists of a matrix of 16×4 parameters: an example of the RayTrace output is presented in *Appendix E* together with an example of the specification and variation files.

4.1.2 IOTA_misalignment

I wrote the program *IOTA_misalignment*, in order to calculate the wavefront distortion, starting from the parameter matrix of the aberration coefficient given by RayTrace, and then to simulate the interferogram obtained when the distorted wavefront interferes with a perfect plane wavefront. In what follows I briefly describe the main steps of the program.

After asking for the input file, the program asks for different values of $DELTA$, which are the actual perturbation (i.e. translations or rotations) we want to simulate, and for the distance from the primary mirror (where the aberrated wavefront exits the telescope) to the laboratory. The numerical value of the variation $DELTA$ has to be in unit of μm for translations and of μrad for rotations. The distance to the laboratory has to be in cm.

After reading the input file, that for each aberration coefficient (called ABC_j where j = 1, 16) gives the parameters A_j, B_j, C_j, D_j, the program calls the subroutine *wavefront* by which each coefficient is evaluated according to the formula (see also eq.(40))

$$ABC_j = A_j + B_j \cdot DELTA + C_j \cdot DELTA^2 + D_j \cdot DELTA^3 \tag{41}$$

where j = 1, 16.
Then, for each point (ξ, η)[16] on the end plane of the optical system, the $\Delta(OPL)$ is cal-

[15] As explained in *section 1.* this equivals to an approximation to the sixth-order in the wavefront

[16] To avoid confusion, the coordinate notation (x, y, z) is used only for the coordinates of the telescope

culated by means of the formula:

$$\begin{aligned}\Delta(OPL)(\xi,\eta) &= ABC_1 + ABC_2\xi + ABC_3\eta + ABC_4(\xi^2+\eta^2) + \\ &\quad ABC_5(\xi^2+\eta^2)^2 + ABC_6(\xi^2-\eta^2) + ABC_7\xi\eta + ABC_8(\xi^2+\eta^2)\xi + \\ &\quad ABC_9(\xi^2+\eta^2)\eta + ABC_{10}\xi^3 + ABC_{11}\eta^3 + ABC_{12}(\xi^2+\eta^2)(\xi^2-\eta^2) + \\ &\quad ABC_{13}(\xi^2+\eta^2)\xi\eta + ABC_{14}(\xi^2+\eta^2)^2\xi + \\ &\quad ABC_{15}(\xi^2+\eta^2)^2\eta + ABC_{16}(\xi^2+\eta^2)^3. \end{aligned} \tag{42}$$

This $\Delta(OPL)$ represents, for each point (ξ,η), the height of the wavefront with respect to a perfect plane wavefront, i.e. the wavefront distortion. By multiplying $\Delta(OPL)$ by $\frac{2\pi}{\lambda}$ we get the phase difference between the aberrated and the reference wavefront, so that we can calculate the modulation term of the interference pattern which results from the interference of the two wavefronts. The modulation is given by

$$f(\xi,\eta) = \frac{1}{2}\left[1 + sin\left(\frac{2\pi\Delta(OPL)(\xi,\eta)}{\lambda}\right)\right], \tag{43}$$

which is the function that is plotted by means of appropriate graphics subroutines.

To compute the position of the beam as it is seen in the laboratory, each ray is propagated from the end surface to the laboratory according to the formulas

$$\begin{aligned}\xi_{lab} &= \xi - \frac{\partial\Delta(OPL)}{\partial\xi}\cdot d \\ \eta_{lab} &= \eta - \frac{\partial\Delta(OPL)}{\partial\eta}\cdot d \end{aligned} \tag{44}$$

where d is the distance to the laboratory.

Eventually, the outputs from the program are the interferogram and the intensity distribution of the beam due to a specific perturbation of magnitude $DELTA$, which affects one of the elements of the optical system as specified in the variation file of RayTrace.

4.1.2.1 Rms evaluation of the calculated wavefront

For each perturbation the program calculates the wavefront rms, which is usually the quantity of interest for evaluating the degradation of the optical performance of the optical system. In particular, we used this quantity to estimate the visibility loss factor, which is assumed to be given by the Strehl ratio associated with the wavefront rms[17].

In order to evaluate the mean deviation of the aberrated wavefront from a plane wavefront, it is necessary to find the best fit plane to the actual wavefront. I will show that,

components.

[17]See section *Instrument visibility for IOTA* and *Appendix I* for the plots of the Strehl ratio as a function of a given misalignment error.

since this plane is given by the first three terms in eq.(42), the rms value can be calculated from the aberration coefficients we already have without performing a new least square best fit.

Eq.(42) gives the position of each point of the wavefront with respect to an arbitrary endplane of the optical system. Each coefficient represents a specific deviation of the wavefront from such an endplane. For example the first coefficient, A_1 simply represents the offset of the wavefront from the endplane. Conversely, we can look at eq.(42) as a plane wave to which other parts, the number of which depends on the order of the approximation to which we calculate the wavefront aberration, are added to give the actual wavefront. The first three terms of eq.(42) represent such a plane wave, that in general is offset (A_1) and tilted (A_2, A_3) with respect to the endplane. Since any coefficient A_i is the result of a best fit to the expression of the true wavefront, we see that

$$\zeta(\xi,\eta) = A_1 + A_2\xi + A_3\eta \tag{45}$$

is actually the best fit plane to the output wavefront.

By subtracting eq.(45) from eq.(42) we get, in each point, the deviation $q(\xi,\eta)$ from the best fit plane wave, and then these values are used to calculate the mean deviation of the aberrated wavefront from a plane wave. In short, if we write $\Delta(OPL)$ as:

$$\Delta(OPL)(\xi,\eta) = A_1 + A_2\xi + A_3\eta + q(\xi,\eta) \tag{46}$$

then, the standard deviation with respect to the plane $A_1 + A_2\xi + A_3\eta$ is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} q_i^2(\xi,\eta)}{(N-3)}} \tag{47}$$

where N is the number of points where the $\Delta(OPL)(\xi,\eta)$ is calculated, and σ is the quantity that characterizes the mean deviation of the aberrated wavefront from the "ideal" wavefront.

In the actual operation of IOTA, the outgoing wavefront is directed to the detector by means of an active mirror placed after the telescope, that removes the tilt of the wavefront with respect to the detector plane. This, in general, means that the term $q(\xi,\eta)$ in eq.(47) does not exactly represent what is seen by the detector[18]. However, I show that the error made in assuming $q(\xi,\eta)$ as the true wavefront height, scales as the square of the wavefront tilt and that this error is negligible for the purpose of our evaluation.

In fig. TA.11 the aberrated wavefront is shown with respect to the reference endplane and to the best fit plane which is assumed to be parallel to the detector plane. By setting $A_1 = A_2 = A_3 = 0$ the residual quantity obtained gives the distance of the wavefront from the best fit plane in a direction which is normal to the end plane and not to the detector plane.

[18]See fig. TA.11: the wavefront height seen by the detector differs by a factor $cos\beta$ from the wavefront height $q(\xi,\eta)$

fig. TA.11 Aberrated wavefront and best fit plane to it. The best fit plane is parallel to the detector plane. The endplane is the plane from which it is calculated the quantity $\Delta OPL(\xi, \eta)$.

With reference to fig. TA.12 let be w the calculated wavefront height and W the true wavefront height, as seen by the detector. We get

$$\begin{aligned} W cos\beta &= w \\ \Delta W &= W - w = W(1 - cos\beta) \simeq \frac{1}{2} W \beta^2 \\ \frac{\Delta W}{W} &\simeq \frac{1}{2}\beta^2 \end{aligned} \tag{48}$$

where β is the tilt angle of the outcoming wavefront.

In conclusion, we see that the wavefront height represented by $q(\xi, \eta)$ differs from the wavefront height that actually is seen by the detector by a factor $cos\beta$, and the error made by using $q(\xi, \eta)$ instead of the actual value W is ΔW. However the error ΔW can be considered negligible for the cases of our interest. In section *3.3.2* we found that when the misalignment consists in a rotation, θ, of one of the optical elements, the wavefront tilt is given by $\beta \sim m\theta$, and when the misalignment is any lateral translation, Δl, $\beta \sim m\frac{\Delta l}{f}$. For IOTA we have $m = 10$, $f \sim 100\ cm$, for the primary and $f \sim 10\ cm$ for the secondary. Even for variations of the order of 100 μrad and of 100 μm, ΔW is of the order $10^{-4} \cdot W$, when the variation is a secondary lateral displacement, or of $10^{-6} \cdot W$, when a rotation or a lateral displacement of the primary are considered.

fig. TA.12 true wavefront height, W, and calculated wavefront height, w; β is the tilt angle of the aberrated wavefront.

4.2 Coordinates system and some tests

RayTrace uses a right-handed reference system where vectors' directions are defined by means of angles Θ and Φ defined according to fig. TA.13

$\Phi = 0°$, $\Theta = 90° \Rightarrow +$ x-axis
$\Phi = 90°$, $\Theta = 90° \Rightarrow +$ y-axis
$\Theta = 0° \Rightarrow +$ z-axis
$\Theta = 180° \Rightarrow -$ z-axis

fig. TA.13 Reference system used in RayTrace.

With respect to this reference system, the IOTA optical system has the optical axis along the z-axis, and the origin of the coordinates system coincides with the "vertex" of the flat (see fig. TA.3). As a consequence, vertical shifts are along the x-axis while lateral shifts are along y-axis. As far as rotations are concerned, a positive rotation is defined to be counterclockwise (CCW), a negative rotation is clockwise (CW).

As a simple test for the *IOTA_misalignment* program, I simulated an upwards displacement of the secondary by $+\Delta x$ and verified that it causes a shift down of the beam in the laboratory, as equivalently does a negative rotation of the flat by $\alpha = \Delta x / f_p$ (where f_p is the primary focal length).

Similarly, I considered a rotation of the secondary around y-axis, and verified that a negative rotation by $\theta = \Delta x / 2 \cdot f_s$ causes a shift of the beam in the laboratory comparable to that caused by a shifting it upwards by $+\Delta x$.

The results of these tests are summarized in tab. TA.5. In the second and third columns $\Delta\theta$ and $TILT$ represent the rotation of the secondary and the tilt of the flat equivalent to a shift Δx of the secondary, calculated according to the paraxial approximation. The range of beam deflection values calculated for the three different variations is given in the fourth column.

tab. TA.5 Beam deflection resulting from the simulations of a lateral shift of the secondary, Δx, tilt of the secondary, $\Delta\theta$, and tilt of the flat, TILT.

Δx	$\Delta\theta$	TILT	Beam Deflection
$+$ 100 μm	$-$ 438.6 μrad	$-$ 87.5 μrad	-6.6 to -6.9 cm
$-$ 100 μm	$+$ 438.6 μrad	$+$ 87.5 μrad	$+6.5$ to $+6.8$ cm

A second test was performed to verify if an appropriate rotation either of the secondary or of the flat, as calculated in the limit of the paraxial approximation, compensates for the deflection due to a secondary shift. The simulation performed by shifting the secondary by 100 μm, shows a good restoration of the beam position in the laboratory, both when the secondary is tilted by -438.6 μrad and the flat by -87.5 μrad. Indirectly this kind of test suggests that the simulation program could be used to find the best compensation that corrects for any given misalignment: starting from the correction position computed in the limit of the paraxial approximation, we either rotate or translate the mirror to "fine tune" the correction, until a perfect restoration is found.

Another test was performed to investigate how the lateral displacements of the primary and of the secondary mirror are related. I considered both an upwards displacement of the primary by 500 μm and a downwards displacement the secondary by the same distance. In both cases the beam in the laboratory is shifted upwards by about 34.5 cm, which is of the order of the paraxial prediction of a shift of 35.8 cm. As far as rotations are concerned, from eq.(16) we have that, assuming the same lateral displacement, Δx, for both primary and secondary,

$$\begin{aligned} \theta_S &= \frac{\Delta x}{2 \cdot f_S} \\ \theta_P &= \frac{\Delta x}{2 \cdot f_P} \\ &= \frac{\Delta x}{2 \cdot f_S} \cdot \frac{f_S}{f_P} \\ \theta_P &= \frac{1}{m_{PS}} \cdot \theta_S \end{aligned}$$

from which we get that, as far as the displacement of the beam is concerned, a positive rotation of the primary by θ_P equivals to a negative rotation of the secondary by $m_{PS} \cdot \theta_P$. I simulated a rotation of the primary by $+40$ μrad and then that of the secondary by -400 μrad and got in both cases a displacement of about 6.3 cm.
The results of these tests are summarize in tab. TA.6

tab. TA.6 Beam deflection resulting from the simulation of independent primary and secondary mirror lateral displacement, Δx_P and Δx_s, and from independent primary and secondary mirror rotation, $\Delta\theta_P$ and $\Delta\theta_s$.

Δx_P	Δx_s	Beam Deflection
$+$ 500 μm	$-$ 500 μm	$\sim 34.5\ cm$
$\Delta\theta_P$	$\Delta\theta_s$	Beam Deflection
$+40\ \mu rad$	$-400\ \mu rad$	$\sim 6.3\ cm$

Then I ran the program to find the actual tilt that restores the beam in the laboratory when the primary is laterally displaced. This angle is exactly equal to the tilt angle given by eq.(16) for $\Delta x \leq 100\ \mu m$, which means that, up to this value of lateral displacement, the paraxial approximation is accurate enough to calculate the amount of tilt to be used to correct for the displacement. For larger shifts, the rotation given by eq.(16) does not perfectly restore the beam. In particular for displacements larger of about 200 μm, even by using the angle rotation that gives a coma-free system (see following *section 4.3.7*), the result did not improve. This is due to the fact that there are aberrations that cannot simply be corrected by tilting the mirror (this is a perfect solution only for a spherical surface). This correction criterion is, once again, derived according to the paraxial approximation, so that it is very likely that it is not valid for the aberrations of higher order which are considered in the program.

Finally, I verified that the number of fringes in the simulated interferograms agrees with that predicted by the theory. In particular, from eq.(36) we get that a secondary displacement of 1 μm causes a wavefront error of about 1.26 λ (where $\lambda = 0.6328\ \mu m$ for the laser source used when the telescope is operated in the autocollimation mode). From the interferograms obtained when the secondary is shifted by 1, 3, 10 μm we get 1, 3, 11 clear fringes, in agreement with the theoretical values of 1.3, 3.8, 12.6 for the number of expected fringes. The same test was performed for sagittal displacement: from eq.(34), for a shift by 50 μm along z-axis, we expect to see 3.2 fringes. In the simulation the interferogram shows 3 clear fringes.

4.3 Results of the simulation

The following sections deal with specific cases which are of interest during the alignment operation. In sections from *4.3.1* to *4.3.5* I analyze the cases in which, starting from a system perfectly aligned, the secondary or/and the primary positions are changed. In *4.3.6* I present the procedure to get to the initial alignment of the telescope and in *4.3.7* I analyze the result of the simulation of this operation. In *4.3.8* I present the results of the simulation performed to find the field of view of the telescope, which is defined as the

maximum source off-axis angle for which the wavefront *rms* is less than $\lambda/8$. I considered in this case an error of $\lambda/8$, instead of $\lambda/4$, because (as explained in *section 3.5.3*), I adopted the $\lambda/4$ limit to define the maximum tolerable wavefront error when the telescope is operated in autocollimation mode. Since for the evaluation of the telescope field of view the simulation is run with the telescope in normal mode, the maximum wavefront error has to be a factor $\frac{1}{2}$ smaller.

As far as the results of the simulation when the telescope is in autocollimation mode is concerned, I also took into account that the flat can be tilted in order to remove the wavefront tilt with respect to the reference plane at the exit of the system (this plane is parallel to the detector plane). In particular, all the interferograms that are collected in the Atlas presented in one of the appendices, are obtained in the tilt-corrected configuration. This implies that they show only fringes which are due to the actual wavefront aberrations and not those due to a tilt of the wavefront with respect to the detector plane.

4.3.0 Summary of the misalignment conditions examined by means of the simulation

In order to analyze a significative range of possible misalignment conditions, I considered seven different initial configurations of the telescope, and a set of perturbations that can be applied to any of the initial configuration. As a whole, I analyzed 12 different combinations.

The seven initial configurations are:

1) optical system perfectly aligned;
2) optical system perfectly aligned but for the secondary which is laterally shifted;
3) optical system perfectly aligned but for the secondary which is laterally shifted and tilted;
4) optical system perfectly aligned but for the primary which is laterally shifted;
5) primary laterally shifted and tilted and secondary tilted by the same angle;
6) both primary and secondary laterally shifted and tilted so that their vertex are aligned on the same axis;
7) system perfectly aligned, source off-axis.

The combinations I simulated and analyzed are the following. For most of them I indicate, in parenthesis, the reference to the paragraph where they are mentioned and related results are presented. Note that for the cases where the system starts from perfected alignment, the analysis of each single perturbation was mainly used to verify the equivalence between a given lateral displacement and the correspondent rotation given by the paraxial approximation.

The 12 combinations are:

1) secondary mirror translations in all directions starting from a perfect alignment (*4.3.1 and 4.3.2)*;
2) secondary mirror rotations around y-axis and x-axis starting from a perfect alignment;

3) primary mirror translations in all directions starting from a perfect alignment;
4) primary mirror rotations around y-axis and x-axis starting from a perfect alignment;
5) secondary mirror rotations around y-axis, when the mirror is shifted along x-axis *(4.3.1)*;
6) secondary mirror translations along z-axis, when the mirror is shifted along x-axis *(4.3.2)*;
7) secondary mirror translations along z-axis when the mirror is shifted along x-axis and tilted around y-axis *(4.3.1 and 4.3.3)*;
8) primary mirror rotations around y-axis when the mirror is shifted along x-axis *(4.3.4)*;
9) secondary mirror translations along x-axis when primary mirror is shifted along x-axis *(4.3.4)*;
10) secondary mirror translations along x-axis when primary mirror is shifted and tilted *(4.3.5)*;
11) secondary mirror translations along z-axis when both primary and secondary mirror are shifted and tilted *(4.3.6 and 4.3.7)*;
12) secondary mirror translations along z-axis when the source is off-axis (*4.3.8).*

4.3.1 Secondary mirror laterally displaced: effects of rotation of the mirror

Runs of the program were performed with the secondary displaced upwards by Δx in the range 1 - 100 μm and tilt angle both zero and equal to that for which the mirror is normal to the primary mirror axis at their intersection point, $\theta_\perp$[19].

When $\theta = 0$ fringes appear in the yx plane, parallel to the y-axis. Their number increases as Δx increases, going from about 1 for $\Delta x = 1$ μm to about 11 for $\Delta x = 10$ μm. Starting from $\Delta x = 30$ μm the interferograms are no more reliable to give information on the number of fringes, because of sampling limitations in the routine that builds the interferograms. These fringes are mainly due to the tilt of the wavefront with respect to the plane of the detector (the reference plane in the simulation). By tilting the flat the wavefront tilt can be removed and the interferograms improve. In this case for $\Delta x = 100$ μm the wavefront rms is of the order of $\lambda/4$.

When $\theta = \theta_\perp$ and $\Delta x < 10$ μm no aberration are observed. For $\Delta x \geq 30$ μm coma effects and other wavefront errors, negligible for very small displacements, show up.

As far as the alignment operations are concerned, let us assume that the primary and the flat are perfectly aligned. Then if less than about 11 fringes are detected, this may mean that the secondary is laterally displaced by less than 10 μm. By tilting the secondary both CW and CCW and observing the number of fringes, we can deduce the sense of the displacement. For example, if a CW tilt increases the number of fringes, then the shift is upwards and, to restore both beam and wavefront, it is necessary to make the opposite rotation.

If more than 11 fringes are detected, the displacement is larger than 10 μm but a rotation of the mirror can almost eliminate them. However in this case it may be useful to correct for the lateral position of the mirror, first. If the interferogram gets worse and the beam in

[19] That is, $\theta_\perp = \frac{\Delta x}{2f}$, see *section 3.3.3* for an explanation of this definition.

the laboratory goes even further from the central position, this means that the correction was operated in the wrong sense. Try in the opposite one.

4.3.2 Secondary mirror laterally displaced: effects of translations of the mirror along the z-axis

Several sets of interferograms were obtained running the program for different combinations of lateral and sagittal (translation along the z-axis) displacements. Here we consider $\Delta x = 1\ \mu m$ and Δz range from -40 to $+40\ \mu m$ and assume that the primary mirror and flat are perfectly aligned. When $\Delta z = 0$ the interferogram shows one fringe parallel to y-axis, which is due to the tilt of the wavefront caused by the mirror lateral shift. By moving the secondary either towards or away from the primary the fringes become curved and, for $\Delta z \geq 20\ \mu m$, their number increases. It is interesting to note the appearance of a characteristic feature, that we named *eye-ball* feature, in the interferogram, which changes position according to the sense of the sagittal displacement. Moving the mirror back and forward along the z-axis, going through the "on focus" point, the *eye-ball* appears at the bottom of the interferogram figure, then the fringes become more and more straight as the mirror approaches the on focus point[20], then start curving again but with opposite curvature and the *eye-ball* appears at the top of the screen. This trend is observed for all the lateral displacements considered, although the characteristic *eye-ball* becomes evident for different values of Δz. This is because the number of fringes increases proportional both to Δx and to Δz, and by different factors. From eq.(34) and eq.(36) we get

$$\frac{w_x}{w_z} = 20 \cdot \frac{\Delta x}{\Delta z}$$

which means that in order to get the same number of fringes either by a Δx displacement or by a Δz translation, it has to be $\Delta z = 20\ \Delta x$. For example, from the interferogram obtained when $\Delta x = 1$ and $\Delta z = 40\ \mu m$, we get 2 fringes (which agrees with the expected ~ 2.5 fringes when only $\Delta z = 40\ \mu m$ is considered) shifted upwards with respect to the center of the figure: the two fringes are due to the defocus error while the shift is due to the mirror lateral displacement.

From a comparison of different cases, we also notice that the *eye-ball* shows up at larger Δz for larger Δx and also that, keeping Δx fixed, the more Δz increases, so that $w_z > w_x$, the more difficult is to detect the *eye-ball* asymmetry that eventually disappears. As a general rule of thumb we can expect to see the *eye-ball* feature when defocus becomes the major cause of wavefront distortion but the effect of lateral displacement is not negligible yet, i.e.

$$\Delta z \geq 20 \cdot \Delta x.$$

As the Δx gets larger, the interferograms get less clear, but the *eye-ball* can still be observed. This feature can be used as a sign to detect the lateral misalignment of the

[20] This is the position of the secondary mirror for which the focal point of the primary and secondary mirrors are perfectly coincident.

secondary, because if it was aligned along the primary axis, the fringes would be concentric to the center of the figure. An evaluation of the magnitude of the misalignment is obtained counting the number of fringes when the mirror passes through the "on focus" position.

4.3.3 Secondary laterally displaced and tilted: effects of sagittal translations of the mirror

For this set of simulations the secondary is laterally shifted by different amounts and tilted by the correspondent angle $\theta_\perp$ that restores the beam in the laboratory. Then, for each shift and tilt combination, the secondary is moved along the z-axis. For small lateral displacements, the fringes observed when the secondary mirror is moved back and forward from the focal point, resemble pretty well those obtained when the mirror is on axis. However, starting from $\Delta x = 30\ \mu m$, and for sagittal shifts within about 10 μm from the on focus position the fringes are no more concentric to the center of the figure. Moving from back to forward passing through the on focus position, the asymmetry moves from the bottom to the top of the figure, a behaviour similar to that described in *4.3.2*. The aberration effect due to the lateral displacement that the tilt has not corrected, added to defocus, causes the observed asymmetrical feature.

Since by moving the mirror out of the on focus position, we observe similar interferogram features both when the secondary is only displaced, and when it is displaced and tilted, we may wonder how we can discriminate between the two situations. One experimental procedure that I would suggest is the following. If the mirror is just shifted, we expect to see the *eye-ball* when Δx is small[21]. By adjusting the lateral position by few microns, we should be able to place the mirror on axis. At this point, for any sagittal translation, no asymmetry should be observed and, when passing through the on focus point, the fringes should disappear.

If the mirror was also tilted, a lateral shift by few microns would not remove the asymmetry in the interferogram because the tilt of the wavefront caused by the mirror rotation is still present. Also, in this case, fringes should be observed even when passing through the on focus position.

4.3.4 Primary mirror laterally displaced: effects of secondary mirror transverse shift

Until now the primary mirror has been assumed in perfect alignment along the focal axis, which is not the most likely situation. Mainly because of the error in the position of its center, the primary mirror may behave to the rest of the system, which is supposed to be aligned along the primary mirror axis, as if it was laterally displaced. Let us consider a transversal shift of 250 μm. The interferogram is formed by a large number of fringes and, actually, the figure appears to be undersampled (as said before, tilting the flat would

[21] When Δx is quite large the parallel fringes due to the wavefront tilt prevail, and we should first correct for the large lateral shift.

remove most of the fringes, and put in evidence the effects on both interferogram and beam shift due to coma. See *section 4.3.7*). If the secondary mirror is moved along the x-axis, no significative changes are observed until Δx_S is of the order of magnitude of the primary displacement. In the range $240 \div 260$ μm, we still observe straight fringes parallel to the y-axis. Only when $\Delta x_S = 250$ μm, the fringes disappear and the beam in the laboratory is perfectly back on itself. Moving the secondary by 1 μm both up and downwards makes immediately fringes show up.

I then verified that in order to compensate for the beam displacement we also either rotate the primary CCW by its $\theta_\perp$ or tilt the secondary CW by an angle 10 times larger than the $\theta_\perp$ of the primary mirror.

4.3.5 Combination of primary and secondary lateral displacement and rotation

First it is considered the case when the primary is both shifted and tilted and the secondary is allowed to translate and rotate.

When the tilt angle is $\theta_\perp$ any variation in the position of the secondary mirror produces a beam shift of the order of that observed when the primary mirror is aligned and only the secondary is misaligned. However for large misalignments (≥ 200 μm) of the primary mirror and corresponding $\theta_\perp$, the interferogram is not exactly as that obtained for misalignments of the secondary only. This is because the primary mirror shift-and-tilt combination does not restore the wavefront perfectly, so that the result is due to a combination of the wavefront error left by the unperfect compensation and the new introduced by the secondary misalignment.

Then I ran the simulation for some combinations of lateral misalignment and " random" tilt of the primary and/or of the secondary mirrors. As seen in *3.4.1* any combination, expressed in terms of a system misalignment, Δ_{SYS}, may also be considered in terms of an equivalent lateral shift of the secondary. Any further variation of the secondary will produce an interferogram similar to those observed as the starting misalignment Δ_{SYS} was carried by the secondary only. In particular, when Δx_{SYS} is of the order of some μm, the interferogram shows the asymmetric *eye-ball* feature when the secondary mirror is displaced along the z-axis. The *eye-ball* moves from the top to the bottom of the screen as the secondary travels towards the primary through the "on focus" point. Hence, this feature cannot be used as a unique signature of a lateral displacement of the secondary. However, if we are able to determine the amount of Δx_{SYS}, by following the technique suggested in *4.3.3* for example, we can improve the overall system performance by displacing the secondary by $-\Delta x_{SYS}$.

Also, by measuring the beam displacement in the laboratory, we can evaluate how much we have to translate the secondary, or rotate the primary, in order to get the return beam on itself. According to *3.4.3*, the beam shift can be written as a function of the equivalent secondary shift, $\Delta x_{S,COMP}$, or the equivalent primary tilt $\Delta\theta_{P,COMP}$ as

$$\Delta x_{beam}\ (cm) = 0.072 \cdot \Delta x_{S,COMP}\ (\mu m)$$

$$\Delta x_{beam}\ (cm) = 0.00072 \cdot 2 f_P \cdot \Delta\theta_{P,COMP}\ (\mu rad) \tag{49}$$

where f_P is in cm.
For example if, for a combination of primary and secondary misalignments, the beam appears shifted upward by about 1 cm, moving the secondary upward by about 14 μm or tilting the primary CCW by about 6 μrad would restore the beam position in the laboratory. These relations hold for the coarse restoration of both beam and interferogram, because they are derived in the paraxial approximation. They give the exact compensation shift and tilt for small beam deviations while, for large beam deviations, i.e. large shift and/or tilt of the mirrors, these values have to be corrected.

However, at this point we are not able to tell anything about the actual misaligned position of any mirror. By looking at the interferogram changes when we translate and rotate the secondary mirror, we can guess about the primary mirror misalignment.

For example, if the interferogram changes significantly by small rotation or lateral shift of the secondary, probably the primary mirror is either on axis or is both shifted and rotated. In the latter case primary mirror shift and tilt effects partly compensate. If either a large rotation (for example $\sim$ 1000 μrad) or large lateral shift ($\geq$ 100 μm) of the secondary mirror is necessary to produce any visible change in the interferogram, then probably the primary mirror is either rotated or shifted by relatively large amounts ($\sim$ 100 μm , or $\sim$ 100 μrad): the effects due to the primary mirror misalignment tend to hide those due to the secondary mirror.

Different combinations of secondary and primary mirror misalignments that give raise at the same beam displacement and to similar interferograms are listed in *Appendix D*.

4.3.6 Initial telescope alignment

The aim of this operation is to position all the mirrors of the system on a line so that the surfaces of the mirrors are normal, at their vertices, to this line. When the system is in this configuration we say that it is perfectly aligned.

In what follows I first describe the ideal process that leads to a perfect alignment. Then I present two situations of not-perfect alignment which are more likely to occur, as a consequence of the error with which the position of the centers of the primary and secondary mirrors are known.

Perfect alignment

The initial configuration is with the secondary mirror off. The search for the center of curvature of the primary mirror is performed by placing an object (cross-hair) in proximity of the calculated center of curvature[22], and adjusting its position in (x, y, z) until its image exactly coincides with the position of the object. The object coordinates are taken as those of the center of curvature, and the axis of the primary mirror is defined

[22]With respect to the center of the parabola the center of curvature is at a distance 2 times the focal length of the mirror.

as the line that joins this point and the vertex of the mirror. At this point the vertex of the secondary mirror is positioned on this line so that the mirror surface is normal to the axis at the vertex.

The final result is that the two mirrors are perfectly aligned: their vertexes are positioned on the same axis and their surfaces are normal to that axis.

However, due to errors in the definition of the position of the center of the primary mirror and the center of the secondary mirror, the line on which the secondary mirror is positioned may not exactly coincide with the actual axis of the system (i.e. the axis of the primary mirror).

Alignment when the position of the primary mirror is not accurately known

a) With reference to fig. TA.14 let us assume that the maximum error in the position of the center of the primary mirror is $\Delta x_{P,max}$.

From fig. TA.14 we see that, the secondary mirror is placed at a distance Δx_S from the actual axis and, in order to be normal to the line that is assumed to be the axis of the primary mirror, it is also tilted with respect to the normal to the actual axis. This tilt angle is $\Delta\alpha_P$, that is the angle corresponding to the shift $\Delta x_{P,max}$. However in this configuration the beam will not be back on itself, so that we will adjust the tilt of the secondary until the beam position is restored. The tilt angle for which the beam is back on itself is, in the paraxial approximation, the tilt angle by which the secondary is normal to the actual axis in their point of intersection. This angle is given by $\Delta\theta_S = \frac{\Delta x_S}{2f_S}$, where in this case the lateral displacement of the secondary is then given by

$$\begin{aligned} \Delta x_S &= [2f_P - (f_P - f_S)] \cdot \Delta\alpha_P \\ &= \left(1 + \frac{f_S}{f_P}\right) \cdot \frac{\Delta x_{P,max}}{2} \end{aligned} \tag{50}$$

The result of the alignment operation is that, even if the vertex of the secondary mirror is not positioned on the actual axis of the system, the two mirrors are both normal to the axis of the system at the point where it intersects their surfaces.

b) In case *a)* it is implicitly assumed that the primary mirror is perfectly aligned with the axis of the system. However a more likely situation is that, since when the center of curvature was determined, the mirror was actually displaced from the axis and tilted by the amount $\theta_\perp$, so that there would not be evidence of its misalignment[23]. In this case the secondary mirror will be positioned on the line which joins the vertex of primary mirror and the center of curvature, and in order to have the beam back on itself in the laboratory, it will be tilted so that it is perpendicular to the actual axis at the point where they intersect (see fig. TA.15).

[23]The center of curvature does not change its position if the primary mirror is both shifted by Δx_P and tilted by $\frac{\Delta x_P}{2f_P}$.

fig. TA.14 Misalignment of the secondary mirror due to uncertainty in the position of the center of the primary mirror. $\Delta x_{P,max}$ is the error in the position of the center, Δx_S and $\Delta \alpha_P$ are the amount by which the secondary mirror has to be shifted and tilted in order to position it normal, at its vertex, with respect to the line joining the center of the primary mirror and the center of curvature.

As result of the alignment operation the two mirrors are displaced with respect to the actual axis of the system, but they are perpendicular to it in the points where the axis intersects their surfaces.

I ran *IOTA_misalignment* to simulate the two cases of non-perfect alignment. When the error in the position of the center of the primary mirror is of the order of few hundreds of microns, by means of the shift-tilt compensation either of the secondary mirror (case *a*), or of both the primary and secondary mirrors (case *b*), the beam position is restored in the laboratory and the interferogram shows almost no aberration. However, for errors in the position of the center larger than about 500 μm, the shift-tilt compensation does not completely eliminate the wavefront aberration. This result is particularly evident when I simulate a translation of the secondary mirror along the z-axis and the characteristic *eye-ball* (see *section 4.3.2*) shows up. Also, the position of the beam in the laboratory is not perfectly restored.

fig. TA.15 Both primary and secondary mirrors are displaced with respect to the axis of the optical system. They are tilted by the angle $\theta_{S,\perp}$ and $\theta_{P,\perp}$ so that their surfaces are normal to the axis at their intersection points

4.3.7 Residual uncertainty on the telescope alignment

In this section I analyze in some detail how the residual aberration affects the accuracy of the alignment. I first show which is the cause of the residual aberration, and then I derive a relation between the wavefront *rms* and the error in the position of the center of the primary mirror.

4.3.7.1 Secondary shifted and then tilted to have it normal to the actual primary axis: why the beam is not perfectly back on itself

The beam-compressor used in IOTA belongs to the category of the *classical two-mirror telescopes*, i.e. those which are characterized for having a paraboloidal primary. The secondary mirror may be one of any of those that make the system to have zero spherical aberration. This requires that the conic constant of the secondary mirror, K_S, be

$$K_S = -\left(\frac{m+1}{m-1}\right)^2$$

where $m = \frac{\frac{R_S}{R_P}}{(\frac{R_S}{R_P}-\frac{D_S}{D_P})}$, $R_{S,P}$ are the radii of curvature at the vertexes, and $D_{S,P}$ are the mirrors diameters (*Schroeder (1987)*).

Schroeder (1987) shows that for a centered (i.e. perfectly aligned) classical two-mirror

telescope coma is exactly the same as that of a single paraboloid of the same focal ratio, while astigmatism is larger. However, in general, coma is responsible for setting the limit on the angular misalignment, since, for small angle, it dominates the effects due to astigmatism. For the particular case in which the secondary is a paraboloid, $K_S = -1$, and the focal points of the two mirrors coincide, the system is an afocal reducer. For this system

$$\frac{D_S}{D_P} = \frac{R_S}{R_P} = \frac{f_S}{f_P} \tag{51}$$

According to the aberration coefficient listed in Tables 5.4 and 5.6 of *Schroeder (1987)*, the afocal reducer is free of spherical aberration and has the classical third-order coma and astigmatism coefficients equal to zero.

However, starting from a centered afocal reducer any decenter or tilt of one of the mirrors can as well introduce aberration terms, primary coma and, for large angle, astigmatism. In general, for a classical two-mirror telescope, the coma coefficient, due to secondary misalignment, can be written as the sum of two independent terms, one function of decenter and one of tilt,

$$\begin{aligned} COMA(mis) &= COMA(dec) + COMA(tilt) \\ COMA(dec) &= \frac{\Delta l}{R_S^3}[K_S - \left(\frac{m+1}{m-1}\right)] \\ COMA(tilt) &= -\frac{\alpha}{R_S^2} \cdot \left(\frac{m+1}{m-1}\right) \end{aligned} \tag{52}$$

We notice that coma due to tilt introduces an aberration term that is independent of the conic constant and proportional to the tilt angle. Setting the global COMA(mis) to zero allows us to find for any lateral displacement Δl, a tilt angle for which the system is coma free. This angle is given by

$$\alpha^* = \frac{\Delta l}{R_S}[K_S \left(\frac{m-1}{m+1}\right) - 1] \tag{53}$$

and in general will depend on the optical parameters of the system.

The tilt angle α^* for an afocal reducer will be obtained by setting to zero the proper equation for COMA(mis) that is given by

$$COMA_{a.r}(mis) = \frac{1}{R_S^2} \cdot \left(2\frac{\Delta l}{R_S} + \alpha\right) \tag{54}$$

and $\alpha^* = \frac{\Delta l}{f_S}$ in this case. Since $\alpha^* = 2\theta_\perp$, any time we adjust (by shift and appropriate tilt) the secondary to have its surface normal to the primary axis at the intersection point, we introduce coma due to both shift and tilt.

According to *Schroeder (1987)* the angular coma in the tangential direction is given by

$$\theta_{coma} = 3 \cdot \left(\frac{h_S}{h_P}\right)^3 \cdot COMA_{a.r}(mis) \cdot h_P^2 \tag{55}$$

where now the tilt angle in $COMA_{a.r}(mis)$ is calculated by using $\alpha = \Delta l/2f_S$, and h_P is the height of the ray at the primary mirror. A ray entering the system parallel to the direction of the primary axis will exit tilted with respect to the entering direction by θ_{coma}, [24] a quantity proportional to the squared height of the ray. Only the ray which travels along the axis will not be affected by any deviation.

For the IOTA telescope the factors in the previous equation are given by

$$\frac{h_S}{h_P} = \frac{f_S}{f_P}$$

$$COMA_{a.r}(mis) = 3 \cdot \frac{\Delta l}{R_S^3}$$

$$h_P = \frac{D_P}{2}$$

where I chose to consider the marginal ray of the beam, so that

$$\theta_{coma} = \frac{9}{8} \cdot \frac{1}{m_{PS}} \cdot \frac{1}{F_\#^2} \cdot \frac{\Delta l}{2f_S} \tag{56}$$

and θ_{coma} results to be 0.018 times the angle by which the secondary is tilted[25].

In the laboratory, assuming a distance from the telescope of 4100 cm, the beam will appear laterally shifted by about

$$\Delta_{beam}\ (\mu m) = 3.23 \cdot \Delta l\ (\mu m) \tag{57}$$

This beam shift will not be detected for small displacement of the secondary, however for $\Delta l \geq 305\ \mu m$ it will be $\Delta_{beam} \geq 1\ mm$. In particular, I simulated the cases with $\Delta l = 275\ and\ 550\ \mu m$. The expected beam shift is about 0.1 and 0.2 cm respectively, and shifts of the order of 0.2 and 0.4 were actually obtained from the simulated intensity plots. An analysis of the Raytrace outputs shows that for large displacements (i.e. large tilts) the coefficient for astigmatism becomes important, and can be considered responsible for the increased beam shift. This also agree with a study performed by *Bhatia (1995)* that shows that, even when the secondary tilt is chosen equal to α^* astigmatism introduces a non-negligible beam shift in the plane of the image.

[24] The angular aberration here considered is equivalent to the angular difference between the raypath with aberration and without aberration. For an afocal reducer rays parallel to the primary axis should come out still parallel to it. If aberration is present the angular deviation between the direction of the incoming and outgoing rays will be given by the angular aberration as formulated by Schroeder.

[25] θ_{coma} given in eq.(55) has been multiply by 2 to take into account that the telescope is operated in autocollimation mode.

As far as the wavefront is concerned, the quantity COMA(mis), which is the aberration coefficient in the formula that gives the deviation of the actual wavefront from a perfect one, introduces a wavefront error which increases proportionally to the error in the position of the center of the primary. This linear relation is shown in fig. TA.16.

fig. TA.16 Non-perfect alignment: wavefront error as a function of the error in the position of the center of the primary mirror, when the secondary mirror is shifted and tilted by $\theta_\perp$ (see *3.3.3*)

4.3.7.2 Non-perfect alignment: both primary and secondary mirrors are shifted and tilted.

The results of the simulation performed for this configuration, show that the system has a better performance with respect to the case in which only the secondary is shifted and tilted. In particular, a detectable shift in the beam position appears only starting from a displacement in the position of the primary mirror of 1000 μm (which corresponds to a secondary shift of 550 μm). Also the wavefront error is improved, and the linear dependence on the primary mirror displacement, is shown in fig. TA.17.

In summary we see that because of the error in the position of the primary center, the secondary will be positioned shifted and tilted with respect to the actual axis of the system. In general the angle by which the secondary is tilted is not that for which the system would be coma-free. However, for small shifts of the secondary ($< 100\mu m$) there is a range of angular position, from the condition of coma-free ($2\theta_\perp$) to the condition of secondary normal to the actual axis at the intersection point ($\theta_\perp$), for which the resulting interferograms and intensity plots do not have detectable differences. For larger shifts, missing the condition of coma free will be made evident by the *eye-ball* feature showing

up when the secondary is displaced along the z-axis.

fig. TA.17 Non-perfect alignment: wavefront error as a function of the primary mirror misalignment when both primary and secondary mirrors are shifted and tilted by $\theta_{\perp,P}$ and $\theta_{\perp,S}$, respectively

The situation improves if, keeping the secondary mirror fixed in this position, the primary mirror is shifted to compensate for the error in the position of its center, and then tilted so that it is normal to the axis of the system.

From the result of the simulation with both mirrors misaligned but perpendicular to the axis of the system, I found that the relation between wavefront rms and error in the position of the center of the primary, Δl, is given

$$rms(\Delta l) = 3.15 \cdot 10^{-4} \cdot \Delta l \tag{58}$$

Now, if according to the Rayleigh's criterion, we assume that the maximum error in the wavefront must not exceed $\lambda/4$, then we can use eq.(58) to calculate the maximum error in the position of the center of the primary that our system can tolerate. This is given by

$$\Delta l_{max} = 800 \cdot \lambda \tag{59}$$

Tab. TA.7 shows values of Δl_{max} calculated for the central wavelengths in the K-, V-, and UV-band.

tab. TA.7 Maximum error in the position of the center of the primary mirror, Δl_{max}, as a function of different wavelength.

	$\lambda(\mu m)$	$\Delta l_{max}(\mu m)$
K-band	2.2	~ 1760
V-band	0.55	~ 440
UV-band	0.15	~ 120

When IOTA is operated in the K-band the constraints on the accuracy of the position of the center are very much relaxed, since we assume the uncertainty in the primary center position to be less than 1 mm. However when it is operated in the visible, to satisfy the Rayleigh criterion we require an accuracy in the center position better than 0.5 mm.

Obviously, the constrains for a similar instrument to be operated in the UV are much stricter.

4.3.8 Telescope's field of view

An afocal reducer is free from spherical aberration and, at the classical third-order approximation, also from coma and astigmatism, but not from field curvature (*Schroeder, 1987*). In order to estimate the effects due to field curvature for the IOTA beam-compressor, I ran the simulation assuming the system perfectly aligned and the source off-axis, by different angular quantities. Then a repeated the same simulation with the addition of a translation of the secondary mirror along the z-axis.

The eventual goal of the simulation is to determine what I called the "field of view" of the telescope, which is the maximum off-axis position for which the wavefront aberration introduced by the telescope still satisfies the Rayleigh criterion. This is not the field of view of the instrument, which is determined by the field stop of the system, which in most cases is given by the edge of the detector. For the calculation of the field of view of an interferometer, and specific evaluation for IOTA (about 2 *arcsec*), see *Traub, 1990.*

4.3.8.1 Perfect system: source off-axis

When only the source off-axis is considered, the simulation shows that for angles up to 100 μrad the total aberration is so small that no evidence of it appears on the interferogram (the wavefront rms $\sim 0.64 \cdot 10^{-3}$ μm).

By gradually increasing the off-axis angle the effect of a slight field curvature starts to show up ($\simeq 300\,\mu rad$) in the interferograms. By analyzing the relation between rms and $\theta_{off-axis}$, I found that the wavefront rms increases $\propto \theta^2_{off-axis}$. However, the constant factor in the relation

$$rms(\theta_{off-axis}) = const. \cdot \theta^2_{off-axis}$$

is not uniquely determined. For angles closer to the axis, the relation is given by

$$rms(\theta_{off-axis}) = 5.5 \cdot 10^{-8} \cdot \theta^2_{off-axis} \quad 100 < \theta_{off-axis} < 500\ \mu rad \tag{60}$$

whereas, by optimizing the relation for a range centered on $\theta_{off-axis} = 850\ \mu rad$, I found

$$rms(\theta_{off-axis}) = 4.75 \cdot 10^{-8} \cdot \theta^2_{off-axis} \quad \theta_{off-axis} \geq 500\ \mu rad \tag{61}$$

I used this last relation to calculate the field of view available to the IOTA telescopes. It is calculated as the source off-axis angle for which the wavefront rms is $\lambda/8$, and it is given by

$$\theta = 1.16 \cdot 10^3 \cdot \lambda^{1/2} \tag{62}$$

Where λ is in μm and θ is in μrad. Tab. TA.8 presents the available telescope field of view for operation in the K-, V-, and UV-band.

tab. TA.8 Field of view for the IOTA telescopes as a function of different wavelength bands.

	$\lambda(\mu m)$	$\theta\,(\mu rad)$	$\theta\,(deg)$
K-band	2.2	2550	0.15
V-band	0.55	1250	0.07
UV-band	0.15	628	0.04

For the actual operation of IOTA we can assume a maximum telescope field of view of about 0.07°.

4.3.8.2 Source off-axis: effects of sagittal displacement of the secondary mirror

A second set of simulation runs was performed with the secondary displaced along the z-axis.

From previous simulation we saw that simple translations of the secondary along the z-axis when the source is on-axis, produce field curvature aberration. Then, in this case, we would expect that by moving the secondary back and forward along the *z-axis* we will either improve or make worse the wavefront rms, depending on the direction of the wavefront curvature introduced by defocusing the system. The appropriate combination of the effects due to off-axis position and translation along the z-axis should minimize the wavefront error.

fig. TA.18 shows the results from the simulation with z-axis translation of $\pm 5\ \mu m$ compared with the curve for the perfect system.

fig. TA.18 *rms* as a function of field angle when the telescope is perfectly aligned, and when the secondary mirror suffers from both a positive and a negative sagittal displacement

Both curves have the same *rms* value when the source is on-axis, because the *rms* value depends only on the despace introduced between the foci of the two mirrors. As the telescope field angle increases the curve with negative defocus keeps increasing too, following $\sim \theta^2_{off-axis}$ as it can be derived from the RayTrace output file, where all the parameters of the coefficient which gives the field curvature aberration, have the same sign and contribute to increase the value of the coefficient. On the contrary, the curve with positive defocus starts decreasing towards a minimum value reached when $\theta_{off-axis} \simeq 1300\ \mu rad$. In this case the first (constant) and the third parameter of the aberration coefficient have opposite signs. Thus, when the contribution of the third term is comparable to that of the first one, the value of the aberration coefficient begins to reduce, until the point of minimum *rms* is reached. After this point the curve constantly increases as $\sim \theta^2_{off-axis}$, as in the case of the negative defocus.

When the off-axis angles are less than 900 μrad it is always preferable to work on focus in order to minimize the wavefront *rms*. However, for off-axis angle greater than 900 μrad, operation of the instrument out of focus could be preferred because it produces a

significative improvement of the wavefront rms.

Tab. TA.9 presents, for the K-, V-, and UV-band, the improvement, in terms of larger field of view available, obtained with a positive sagittal displacement with respect to the cases with perfect system and negative sagittal displacement.

tab. TA.9 IOTA telescope field of view. Results are for a perfect system ($\Delta z = 0$) and system affected by both positive and negative sagittal displacement ($\Delta z = \pm 5\ \mu m$), for different wavelength bands.

	$\lambda(\mu m)$	$\theta_{\Delta z=0}$	$\theta_{\Delta z=-5}$	$\theta_{\Delta z=+5}$
K-band	2.2	~ 2550	~ 2500	~ 3200
V-band	0.55	1250	$\sim$650	$\sim$1900
UV-band	0.15	650	-	-

In the UV the condition $rms < \lambda/8$ can be met only if the system is on focus (see fig. TA.17).

Beamsplitter

0. Introduction

An optical interferometer consists of separate light-collecting elements, each collecting a section of the wavefront from a star, and a means of bringing the sections of wavefront together. The most common beam combination method is to illuminate a beamsplitter with one input beam from each side of it. A beamsplitter is a flat, transparent plate which is coated on one side to give 50% transmission and 50% reflection (in principle). In the present context it might more properly be called a beamcombiner, but we shall use the traditional nomenclature. The two output beams, each containing contributions from both inputs, are then focused onto detectors whose output eventually produces interferometric fringes given by variation of intensity as a function either of time or wavelength.

In this section I present a description of the characteristics of a beamsplitter to be used for astronomical interferometry, which are eventually summarized by means of four figures of merit. Most of what is presented at this regard derives from the only, to my knowledge, documented study on this subject. It is a study by James D. Phillips who has been developing theoretical formulations and experimental testing specifically for an astronomical beamsplitter (in particular his work developed within the design project of a space interferometer, POINTS) and with whom I was able to have very interesting and stimulating conversations.

The second part of this section deals specifically with polarization effects introduced by the beamsplitter, and represents the original work by the author on this subject. It presents the analytical formula for the intensity of the output beams when phase shifts due to polarization are present, and the evaluation of the loss in visibility caused by these phase shifts. I also present the result of the computer program that I implemented in order to simulate interferometric fringes affected by phase shifts due to polarization.

1. General considerations

Whenever light travels from one medium into another of different optical properties, part of the light is reflected and part is transmitted. For many uncoated optical glasses, the reflected light typically represents a few percent of the incident radiation, so that for optical designs using more than a few optical components, losses in transmitted light level can accumulated rapidly. According to the optical component application, it is generally required that the reflected portion of the incident light approaches 0% for transmitting optics (lenses), and 100% for reflective optics (mirrors), or it is at some fixed intermediate value for partial reflectors, as it is the case for a beamsplitter.

In principle, the surface of any optical element can be coated with thin layers of various materials (thin films) in order to ensure the desired reflection/transmission ratio. This ratio depends on the nature of the material from which the optic is fabricated as well as the wavelength of the incident light and the angle of incidence. Also, there is a polar-

ization dependence to this ratio when the angle of incidence, with respect to the surface normal, is not 0°.

As far as partial reflectors are concerned, the simplest that can be conceived is a boundary between two transparent media. The surface of a glass plate in air forms a beamsplitter, whose reflection and transmission properties could be tuned by selecting an appropriate coating. However, when used in an interferometric device, a beamsplitter consisting of a coating on a supporting plate introduces an asymmetry[26] unless it is balanced by a compensating plate put next to the side where the reflecting coating is found. When this plate does not match the support of the beamsplitter in both material and thickness, the asymmetries in path length introduced by the beamsplitter are not completely compensated.

A completely symmetrical beamsplitter would eliminate this problem. Such a beamsplitter can be an unsupported dielectric film, but this solution is used mainly for application at long wavelength[27] (for example Melinex or Mylar sheets are used in the far infrared *Steel, 1983*). Otherwise, the dielectric or metal coating can be sandwiched between two identical plates or two identical prisms. For complete symmetry, this sandwich should be made without a cement and the two plates should be closely matched in thickness. Nowadays, independently of the solution adopted, dielectric beamsplitters are commonly realized with multilayer coatings (up to ~ 100 layers can be present) whose number of layers and composition are selected in order to meet the reflection and transmission requirements imposed by the application.

The ideal partially reflecting beamsplitter, to be used in an optical stellar interferometer, would transmit and reflect one half of the incident radiation with a constant phase change, whatever its polarization. Also, it would remain close to this ideal if used over a range of frequencies, and it would introduce no phase change that varied with frequency. However, in practice, since all reflecting surfaces used at other than normal incidence have properties that vary with the polarization of the radiation, and any substrate has frequency dependent response, a real beamsplitter may depart far from this ideal.

In particular, the last paragraph of the following section presents four parameters, called figures of merit, which are used to describe how much a given beamsplitter departs from this ideal behaviour.

2. Definitions, constraints and figures of merit for the beamsplitter

In this section are presented the quantities and parameters by means of which beamsplitter properties are described.

[26] This is because one of the reflected beams travels across the plate twice and the other does not cross it at, while both transmitted beams travel across the plate once.

[27] A pellicle beamsplitter is made of a high tensile strength elastic membrane, stretched like a drumhead and bonded to the edge of a frame. These kind of pellicles have a thickness of few microns, are easily deformed by stressing the mounting frame, and are sensitive to acoustical disturbances.

2.1 Definitions

To fully specify a beamsplitter's reflective and transmissive properties at a given wavelength requires 16 parameters, 8 for each polarization state of the incoming light [28]. These parameters, for one polarization only, are presented in tab. BS.1

tab. BS.1 Beamsplitter parameters by means of which its reflective and transmissive properties are specified. One polarization only

T_+, T_-	Transmittance
R_+, R_-	Reflectance
ψ_+, ψ_-	Phase shift of the transmitted wave
ρ_+, ρ_-	Phase shift of the reflected wave

where quantities referred to the right side of the beamsplitter are labelled with "+" and those related to the left side are labelled with "−". R_+, R_-, T_+, and T_- are the intensity reflection and transmission coefficients, referred to as reflectance and transmittance, and the phase shift parameters are defined according to fig. BS.1.

Fig. BS.1 shows the schematic of a beamsplitter (in this case a sandwich beamsplitter), which may have not symmetry and may have loss. Let us consider two incident beams, ray 1 and ray 2, one for each side. Here it is assumed that the transmitted ray leaves the beamsplitter unchanged in direction, and the reflected ray emerges in the same direction as from a simple mirror at the reference plane, so that contributions from ray 1 and ray 2 to the two outgoing wavefronts are parallel. If the beamsplitter has mirror symmetry, the reference plane is at its plane of symmetry. If it does not, the reference plane may be chosen arbitrarily.

[28] Polarization at a beamsplitter operating at non-normal incidence is usually described in terms of a pair of states with orthogonal linear polarization. Defining the *plane of incidence* to include the beam propagation vector and the surface normal, the chosen states have their electric field perpendicular to and parallel to this plane, and are labelled s and p respectively.

fig. BS.1 Beamsplitter phase definition

Let us consider the transmitted and reflected beams on the right side. In the absence of the beamsplitter, wavefronts from 1 and 2 that pass O at the same time would arrive at W_+ at a time t_0. When the beamsplitter is put in, at t_0 the wavefronts will instead be at W_{T+} and W_{R+}, respectively, delayed in phase with respect to W_+ by ψ_+ and ρ_+. By analogy, ψ_- and ρ_- are the phase shifts on the left side. Except for differences in different polarizations, all phase effects of the beamsplitter, on transmitted and reflected beams, are summarized in these four parameters.

According to *Phillips and Hickey (1995)* the *beamsplitter phase* is defined as

$$\phi_\pm = \rho_\pm - \psi_\pm \tag{63}$$

It represents the OPD, between the two interfering beams, introduced by the beamsplitter, expressed as radians of phase at a particular optical frequency; it is the critical quantity to be considered for interferometric applications. The other quantity of interest is the intensity of the outgoing beams. Following the notations of fig. BS.1, and the definition in eq.(63), the intensity leaving each of the two ports of the beamsplitter is given by

$$I_\pm = \frac{I_0}{2}[R_\pm + T_\pm + 2\sqrt{R_\pm T_\pm}cos(\theta \pm \phi_\pm)] \tag{64}$$

where $\frac{I_0}{2}$ is the intensity of the incoming beam in one polarization and θ is the phase difference between the incoming beams, ray 1 and ray 2[29].

[29]We make use of the quadratic law that gives the intensity, I, of an electric field, $Ee^{i\gamma}$,

2.2 Constraints

Phillips derived the constraints (we could also say the properties of) on the parameters which describe the reflected and transmitted beams, arising from requiring that the beamsplitter satisfies four particular conditions. The constraints were derived both when the conditions are considered separately and in combination. The conditions are denoted as

- No constraints (N);
- Rotational invariance (R), which is the invariance under 180° rotation about the surface normal;
- Mirror symmetry (M), which applies to a sandwich beamsplitter, where it is assumed to be simmetric with respect to the mid-plane of the coating;
- Losslessness (L), which assumes perfect conservation of energy.

The results from applying the previuos conditions are shown in tab. BS.2

$$I = \left|Ee^{i\gamma}\right|^2 = (Ee^{i\gamma}) \cdot (Ee^{i\gamma})^*$$

where E is the amplitude and γ is the phase of the field, and $*$ means complex conjugation.

According to this definition, at each port of a beamsplitter we have

$$I_\pm = \left| \sqrt{\frac{I_o}{2} R_\pm}\, e^{i(\theta+\rho_\pm)} + \sqrt{\frac{I_o}{2} T_\pm}\, e^{i\psi_\pm} \right|^2$$

so that we get

$$I_\pm = \frac{I_o}{2}\left[R_\pm + T_\pm + \sqrt{R_\pm T_\pm}\, e^{i(\theta+\rho_\pm-\psi_\pm)} + \sqrt{R_\pm T_\pm}\, e^{-i(\theta+\rho_\pm-\psi_\pm)}\right]$$

$$I_\pm = \frac{I_o}{2}\left[R_\pm + T_\pm + 2\sqrt{R_\pm T_\pm}\, cos(\theta+\rho_\pm-\psi_\pm)\right]$$

which, using definition (72), gives eq.(64).

tab. BS.2 Conditions upon a beamsplitter and resulting constrain sets

	Conditions	Constrain set		
(N)	No conditions	no constraints		
(R)	Rotationally invariant	T_+	=	T_-
		ψ_+	=	ψ_-
(M)	Mirror-symmetric	R_+	=	R_-
		ρ_+	=	ρ_-
		T_+	=	T_-
		ψ_+	=	ψ_-
(L)	Lossless	R_+	=	R_-
		T_+	=	T_-
		$R_+ + T_+$	=	1
		$\phi_+ + \phi_-$	=	$(2k+1)\pi$
(RL)	Rotationally invariant and Lossless	R_+	=	R_-
		T_+	=	T_-
		$R_+ + T_+$	=	1
		ψ_+	=	ψ_-
		$\phi_+ + \phi_-$	=	$(2k+1)\pi$
(ML) and (RML)	Mirror-symmetric and Lossless	R_+	=	R_-
		T_+	=	T_-
		$R_+ + T_+$	=	1
		ρ_+	=	ρ_-
		ψ_+	=	ψ_-
		ϕ_+	=	$\left(k+\frac{1}{2}\right)\pi$
		ϕ_-	=	$\left(k+\frac{1}{2}\right)\pi$

From Table II of *Phillips, (TM94-12)*

The constraints in tab. BS.2 show that the most desiderable beamsplitter would be one that is both mirror-symmetric and lossless, in short defined as RML, because (*Phillips, (TM94-12)*) the combination of the two conditions M and L automatically implies that it is also rotationally invariant. In this case the important constraint, for interferometric applications, is that on phase differences, namely

$$\phi_+ = \phi_- = \left(k + \frac{1}{2}\right)\pi \tag{65}$$

because it says that the beamsplitter contribution to the fringe phase (remember that in eq.(64) it is the quantity $\phi_\pm$ which appears in the argument of the cosine term) is independent of frequency[30].

In the condition RML there is one free parameter (there are 8 unknowns and 7 equations), that we can decide to be the reflectance. The choice of its value determines the characteristics of the beamsplitter and may be selected in order to meet the requirements for application in an interferometer.

However, a more likely situation is that for a beamsplitter which is lossy and not rotationally invariant (and asymmetric, if it is not a sandwich beamsplitter), i.e. conditions RML are not met. In this case four independent parameters, called *figures of merit*, can be defined and according to them the suitability of the beamsplitter for interferometric applications is determined.

2.3 Figures of merit

a) Equality of reflectance and transmittance

High fringe contrast requires, among other things, that reflectance and transmittance match. From eq.(64) and according to the standard definition of visibility ($V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$), we get

$$V = \frac{2 \cdot \sqrt{R_\pm T_\pm}}{R_\pm + T_\pm}$$

so that the highest visibility requires that the two interfering beams have the same intensity ($R_\pm = T_\pm$)[31]. In terms of figure of merit this is expressed by

[30] The relations on phase differences for L and RL conditions, say that for any frequency there is a combination of ψ_+, ψ_-, ρ_+, ρ_-, such that the sum of the ϕ_+ and ϕ_- beamsplitter phases is constant, but each of these phases can be considered an independent parameter. This implies that ϕ_+ and ϕ_- can vary with frequency.

On the contrary, the relation for the condition RML says that each of the two beamsplitter phases is always equal to a constant quantity, and this can happen only if ϕ_+ and ϕ_- are frequency independent ($\psi_\pm$ and $\rho_\pm$ can be frequency dependent but change in the same way so that this dependence cancel when we consider ϕ_+ and ϕ_-.

[31] From eq.(64) we have

$$I_{max} = \left[R_\pm + T_\pm + 2\sqrt{R_\pm T_\pm}\right]$$

and

$$I_{min} = \left[R_\pm + T_\pm - 2\sqrt{R_\pm T_\pm}\right]$$

so that

$$V = \frac{2 \cdot \sqrt{R_\pm T_\pm}}{R_\pm + T_\pm}$$

$$\mu_{\pm} = R_{\pm} - T_{\pm} \tag{66}$$

so that small values of $\mid \mu \mid$ give high values of visibility (which means also shorter integration time).

b) Loss

Regardless if condition a) is met, the beamsplitter can absorb and dissipate energy, so that the intensity of the emerging beams is reduced. The loss can be expressed by

$$l = 1 - (R_{\pm} + T_{\pm}) \tag{67}$$

where $R_{\pm}$ and $T_{\pm}$ give the fraction of the intensity of the incident beam which is reflected and transmitted and are quantities which are always between 0 and 1.

l should be kept as small as possible, and usually this is possible with dielectric beamsplitters.

c) Beamsplitter phase variation with optical frequency

The actual phase of the detected fringe pattern is the sum of the starlight phase (given by $\theta = 2\pi\alpha B\nu/c$, α is the stellar offset, B is the baseline length), and the phase shift added by refractive and reflective optics (as analyzed in other sections), and the beamsplitter. In this section we neglect the contribution due to the optics preceeding the beamsplitter, and concentrate on the beamsplitter contributions which are given by ϕ_+ on the right side and ϕ_- on the left side. From tab. BS.2 we see that, unless a beamsplitter is both lossless and symmetric, only the sum of ϕ_+ and ϕ_- is constrained ($\phi_+ + \phi_- = (2k+1)\pi$)[32]. Because

---

and in order to have the constrain on $R_{\pm}$ and $T_{\pm}$ for maximum of visibility, we impose $V = 1$, so that we get

$$\left(\sqrt{R_{\pm}} - \sqrt{T_{\pm}}\right)^2 = 0$$

which implies the condition $R_{\pm} = T_{\pm}$.

[32]For a lossless beamsplitter the conservation of energy implies that $I = I_0$, where I_0 is the total incoming intensity (given by both ray 1 and ray 2) and I is given by the sum of the intensity at the two output ports, i.e.

$$I = I_0 \left[1 + \sqrt{R_+ T_+} cos(\theta + \phi_+) + \sqrt{R_- T_-} cos(\theta + \phi_-)\right].$$

The condition that $I = I_0$ requires that the sum of the second and third terms in square brakets, which are due to interference, vanish for any θ. For this to happen two conditions have to be satisfied. The first one is a constrain on the intensity coefficients, namely: $R_+T_+ = R_-T_-$, which is a necessary but not sufficient condition. The constrain to be also satisfied is that on the phase difference. Making use of $R_+T_+ = R_-T_-$ we have

$$I = I_0 \left\{1 + 2\sqrt{RT}\left[cos\left(\theta + \frac{\phi_+ - \phi_-}{2}\right) cos\left(\frac{\phi_+ + \phi_-}{2}\right)\right]\right\}$$

so that the second necessary condition is that $\phi_+ + \phi_- = (2k+1)\pi$.

their values, separately, can vary arbitrarily, the beamsplitter makes a direct contribution to the fringe phase. To clarify this point, let us consider a lossless and initially mirror symmetric beamsplitter, with phases $\phi_{\pm} = \frac{\pi}{2}$. Then assume that a phase delay Δ is added on one side of the coating (for ex. by adding a layer or for deposit of impurity, or for thermal changes that affect the two side differently), so that the simmetry condition is not satisfied any more. Both ψ_+ and ψ_- will increase by Δ while ρ_+ will increase by 2Δ and ρ_- will not increase. As a whole, ϕ_+ will increase by Δ and ϕ_- will decrease by Δ, but the sum $\phi_+ + \phi_-$ remains the same. With reference to eq.(65) we see that, without violation of symmetry, the fringe phases on the + and − sides are

$$\begin{aligned} \theta + \phi_+ &= \theta + \frac{\pi}{2} \\ \theta + \phi_- &= \theta - \frac{\pi}{2} \end{aligned}$$

because in one case the stellar phase θ is carried by the reflected beam and in the other case by the transmitted beam[33]. With the violation of symmetry, $\Delta\phi_+ = \Delta$ and $\Delta\phi_- = -\Delta$, so that those phases have to be replaced by

$$\begin{aligned} \theta + \phi_+ &= \theta + \frac{\pi}{2} + \Delta \\ \theta + \phi_- &= \theta - \frac{\pi}{2} + \Delta \end{aligned}$$

and this is equivalent to replace the starlight phase θ with $\theta + \Delta$.

From $\theta = 2\pi\alpha B\nu/c$ we see that the expected variation of starlight with frequency is linear. If the beamsplitter introduces a phase variation, Δ, which is also linear with frequency, it will do no harm, as far as it is constant in time, since it can be modelled out during calibration (What a linear Δ does, is to add to the stellar phase term, which implies a shift of the apparent angle α on the sky. This could do harm, if not modelled out , in astrometric measurements). A worse case is when the phase variation, Δ, is not linear with frequency, because it cannot be modelled out and, in the case of astrometric measurements, the modelling required for estimation of the stellar offset from the fringe pattern may result in systematic error.

When Δ is not linear, we consider the deviation of the beamsplitter phase about a best-fit

[33] On the right side we have

$$I_+ = \left| R_+ e^{i(\theta+\rho_+)} + T_+ e^{i\psi_+} \right|^2 \to cos(\theta + \phi_+)$$

on the left side we have

$$I_- = \left| R_- e^{i\rho_+} + T_- e^{i(\theta+\psi_-)} \right|^2 \to cos(-\theta + \phi_-) = cos(\theta - \phi_-)$$

and we adopt the convention of keeping the same sign on both sides for the stellar phase.

straight line (which represents the ideal linear dependence on frequency), and call this deviation the *residual phase*. During the study for the optimization of the beamsplitter design, the aim should be to keep this quantity as small as possible.

The figure of merit to be considered to characterize the frequency dependence of the beamsplitter should be the beamsplitter phase as a function of frequency, $\phi_{\pm}(\nu)$. However, when it is known that the phase varies non-linearly with frequency, the residual phase is used as well.

d) Polarization

Until now we have made no distinction between the two polarization components of the incoming light, when considering the beamsplitter phase. However, I show (see following sections) that difference in phase variation for different polarizations may introduce loss in the visibility, as well as errors in astrometric measurements (the apparent position of targets whose light is polarized will vary as the orientation of the interferometer with respect to the target varies). In order to consider the polarizations effects of a beamsplitter, we define the quantity

$$\Delta\phi_{sp} = \phi_{s\pm} - \phi_{p\pm} \tag{68}$$

as the polarization figure of merit. When studying the optimization of a beamsplitter, we should aim to keep this quantity as small as possible.

3. Polarization effects

In this section I present the effects on different polarization states, of both the transmitted and reflected beam, caused by the beamsplitter interface. In particular I analyze how different phase variations for the two polarizations affect the final interference pattern.

3.1 Intensity and phase of the polarized components of transmitted and reflected beams

When a plane wave falls onto a boundary between to homogeneous media of different optical properties, it is split into two waves: a transmitted wave proceeding into the second medium and a reflected wave propagated back into the first medium. The intensity of reflected and refracted beams of unpolarized light striking, at an angle of incidence α, an optical medium having refraction index n', are given by Fresnel's equations. In particular, the Fresnel's equations give the intensity of these beams according to their polarization component, parallel (*p-polarized*) and normal (*s-polarized*) to the plane of the incident beam. The Fresnel's equation are

$$\begin{aligned} T_p &= \frac{2\,n\,cos\alpha_i}{n'\,cos\alpha_i + n\,cos\alpha_t} \cdot I_p \\ T_s &= \frac{2\,n\,cos\alpha_i}{n\,cos\alpha_i + n'\,cos\alpha_t} \cdot I_s \end{aligned}$$

$$R_p = \frac{n^{'} cos\alpha_i - n\, cos\alpha_t}{n^{'} cos\alpha_i + n\, cos\alpha_t} \cdot I_p$$

$$R_s = \frac{n\, cos\alpha_i - n^{'} cos\alpha_t}{n\, cos\alpha_i + n^{'} cos\alpha_t} \cdot I_s$$

where I_p and I_s are the p- and s-polarization component of the incident beam, and α_i, α_t the angles the incident and transmitted beams form with respect to the normal to the surface.

As far as the phase of each component of the reflected or transmitted wave is concerned, it can be shown (*Born and Wolf, 1980*) that for a non-absorbing single boundary, the phase of each component is either equal to the phase of the corresponding component of the incident wave or differs from it by π. The phases of the two components of the transmitted wave are equal to those of the respective components of the incident wave. In the case of the the reflected wave, however, the phase depends on the relative magnitude of the media index of refraction and on the incidence angle.

If the second medium is optically denser than the first ($n' > n$), the s-component phase differs by π from the incident s-component. The p-component has the same phase until the sum of the incidence and transmission angles is less than $\frac{\pi}{2}$, then the phases differ by π. The opposite happens when $n' < n$. The s-component has the same phase as the incident one, while the parallel component differs in phase by π as far as $\alpha_i + \alpha_t$ is less than $\frac{\pi}{2}$. Then the two components have the same phase. As far as the latter case is concerned, the phase shift is not constant for incident angles larger than the *critical* angle (for which $\alpha_t = \frac{\pi}{2}$, case of internal reflection). For these angles the phase of both components starts increasing according to (different) non-linear laws.

When the boundary surface is an absorbing one (i.e. a metal, for ex. the coating of a mirror) the p- and s-component undergo changes in phase that are more complicated and their difference varies between 0 and π depending on the incident angle (see Born and Wolf page 619).

3.1.1 Polarization effects on reflected and transmitted beams

Let us now consider a dielectric beamsplitter for application in a stellar interferometer. It is requested to split each incident beam into two beams of equal intensity, one transmitted and one reflected, without energy loss. For this purpose a multilayer partially reflecting dielectric coating is normally used. A multilayer coating is a succession of thin plane-parallel films (sometimes over 100 individual layers) whose thickness can be controlled with very high accuracy. Each layer is influenced by the optical properties of the layer next to it, and the properties of the last layer are eventually influenced by the environment (either the substrate or air or vacuum). By an appropriate choice of the film material, number and thickness of the layers, the reflection/transmission ratio can be optimized for several sets of conditions (wavelength and angle of incidence) or optimized over a particular range of conditions. In general, the performance of dielectric coatings is more

wavelength sensitive than that of metallic coatings, in addition to the fact that the ratio of transmitted and reflected intensities may be quite different for the s- and p-polarization components of the incident beam, depending on the angle of incidence. This last aspect is a serious drawback for a beamsplitter, hence it is necessary to optimize the multilayer coating taking into account the constraints on the incidence angle. Performances with the reflectance and transmittance of the two components matched to better than 5% are usually achievable. As an example, in fig. BS.2 it is shown the reflectance curve, as a function of wavelength, for the beamsplitter used in the optical branch of IOTA. This beamsplitter has been optimized for an incidence angle of 10°, and certainly a different optimization would have been required for a different incidence angle.

fig. BS.2 Reflectance curve for s- and p-polarization components of the incident beam for the IOTA beamsplitter. The incidence angle is 10°

3.1.2 Polarization effects on the phase shift between reflected and transmitted beams

Not only the reflection and transmission properties are different for the two polarizations, the phase shifts also are different (and wavelength dependent), and are peculiar to each particular multilayer.

In what follows I analyze the effect of the polarization phase shift on the intensity of the beam resulting from interference, and I show that it affects the phase of the visibility modulation term as well as the modulus of the visibility of the interference fringes.

With reference to fig. BS.3. let us consider two beams, B_1 and B_2, whose reflected (R_1) and transmitted (T_2) parts are to interfere. Assuming that they have same amplitudes, let $\alpha_{i,j}$ ($i = s, p$; $j = 1, 2$) be the phase for the p- and s-polarization of the two beams, due

to the reflections experienced through the whole light path preceeding the beamsplitter. Taking into account only the phase terms, we can write:

$$\begin{aligned} B_1 &= (e^{i\alpha_{p,1}}, e^{i\alpha_{s,1}})e^{i\theta} \\ B_2 &= (e^{i\alpha_{p,2}}, e^{i\alpha_{s,2}}) \end{aligned} \tag{69}$$

where θ is the stellar phase (i.e., $\theta = 2\pi x/\lambda$ where $x = Bsin\alpha$ and α is the stellar offset with respect to the normal to the baseline).

fig. BS.3 Beams from the two arms of the interferometer at the beamsplitter interface: R_1, reflected beam from B_1, and T_2, transmitted beam from B_2, interfere

After interaction with the beamsplitter, both the reflected and the transmitted beams on one side of the beamsplitter, say the right side, will show a variation in the phase of the p- and s-polarization component, expressed by $\beta_{i,j}$ $(i = s, p\ ;\ j = 1, 2)$. Assuming a lossless beamsplitter we still have equal amplitude for the reflected and transmitted beams, and we can write:

$$\begin{aligned} R_1 &= (e^{i(\alpha_{p,1}+\beta_{p,1})}, e^{i(\alpha_{s,1}+\beta_{s,1})})e^{i\theta} \\ T_2 &= (e^{i(\alpha_{p,2}+\beta_{p,2})}, e^{i(\alpha_{s,2}+\beta_{s,2})}) \end{aligned} \tag{70}$$

where R_1 and T_2 are the two beams on the right side that are going to interfere. The resulting intensity is given by

$$I = |R_1 + T_2|^2 \tag{71}$$

so that by adding the two vectors in (70) and calculating the squared modulus, we get

$$\begin{aligned} I &= 4 + 2cos(\alpha_{p,1} + \beta_{p,1} - \alpha_{p,2} - \beta_{p,2} + \theta) + 2cos(\alpha_{s,1} + \beta_{s,1} - \alpha_{s,2} - \beta_{s,2} + \theta) \\ &= 4 + 2\left[cos(\Delta\alpha_p + \Delta\beta_p + \theta) + cos(\Delta\alpha_s + \Delta\beta_s + \theta)\right] \end{aligned} \quad (72)$$

where $\Delta\alpha_i = \alpha_{i,1} - \alpha_{i,2}$, and $\Delta\beta_i = \beta_{i,1} - \beta_{i,2}$, and $i = s, p$.

This equation shows that the resulting intensity is due to two different interference patterns (one for the p- and one for the s-polarization) which do not perfectly overlap. They are shifted with respect to the stellar phase by different amounts (namely, $(\Delta\alpha_p + \Delta\beta_p)$ and $(\Delta\alpha_s + \Delta\beta_s)$) so that the fringe visibility is in general reduced.

We can rewrite eq.(72) as

$$I = 4\left[1 + cos\left(\frac{\Delta\alpha_p + \Delta\beta_p + \Delta\alpha_s + \Delta\beta_s}{2} + \theta\right) cos\left(\frac{\Delta\alpha_p + \Delta\beta_p - \Delta\alpha_s - \Delta\beta_s}{2}\right)\right] \quad (73)$$

where the two cosines are the factors which multiply the modulus of the visibility ($|V| = 1$).

This shows that the information on the stellar phase is affected by a systematic shift equal to one half of the sum of all the polarization phase differences, while the visibility deterioration depends on the difference between the overall p- and s-polarization phases.

As far as the phase shift due to the multiple reflections is concerned [34], a sufficient (but not necessary) condition for the two beams to have the same phase shift is that, at any given stage in each beam, the mirrors have the same optical type and be oriented in the same direction (*direction cosine rule, Traub, 1988*). In general, with an appropriate design, it is possible to make the cosine rule satisfied everywhere except in the beam-combination area, the last stage of the interferometer, where an asymmetry occurs in the act of reflecting one of the beams toward the beamsplitter. However, for incidence angles relatively small (i.e. closer to 20° than to 45°) the difference $\Delta\alpha_p - \Delta\alpha_s$ can be kept small enough (of the order of 5° (*Traub, 1988*)) so that it introduces an almost negligible visibility deterioration ($\sim 0.1\%$ loss of visibility).

Let us now consider the phase shift due to the beamsplitter. In this case it is not possible to find a general rule by which the phase shift can be kept close to a minimum value, because of the asymmetry of the splitting and combining operation. Also, the phase shift between the two components is characteristic of each multilayer coating and can only be experimentally determined. However, by accurate simulation analysis, an optimization of the multilayer coating might be found, so that the phase difference is sufficiently small. From eq.(73) it could be suggested that the optimization aim to achieve $\Delta\beta_p - \Delta\beta_s$ equal to $-(\Delta\alpha_p - \Delta\alpha_s)$, over a range of wavelengths, so that the visibility is almost free from

[34] Note that it can be written
$\Delta\alpha_p - \Delta\alpha_s = \alpha_{p,1} - \alpha_{p,2} - \alpha_{s,1} + \alpha_{s,2} = (\alpha_{p,1} - \alpha_{s,1}) - (\alpha_{p,2} - \alpha_{s,2}) = \Delta\alpha_1 - \Delta\alpha_2$
where $\Delta\alpha_1$ and $\Delta\alpha_2$ are the phase difference between the two polarization of B_1 and B_2.

deterioration due to polarization phase shift. In this case the offset affecting the stellar phase would be also a priori determined, and given by $(\Delta\alpha_s + \Delta\beta_s)$ (or equivalently by $(\Delta\alpha_p + \Delta\beta_p)$).

Optimization studies performed for POINTS showed that the magnitude of the s-p phase difference may be kept, smaller than 2.5°, in the range 0.3 - 0.9 μm, even for a lossy beamsplitter, if the incidence angle is of the order of 15° *(Phillips, 1994)*.

4. The IOTA beamsplitter

For the optical table of IOTA the beam-combiner element consists of a beamsplitter plus a compensator plate. Both the beamsplitter substrate and the compensator plate are made of a BK-7 plate, 14.0 mm thick. The compensator has antireflection coating on each side; the beamsplitter has antireflecting coating on one side, and on the other it is coated by a dielectric film optimized to give, at incidence angle of 10°, $50 \pm 5\%$ reflectance and transmittance, in the wavelength range from 0.5 to 0.8 μm.

The manufacturer provided a table of calculated phase shift for both polarization components, at different wavelengths between 0.5 and 0.8 μm, which is reported in a *Appendix F*. The phase variation of each component is strongly wavelength dependent with values between 18.41° and 358.50°. However, the difference in the phase variation of the two polarizations is small, and $\Delta\beta_p - \Delta\beta_s$ is always less than 1°. The optimization of the coating is to be very good, at least according to the calculated values. The overall deterioration of visibility due to polarization effects can be calculated by mean of eq.(73), where it is assumed that, for incidence angle $\sim 10°$, the largest $\Delta\alpha_p - \Delta\alpha_s$ is 14° (*Traub, 1988*), and the largest $\Delta\beta_p - \Delta\beta_s$ is 1°. The associated loss of visibility is of about 0.85 %.

4.1 Fringe pattern simulation for polarization effects

As a part of a program for the simulation of the expected output for the PAPA camera, the detector initially used for the visible observations with IOTA, I performed a simulation showing the effect of the phase shift due to polarization on the visibility. The function that is plotted is given by

$$V(\sigma) = \left| \frac{2J_1(\pi\sigma\theta B)}{\pi\sigma\theta B} \right| * cos\left(2\pi\sigma x + \frac{\Delta p + \Delta s}{2}\right) * cos\left(\frac{\Delta p - \Delta s}{2}\right) \qquad (74)$$

where the modulus term gives the visibility for a circular source of uniform brightness and angular diameter θ, B is the projected baseline, x is the delay line position with respect to the zero-path-difference point, Δp and Δs are the overall differences between the phase value of the reflected and transmitted beam for the p- and s- polarization, respectively, as a function of σ $(\sigma = 1/\lambda)$.

An example of this simulation, based on the calculated phase shift for the IOTA beamsplitter, is presented in fig. BS.4, where for comparison the visibility curve is shown as a function of wavelength, between 0.5 μm (channel number 202) a and 0.8 μm (channel number 24). The three panels are: a) a perfect system and 0 delay; b) perfect system

with non-zero delay (the delay amount can be determined a priori according on how many fringes we want to have in the interferogram); c) system affected by polarization with polarization phase variation dependent on the wavelength, and non-zero delay. For comparison the case with no polarization effects is shown too, it is represented by the solid curve.

The main effect due to polarization that can be observed in these plots is a shift in the fringe position, i.e. the phase variation added to the stellar phase θ in eq.(73). This is an effect that has to be considered in case of astrometric measurements, but, per se, it does no harm on visibility measurements, which are of concern for IOTA. The calculated visibility loss is of the order of 0.85 %, too small to show up in these plots. Nevertheless this contribution has to be considered in the calculation of the overall visibility deterioration as it is shown in the section *Instrument visibility for IOTA*.

fig. BS.4 Visibility curve as a function of wavelength between 0.5 μm (channel 202) and 0.8 μm (channel 24). a) Perfect system and 0 delay; b) perfect system with non-zero delay; c) system affected by polarization phase variation dependent on the wavelength, and non-zero delay (dashed curve).

Thermal effects

0. Introduction

Il problema degli effetti della variazione dell'ambiente termico su un interferometro stellare riveste una notevole importanza, dal momento che anche minime variazioni di temperatura possono essere responsabili dell'introduzione di errori sul fronte d'onda in grado di alterare la figura interferometrica prodotta dallo strumento. Una conoscenza di questi effetti, e possibilmente delle caratteristiche termiche del sito dove lo strumento dovrá operare, é indispensabile al momento della progettazione, al fine di ridurre al minimo l'impatto dell'ambiente termico sulla sensibilitá dello strumento.

Nel caso particolare di IOTA tale tipo di analisi non é stata effettuata (ad eccezione dello studio del comportamento della componente in Neoceram presente nella struttura di supporto che collega primario e secondario, si veda oltre) al momento della progettazione. Pertanto i risultati dell'analisi che presento in questa sezione sono stati utilizzati per una valutazione a posteriori delle conseguenze sulla visibilitá delle frange interferometriche dovute alle variazioni termiche a cui IOTA é soggetto.

La prima parte del capitolo si occupa degli effetti sulle componenti ottiche vere e proprie del sistema: sia ottiche riflettenti che rifrattive. Ogni variazione termica produce un'alterazione delle caratteristiche fisiche dei materiali delle ottiche che si ripercuotono sulla qualitá del fronte d'onda che interagisce con esse, al momento della riflessione o della trasmissione.

La seconda parte invece prende in considerazione gli effetti sulle strutture di supporto dell'insieme specchio primario piú secondario. In questo caso la variazione termica puó causare una perturbazione nell'allineamento del telescopio e conseguentemente introdurre un errore sul fronte fronte d'onda.

1. Thermal effects affecting the optics of the system

The thermal sensitivity of an optical system can place tight demands on the thermal control system of an interferometer. In the following it is presented how thermal changes in the reflecting and transmitting optics of the system affect the wavefront, by introducing a variation in the optical path, δOPL, traveled by the interfering beams.

When the thermal-induced δOPL is the same across the whole beam, the main effect on the interference fringe consists of a shift in the position of the white-light point. This is of concern when considering astrometric measurements. If the δOPL is different in different areas within the beam section, the wavefront results distorted, and a wavefront error (in the following section it will be indicated by δw_z) is introduced that causes a reduction of visibility in the interference fringes. This is the case of particular concern for IOTA.

In section 1. the thermal problem is addressed in relation with the effects on the mirror substrates. In section 2. it is addressed in relation to the effect on the transmitting optics.

1.1 Substrates

A mirror shape directly affects the wavefront of a beam that reflects from it, and for a given surface perturbation the wavefront perturbation can be as much as twice large. This is because wavefront changes are given by the "round-trip" delay occurring at reflection; this delay is the surface change projected along the beam direction and actually depends on the angle of incidence.

Let us consider focusing mirrors, in particular paraboloids, which are typically used in afocal systems, as the beam reducer of IOTA for example. For these systems the separation between the focusing and recollimating elements must be precisely equal to the sum of their focal lengths. Small variations of these lengths result in a change of the wavefront curvature and hence of the optical path traveled by the beam.

In a paraboloidal mirror the focal length, f, and the curvature at the vertex of the mirror, $1/R$, are simply related by $1/f = 2/R$, so that if the focal length suffers a differential change Δf, the curvature will suffer a differential change

$$\Delta\left(\frac{1}{R}\right) = \frac{1}{2}\frac{\Delta f}{f^2},$$

and the wavefront sagitta, w_z (i.e. the height of the wavefront with respect to a reference plane, in the direction of the propagation of the wave), will suffer a variation δw_z given by

$$\delta w_z = \frac{\rho^2}{2}\Delta\left(\frac{1}{r}\right) \tag{75}$$

where ρ is the radius of the beam and $1/r$ is the wavefront curvature. Eventually, the wavefront error introduced when one focal length suffers a change Δf, is given by [35]

$$\delta w_z = \frac{\rho^2}{2}\cdot\frac{\Delta f}{f^2} \tag{76}$$

If the substrate of the mirror experiences a temperature change ΔT its linear dimension will suffer a fractional variation $\alpha\Delta T$, where α is the coefficient of thermal expansion for the substrate (it is assumed to be uniform). The focal length will be affected by the same fractional variation[36], so that eventually we get the expression of δw_z as a function of the

[35] In calculation performed for the IOTA telescope when operated in autocollimation mode, we found $\delta w_z = \frac{\Delta f}{4f_{\#}^2}$. For the normal operation we have to considered a variation in the wavefront sagitta $\frac{1}{2}\delta w_z$, from which we obtain $\delta w_z = \frac{\Delta f}{8f^2}\cdot 4\rho^2 = \frac{\rho^2}{2}\cdot\frac{\Delta f}{f^2}$

[36] A fractional linear variation, $\frac{\Delta z}{z} = \alpha\Delta T$, of the surface implies a change in the surface curvature, hence a fractional variation of the focal length:

$$\frac{\Delta z}{z} = \frac{\Delta R}{R} = \frac{\Delta f}{f} = \alpha\Delta T$$

.

temperature change

$$\delta w_z = \frac{\rho^2}{2} \cdot \frac{\alpha \Delta T}{f} \tag{77}$$

As far as a beam compressor is concerned, it is evident that a change in temperature affecting only the secondary mirror implies a smaller error than that due to the primary. Assuming that ρ is the height of the beam at the rim of the mirror, eq.(77) can be written as

$$\delta w_z = \frac{1}{8 f_{\#}^2} \cdot f \alpha \Delta T \tag{78}$$

where $f_{\#}$ is the focal ratio of the system and f is the focal length of the mirror that experiences the temperature variation. If m is the magnification factor of the system ($m = f_P / f_S$), then δw_z due to the primary mirror is m times larger than that due to the secondary.

If both the mirrors (identical substrates are assumed) experience the same ΔT, then the overall focal length variation is given by

$$\Delta f = f_P \left(\frac{m-1}{m} \right) \cdot \alpha \Delta T$$

and eventually the beam travelling through the system is affected by a wavefront error

$$\delta w_z = \frac{1}{8 f_{\#}^2} \cdot f_P \left(\frac{m-1}{m} \right) \cdot \alpha \Delta T \tag{79}$$

If the mirror substrate acquires a thermal gradient $\frac{dT}{dz} = \gamma_\perp$[37], the curvature will change by[38]

$$\Delta \left(\frac{1}{R} \right) = \frac{\Delta R}{R^2} = \frac{\alpha R \Delta T}{R^2} = \alpha \gamma$$

[37] When considering a thermal gradient affecting an optical surface we make a distinction between a thermal gradient perpendicular to the optical surface, indicated by $\gamma_\perp$, and a thermal gradient parallel to the optical surface, indicated by $\gamma_\parallel$.

[38] If the thermal gradient $\gamma = \frac{dT}{z}$, calculated over the distance z, causes a variation Δz along z, because of the homology relation, $\frac{\Delta z}{z} = \frac{\Delta R}{R}$, R undergoes a variation ΔR that can be expressed as due to the thermal gradient $\frac{\Delta T}{R}$ calculated over the distance R. We have

$$\frac{\Delta z}{z} = \frac{dT}{z} \alpha$$

$$\frac{\Delta z}{z} = \frac{\Delta R}{R} = \frac{\Delta T}{R} \alpha$$

so that $\gamma_\perp = \frac{dT}{z} = \frac{\Delta T}{R}$.

and eventually the wavefront sagitta variation will be given by

$$\delta w_z = \rho^2 \alpha \gamma \tag{80}$$

where ρ is the height of the beam at the mirror.

1.2 Transmitting optics

A ray passing obliquely through a plane parallel plate of refractive material emerges parallel to its original path but undergoes a lateral shift s, to which is related a variation in the optical path length (OPL). Reference to fig. T.1 shows that

$$\delta OPL = \left| \frac{d}{cos\theta} - \frac{nd}{cos\theta'} \right| \tag{81}$$

where d is the thickness of the plate, θ is the angle of incidence, n is the refractive index of the plate relative to the surrounding medium. Using Snell's relation $nsin\theta' = sin\theta$, we get

$$\delta OPL = \frac{d}{cos\theta} \left(1 - \frac{n^2 cos\theta}{\sqrt{n^2 - sin^2\theta}} \right) \tag{82}$$

For interferometric purposes, this variation in the OPL has to be the same both for the beams that are to interfere considered before the beamsplitter, and for the reflected and the refracted ray after the beam splitter.

fig. T.1 Variation in OPL for a beam travelling through a plane parallel plane of thickness d and refractive index n.

Assuming incidence angle $\theta \ll 1^{\circ}$[39] the path difference is simply expressed by

[39]This assumption eliminates the dependence of δOPL from the incidence angle, which could be ques-

$$\delta OPL = d(n-1) \tag{83}$$

where n is a function of the wavelength ($n(\lambda) > 1$).[40] In the following two sections we will address separately the effects of thermal changes for the windows and for the beamsplitter.

1.2.1 Windows: δOPL due to absolute temperature variation ΔT and thermal gradients

1.2.1.1 Absolute temperature variation

Any temperature variation will affect δOPL both for the induced change in d by thermal expansion and for the effect on the refractive index. Differentiating eq.(83) with respect to the temperature gives

$$\frac{d\delta OPL}{dT} = d \cdot [\alpha(n(\lambda) - 1) + \beta(\lambda)] \tag{84}$$

where α is the coefficient of linear expansion and β is the coefficient of temperature variation of the refraction index. In general we can express the variation in OPL as a function of d, D, and ΔT

$$\delta OPL = d \cdot D(\lambda) \cdot \Delta T, \tag{85}$$

where D is the term in square brackets calculated, for a λ in the middle of the wavelength range, and ΔT is the temperature variation of interest.

1.2.1.2 Thermal gradient

If a temperature difference arises at the two parallel surfaces of the window, a transverse thermal gradient arises in the transverse direction across the window. In a steady-state condition, the thermal gradient will be simply given by

$$\gamma = \frac{\Delta T}{d_W}$$

where d_W is the thickness of the window. The δOPL due to the thermal gradient will be given by[41]

$$\delta OPL = d^2 \cdot D \cdot \gamma. \tag{86}$$

tionable. However, for many applications it is preferable to have incidence on refractive optics as close to normal as possible, so that this assumption is generally accepted

[40] For the evaluation of the instrumental visibility it will be of interest to consider the error in the plate thickness, Δd. This introduces an additional δOPL proportional to Δd and dependent on the wavelength.

[41] Differentianting eq.(85) with respect to x gives

$$d(\delta OPL) = d \cdot D \cdot \gamma \cdot dx$$

and integration of $d(\delta OPL)$ between 0 and d gives eq.(86).

In the case of temperature variation depending on time, the thermal gradient γ in eq.(86) is obtained by applying the Fourier's law of heat conduction using the expression of the heat transfer rate across the window, and is given by

$$\gamma = \frac{\rho_W \cdot c_W \cdot d_W}{k_W} \cdot \frac{dT}{dt} \tag{87}$$

where ρ_W, c_W, and k_W are the density, specific heat, and coefficient of thermal conduction of the window material.

However, for a more general evaluation we should consider the thermal change affecting the frame of the window and the surrounding structures as well. Because of the small coefficient of thermal conductivity of the material used for the optics, a thermal change occurring at its boundary will gradually affect the whole component, with the main result of creating an internal radial temperature gradient. Because of this, and according to eq.(85), rays transmitted through different areas of the plate will be affected by different δOPL, and this will cause a curvature of the wavefront. For each annular area at a distance r_i from the center of the optic the thermal gradient causes an error in OPL given by

$$\delta_i OPL = d \cdot D \cdot \gamma r_i$$

where γ is the radial thermal gradient, so the wavefront error will be given by

$$\delta w_z = d \cdot D \cdot \gamma r \tag{88}$$

where r is the radius of the beam.

In order to estimate the magnitude of the induced thermal gradient, we have to take into account the presence of the metallic frame that supports the optic and that is the part directly exposed to the thermal variations.

First, let consider a steady-state condition. Assuming perfect thermal contact between frame and external environment and between frame and glass, the heat transfer rate at the boundary between glass and frame is, applying the Fourier's law for heat conduction (*Poulikakos, 1994*)

$$\begin{aligned} \dot{Q}_W &= \dot{Q}_F \\ k_W \cdot (2\pi r_W L) \cdot \frac{dT}{dr} &= \frac{2\pi k_F L \cdot (T_{in} - T_{out})}{ln\frac{r_W + t_F}{r_W}} \end{aligned} \tag{89}$$

where k_W and k_F are the coefficient of thermal conductivity for the window material and the frame material, r_W is the window radius, t_F is the frame thickness, T_{in} is the temperature of glass+frame before the thermal change (which is also the temperature of the inner part of the window at the moment of the thermal change), T_{out} is the temperature

at the external frame boundary, so that

$$T_{out} - T_{in} = \Delta T$$

is the external temperature variation affecting the whole system. From the previous equation we find that the radial thermal gradient within the window is given by

$$\gamma = \frac{dT}{dr} = \frac{k_F}{k_W} \cdot \Delta T \cdot \frac{1}{r_W} \cdot ln^{-1}\left(\frac{r_W}{r_W + t_F}\right). \tag{90}$$

Let consider now the condition in which the external temperature varies with time (which is the case we are interested in when considering the effect of the diurnal temperature variation). The heat transfer rate for the frame is given by

$$\dot{Q}_F = M_F c_F \cdot \frac{dT}{dt}$$

where M_F is the mass of the frame and c_F is the frame specific heat. Equating the heat rates at the glass-frame boundary[42], and solving for the thermal gradient within the window, we get

$$\gamma = \frac{M_F \cdot c_F}{k_W \cdot 2\pi r_W L} \cdot \frac{dT}{dt} \tag{91}$$

1.2.2 Beamsplitter: δOPD[43] due to absolute temperature variation ΔT and thermal gradients

1.2.2.1 Absolute temperature variation

In the case of the beamsplitter we start assuming that there are no local temperature differences within it, so that it is to be considered only the effect of an absolute temperature variation when there is a lack of symmetry in its linear dimensions.[44] Starting from

[42]This implies to assume that the heat flows across the frame much quicker than across the glass, so that we assume that the heat transferred by the frame is accumulated at the window borders, and that this induces an internal thermal gradient across the window.

[43]When considering optical path variation occurring at the beamsplitter, we introduce the quantity OPD instead of OPL. OPD represents the difference in the OPL travelled by the two beams.

[44]Since both interfering beams pass across the same optical element, they should experience the same δOPD, if something causes an alteration of the beamsplitter length. The whole effect would be a shift in the position of the interferometric fringe, and, in principle, the visibility does not suffer any degradation. However, because δOPD depends on the wavelength, by means of $D(\lambda)$, for large ΔT we might expect the visibility to be affected too, as a function of the mean error on the optical path given by

$$\delta OPD_{ave} = d \cdot \Delta T \cdot \left(\frac{D_{max} - D_{min}}{2}\right).$$

For example, in the case of BK 7, the material used as substrate for the IOTA beamsplitter, D is dominated by the thermal coefficient of refraction. Its variation, for a wavelength range from 400 to 1000 nm is of about $10^{-6}\,{}^{\circ}C^{-1}$ *(Musikant, 1985).*

a symmetric design (i.e. a sandwich beamsplitter), let us consider the relative thickness difference (between the right and left sides with respect to the partial reflecting coating) Δd.

As explained in the section *Beamsplitter*, and with reference to fig. BS.1, on the "+" side the reflected beam will suffer twice and the transmitted beam once the phase shift due to Δd; on the "−" side, the reflected beam does not suffer any phase change while the transmitted crosses Δd only once. On both sides a phase change $\propto \Delta d$ is introduced, which is what causes a systematic error in the determination of the stellar phase.[45]

When facing the problem of the design of a beamsplitter, the substrate should be chosen according to its minimum sensitivity to environmental conditions. In this case, given Δd and a maximum ΔT, the accuracy goal in the determination of the star position sets the constrains on the parameter D. In formula

$$D \leq \frac{\delta OPD}{\Delta d \cdot \Delta T}.$$

Viceversa, the selection of a given substrate will set the constrains for the maximum temperature variation at the beamsplitter, i.e. sets the constrains for the temperature control system of the interferometer.

[45] The stellar phase is given by

$$\frac{2\pi}{\lambda}\mathbf{B}\cdot\mathbf{s}$$

where $\mathbf{B}\cdot\mathbf{s} = B\,sin\alpha \simeq B\cdot\alpha = OPD$ between the interfering beams. The extra δOPD due to Δd is a cause of limited accuracy in the determination of the angular star position, α.

fig. T.2 δOPD for a sandwich beamsplitter affected by a thermal gradient $\gamma_\perp$.

1.2.2.2 Thermal gradient

Consider now the effect of a thermal gradient $\gamma_\perp$ across the beam splitter. With reference to fig. T.2 we see that, because the left and right parts of the sandwich, which were initially at the same temperature, are now at different temperatures, a δOPD is introduced between beam 1 and beam 2. In fact the travel (once) the same distance $d/2$ on two opposite side of the beamsplitter, which are at temperatures which differ, in average, by $\Delta T/2$, being ΔT the temperature difference at the two extremes of the beamsplitter. For each corresponding dx_L and dx_R, which underwent the temperature change, the average $\Delta T_{LR} = \Delta T/2$, will cause a $d(\delta OPD)$ between the two beams given by

$$d(\delta OPD) = D \cdot \frac{\Delta T}{2} \cdot dx = D \cdot \frac{\gamma_\perp d}{2} \cdot dx \tag{92}$$

so that the total δOPD after travelling $d/2$ is

$$\delta OPD = \int_0^{d/2} \frac{d}{2} \cdot D \cdot \gamma_\perp \cdot dx = \frac{d^2}{4} \cdot D \cdot \gamma_\perp \tag{93}$$

The same arguments are valid if the beamsplitter consists of a substrate plate and a separate compensator. However, the effect due to the thermal gradient may be larger, depending on the linear separation between substrate plate and compensator. With reference to fig. T.3 we see that the difference in temperature[46] between dx_L and dx_R is now given by

[46] assuming that $\gamma_\perp$ is the same for both the beamsplitter substrate and the air gap between beamsplitter and compensator

$$T_L - T_R = \frac{dT}{dx}(d+l) = \gamma_\perp \cdot (d+l)$$

where l is the distance between plate and compensator. Integrating on the whole thickness d gives

$$\delta OPD = \int_0^d (d+l) \cdot D \cdot \gamma_\perp \cdot dx = d^2 \cdot D \cdot \gamma_\perp + d \cdot l \cdot D \cdot \gamma_\perp \tag{94}$$

As in the case for the window we consider also the radial thermal gradient induced by the thermal exchange between the beamsplitter substrate and the metallic frame. In this case, rays transmitted through different annular areas of the beamsplitter would suffer different δOPD. This will cause a wavefront error which is calculated according to eq.(88), where d is now the beamsplitter thickness and r is the radius of the beam.

fig. T.3 δOPD for a (plate+compensator) beamsplitter affected by a thermal gradient $\gamma_\perp$.

1.3 Some calculations for the IOTA components

In the previous sections it has been shown how thermal effects can introduce errors in the optical paths of the interfering beams, which eventually lead to a deterioration of the interferogram visibility, or in the case of astrometric measurement, of loss of accuracy in the determination of the object position.

In this section we will focus on the optical elements of IOTA and of the effects on the optical paths due to the typical thermal variations at the IOTA site.

1.3.1 Thermal conditions at the IOTA site

The source for the information about the weather, and specifically the temperature pro-

file, at the IOTA site, is the *SAO Special Report 345*. This is a quite old compilation of weather data collected at Mt. Hopkins from 1968 to 1971 (apparently there is no record of other compilations like this made in more recent years). However, due to the long period of data recording, their statistical information should be accurate enough for the purpose of our evaluation.

Temperature measurements were taken hourly and the estimated measurement accuracy was $\pm 1^{\circ}F$. Within the weather pattern description, specific information on the temperature range is given according to the different seasons of the year. For our calculation we are mainly interested in average temperature variations as a function of time, both during the night and in the hours immediately before the time of starting observations. For the latter we consider the period of time which goes from sunset to the time of the astronomical twilight (namely when the sun is 18° below the horizon).

Making use of *fig.3, Monthly average hourly temperature at knoll # 2, 1968 to 1971* of the SAO report, and the average sunset, twilight and sunrise time calculated for the year 1995, I extracted the information reported in tab. T.1. I refer to *Appendix G* for a more complete information.

tab. T.1 Seasonal temperature profile at the IOTA site: average value at sunset (SS), twilight (TWL), and sunrise (SR) in degree C; ΔT between sunset and twilight in degree K, and ΔT per hour during the night in degree K per hour.

Season	sunset	twilight	sunrise	ΔT SS to TWL	ΔT rate SS to SR
	$^{\circ}C$	$^{\circ}C$	$^{\circ}C$	K	Kh^{-1}
Winter	22.23	21.5	19.44	0.73	0.21
Spring	25.55	24.44	20.83	1.11	0.42
Summer	38.33	36.11	30.55	2.22	0.74
Fall	27.77	26.94	25.33	0.83	0.20

The average hourly temperature variations show a quite large range of values, from the lowest during the fall to the highest in the summer. For the following estimation we will take into account the summer rates because they reflect in particular the temperature trade in the month of June (which results, in average, to be the hottest month). This is, of the summer months, probably the only one during which we can expect to have almost

full time of observation (because of the storm season starting in July).

1.3.2 Mirror substrates

1.3.2.1 Absolute ΔT

Both mirrors of the IOTA telescopes are in zerodur, and for their batches the coefficient of thermal expansion was measured to be $\alpha = 0.007 \cdot 10^{-6}\ ^{\circ}C^{-1}$. Considering a mean temperature variation, for the hottest month, of about 8 K we get a relative despace between the two foci $\Delta f = 0.57\ \mu m$ which causes a δw_z of $1.12 \cdot 10^{-2} \mu m$ that implies a Strehl factor[47] of $S = 0.982$ calculated over the whole night. In tab. T.2 it is presented the despace error, Δf, the wavefront error, δw_z, and the corresponding Strehl ratio for the average temperature variation during the whole night, in one hour, and in the period between sunset and twilight.

tab. T.2 Despace error, Δf, wavefront error, δw_z, and Strehl ratio relative to the average temperature variation during the whole night (a), one hour (b), and the period between sunset and twilight (c).

	ΔT K	Δf μm	δw_z μm	S
(a)	$\simeq 8$	0.57	0.0112	0.982
(b)	$\simeq 0.74$	0.05	0.0010	0.999
(c)	$\simeq 2.2$	0.16	0.0032	0.998

1.3.2.2 Thermal gradient

A thermal gradient may arise because of the contact of the zerodur mirror support and substrate with the steel screw which is used to control the position of the primary.

Applying the Fourier's law of thermal conduction to the surface of contact between the two materials, we get the expression for the thermal gradient induced within the zerodur structure

$$\gamma_{\perp} = \frac{M_S \cdot c_S \cdot \frac{\Delta T}{\Delta t}}{k_Z \cdot A}$$

where M_S and c_S are the screw mass and specific heat, k_Z is the thermal conductivity of zerodur ($1.64\ W\,cm^{-1}\,K^{-1}$) and A is the area of the contact surface. For the average

[47] The Strehl ratio is defined by

$$S = e^{-(\frac{2\pi}{\lambda}\sigma_w)^2}$$

where σ_w is the wavefront *rms*. This quantity is used as an estimator of the visibility loss due to the wavefront error. see *section 0.1* in *Instrument visibility for IOTA*

temperature change in one hour of observation the induced thermal gradient is of the order of $0.03\,mK\ per\ cm$ and the associated δw_z is negligible. This result is mainly due to the very low value for the coefficient of thermal expansion of zerodur.

However, in case a temperature variation of the order of $0.75\ K$ showed up at the contact surface, the mean thermal gradient would be given by $\frac{\Delta T}{d}$ and δw_z would be of the order of $0.033\ \mu m$ which would cause a visibility reduction of about 13 %.

1.3.3 Windows

1.3.3.1 Absolute ΔT

For IOTA the windows are made of fused silica[48] for which D calculated for $\lambda = 0.5461\ \mu m$ is about $13 \cdot 10^{-6}\ {}^{\circ}C^{-1}$, and the window thickness is $d_W = 1.27\,cm$. If an absolute variation of the temperature of the windows occurs, the effect on OPL are quite large, and given by

$$\delta OPL = 0.165\ \mu m\ K^{-1}.$$

For temperature changes less than $10^{-1}\ K$, which are out of our control (above all for the external window), we may expected the position of the withe-light point to change by about $0.02\ \mu m$, which needs to be compensated by moving the fine delay line.

However, a more likely situation is probably that in which the thermal behaviour of the window depends on the presence of internal temperature gradients, for example because the window interacts with the metallic frame which is directly affected by the diurnal temperature variation.

1.3.3.2 Thermal gradient

Let assume that a thermal change in time creates a temperature difference between the two parallel surfaces of the window. Applying the Fourier's law to the window the internal thermal gradient is given by

$$\gamma_{\perp} = \frac{\rho_W \cdot c_W \cdot d_W}{k_W} \left(\frac{\Delta T}{\Delta t} \right).$$

In tab. T.3 are presented the result obtained for the thermal gradient and the δOPL when the mean temperature variation during the night and in the period between sunset and twilight are considered.

[48] See *Appendix H*

tab. T.3 Transverse thermal gradient, $\gamma_\perp$, and δOPL for mean temperature changes during the night and in the period between sunset and twilight.

$\frac{\Delta T}{\Delta t}$ $K\,s^{-1}$	$\gamma_\perp$ $K\,cm^{-1}$	δOPL μm
$2 \cdot 10^{-4}$	0.03	0.006
$4 \cdot 10^{-4}$	0.06	0.012

Let us now consider the radial[49] thermal gradient due to the interaction with the aluminum frame. In the case of an absolute ΔT arises, $\gamma_\parallel$ is calculated according to the expression valid for the steady-state condition (eq.(90)). It results to be $\gamma_\parallel \simeq 159 \cdot \Delta T\ K\,cm^{-1}cm$ which, for $\Delta T = 1\ mK$ implies $\gamma_\parallel \simeq 0.16\ K\,cm^{-1}$. This value of thermal gradient would imply an rms wavefront error across the beam[50] of the order of 0.04 μm and a loss of visibility of about 20%.

However, in general we would expected the thermal change to be due to the diurnal temperature variation rate. In this case the thermal gradient has to be calculated according to eq.(91) and we get

$$\gamma_\parallel = 262 \cdot \frac{\Delta T}{\Delta t}\ \frac{K}{cm}.$$

In tab. T.4 it is shown the value for the radial thermal gradient, δw_z across the beam, and the Strehl ratio, for the temperature rate associate with the mean temperature variation during the night, (a), and the mean temperature variation in the period between sunset and sunrise, (b).

tab. T.4 Radial thermal gradient, $\gamma_\parallel$, δw_z across the beam, and Strehl ratio, for the temperature rate due to the mean temperature change during the night, (a), and in the period between sunset and twilight, (b)

	$\frac{\Delta T}{\Delta t}$ $K\,s^{-1}$	$\gamma_\parallel$ $K\,cm^{-1}$	δw_z μm	S
(a)	$2 \cdot 10^{-4}$	0.05	0.018	0.980
(b)	$4 \cdot 10^{-4}$	0.10	0.037	0.910

[49] According to the definition at the beginning of the section, this thermal gradient is indicated by $\gamma_\parallel$.

[50] The rms error is calculated as

$$\frac{\delta w_{z,max}}{\sqrt{2}}$$

where $\delta w_{z,max}$ is the maximum variation of the optical path across the beam, calculated according to eq.(88) with $r = r_{beam}$.

The effect, in both cases, is quite large. However we have to consider that this result applies when assuming that the heat transferred by the frame in the time dt is accumulated at the window borders. This holds because the coefficient of thermal conduction of the frame is quite larger than that of the window substrate, but also if we assume that the heat transfer is calculated over short periods of time (because for long periods we should consider that heat flows across the substrate, too). Besides, when considering the effects of the thermal variation during an integration time (10 ms), which is the case of real interest, the induced thermal gradient is 100 times smaller than that given in tab. T.4 and the resulting Strehl ratio is $\simeq 1$.

1.3.4 Beamsplitter

1.3.4.1 Absolute ΔT

The beamsplitter used on the visible table of IOTA is made of BK 7[51], for which the quantity D equals $6.68 \cdot 10^{-6}\,/K$. The substrate thickness is 1.4 cm, while the manufacturing error for the substrate is of the order of 7.62 μm. For this value of Δd, the δOPD due to an absolute thermal variation ΔT is of $5.1 \cdot 10^{-5} \mu m\ per\ 1\ K$, a negligible error for our measurements. For a temperature change involving only either the beamsplitter or the compensator plate, eq.(85) has to be use where d is the plate thickness. The corresponding ΔOPD between the two interfering beams would be of the order of 0.094 $\mu m\ per\ 1\ K$.

1.3.4.2 Thermal gradient

The beam combiner used for IOTA consists of the beamsplitter in itself plus a compensator plate. For these calculation we hence make use of eq.(94). In the case of a transverse thermal gradient, being the distance in between the beamsplitter and the compensator plate of the order of 1.7 cm, we get a δOPD of the order of 0.3 $\mu m\ (K/cm)^{-1}$. For a temperature variation of $10^{-4}\ K$ (i.e. of the order of the diurnal temperature variation, per second, experienced by the external windows) arisen on one side of the system beamsplitter+compensator, the corresponding δOPD, that is the change in the position of the white-light point, would be of the order of $9.6 \cdot 10^{-6}\ \mu m$.

Also for the beamsplitter the more likely situation is that in which a radial thermal gradient is induced by mean of the interaction of the substrate with the metallic frame. In the laboratory we will considered only the steady state condition for the temperature change, hence the value of the induced thermal gradient is calculated using eq.(90). The coefficient of thermal conductivity of BK 7 is very small, 0.01114 $W\,cm^{-1}\,K^{-1}$, so that the choice of the material of the frame is very important in determining the magnitude of the induced radial thermal gradient.

For an aluminum frame ($K_{Al} = 2.4\ W\,cm^{-1}\,K^{-1}$) 1.27 cm thick, and a beam splitter radius of 3.8 cm, we get $\gamma \simeq 196 \cdot \Delta T\ K\,cm^{-1}$ that for ΔT of the order of $10^{-4}\ K$ would imply $S = 0.998$ and for ΔT of the order of $10^{-3}\ K$ would imply $S = 0.890$. This is a large effect if we cannot control the temperature in the laboratory at the $10^{-4}\ K$ level.

[51] See *Appendix H*

However, we have to remember that for this calculation we assume a supporting frame consisting in a continuous metallic ring, as it was the case for the windows. In reality, for the IOTA beamsplitter, the support consists of three separated aluminum bolts whose total contact area with the beamsplitter substrate is smaller than that of a continuous ring so that we would expect a smaller value for the induced thermal gradient.

For a steel frame the situation would be quite better, being the coefficient of thermal conductivity equal to $0.59\ W\,cm^{-1}\,K^{-1}$. The corresponding thermal gradient is of the order of $48 \cdot \Delta T\ K\,cm^{-1}$, so that for $\Delta T = 10^{-4}\ K$ we have $S = 0.999$, while for $\Delta T = 10^{-3}\ K$ we have $S = 0.993$.

2. Thermal effects as causes of telescope misalignments

Within the investigation of the causes for the displacement of the telescope optics, we consider the effect of thermal variations affecting the components of the telescope. Specific calculation are carried out with reference to the IOTA telescope specifications and environmental conditions.

A common practise is to align the telescope optics operating the telescope in autocollimation mode, so that the light source is within the laboratory and it is not necessary to open the dome to use starlight. The environment with the dome closed is quite different from the outside one so that, when the dome is opened to start the observation, we immediately observe a deviation in the position of the beam as well as a distortion of the interferogram. This is because both the optical components of the telescopes and the wavefront itself adjust to the new conditions. Here we will analyze mainly the effects of temperature changes on the optical components, and only briefly the effects induced on the wavefront by air-temperature gradients.

2.1 Thermal effects on the secondary mirror: a description of the secondary assembly

Because of temperature change since the period when the telescope was aligned, we may expect the telescope to undergo some departures from collimation. In order to have an estimation of this loss of collimation, here we address the effect of thermal changes on the structure that supports the secondary mirror, hereafter called the secondary assembly.

The secondary assembly is supported by a heavy perforated tube cantilevered out from the primary plate. The assembly consists of three spider vanes connected to a hub that supports a compact three-axis stage to which the secondary mirror is mounted. The outer ends of the spider are not fastened directly to the support tube, but are connected to it by flexures that permit movement along the optical axis, but give stiff radial support. The ends of the spider are spring-loaded axially against rods of low-thermal-expansion material (Neoceram[52], made by Nippon Electric Co), whose other ends are supported at the primary flange. These rods, which have a small negative coefficient of thermal

[52]See *Appendix H*

expansion, are combined with metal parts (both stainless steel and aluminum) to make a very-low-expansion axial path between primary and secondary. The combination of this with rigid, nearly-symmetrical radial support, allows the telescopes to remain in collimation when the temperature changes. The purpose of the following paragraphs is to present an evaluation of the thermal effects on some of the components of the secondary assembly.

2.2 Telescope misalignment due to the secondary assembly response to thermal changes

As far as the secondary assembly response to temperature changes is concerned, two distinct components have to be considered: the whole supporting structure (hereafter indicated as the spacer, one for each spider, see also fig. T.7) that links the three single spiders to the primary mirror by screws, rods, and clamps; and the three spiders themselves.

It is evident that an accurate analysis would require consideration of the interaction between spiders and spacers in the presence of a thermal gradient across the transverse section of the primary-secondary system.

However for the estimation of our interest (i.e. the order of magnitude of the wavefront error induced by thermal effects), it will be sufficient to analyze the case for the spiders and the spacers independently and then combine their effects for some particular conditions.

First, we consider the spiders. If they all are affected by the same temperature variation we would expect their lengths to vary by the same amount and the position of the secondary should suffer neither transverse nor angular variation. If a thermal gradient exists between upper and lower spiders, the secondary will be eventually affected by some misalignments.

Later, we consider a temperature change affecting the spacers, and the resulting motion of the spider-end. There will be effects on the secondary alignment both when the three spacers suffer the same thermal change and in the case of a thermal gradient between upper and lower spacers.

For the situation in which a thermal gradient is present, a combination of the spacer and spider effects will be proposed.

2.2.1 Spiders' thermal expansion

With reference to fig. T.4, suppose that the upper spiders are cooler than the lower spider. If the telescope was aligned correctly when ΔT was zero, then for $\Delta T \neq 0$ there will be a shift upwards of the secondary, along the axis of the temperature gradient. This will cause the beam to deflect downward in the telescope and in the laboratory.

According to the law of the thermal linear expansion, for a temperature difference ΔT, the variation Δl in the length of the cooler spider is given by

$$\Delta l_{spider} = \alpha \cdot l_{spider} \cdot \Delta T \tag{95}$$

where α is the coefficient of linear expansion, and l_{spider} is the spider length. Typical values for the coefficient of linear expansion for materials which could be used in spider manufacturing are given in tab. T.5.

tab. T.5 Coefficient of linear expansion for materials typically used in spider manufacturing.

Material	α $\left(10^{-6}\ {}^{\circ}C^{-1}\right)$
aluminum	23.2
steel	12.7
stainless steel	$5.5 \div 6.2$
invar	0.9

fig. T.4 Secondary spiders affected by thermal gradient

In the laboratory the beam would be shifted by

$$\theta_{lab} = \frac{\Delta x}{f_S} = \frac{\Delta l_{spider}}{f_S}$$

which is of the order of 14 μrad per $1\,K\,\Delta T$, which could be generated by a mean thermal gradient within the spider of the order of 0.04 Kcm^{-1}, for the IOTA telescopes.

2.2.2 Spider-end axial motion due to thermal changes

With reference to fig. T.5 we define l_{spider} the length of the spider, l_{spacer}, the axial distance between the primary plane and the annular support of the spiders, l_{stage}, the length of the stage that is attached to the secondary and to which the three spiders are connected. From the fig. T.5 we see that

$$l_{spacer} \; = \; f_P \; - \; f_S \; + \; l_{stage} \; = \; 0.9 f_P \; + \; l_{stage}.$$

We start by considering the effects on the secondary position when one spacer suffer a temperature change ΔT.

fig. T.5 Change in the position of the secondary mirror when one of the spacers is affected by temperature variation

Let us assume, for example, that the bottom spider spacer shrinks and that the spider-end moves by z_{spider}, causing the secondary to tilt CCW, translate axially toward the primary, and translate radially away from the spacer, i.e., upwards. Consequently the secondary vertex will be affected by three coordinate changes in θ, z, and x. Let θ_v be the angular motion of the normal to the surface at the vertex. From fig. T.5 we get

$$\begin{aligned} tan\theta_v &= \frac{z_{spider}}{1.5 \cdot l_{spider}} \\ \theta_v &\simeq \frac{z_{spider}}{1.5 \cdot l_{spider}}, \\ z_v &= \frac{1}{2} l_{spider} \cdot tan\theta_v - l_{stage} \cdot (1 - cos\theta_v) + k x_v^2 \end{aligned} \tag{96}$$

$$
\begin{aligned}
z_v &\simeq \frac{1}{2} l_{spider} \cdot \theta_v \\
z_v &\simeq \frac{1}{3} \cdot z_{spider} \qquad (97)
\end{aligned}
$$

$$
\begin{aligned}
x_v &= \frac{1}{2} l_{spider} \cdot (1 - cos\theta_v) + l_{stage} \cdot sin\theta_v \\
x_v &\simeq l_{stage} \cdot \theta_v \\
x_v &\simeq \left(\frac{l_{stage}}{1.5 \cdot l_{spider}} \right) \cdot z_{spider}. \qquad (98)
\end{aligned}
$$

The radial distance from the vertex of a secondary tilted by θ_S to the "normal point", the point which reflects an incident plane wave from the laboratory back on itself, is given by x_N where

$$x_N \; = \; 2 \cdot f_S \cdot \theta_S \; = \; 2 \cdot f_S \cdot \theta_v$$

because in this case $\theta_S \; = \; \theta_v$. Since $2 \cdot f_S \; > \; l_{stage}$ from the second of eqq.(98) we see that $x_N \; > \; x_v$, which means that an additional translation $(x_N \; - \; x_v)$ is necessary to get the reflected beam back on itself in the laboratory (see fig. T.6).

fig. T.6 Radial distance between the vertex of the tilted secondary to the normal point.

From fig. T.6 we see that $x_N \; > \; x_v$ and the incident chief ray is reflected upward, by an angle θ_c which is essentially the same angle that would result from a downward motion of the secondary by the distance $x_{Nv} \equiv x_N - x_v$ (it is exactly true for a spherical secondary).

The additional translation gives

$$x_{Nv} \equiv x_N - x_v = (2 \cdot f_S - l_{stage}) \cdot \theta_v = \frac{(2 \cdot f_S - l_{stage}) \cdot z_{spider}}{1.5 l_{spider}} \tag{99}$$

The tilt of the beam returned to the laboratory will be, according to eq.(13) in the section *Telescope alignment*,

$$\theta_{lab} = \frac{m \cdot x_{Nv}}{f_P} = \frac{m}{f_P} \cdot \frac{(2 \cdot f_S - l_{stage})}{1.5 l_{spider}} \cdot z_{spider}. \tag{100}$$

The whole picture is then that, for the downward spacer shrinking, the secondary rotates CCW, which is equivalent to a shift downward of the secondary itself, so that the return beam in the laboratory results displaced upward. To compensate for this beam shift, it is necessary to shift the secondary upward by the distance x_{Nv} between the normal point and the actual position of the secondary vertex due to the rotation.

In case the two upper spacers are affected by the same ΔT variation with respect to the lower spacer, the secondary would rotate by the same amount in two opposite direction around the *x-axis*, so that it will suffer no variation in the angular position. Also the resulting vertex shift in the transverse direction, along the *y-axis*, is null. On the contrary, there will be no compensation for the axial motion of the secondary, and the vertex will be affected by (see fig. T.5)

$$z_v \simeq l_{spider} \cdot \theta_v = \frac{2}{3} \Delta z_{spider} = \frac{2}{3} \Delta l_{spacer}.$$

Finally, if the three spacers all suffer the same ΔT, the secondary will experience an axial motion which is given by the length variation of the spacer itself.

2.2.3 Thermal gradient: secondary affected by both spider and spacer length variation

Let us consider the simplified case where both the bottom spacer and the bottom spider undergo the same ΔT variation while the upper spacers and spiders do not. Because of the shrinking of the spider, the shrinking of the spacer will cause a larger angular motion and consequently the secondary vertex will experience also a large variation in the transverse and axial position. In short the vertex coordinates change by

$$\theta_v \simeq \frac{z_{spider}}{(1.5 + \alpha \Delta T) \cdot l_{spider}} \simeq \frac{z_{spider}}{1.5 \cdot l_{spider}} \tag{101}$$

$$z_v \simeq \frac{1}{2 \cdot (1.5 + \alpha \Delta T)} \cdot z_{spider} \simeq \frac{1}{3} \cdot z_{spider} \tag{102}$$

$$x_v \simeq \left(\frac{l_{stage}}{(1.5 + \alpha \Delta T) \cdot l_{spider}} \right) \cdot z_{spider} \simeq \frac{l_{stage}}{1.5 \cdot l_{spider}} \cdot z_{spider} \tag{103}$$

where it is taken into account the fact that $\alpha \sim 10^{-6}\ ^\circ C$.

2.3 Estimation of the uncompensated spacer length variation

In order to make an evaluation of the wavefront error induced by thermal variation, it is necessary, as it is shown by the previous equations, to estimate the spider-end axial motion. i.e. how much of the spacer length variation is not compensated.

With reference to fig. T.7, the uncompensated thermal path, Δl_{unc}, is calculated considering the thermal length variation of the components of the telescope assembly support from the back-end of the primary to the spider-end. Δz_{spacer} is calculated separately taking into account the components from the spider-end to the back-end of the secondary. In the following section, we will see how these quantities combine together in the expression for the wavefront error.

The linear dimensions of the assembly components are given in tab. T.6.

fig. T.7 Schematic of the primary-secondary support assembly

tab. T.6 Linear dimension of the telescope assembly components

			inch	*cm*
d_1	primary substrate and block		3.3	8.38
d_2	steel screw		1.125	2.85
d_3	primary support	(1.0 - 0.750) =	0.250	0.63
d_4	aluminum clamp		2.625	6.67
d_5	stainless steel rod		5.375	13.65
d_6	neoceram rod		42.5	107.95
d_7	aluminum screw		5.125	13.02
d_8	spider thickness		1.0	2.54
d_9	secondary stage support		0.45	1.14
d_{10}	secondary stage		3.05	7.75
d_{11}	secondary thickness		0.5	1.27

The spider-end motion is given by Δl_{unc}, that is

$$\Delta l_{unc} = (-d_1\alpha_Z - d_2\alpha_S - d_3\alpha_{SS} - d_4\alpha_{Al} + d_5\alpha_{SS} + d_6\alpha_{NC} + d_7\alpha_{Al}) \cdot \Delta T \; ; \qquad (104)$$

the overall variation due to the secondary stage and secondary mirror is given by

$$\Delta z_{stage} = (d_8\alpha_{SS} + d_9\alpha_{SS} + d_{10}\alpha_S + d_{11}\alpha_Z) \cdot \Delta T \qquad (105)$$

where Z stands for zerodur[53], S for steel, SS for stainless steel, Al for aluminum, and NC for neoceram.

2.4 Wavefront error

For the wavefront error estimation, we consider again the three cases of independent spiders and spacers effects, and combination of the effects of a thermal gradient.

2.4.1 Wavefront error due to spiders' thermal expansion

In *section 2.2.1* it is shown that the transverse motion of the secondary, Δx, is proportional to the thermal change suffered by one of the spiders; then, the related wavefront error is given by

$$w_{\Delta x} = \frac{\Delta x}{f_\#} = \frac{\alpha \cdot l_{spider} \cdot \Delta T}{f_\#} \qquad (106)$$

where ΔT is actually the temperature difference at the two ends of the spider, hence $w_{\Delta x}$ is proportional to γl^2_{spider}, and γ is the mean thermal gradient within the spider.

[53] See *Appendix H*

2.4.2 Wavefront error due to spider-end axial motion (uncompensated spacer length variation)

In order to calculate the wavefront error due to misalignment of the secondary for the axial motion of the spiders, we use eqq.(96), (97), and (98). The linear and angular variation to use in these equations are : $\Delta z = z_v$, $\Delta x = x_{Nv}$ (because it has to be considered the displacement of the vertex with respect to the position where the chief ray would be reflected exactly on itself, and not simply the displacement due to the shrinking of the spacer), $\Delta\theta_x = \theta_v$, and we will consider the series expansion of θ_v up to the third power in z_{spider}, so that we have:

$$w_{\Delta z} = \frac{\Delta z}{8 \cdot f_\#^2} \simeq \frac{1}{8 \cdot f_\#^2} \cdot \frac{1}{2} l_{spider} \left[\frac{z_{spider}}{1.5 l_{spider}} - \frac{1}{3} \left(\frac{z_{spider}}{1.5 l_{spider}} \right)^3 \right] \tag{107}$$

$$w_{\Delta x} = \frac{\Delta x}{f_\#} \simeq \frac{l_{stage}}{f_\#} \cdot \left[\frac{z_{spider}}{1.5 l_{spider}} - \frac{1}{3} \left(\frac{z_{spider}}{1.5 l_{spider}} \right)^3 \right] \tag{108}$$

$$w_{\Delta\theta} = \frac{2 \cdot f_S}{f_\#} \cdot \Delta\theta_x \simeq \frac{2 \cdot f_S}{f_\#} \left[\frac{z_{spider}}{1.5 l_{spider}} - \frac{1}{3} \left(\frac{z_{spider}}{1.5 l_{spider}} \right)^3 \right]. \tag{109}$$

The wavefront error associated with the linear term in z_{spider}, which corresponds to a tilt of the plane wavefront, is actually removed by both adjusting the siderostat and the action of the piezo-mirror. The wavefront error left is that due to the term in z_{spider} to the third power. For a general wavefront error calculation, in any of the following cases, we assume that the errors add quadratically.

If only one spacer is affected by thermal change the squared mean error on the wavefront is given by

$$w^2 = w_{\Delta x}^2 + w_{\Delta z}^2 + w_{\Delta\theta}^2 \tag{110}$$

where w_i^2 are given by eqq.(107), (108), (109) after removing the linear term.

If all three spacers are affected by the same ΔT, it is $z_v = z_{spider}$; however in this case also the secondary stage is affected by this ΔT so that the effective z_v is given by

$$z_v = z_{spider} - \Delta z_{stage}.$$

The squared mean wavefront error is given by

$$w^2 = w_{\Delta z}^2 = \left(\frac{z_{spider} - \Delta z_{stage}}{8 \cdot f_\#^2} \right)^2 \tag{111}$$

2.4.3 Wavefront error due to thermal gradient affecting both spiders and spacers

For sake of simplicity let assume that the two upper spacers are affected by the same Δl_{unc}

and that the corresponding spiders are affected by the same internal thermal gradient. When considering the effect of Δl_{unc}, the transverse motion of the vertex along the *y-axis* and its rotation around the *x-axis* will nullify each other. The vertex will move axially by an amount proportional to Δl_{unc}. The length variation of the spiders will produce a transverse motion of the secondary vertex along the *x-axis* proportional to $\frac{1}{2} l_{spider}$.

For the bottom spacer we can consider the vertex variation as calculated in *section 2.2.2*, where instead of $\frac{1}{2} l_{spider}$, and l_{spider} we should consider $\frac{1}{2} l_{spider} \cdot (1 + \alpha \Delta T)$ and $l_{spider} \cdot (1 + \alpha \Delta T')$ where ΔT and $\Delta T'$ are given by the appropriate thermal gradient within the spider. As far as the secondary stage is concerned, it is assumed to be found at an intermediate temperature. Its thermal response will affect z_v by an amount proportional to $\frac{1}{2}(\Delta T + \Delta T')$.

Finally, to calculate the overall vertex coordinate variation, the subscript *"top"* will be given to all quantities related to the upper spiders and spacers and the subscript *"bot"* to all quantities related to the lower spider and spacer, so that we get

$$\begin{aligned} \theta_v &= \Delta\theta_{bot} \\ x_v &= \Delta x_{bot} \\ z_v &= \Delta z_{top} + \Delta z_{bot} \end{aligned}$$

where Δz_{top} and Δz_{bot} are the contribution to Δz_v by the *top* and *bottom* spacers.
At the third order approximation the coordinate variations for the vertex are given by

$$\begin{aligned} \Delta\theta &= \left(\frac{\Delta l_{bot} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right) \\ &- \frac{1}{3} \left(\frac{\Delta l_{bot} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right)^3 \end{aligned} \tag{112}$$

$$\begin{aligned} \Delta x &= l_{stage} \cdot \left(\frac{\Delta l_{bot} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right) \\ &- \frac{l_{stage}}{3} \left(\frac{\Delta l_{bot} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right)^3 \end{aligned} \tag{113}$$

$$\begin{aligned} \Delta z &= l_{spider} \cdot (1 + \alpha \Delta T_{bot}) \left(\frac{\Delta l_{top} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right) \\ &- \frac{l_{spider} \cdot (1 + \alpha \Delta T_{bot})}{3} \left(\frac{\Delta l_{top} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right)^3 \\ &+ \frac{1}{2} l_{spider} \cdot (1 + \alpha \Delta T_{top}) \left(\frac{\Delta l_{bot} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right) \\ &- \frac{\frac{1}{2} l_{spider} \cdot (1 + \alpha \Delta T_{top})}{3} \left(\frac{\Delta l_{bot} - \Delta z_{stage}}{[1.5 + \alpha(\frac{1}{2}\Delta T_{top} + \Delta T_{bot})] \cdot l_{spider}} \right)^3 . \end{aligned} \tag{114}$$

For the evaluation of the wavefront error we will used the previous expressions in the formulas

$$\begin{aligned} w_{\Delta\theta} &= \frac{2 \cdot f_S}{f_\#} \cdot \Delta\theta \\ w_{\Delta x} &= \frac{1}{f_\#} \cdot \Delta x \\ w_{\Delta z} &= \frac{1}{8 \cdot f_\#^2} \cdot \Delta z \end{aligned}$$

where, in case of tilt removed, only the third power term will be considered, while if the tilt is not removed the first term prevails.

The overall wavefront error is eventually obtained adding quadratically the three contributions

$$w^2 = w_{\Delta x}^2 + w_{\Delta z}^2 + w_{\Delta\theta}^2.$$

2.5 Numerical evaluations

In this section the wavefront error for the cases analyzed in *2.2* are calculated as a function of temperature, assuming that the tilt is both removed by reorienting the flat and by the action of the fast-guiding mirror, and that tilt is not removed. This second case is of particular interest for the evaluation of the beam deflection in the laboratory due to thermal effects. In particular, we refer to data from Spring 1994, when, having aligned the system with the dome closed, in few minutes after the opening of the dome the beam shifted of about $\frac{3}{4}$ of the beam size (i.e. about 3.4 cm).

Finally, the thermal expansion due to the uncompensated part of the spacer and the stage plus secondary mirror are calculated according to eqq.(104) and (105) and result respectively

$$\Delta l_{unc} = 1.17\ \mu m\ K^{-1}$$

and

$$\Delta z_{stage} = 1.22\ \mu m\ K^{-1}.$$

These are the quantities to be used in equations (105) to (112), where $\Delta l_{unc} = z_{spider}$. In particular, when the three spiders and the secondary stage are affected by the same ΔT, the secondary vertex moves along the z-axis by $z_v = |\Delta l_{unc} - \Delta z_{stage}|$, i.e. $z_v = 0.05\ \mu m\ per\ 1\ K$, which represents the accuracy in thermal compensation for the IOTA primary-secondary support assembly.

For numerical evaluation related to this case, see second part of *section 2.5.2*.

2.5.1 Spiders' thermal expansion

The spiders which support the secondary mirror of the IOTA telescopes, are stainless rods 26.5 cm long and 5 cm in cross section. According to eq.(95), if on e of the spiders undergoes a temperature change of 1 K, its length varies of about 1.55 μm which in turn causes a beam deflection in the laboratory of the order of 0.055 cm, or 0.111 cm when the telescope is operated in autocollimation mode.

In order to observe a beam shift in the laboratory of 3.4 cm, in autocollimation mode, the temperature variation at the opening of the dome should have been of about 30 $^{\circ}C$. Even if we can expect very large temperature differences between the air inside of the dome (that heated up during the entire day, being exposed for many hours to temperatures 10 to 15 $^{\circ}C$ higher than the external temperature at the opening of the dome) and the air outside of the dome when this was opened, this seems a too large temperature variation. However, remember that not only the spiders are affected by the temperature change and are responsible for the beam shift. It is very likely that this is the result of the combination of the effects due to the spacers, and probably also to the metallic structure supporting the telescope.

A variation in the length of one spider causes a lateral motion of the secondary mirror and hence a wavefront error. This is of the order of 0.62 μm per 1 K, and it is mainly due to tilt of the wavefront. By simulating this simple case with the program *IOTA_misalignment*, it is shown that this wavefront error is about 400 times larger than that obtained with tilt removed. The corresponding visibility loss when tilt is removed is of about 0.1 %.

2.5.2 Spider-end axial motion: only one spacer affected by thermal variation

The variations in the coordinates of the secondary mirror vertex are calculated according to eqq.(96), (97), and (98), where z_{spider} is given by Δl_{unc} per 1 K of temperature variation, and l_{stage} is obtained by adding d_8 and d_9 and is about 8.89 cm. We get

$$\begin{aligned} \Delta\theta_v &\simeq 3\ \mu rad \\ \Delta x_v &\simeq 0.26\ \mu m \\ \Delta z_v &\simeq 0.4\ \mu m \end{aligned}$$

The mirror rotation is mainly responsible of the beam shift in the laboratory, if tilt is not removed. In this case, the beam would move by about 0.025 cm per 1 K.

To take into account also the effect due to spider length change we should use eq.(95), and eqq.(101), (102), and (103) for the effect due to the spacer. In particular, if we assume that the thermal change for both spider and spacer, immediately following the opening of the dome, is about 20 $^{\circ}C$, we will have for the beam a lateral shift of about 1.6 cm which becomes 3.2 cm when the system is set in autocollimation mode. This is a likely temperature difference between inside the dome and outside, if no special care was taken to slowly thermalize the air inside the dome before opening it, and would explain the large beam shift which has been documented.

The total wavefront error due to the spacer only is calculated according to eqq.(107), (108), and (109). If tilt was not removed we would have

$$\begin{aligned} w_{\Delta z} &\simeq 8 \cdot 10^{-3}\ \mu m \\ w_{\Delta x} &\simeq 0.11\ \mu m \\ w_{\Delta \theta} &\simeq 0.27\ \mu m \end{aligned}$$

and the total wavefront rms would result to be 0.29 μm i.e. of the order of $\lambda/2$, per 1 K. However, in general we expect the wavefront tilt to be removed, so that the total wavefront rms results of the order of 10^{-12} μm, which is absolutely negligible.

If the three spacers and the secondary mirror stage are affected by the same temperature change, the wavefront rms error, calculated using eq.(111), will be of the order of 0.0011 μm which implies a visibility loss of the order of 0.02 %.

If we consider the temperature variation between sunset (the time when the dome is opened) and twilight (when we can start the observations) the despace error causes a wavefront rms error of 0.0024 μm that, if it is not corrected for, is responsible for a loss in visibility of 0.08 %. Finally, considering the temperature variation during the whole night, the despace error would cause the instrument visibility to be, at the end of the night, 1% lower than at the beginning of the observation period.

2.5.3 Thermal gradient affecting both spiders and spacers

For this evaluation it is assumed that all the components undergo the temperature variation occurring between sunset and twilight, $\Delta T = 2.22\,K$, and that gradually a temperature difference arises between the *top* and *bottom* spacers[54] Let us assume this difference to be ΔT of 1 K, which is equivalent to a thermal gradient of the order of 0.025 $K\,cm^{-1}$. The temperature variation considered for the three elements involved in the calculation are

$$\begin{aligned} \Delta T_{top} &= -1.85\ K \\ \Delta T_{stage} &= -2.22\ K \\ \Delta T_{bot} &= -2.85\ K \end{aligned}$$

The corresponding variation in the coordinates of the secondary mirror vertex, when tilt is not removed, calculated according to eqq.(112), (113), and (114), result to be

$$\begin{aligned} \Delta\theta_v &= 1.60\ \mu rad \\ \Delta x_v &= 0.14\ \mu m \\ \Delta z_v &= 0.11\ \mu m \end{aligned}$$

[54]This is because for a long enough period of time the dome is opened, during which the whole system has time to thermalize, the upper layers of the atmosphere within the dome tend to be warmer than the lower ones. On the contrary, at the moment immediately after the opening of the dome the upper layers are colder than the lower ones; this case is presented in more details in *section 2.5.2.*

In particular the tilt of the mirror would cause the beam in the laboratory to shift by about 0.013 cm.

There is a clear difference with respect to the case presented in *2.5.2*. In that case we assumed that only one spacer was affected by the large temperature variation occurring at the opening of the dome. This is the situation shown in fig. T.5. Because only one spacer pulls the mirror, it suffers a quite large rotation.

In the case presented in this section all three spacers and the stage are affected by a thermal change, which eventually is not the same for all of them, because of the thermal gradient. However, the rotations and translations induced by each spacer partially compensate each other, so that the final mirror tilt is less than that expected when only one spacer shrinks[55]. The related wavefront errors are

$$\begin{aligned} w_{\Delta z} &\simeq 0.002\ \mu m \\ w_{\Delta x} &\simeq 0.057\ \mu m \\ w_{\Delta \theta} &\simeq 0.147\ \mu m \end{aligned}$$

and the total wavefront rms error is of 0.158 μm, which is of the order of $\lambda/4$ in the visible. Although this would still allow the formation of interferometric fringes, the Strehl ratio[56] for this wavefront error would be about 0.04. For values less than about 0.4, the Strehl ratio over estimates the loss in visibility (*ten Brummelaar, 1995*). However it is clear that, if tilt was not removed, the detection of fringes would be quite difficult.
When tilt is removed the overall wavefront rms error is of the order of $2 \cdot 10^{-13}\ \mu m$, which is a completely negligible quantity.

2.6 Ray bending due to an air-temperature gradient

We start by considering the general case of the deflection of the line of sight by a transverse air gradient.

The temperature gradient introduces a gradient in the air index of refraction n, since it increases for denser, i.e. colder, medium layers. Because of that the wavefronts tilt towards colder medium, according to Snell's law ($n_i sin\theta_i = n_t sin\theta_t$). With reference to fig. T.8 consider a layer height $\Delta h = h_f - h_i$ where the vertical refractive index gradient is given by $\frac{\partial n}{\partial x}$.

[55] For example, let us assume that the $\Delta T = 1\ K$ between *top* and *bottom* spacers is due to a temperature variation of the upper spacer only ($\Delta T = 1\ K$ is concentrated at the *top* of the assembly). In this case we would expect, according to *2.5.2*, a beam shift in the laboratory of about $1.6/20 = 0.08\ cm$.

On the contrary, assuming that all the three spacers are affected by the same temperature variation corrected for the thermal gradient ($\Delta T = 1\ K$ is distributed between *top* and *bottom* part of the assembly), the beam shift in the laboratory results to be about 6 times less large.

[56] As it is explained in *section 0.1 of Instrument visibility for IOTA* the Strehl ratio is an estimator of visibility loss, where no loss corresponds to Strehl ratio equal 1

fig. T.8 Wavefront bending due to air-temperatur gradient

We want to calculate the vertical deviation of the wavefront from the z-direction, along which the wave is supposed to travel.

The speed of the wave in the medium is given by

$$v(z) \;=\; \frac{c}{n(z)}$$

so that the wave travels a distance d(x) given by

$$d(z) \;=\; v(z)\cdot t \;=\; \frac{c\cdot t}{n(z)}$$

With reference to fig. T.8 we assume that the upper layers are colder than the lower ones[57], so that the wavefront is tilted by an angle θ, which is also the angle by which a ray perpendicular to the wavefront is bent upwards, and it is given by

$$\begin{aligned}\theta &= \frac{-d(x+\Delta x)+d(x)}{\Delta x}\\ &= -\frac{\Delta d}{\Delta x} \;\rightarrow\; -\frac{\partial d}{\partial x}\end{aligned}$$

[57] This is because in the last part of this section we are interested in the particular case when the dome is just opened and the air inside the dome is warmer than the air outside. The air inside the dome will change to an inverted gradient, with the upper layers colder than the lower ones.

It is to be remember, however, that in general, during the observations the situation is the opposite, with upper layers warmer than the lower layers and we would expect that the index of refraction has smaller values going towards higher layers.

$$
\begin{aligned}
&= \frac{1}{n^2} \cdot \frac{\partial d}{\partial x} \cdot ct \\
&\simeq \frac{z}{n_{ave}} \cdot \frac{\partial d}{\partial x}
\end{aligned} \tag{115}
$$

where the increment in d is negative because, as x increases, the ray travels less distance along z, and n_{ave} is the average index of refraction across the layer.

After having travelled a distance Δz, a ray will be vertically displaced by

$$\Delta h \simeq \theta \cdot \Delta z$$

so that the cumulative displacement is given by

$$
\begin{aligned}
h_f - h_i &= \int_{initial}^{final} \Delta h \\
&= \int_{initial}^{final} \theta \cdot dz \\
&\simeq \frac{\partial n}{\partial z} \cdot \frac{1}{n} \int_0^z z' dz' \\
&= \frac{z^2}{2n} \cdot \frac{\partial n}{\partial z}
\end{aligned} \tag{116}
$$

For a gas, n is close to unity and we get that the quantity $(n^2 - 1)$ is proportional to the number of molecules (*Born and Wolf, 1980*) [58], hence it varies as the inverse of the temperature T, expressed in Kelvin degrees. From

$$(n^2 - 1) = const. \cdot T^{-1}$$

we get

$$
\begin{aligned}
2n \cdot \frac{\partial n}{\partial z} &= const. \cdot (-\frac{1}{T^2} \cdot \frac{\partial T}{\partial x}) \\
\frac{\partial n}{\partial z} &= -\frac{n^2 - 1}{2n} \cdot \frac{1}{T} \cdot \frac{\partial T}{\partial x}
\end{aligned} \tag{117}
$$

[58] From eq.(32) on page 92 we have

$$n^2 - 1 = 4\pi N \alpha$$

where α represents the mean polarizability of the gas, and N is the number of molecules per volume unit. Since

$$N = \frac{N_A\, p}{RT}$$

where N_A is the Avogadro number, p is the gas pressure, and R is the universal gas constant, we have $n^2 - 1 \propto T^{-1}$.

and eventually we get that the angle of deflection, θ, and the vertical displacement, h, are given, in terms of the transverse temperature gradient[59], by

$$\theta = -\frac{n^2-1}{2n} \cdot \frac{1}{T} \cdot \frac{\partial T}{\partial x} \cdot z \tag{118}$$

$$h = -\frac{n^2-1}{2n} \cdot \frac{1}{T} \cdot \frac{\partial T}{\partial x} \cdot \frac{z^2}{2} \tag{119}$$

setting $h_{initial} = 0$.

As an example let us now consider propagation along the long arm of IOTA and the typical observed deflection, when comparing the line-of-sight changes between a daytime case with trapped hot air in the tube of the delay line ($\frac{dT}{dx} \neq 0$), versus the night-time case when the air is of uniform cool temperature ($\frac{dT}{dx} = 0$).

With an air-filled long delay line 30 m in length, assuming $\frac{dT}{dx} = 0.1K\,cm^{-1}$ in the tube during the day, $n \simeq 1.0003$, and $T = 300\ K$, the calculated deflection (according to eq.(119)) is of the order of $-4\ mm$ (i.e. during the day upper layers are warmer so that the beam is bent downward with respect the night uniform temperature case), which is typical of the observed deflection values.

Let us considered again the deflection of the beam in the laboratory observed during the tests made in 1994. After opening the dome, in about 15 minutes the geometric beam moved by about 3/4 of the beam size downward along the x-direction. Could this deflection be due to a transverse variation of the index of refraction within the telescope?

Let us consider eq.(118) which gives the tilt angle for the wavefront, or the deflection angle for the ray, as a function of the transverse temperature gradient, the average refractive index, and the distance traveled by the beam, whose most sensitive path is in the expanded beam, between the primary and the flat. The distance travelled by the expanded beam in the telescope is given by

$$z \simeq 2 \cdot (2.5 \cdot f_P) \simeq 5.71\ m,$$

and then we assume $n \simeq 1.003$ and $T \simeq 300\ K$.

A deflection by 3/4 of the beam size is equivalent to an angular deflection of about 836 μrad in the laboratory, which is equivalent to a tilt in the expanded beam path of about 83.6 μrad, when considering the magnification factor of the telescope. The deflection of the beam in the laboratory is then related to the gradient of temperature by

$$\frac{\partial T}{\partial x} = -\frac{1}{m} \cdot \theta_{lab} \cdot \frac{2n}{n^2-1} \cdot \frac{T}{z} \tag{120}$$

where $\theta_{lab} = 836\ \mu rad$ and $m = 10$ so that

[59] These equations are of general validity, i.e. when the thermal gradient is given with its correct sign they say in which direction, with respect to the line of sight, the beam is bent.

$$\frac{\partial T}{\partial x} \simeq -14.7\, K\, m^{-1}.$$

Assuming that the distance across the expanded beam is given by the diameter of the primary, $d_P \simeq 0.45\, m$, the temperature difference across the beam is

$$\Delta T \simeq \frac{\partial T}{\partial x} \cdot d_P$$
$$\Delta T \simeq -6.6^{\circ} C.$$

A negative variation of this magnitude would have result from the fact that the temperature, initially warmer on top because the dome was closed for the all day, changed to a inverted gradient when the dome was opened. However this inverted gradient could explain only the magnitude of the shift observed, but not the sign. In fact cooling on top would bend the rays upwards both in the telescope and in the laboratory, and not downward in the x-direction, as we observed. This means that the thermal effects on the primary-secondary support assembly are mainly responsible for the beam deflection in the laboratory. As shown in *section 2.5.2*, the sudden cooling of the upper spacer makes it shrink and consequently the secondary mirror tilts CW so that the beam in the laboratory results shifted downward.

Mechanical stability for IOTA

0. Introduction

I present here a list of some of the sources of mechanical instability, i.e. of movement of any component relative to the others, we can expect to have for IOTA. (Most of the information in this section were taken from *Carleton, 1990*). In most cases however, we present only a description of the phenomena and a rough evaluation their effects. A more accurate evaluation of the wavefront error induced by each effect would require an extensive engineering analysis, involving a study of the propagation of vibrations within the whole interferometer structure. However, experience with mechanical structures has often shown that calculations are inadequate to represent reality, so it would be naive to put too much faith in mathematical simulations of vibrations in structures.

The analysis has to start from a measurement (or simulation) of the power spectrum of the addressed vibrational phenomena in order to apply it to the behaviour of the different components of the interferometer structure, taking into account the resonance frequencies of each component. To be rigorous, this kind of analysis should have been performed during the design of the instrument. Actually what we do now is to evaluate the effect of the disturbances and, in case some of them result to produce intolerable degradation to the fringe visibility, counteracting actions are studied to eliminate them.

Because IOTA is not intended to be an astrometric instrument, the long-term mechanical stability requirements are very much relaxed. Our main concern is about the short-term (high frequency) disturbances which can affect the instrumental visibility during the acquisition of one fringe. In particular this will be important when operating in the faint limit. In this case we will have only a very few photons per atmospheric coherence time, and we will operate in open loop, i.e. we will not be able to compensate with active systems for path length variation occurring during the integration time.

1. Delay line

The delay line is on one hand the means for correcting path length disturbances originating elsewhere in the system, and on the other hand is itself a potential source of disturbance. As far as the long delay-line is concerned, this operates in a slew-and-clamp mode, so that should not introduce any variation in the optical path during a fringe acquisition. Fine adjustment and tracking are provided by a carriage put on an air-bearing on a polished granite beam. The use of an air-bearing support and brushless linear motor eliminates much of the vibration usually associated with mechanical devices. The high-frequency ($> 100\ Hz$) noise that this carriage introduces is at sufficiently low levels that we do not need to introduce passive or active compensation for that. At lower frequencies ($< 40\ Hz$), the noise of the carriage is quite negligible and it is possible to command it to compensate for path length variations introduced by the atmosphere or by other system components. At all frequencies, the noise of the carriage is smaller than the noise due to the turbulence of the atmosphere (see fig. MS.1).

fig. MS.1 Power spectrum of the Anorad air-bearing carriage moving at 2.5 $mm\,s^{-1}$ in a noisy environment. Peaks at 25-50 Hz are from laboratory noise. The straight line shown is the asymptotic atmospheric disturbance (*Carleton, 1990*)

While the air-bearing carriage is fundamental in providing a very smooth motion of the carriage, it does introduced the problem of disposing of the air discharged into the vacuum chamber. This vacuum is only of the order of 1 $mbar$, which can be easily maintained against the air-discharge by modest-sized mechanical pumps. The question is whether the vibration generated by the pumps will, in turn, affect the system.

2. Vacuum pumps

In the early phase of the instrument assembling, it was verified that vibration from the vacuum pumps did not affect interference fringes generated in laser light sent out to cube corners on top of the stovepipes and returned inside to the optical table. However when the path was extended to the siderostat in autocollimation mode, there was an evident degradation of the path length stability. There are four possible ways by which the pump induced vibration might work an effect: coupling through the ground, acoustic coupling through the air to the telescope structures, coupling through the walls of the pumping line, and acoustic coupling through the stream of gas in the pumping line. The first of these seemed to be the most likely candidate for causing the observed visibility degradation. The solution adopted to eliminate coupling through ground, was to provide a platform that holds all of the pumps and the air compressor on a concrete slab that is suspended on four springs, such that its principal modes have proper frequencies that give a large attenuation at the typical pump-vibration frequencies.

In case this should not be enough to counteract the effects induced by the pump vibrations, other solutions like acoustic-isolation walls inside the pump shed, sand beds for the

pumping line and baffles in the line may be studied.

2.1 Calculation of the isolation efficiency of the spring system

When a harmonic disturbing force $P sin\, 2\pi f t$ (in our case due to the pumps+slab load) acts on a spring system of spring constant k, the amplitude of the forced vibration is given by

$$x = \frac{P}{k} \frac{1}{\sqrt{\left[1 - \left(\frac{f}{f_o}\right)^2\right]^2 + 4\gamma^2 \left(\frac{f}{f_o}\right)^2}}. \tag{121}$$

The first factor on the right side is the *zero frequency deflection*, i.e. the static deflection which would result if a steady force of magnitude P was applied to the spring of stiffness k. The second factor is due to the dynamical conditions and is called the *magnification factor*. f is the frequency of the disturbing force, f_o is the undamped natural frequency of the spring, γ is the damping factor due to resisting forces proportional to the velocity of the system.

As we have directly experimented, vibrations originating from machines or other sources are in general transmitted to the neighboring structure. To reduce transmitted vibrations, isolators in the form of springs are frequently used (as well as rubber mounts and cork padding). Of our interest is the ratio of the transmitted to the disturbing force (which equals the ratio between the amplitude of motion induced by the transmitted and the disturbing forces), this ratio being designated as *transmissibility*. Assuming that the isolator can be represented by a spring and a dashpot, the force transmitted is the vector sum of the spring and dashpot force which is

$$F_{tr} = kx \cdot \sqrt{1 + \left(2\gamma \frac{f}{f_o}\right)^2}$$

Expressing the disturbing force as a function of the spring displacement, according to eq.(121) we get

$$\frac{F_{tr}}{P} = \frac{\sqrt{1 + 4\gamma^2 \left(\frac{f}{f_o}\right)^2}}{\sqrt{\left[1 - \left(\frac{f}{f_o}\right)^2\right]^2 + 4\gamma^2 \left(\frac{f}{f_o}\right)^2}}$$

The above ratio decreases for increasing values of f/f_o, thus for an isolator to perform its function the frequency f_o of the supporting structure must be small in comparison to the frequency f of the disturbing force. (*Baumeister and Marks, 1967*)

In the isolation system adopted for the IOTA pumps, we do not make use of any dashpot, and assuming the damping effect due to air friction negligible, the *isolation factor* which relates the transmitted disturbance to the induced disturbance is given by

$$R = \frac{1}{1 - \left(\frac{f}{f_o}\right)^2}. \tag{122}$$

The four springs chosen for our isolation system have a stiffness constant k of 238 $N\,cm^{-1}$ and a maximum load of 3015 N. The total load applied to the four springs is about 9466 N so that each spring is loaded by about 2367 N. Hence, the deflection, d, that each spring undergoes is 9.94 cm. The frequency of the spring motion for this load is given by

$$f_o = \frac{1}{2\pi}\sqrt{\frac{g}{d}}$$

and equals to 1.58 Hz. Since the frequency of vibration of the pumps is 30 Hz, the isolation factor calculated according to eq.(122) is of the order of 0.003.

For the computation of the effect of the transmitted vibration on the wavefront of the interfering beams, we assume that, because without isolation system the interfering fringes vanished most of the time, but not always, the vibration amplitude caused a wavefront error of $\lambda/2$ or slightly larger. With the isolation system, the wavefront error should be reduced by the same isolation factor as the vibration amplitude. This means that we can assume a wavefront error due to the pumps vibration of the order of $8 \cdot 10^{-4}$ μm in the visible.

3. Mirror drives

Other active elements in our system are the two-axis yoke mount for the siderostat flat and the piezo-driven fast guiding mirror. These drives both have effective pivot points that lie essentially at the intersection of the system optical axis with the mirror surface, so that there is negligible coupling between rotational motion and optical path length. In particular, the chief ray of the system would not suffer any path length variation whatever rotation the two mirrors undergo.

The procedure for system alignment is intended to maintain the optical axis at the correct location to ensure this lack of coupling. Also we need to make sure that the high-bandwidth (to 100 Hz) drive of the guiding mirror cannot excite appreciable vibration in the beam-compressor telescope, near which it is mounted. This is accomplished by mounting the drive on a stiff and heavy structure with high damping properties.

4. Ground vibration

The effect of ground vibrations is considered a fundamental source of possible path-length fluctuations in our instrument. Ground vibration arise both from natural background and from human activities in the site area. Seismological measurements were made of both vertical and horizontal motion which result in comparable spectra as a function of the frequency, (range $0.5 - 100$ Hz). Mt. Hopkins is not a seismically active area, and the disturbance from human activity comes mainly from the ventilation system and machinery in the nearby dormitory and common building, and from the MMT (Multi Mirror Telescope) building, whose slewing is a potential source of vibration in the $0.5 - 5$ Hz

range. This source is measurable but it is comparable with the natural background. Vehicle movement can be detected over about 500 m distance, and the effect becomes many times the background at close approach.

In conclusion the ground seems essentially benign at the IOTA site, in comparison for example with Mauna Kea, that was the first site for which measurement of seismic noise were performed for the purpose of designing an astronomical facility. With reference to fig. 3. in *Carleton, (1990)* we see that the background noise at Mt. Hopkins is lower than at Mauna Kea (due to the lower level of seismic and human activity). Also we see that the rms vertical displacement has a peak of the order $10^{-2}\mu m/Hz^{1/2}$ in a range centered at about $1.5\,Hz$.

Wavefront error due to deformation of the primary mirror's surface by gravity

0. Introduction

Deformation of the mirrors is mainly due to thermal effects and gravity load. For the thermal effect the reader is referred to section *Thermal effects*. In this section we present a few steps by which it is possible to perform an estimation of the wavefront error due to the load applied to the primary mirror.

1. Bending of uniform-thickness plates with circular boundaries

For the sake of our evaluation, we will assume that the behaviour of the parabolic substrate of the primary is well simulated by that of a flat circular plate. It is also assumed that the plate is uniform in thickness and of homogeneous isotropic material, and that it is nowhere stressed beyond the elastic limit.

For a uniformly distributed load, q, the deflection at the center of the plate is given by *(Roark and Young, 1975)*

$$\delta y = K_y \cdot \frac{q}{D} \cdot r^4 \tag{123}$$

where K_y is a constant which depends on the location of the start of the distributed load (i.e. on the radial position of the supporting points), q is the load per unit area, r is the plate radius and D is given by

$$D = \frac{E \cdot d^3}{12(1-\nu^2)}$$

where E is the modulus of elasticity, ν is the Poisson's ratio, and d is the plate thickness.

For a plate with N support points the rms error of the plate surface is given by

$$y_{rms} = \gamma \cdot \frac{q}{D} \cdot \left(\frac{A}{N}\right)^2 \tag{124}$$

where A is the plate area, and γ a constant related to K_y.

For support points at a constant radius, r_o, supporting a circular plate the minimum deflection is achieved at an *optimum* radius given by $r_o = \beta r$, where β depends on the number of supporting points.

2. Plate deflection for the IOTA primary mirror

In IOTA the primary mirror is made of zerodur and is supported by mean of three supporting points at the optimum radius given by $\beta = 0.681$. For this β the value of γ is $2.93 \cdot 10^{-3}$. According to eq.(124) we calculate the rms error of the surface plate as

$$y_{rms} = \gamma \cdot \frac{\rho}{\frac{E \cdot d^2}{12(1-\nu^2)}} \cdot \left(\frac{\pi r^2}{N}\right)^2 \tag{125}$$

where ρ is the material density. For zerodur it is $\rho = 2.5\ gcm^{-3}$, $\nu = 0.25$, $E = 91 \cdot 10^3\ Nmm^{-2}$, $d = 7.62\ cm$, so that for full weight the rms error would be equal to $4.63 \cdot 10^{-3}\ \mu m$. Because the IOTA telescopes are mounted at 30° with respect to the horizontal, the effective load on the mirror is one half of the full weight, and eventually the rms error is $2.31 \cdot 10^{-3}\ \mu m$.

1.2 Estimate of the effect of shear

The formula for deflection given in *1.1* takes into account bending stresses only; however, there is always some additional deflection due to shear. Deflection due to shear may be expressed as a function of the deflection due to bending stress by

$$\delta y_{shear} = \alpha \cdot \left(\frac{d}{u}\right)^2 \cdot \delta y_{bend} \tag{126}$$

where u is the effective separation between support points, and α is of the order of 2 when u is calculated by $\frac{r}{N^{1/2}}$.

Eventually the total rms error of the plate surface due to gravitational loading is given by

$$rms_{tot} = y_{rms} \cdot \left(1 + \alpha \cdot \left(\frac{d}{u}\right)^2\right) \tag{127}$$

and for IOTA it results

$$rms_{tot} = \frac{5}{3} \cdot y_{rms}$$

i.e. $rms_{tot} = 3.86 \cdot 10^{-3}\ \mu m$.

Instrument visibility for IOTA

0. Introduction

In order to characterize the efficiency of the IOTA interferometer, we refer to a coherence transfer factor, η, which relates the visibility of an object, V_{obj}, to the measured visibility, V_{meas}, according to

$$V_{meas} = \eta \cdot V_{obj}. \tag{128}$$

A useful, approximate predictor of η is based on the Strehl ratio, S, (such that $\eta \simeq S$), which is also a standard indicator of the performance of classical imaging systems (a definition of Strehl ratio is given in *section 0.1.1*). The justification for the use of S, relative to one aperture, as an estimator of η for a two-aperture interferometer, is mainly due to *ten Brummelaar et al., (1995)*, as it is presented in paragraph *0.1.*

For an interferometric array, η is described in terms of the visibility loss, l, by

$$\eta = 1 - l, \tag{129}$$

so that, eventually, the instrumental Strehl ratio of the system will give the visibility loss due to instrumental causes (i.e. independent of any atmospheric effect).

For a practical evaluation of the coherence transfer function for the IOTA interferometer, I assume that the two quantities S and η are interchangeable, and the Strehl ratio is formulated in terms of the rms deviation of the wavefront, w, according to

$$S = exp\left[-\left(\frac{2\pi}{\lambda}w\right)^2\right] \tag{130}$$

where the argument of the exponential represents the phase error of the wave, which is generally expressed as $\sigma = \frac{2\pi}{\lambda}w$.

0.1 Relation between Strehl ratio and coherence transfer factor

One way to characterize the efficiency of an interferometer is with a coherence transfer factor η, which relates the measured visibility, V_{meas} to the visibility of the object, V_{obj} according to

$$V_{meas} = \eta \cdot V_{obj}. \tag{131}$$

For accurate predictions of the coherence transfer factor simulation will give the best results. However, in situation of most interest to optical interferometry, it is sufficient to make use of an approximate predictor of η based on the Strehl ratio, which is the parameter usually used as an indicator of the performance of a classical imaging system. This possibility was mentioned by several authors but the only ones who, to my knowledge, explicitly justified the use of the Strehl ratio in this role were *ten Brummelaar et*

al. (1995).

0.1.1 Strehl ratio

A simple criterion to define the image quality produced by an optical system, and hence to estimate the amount of aberration that can be tolerated, is by means of the Strehl ratio. This is defined as the ratio of the irradiance at the center of the reference wavefront sphere (with respect to which the aberration is defined) with and without aberration.

Among many approximations for the Strehl ratio which have been investigated, the most accurate one (*Mahajan, 1982*) is given by

$$S \simeq exp(-\sigma^2) \tag{132}$$

where σ is the rms of the phase aberration across the pupil of the imaging system ($\sigma = \frac{2\pi}{w}$, where w is the wavefront rms). *Mahajan (1982)* shows that this expression approximates the Strehl ratio of the system with an error of less than 10 % as long as $S > 0.3$.

Let consider now the Strehl ratio in terms of the optical transfer function of the system. If the optical transfer function of the system is $T(\nu)$ and the optical transfer function of the phase aberration is $B(\nu)$, the intensity of the aberrated image is the Fourier transform of the combined optical transfer function $T(\nu)B(\nu)$, and that of the unaberrated image, is the Fourier transform of $T(\nu)$. The Strehl ratio will be given by the ratio of these transforms at the center of the reference sphere

$$S = \frac{\int T(\nu)B(\nu)d\nu}{\int T(\nu)d\nu}. \tag{133}$$

It is to be noticed that both the approximation given by eq.(131), and the theoretical expression (132), represent the Strehl ratio for an imaging system with one aperture.

0.1.2 Coherence transfer factor

As far as a two-aperture system is concerned, *Tango and Twiss (1980)* have shown that, if the aberration effects are small, the loss of coherence (i.e. of visibility), in the combined beam is given by the root square of

$$\langle \eta^2 \rangle = 1 - (\sigma_{\varphi 1}^2 + \sigma_{\varphi 2}^2) - (\sigma_{\chi 1}^2 + \sigma_{\chi 2}^2) \tag{134}$$

where $\sigma_{\varphi i}$ and $\sigma_{\chi i}$ represent the rms of the fluctuation in phase and log amplitude (which represents effects due to scintillation) at each aperture, and it is assumed that the fluctuations at the two apertures are independent. Assuming also that the phase variations dominate and that the two apertures have the same phase variance, σ_φ, or equivalently the same Strehl ratio, it results that

$$\langle \eta^2 \rangle \simeq 1 - 2\sigma_\varphi^2. \tag{135}$$

Tango and Twiss derive also the expression of the squared coherence transfer factor as a function of the optical transfer function of a single aperture, $T(\nu)$, and the aberration

transfer function, $B(\nu)$. This is given by

$$\langle \eta^2 \rangle = \frac{\int T(\nu) B^2(\nu) d\nu}{\int T(\nu) d\nu} \tag{136}$$

and differs from the theoretical expression of the Strehl ratio only by the appearance of $B^2(\nu)$ instead of $B(\nu)$. The approximation (135) and the theoretical expression (136) of the squared coherence transfer factor are now put in relation with the approximation and theoretical expression for the Strehl ratio given by the eqq.(132) and (133).

As far as the approximated forms are concerned, ten Brummelaar et al. show that, if the phase variance is small, so that the two relations can be expanded by Taylor series, the Strehl ratio can be written as

$$S \approx 1 - \sigma_\varphi^2 + \frac{(\sigma_\varphi^2)^2}{2} + O[(\sigma_\varphi^2)^3], \tag{137}$$

and the coherence factor is given by the square root of $\langle \eta^2 \rangle$, that is by

$$\eta_{rms} \approx 1 - \sigma_\varphi^2 - \frac{(\sigma_\varphi^2)^2}{2} + O[(\sigma_\varphi^2)^3], \tag{138}$$

Hence, if the Strehl ratio is high, to second order, η_{rms} and S can be considered the same.

For a more precise comparison, ten Brummelaar et al. consider the expressions of $T(\nu)$ and $B(\nu)$ for a single circular aperture and for atmospheric turbulence as derived by *Fried (1966)*. The results of this operation is given in fig. IV.1.

The plot shows that, as predicted by the Taylor expansion, Strehl ratio for a single aperture and coherence transfer factor for a two-aperture interferometer, whose single apertures have the same Strehl ratio, are approximately equal for low D/r_o values, where D is the aperture diameter and r_o is the coherence length, and diverge as D/r_o increases.

The plot actually shows two curves, one for tilt-corrected wavefront and one for uncorrected wavefront. Since large-baseline interferometers must employ a tip/tilt servo, to keep the interfering beam parallel, the interesting case is the corrected-wavefront one. For this case the equality between Strehl ratio and coherence transfer factor holds for greater D/r_o values.

fig. IV.1 Coherence loss (solid curves) and Strehl ratio (dashed curves) for uncorrected and tip-tilt corrected wavefronts as a function of D/r_o, from *ten Brummelaar et al. (1995).*

1. Strehl budget allocation for the array

Several factors enter into the instrumental visibility, so that in order to take into account most of them, the visibility loss is described by a Strehl ratio budget allocated among the main array subsystems. Following the scheme adopted for the CHARA[60] Array project (*Technical Report No.15, May 1995*), it considers: S_{opt}, referred to visibility loss due to any optical component, S_{OPD}, referred to visibility loss due to optical path change during an integration time, S_{diff}, referred to visibility loss due to imbalance in diffraction effects in beams, and S_{servo}, referred to visibility loss due to imperfect tilt correction by the servo system. The instrumental Strehl ratio will eventually be given by the product of all these factors. Here we are going to evaluate in more details the factor related to the optical components, and to give some approximate evaluation for the other factors, in order to perform an estimation of the expected instrumental visibility loss for IOTA.

2. Strehl ratio for the optical subsystem

We consider IOTA in the configuration with two collecting elements, each of them consisting of a siderostat plus a beam compressor. Following the beam compressor, each light beam passes through 2 windows, reflects at a number of flats defined by the beam relay configuration, and combines with the other beam at the beamsplitter. Hence, the Strehl budget for the optical components is given by the following factors: S_{sid}, related to the siderostat, S_{tel}, related to the telescope, S_{win}, related to all the windows, S_{flats}, related

[60] The Center for High Resolution Astronomy (CHARA) of Georgia State University will build a facility for optical/infrared multi-telescope interferometry, called the CHARA array. This array will consist of initially five (with a goal of seven) telescopes distributed over an area approximately 350 m across.

to all the the flats used for the relay of the beam, S_{BS}, related to the beamsplitter.

2.1 Strehl ratio for the siderostat

The flatness specification for the siderostat is $\lambda/3$ peak-to-valley wavefront error over the entire clear aperture, and $\lambda/10$ peak-to-valley wavefront error over any 12 cm area within the clear aperture, at $\lambda = 0.633\,\mu m$. For the whole section the conversion between the peak-to-valley error, w_{p-v}, and the wavefront rms, w_{rms}, is assumed to be $w_{p-v} = 4\,w_{rms}$. Making use of this relation, the Strehl ratios derived from the flatness specification are $S_{sid} = 0.760$ and $S_{sid} = 0.975$, respectively.

2.2 Strehl ratio for the telescope

At the moment I have identified 7 factors by which to describe the wavefront error and the consequent visibility loss due to the telescope.

2.2.1 Overall optical performance of primary plus secondary

The optical specification for the telescope, i.e. primary plus secondary, was set to have $\lambda/10$ peak-to-valley wavefront error over any 15 cm diameter patch, and $\lambda/3$ peak-to-valley wavefront error over 90 % of the full aperture, where $\lambda = 0.633\,\mu m$ in both cases, or equivalently if we consider $\lambda = 2.2\,\mu m$, i.e. the case for observation in the infrared band, we have $\lambda/10$ over almost the full aperture.

Assuming that the peak-to-valley wavefront error is 4 times larger than the its rms deviation, the Strehl ratio for a $\lambda/10$ wavefront error is $S_{P-S} = 0.975$. This is the value when considering integration of the signal either on subaperture patches in the visible, or on almost the full aperture in the infrared. It is to be noticed that integration of the signal in the visible on 90 % of the aperture would give $S_{P-S} = 0.760$, a value that by itself is already close to that due to the atmospheric effects.

2.2.2 Alignment: secondary decenter and tilt

The simulation performed for the study of the misalignment effects for the IOTA telescopes, allows one to derive the maximum decenter and tilt of the secondary that could escape detection when we look at the interferograms obtained in autocollimation mode. The secondary undetected decenter is about $5\,\mu m$, and the secondary tilt is less or equal to about $50\,\mu rad$. For these values the Strehl ratio computed by the IOTA_misalignment program, for $\lambda = 0.6328\,\mu m$, are $S_{dec} = 0.999$ and $S_{tilt} = 0.998$. The corresponding values for the visible ($\lambda = 0.5461\,\mu m$[61]) and infrared case ($\lambda = 2.2\,\mu m$) are $S_{dec,VIS} = 0.998$ and $S_{dec,IR} = 0.999$, $S_{tilt,VIS} = 0.997$ and $S_{tilt,IR} = 0.999$ respectively.

[61] The visible wavelength for the Strehl ratio evaluation is assumed to be $\lambda = 0.5461\,\mu m$, in the whole section. This choice is made because most of the wavelength dependent quantities of the optical components examined in this study are given at $\lambda = 0.5461\,\mu m$. For the infrared case, $\lambda = 2.2\,\mu m$ is used because it is the infrared wavelength at which IOTA operates.

2.2.3 Defocus: secondary axial motion

As in the previous case, the misalignment simulation program is used to derive the maximum axial motion of the secondary that cannot be detected by mean of a degradation of the interferograms obtained in autocollimation mode. The maximum undetected error in the axial separation between primary and secondary is $2\,\mu m$. Such value implies $S_{focus} = 0.950$ at $\lambda = 0.6283\,\mu m$, so that it is $S_{focus,VIS} = 0.934$ and $S_{focus,IR} = 0.996$.

2.2.4 Strehl ratio due to the position error of the center of the primary and secondary mirrors

Because of the uncertainty with which the position of the center of the primary and secondary mirrors are known, the initial alignment of the telescope may be obtained with both the primary and the secondary shifted and tilted (see section *4.3.6 and 4.3.7* in *Telescope alignment*), so that the combination of all these misalignments gives a minimum wavefront error. However, there will be a loss of visibility with respect to the case of perfect alignment. The result of the simulation performed using the program *IOTA_misalignment*, suggests that we should be able to correct the alignment, by looking at the interferograms, for displacements larger than that equivalent to an error in the center position of the primary mirror of about 100 μm. This corresponds to a Strehl ratio $S_{pri,c} = 0.976$ at $\lambda = 0.6328\,\mu m$, so that $S_{pri,c} = 0.968$ in the visible and $S_{pri,c} = 0.998$ in the infrared. This term is considered as another multiplying factor to be added to the term for the telescope misalignment, because, even if it actually results from a combination of relative decenter and tilt between the two mirrors, it is equivalent to considering the observation of a source off-axis.

2.2.5 Strehl ratio due to spider and spider-end axial motion

With reference to the section *Thermal effects* we considered the axial motion caused by the temperature change since the alignment of the telescope. This temperature change, ΔT, is that occurring between sunset (when we suppose the telescope to be aligned) and twilight (the time when we can start observing). According to *tab. T.2 in section 1.3.2.1* we assume for this ΔT a mean value of $2.2\ K$. The overall effect for the primary-secondary assembly being affected by the same ΔT, is a despace error which is responsible for a wavefront error given by

$$w^2 = w_{\Delta z}^2 = \left(\frac{\Delta l_{unc} - \Delta z_{stage}}{8 \cdot f_{\#}^2} \right)^2 \tag{139}$$

where Δl_{unc} and Δz_{stage}, which are defined in section *2.3 in Thermal effects*, both depend on ΔT.

For IOTA we get $w = 2.42 \cdot 10^{-3}\ \mu m$ which implies a Strehl ratio $S = 0.999$.

A thermal gradient between upper and lower spacers should not introduce any degradation in visibility, because, if, as it is expected, the wavefront tilt is removed, the remaining wavefront aberrations will be negligible.

Finally, taking into account the mean temperature variation during the whole night ($\Delta T \simeq 8\ K$), the Strehl ratio at the end of the night degrades to $S = 0.989$.

In conclusion, assuming that the despace error due to temperature variation during the night is not compensated for, we have $S = 0.999$ at the beginning of the night and at the end $S = 0.989$, so that we can considered a mean Strehl ratio value during the whole night of the order of $S_{sp} = 0.994$, for observations in the visible range, and $S_{sp,IR} = 0.999$ in the infrared.

In general it is not expected that a large temperature variation will affect only one of the components of the primary-secondary assembly during the observation period. In case this would happen, the major effect would be the introduction of wavefront tilt that should be removed by both the reorienting of the flat and the action of the piezo-mirror that follows the telescope. Thanks to this, the visibility should not be deteriorated at all.

2.2.6 Strehl ratio related to the mirrors' substrate

In section *1.3.2* in *Thermal effects* it is shown that because of the sensitivity of the mirror substrates to temperature changes over the whole night, the relative position of the foci of the primary and secondary mirror suffer a variation $\Delta f = 0.57\ \mu m$. This means that, if no correction is applied, the visibility degrades during the night by about 1.8 %.

For the purpose of the error budget, and for a conservative estimation, we can consider the loss in visibility due to the temperature variation occurring between sunset (when we assume the telescope is aligned) and twilight (starting of observations but compensation has not stabilized yet) we get for the Strehl ratio the value $S_{sub} = 0.998$. For the infrared range, it would be $S_{sub,IR} = 1$. As explain in section *1.3.2* in *Thermal effects*, the effect of thermal gradient induced by the mean diurnal temperature variation is negligible.

2.2.7 Strehl ratio due to the effect of gravitational load on the primary mirror

In the section *Wavefront error due to deformation of the primary mirror's surface by gravity*, I calculate the error on the surface figure of the primary mirror due to the gravity load. According to that estimation, the *rms* error on the wavefront is of the order of $7.72 \cdot 10^{-3}\ \mu m$ and the associated Strehl ratio is $S_{def} = 0.992$. In the infrared it is $S_{def,IR} = 0.999$.

According to the previous calculation the overall Strehl ratio due to the telescope, for the visible band, is $S_{tel} = 0.863$, and for the infrared band it is $S_{tel,IR} = 0.965$. The overall breakdown, for the visible wavelength range, is presented in tab. IV.1

tab. IV.1 - Telescope Strehl ratio breakdown: visible

S_{tel}	S_{P-S}	S_{dec}	S_{tilt}	S_{focus}	$S_{pri,c}$	S_{sp}	S_{sub}	S_{def}
0.863	0.975	0.998	0.997	0.934	0.968	0.994	0.998	0.992

2.3 Strehl ratio due to the windows

In this section we are concerned with the loss in visibility due to error in the transversal dimension of the windows. A difference Δd in the thickness of the parallel plates encountered by the beam, introduces a difference in the OPL which is given by

$$\delta OPL = (n(\lambda) - 1) \cdot \Delta d, \tag{140}$$

which also depends on the wavelength considered[62]. When considering the whole range of wavelength observed, each monochromatic beam travels a different OPL and creates an interference fringe slightly shifted with respect to those at other wavelengths. The effect on the resulting interference fringe may consists in some loss of visibility.

Assuming the wavelength range goes from 458 to 820 nm, for fused silica, the maximum variation in $(n(\lambda) - 1)$ is 0.012. Assuming that $\delta OPL = 0$ for the central wavelength of the range, we can consider an average variation for $(n(\lambda) - 1)$ of 0.003. As far as Δd is concerned, the manufacturer specifies the thickness of the window with an uncertainty of $\pm 5\,\mu m$. Assuming a mean error of 2.5 μm[63], the Strehl ratio calculated for two windows encountered by the beam is $S_{win} = 0.984$.

A variation of $(n(\lambda)-1)$ of the same order as for the visible can be assumed in the infrared band, so that it is also $S_{win,IR} = 0.984$.

2.4 Strehl ratio due to the flats in the relay path

The specified flatness for all the flats is $\lambda/20$ peak-to-peak over an area that equals the expected beam footprint, plus a band of 6-mm width around it. According to this specification the Strehl ratio for each mirror is of 0.975. For the visible case the beams undergo 9 reflections (plus one for only one of them) after leaving the telescope and before combining at the beamsplitter. (For the infrared case, in the actual configuration, the reflections are 10.) The Strehl ratio for all the flats is then $S_{flats} = 0.781$ (for 10 reflections), which shows how much the instrumental visibility can be sensitive to any added reflection. For 11 reflections it would be $S_{flats} = 0.761$

2.5 Strehl ratio due to the beamsplitter

As far as the beamsplitter is concerned, we have to consider the behaviour of one of the beams undergoing one reflection and traversing a parallel plate (either the beamsplitter substrate or the compensator plate). The reflective surface has the same characteristics of the other flats, so that $S_{BS,ref} = 0.975$, while the substrate behaves as a window which in this case is made of BK 7. In this case the mean error in the substrate thickness is 3.81 μm so that $S_{BS,sub} = 0.983$. At the moment data for the beamsplitter used for the infrared are not available. In this case, the Strehl ratio due to the optics, beamsplitter

[62] $n(\lambda)$ is the window index of refraction relative to the surrounding medium. For sake of computation here it is assumed $n_{air}(\lambda) = 1$

[63] It is used the same relation used for the evaluation of the wavefront error due to the flatness error of the mirror, i.e. $w = w_{p-w}/4$)

excluded, is $S_{opt,IR} = 0.704$.

2.5.1 Polarization effects

In section *3.1.2* in *Beamsplitter* I estimated the effects on visibility due to different phase shift for different polarization states. Actually, I considered not only the contribution due to the beamsplitter but also that due to the unmatched reflection necessary for directing one of the two beams to the beam splitter. It was shown that, according to calculated values for the beam splitter performance, the overall visibility loss due to polarization effects is of the order of 0.85%. Since the contribution from the extra reflection is related to the presence of the beamsplitter, we allocate the overall polarization effect to the beamsplitter. In terms of Strehl ratio we have $S_{pol,BS} = 0.991$, so that the total beamsplitter factor results to be $S_{BS} = 0.950$

In tab. IV.2 it is shown the breakdown of the Strehl budget for the optical components, relative to the visible range.

tab. IV.2 - Optical components Strehl ratio breakdown: visible

S_{opt}	S_{sid}	S_{tel}	S_{win}	S_{flats}	S_{BS}
0.614	0.975	0.863	0.984	0.781	0.950

3. Visibility loss due to optical path change during an integration time

In this section we consider random events that can show up during an integration period and introduce a variation in the optical path that cannot be compensated for by the fringe-tracking and delay line control system. In all the following calculations we assume that the integration lasts for τ_o, the coherence time imposed by the atmosphere, i.e. about 10 ms in the visible.

In general, we can address two main causes for this: thermal variations and structural vibrations.

3.1 Thermal effects

As far as windows and the beamsplitter substrate and compensator are concerned, we make reference to sections *2.1.1 and 2.1.2* in *Thermal effects*. It is shown that any thermal change occurring during the time interval δt produces a δOPL given by

$$\delta OPL = d \cdot D(\lambda) \cdot \frac{\Delta T}{\Delta t} \cdot \delta t \tag{141}$$

where D[64] is a quantity which depends on the material, d is the window thickness or the

[64] $D = \alpha \cdot (n(\lambda) - 1) + \beta(\lambda)$ where α is the coefficient of thermal expansion, $n(\lambda)$ is the index of refraction, and $\beta(\lambda)$ is is the coefficient of temperature variation of the index of refraction. (See section 1.2.1.1 in *Thermal effects*).

thickness error in the beamsplitter substrate (or in the compensator plate), and $\frac{\Delta T}{\Delta t}$ is the temperature variation per unit time.

As far as δOPL is the same across the whole beam and for any λ, there is no effects on the visibility. However, because of the variation of D with λ, we might expect some degradation in the visibility.

Actually, the δOPL due to the absolute ΔT due to the diurnal thermal variation, in one τ_o results to be negligible, at any λ, for both windows and beamsplitter, so that the related Strehl ratio is considered equal to 1.

In addition the diurnal thermal variation causes in each window a radial thermal gradient, γ. Because of that, rays which traverse (propagate through) adjacent areas of the window have different optical path lengths. The maximum δOPD is given by

$$\delta OPD_{max} = d \cdot D \cdot \gamma \cdot r \tag{142}$$

where d is the thickness of the window and and r is the radius of the beam, D is a coefficient whose value depends on the material, γ is the radial thermal gradient and is related to the window frame parameters. Assuming an rms error on the wavefront equal to $\frac{1}{\sqrt{2}}\delta OPD_{max}$, produced by a mean temperature variation of $2 \cdot 10^{-6}\,K$ in one τ_o, the related Strehl ratio is 0.999.

For the beamsplitter and the exit windows the thermal variation is mainly related to the electronic components, space heaters and human presence within the laboratory. A precise estimation of these contributions is beyond the scope of this study. A rough estimation would lead to something like $0.4\ Kh^{-1}$, which implies, for both internal windows and beamsplitter, a Strehl factor equivalent to that obtained for the effect of the diurnal change on the external windows as far as the radial gradient is concerned.

For the beamsplitter the δOPD due to a transverse thermal gradient is given by

$$\delta OPD = (d^2 + d \cdot l) \cdot D(\lambda) \cdot \gamma$$

where d is the thickness of the beamsplitter substrate and l is the separation between beamsplitter and compensator plate.

For the thermal gradient that may arise in one τ_o the path error is negligible at any λ.

As far as the infrared case is concerned, the integration time is longer, hence the related temperature variation is larger. Assuming for τ_o a mean value of 50 ms, the Strehl ratio due to a radial thermal gradient at the windows is $S = 0.999$.

The specification data of the infrared beamsplitter are not available at the moment.

As far as the primary mirror substrate is concerned, the δOPD is caused by the diurnal thermal variation in one τ_o of integration (i.e $\Delta T \simeq 2 \cdot 10^{-6}\,K$). Thanks to the very low coefficient of thermal expansion of zerodur such δOPL is completely negligible, as it is the effect due to the thermal gradient induced in the substrate, so that for the sake of

this calculation it is assumed $S = 1$.

Concern may arise if an unexpected ΔT shows up at one of the telescopes. However, for $\Delta T = 1\,K$ and $\gamma = 10^{-3}\,Kcm^{-1}$, affecting the primary mirror in one τ_o, the error introduced on the wavefront is small enough for the Strehl ratio to be of the order of 0.999 for visible wavelengths and of the order of 1 in the infrared range.

As far as the thermal effect due to dimension variation in the primary-secondary assembly is concerned, for the diurnal temperature variation during an integration time, we can be confident that, because of the compensation within the structure and the wavefront tilt correction by the piezo-mirror, the relative Strehl ratio is equal to 1.

In conclusion for expected thermal changes occurring in one τ_o of integration, we will allocate to the visibility budget a Strehl ratio $S_{ther} = 0.994$ which takes into account the effects on both external and internal windows and the beamsplitter.

3.2 Path fluctuations generated by vibration and other path uncertainty

In the section *Mechanical stability for IOTA*, we consider some of the sources of vibration for our instrument. As it was explained there, we are not able to perform an evaluation of the wavefront error induced by the different sources presented. The only exception is for the wavefront error due to vibrations induced by the vacuum pumps, after considering the attenuation obtained by the isolation system. For a mean wavefront error in the visible of $8.25 \cdot 10^{-4}\ \mu m$, the corresponding Strehl ratio is $S = 0.999$. Each of the other sources of disturbance (delay-line motion, mirror drives, ground vibration) results to have a power spectral density at high frequencies ($\geq 100\ Hz$) of the same order or less than that of the atmospheric disturbance, and should not be cause of evident visibility degradation.

Despite the lack of experimental information about disturbance induced by vibrations, in order to estimate an a priori evaluation of the expected Strehl ratio, we refer to the study on path fluctuations in the IOTA instrument presented in *Traub, 1989*. That study also gives information on the efficiency of fringe-tracking when a real-time estimate of the optical path delay, occurring during an integration time, is performed by means of a matched filter photon-counting fringe tracking algorithm. This algorithm was developed and tested for IOTA, with the aim of using the results from its application in a feedback loop to "fine-tune" the delay line position.

For the estimation of the Strehl ratio, we first consider instrument vibrations in the low- and mid-frequency range. For them we assume that the vibration suppression system of IOTA is able to reach an accuracy at the level of the residual tracking error which is that achieved in the computer simulation based on the use of the matched filter algorithm (*Traub et al., 1990*).

From an extrapolation of the data shown in fig.6 of *Traub et al. (1990)*, it can be assumed that the *rms* tracking accuracy will be of about $\lambda/25$, for an integration time τ_o and 100 photon counts per integration. This implies $S_{track,LF} = 0.939$.

The high frequency range ($\geq 100\,Hz$) vibrations are related to either undetected or un-

compensated optical path fluctuations. As far as these fluctuations are concerned, it is assumed that their variation in time do not cause an uncertainty in the visibility larger than 1%. This translates to a *rms* goal of about $\lambda/40$, and implies $S_{vib,HF} = 0.975$.

The overall Strehl ratio due to uncompensated instrument vibrations occurring during an integration time, τ_o is then given by $S_{vib} = 0.915$.

Finally, taking into account both thermal and vibration effects occurring during an integration time, we can estimate for IOTA $S_{OPD} \sim 0.909$. For the infrared band it results to be $S_{OPD,IR} \sim 0.993$.

tab. IV.3 - OPD variation: Strehl ratio breakdown for the visible band

S_{OPD}	S_{therm}	S_{vib}
0.909	0.994	0.915

4. Visibility loss due to diffraction effects

Diffraction effects may be significant over long distances, even for a wide beam. Therefore diffraction effects can seriously degrade the performance of an interferometer[65]. Several authors have calculated visibility loss for specific cases, however no analytic solutions, for beams with aberrations due to atmospheric turbulence, have been published to date. Here we will refer to the estimated visibility losses calculated for the CHARA project, which were obtained by means of numerical calculations performed by applying the rules of diffraction theory in the scalar formulation. In order to adapt those results to the case for IOTA, we took into account that the amount of diffraction effects in the system is proportional to the beam reduction factor squared and to the propagation distance. If it is assumed that $\Delta V \propto m^2 \cdot d$ then we have

$$\frac{\Delta V_{IOTA}}{\Delta V_{CHARA}} = \frac{m^2_{IOTA} \cdot d_{IOTA}}{m^2_{CHARA} \cdot d_{CHARA}},$$

which gives the scaling factor here considered. The propagation distance depends on the array configuration, but in general we can assume that one of the beams travels a distance of the order of the baseline being used (maximum baseline $\approx 40\,m$ for IOTA) and the beam from the other telescope travels via the long delay line a distance up to almost twice the baseline, depending on the declination of the object being viewed.

[65]The first effect of diffraction is loss of signal, which is always detrimental since it eventually affects the magnitude limit of the instrument. However, the loss of visibility resulting form the finite aperture of the instrument is more serious for the instrument performance, and besides it is baseline dependent. Actually, this loss of visibility would not arise if the path length in each arm of the interferometer were exactly the same. But this could only be the case if the interferometer could be always oriented so that the baseline was always normal to the direction of the star. *Tango and Twiss (1974)* have calculated the loss of visibility as a function of the beam size a and the distance travelled by the beam, d, for a perfect plane wavefront only affected by the entrance aperture.

From the CHARA simulation results, we consider those obtained setting the shortest propagation distance equal to 200 m. For the beam reduction factor, we have $m_{IOTA} = 10$ and $m_{CHARA} = 8$, so that the approximate relation between the visibility loss computed for CHARA and the extrapolated one for IOTA is given by

$$\Delta V_{IOTA} \simeq 0.3125 \cdot \Delta V_{CHARA}.$$

From the CHARA simulations we have that, in the visible ($\lambda = 550\,nm$) with very good seeing, $r_o(V) = 20\,cm$, and aperture of 20 cm, the diffraction free (this is obtained setting equal to zero the propagation distances) expected visibility is $V = 0.862$. When diffraction-induced losses are present the visibility is $V = 0.802$, which implies $\Delta V_{chara} = 0.06$.

Adopting the scaling relation we have for IOTA $\Delta V \simeq 0.019$. To calculate the coherence loss factor for IOTA we assume the same diffraction free visibility[66] as for CHARA, so that, from $\eta = \left(1 - \frac{\Delta V}{V}\right)$, we get $\eta = 0.978$.

For somewhat worse seeing conditions, $r_o(V) = 10\,cm$ and aperture diameter of 15 cm, the expected diffraction free visibility is $V = 0.743$, and with diffraction-induced losses is $V = 0.632$. The derived visibility loss for IOTA is $\Delta V \simeq 0.035$, for which we get $\eta = 0.953$.

Being aware of the rough approximation involved in deriving the IOTA values, we could allocate to the diffraction effects a Strehl ratio $S_{diff} = 0.960$, for a conservative approach, or $S_{diff} = 0.980$ for conditions of very good seeing. The latter value is exactly the Strehl ratio we calculate considering only the diffraction effects produced at the entrance aperture on a perfect plane wavefront (see *Tango and Twiss (1974)* for a Rayleigh length[67], of 3200 m).

For the infrared band we should consider the case of good seeing for which the CHARA simulation were performed assuming aperture diameter of the order of r_o (i.e., for CHARA, $r_o = 106\,cm$ and aperture of 100 cm), in order to have observing conditions similar to

[66] As far as the propagation distances are zero and the collecting area (20 cm) are the same, the model used for CHARA can in principle work as well for IOTA. In order to evaluate diffraction effects an atmospheric phase-plate model is constructed out of a finite number of waves. The amplitudes are deterministic and chosen to give the Kolmogorov structure function. Wave directions and phase shifts are randomly chosen. Wavefronts are next propagated through two telescopes and light-combining optics for given pathlength. The two beams are combined in terms of complex amplitude via a beamsplitter, in which the reflected beam is retarded by π radians. Then an estimate of the intensity is given by

$$I = I_0[1 \pm \gamma cos(\phi + r\pi/4)]$$

where ϕ is the random phase shift and $r = 0, 1$, and from which the visibility, γ, eventually is calculated.

[67] The Rayleigh length is defined as the quantity

$$D_R = \frac{\pi\, a^2}{4\,\lambda}$$

where a is the size of the beam.

those expected for IOTA (even if in this case the aperture is 45 cm). With these assumptions the expected visibility is $V = 0.867$ and the derived visibility loss for IOTA is $\Delta V \simeq 0.0081$, and which results in $\eta = 0.991$.

5. Visibility loss due to error in the wavefront tilt correction by the servo control system

The basic function of the tilt correction servo is to keep the interfering beams of the arms of the array parallel. If the difference in beam tilt is too large, losses in signal-to-noise will occur in the visibility measurement of the interferometer. The coherence transfer factor caused by the tilt error when the beams are combined is given by (*Busher, 1988*)

$$\eta = 1 - 1.8\langle(\theta/\theta_0)^2\rangle \tag{143}$$

where θ is the differential tilt angle between two combining beams and θ_0 is the angular radius of the Airy disc formed by the stellar image. Here we consider the Strehl ratio due to residual variance of tilt after the servo system. The two variance contributions examined are that related to the maximum frequency bandwidth of the mirror and that related to the detector noise (the transfer factor allocated to the surface quality of the mirror was already considered in the factor S_{flats}).

5.1 Bandwidth

In almost all tip/tilt servos the bandwidth of the system is limited by the performance of the wobbler mirror itself. In order to estimate the residual tilt variance we assume that the servo removes tilt at all frequencies up to the maximum bandwidth of the mirror, f_m, assumed to be greater than or equal to the frequency associated with the turbulence of the atmosphere, f_o (where $f_o = \frac{2\,v_{perp}}{\pi\,D}$, v_{perp} is the transverse wind speed, and D is the aperture diameter). With this assumption the residual variance of tilt is given by (*CHARA Technical Report No.15, May 1995*)

$$\sigma_\theta^2 = 0.0205 \left(\frac{D}{r_o}\right)^{\frac{5}{3}} \left(\frac{\lambda}{D}\right)^2 \left(\frac{f_m}{f_o}\right)^{-\frac{8}{3}}. \tag{144}$$

The coherence transfer factor can be written as a function of the Airy disc size and of the residual variance of tilt[68] as

$$\eta_{tilt} = 1 - 2.418 \left(\frac{D}{\lambda}\right)^2 \sigma_\theta^2 \tag{145}$$

so that eventually η_{tilt} is given by

$$\eta_{tilt} = 1 - 0.05 \left(\frac{D}{r_o}\right)^{\frac{5}{3}} \left(\frac{f_m}{f_o}\right)^{-\frac{8}{3}}. \tag{146}$$

[68]Because each beam has two axes of tilt, assuming the tilt error are gaussian in distribution, the tilt variance for one beam is given by $\sigma^2_{\theta,beam} = 2\sigma^2_\theta$ where σ^2_θ is the variance for one axis.

For the IOTA site and typical condition of seeing in the visible (at $\lambda = 555\,nm$, $r_o \sim 10\,cm$, $\tau_o \sim 6\,ms$, $v \sim 14\,m\,s^{-1}$ are standard values[69]) let us consider f_o may lie between between 20 and 35 Hz (for $v = 14$ and $v = 25\ m\,s^{-1}$ respectively). Assuming that the maximum bandwidth for the IOTA servo mirror is $f_m = 100\,Hz$, we have that $\eta_{tilt} = 0.991$ and $\eta_{tilt} = 0.962$ for the two values of wind speed considered. It is to be considered that in principle the maximum bandwidth of the servo mirror could be as high as $250\,Hz$. In this case the loss in visibility would be greatly reduced and we would have $\eta_{tilt} = 0.996$ even with a wind speed of $25\,m\,s^{-1}$.

5.2 Detector

The tilt-correction information necessary to guide the servo mirror is provided by reflecting a small part of the beam to a quadrant detector and using its output to detect the wavefront tilt. [70]

The error associated with angular position measurements using quadrant detectors is given by the expression derived by *Tyler and Fried (1982)*

$$\sigma_{qua} = \frac{\frac{3\pi}{16}\left(\frac{\lambda}{D}\right)}{SNR} \tag{147}$$

where a point-like source is assumed, and SNR is the signal-to-noise ratio of the four quadrants summed to act as a single detector

$$SNR = \frac{S}{\sqrt{S + 4\sigma_{read}^2}},$$

where σ_{read} is the read noise of the detector and S is the total signal, both in electrons.

By writing σ_{qua} as a function of the angular size of the Airy disc, the coherence transfer factor is given by

$$\eta_{qua} = 1 - 1.8\left(\frac{0.483}{SNR}\right)^2. \tag{148}$$

The CCD used with the quadrant detector in IOTA is expected to have σ_{read} of about 5 electrons. Assuming SNR = 10 (obtained for visual magnitude 12), it is $\eta_{qua} = 0.995$. Lower values of SNR would imply smaller η_{qua}, for example 0.983 for $SNR = 5$. [71]
The detector used for the first observations with IOTA had a larger read noise, and

[69] This is according to the relation $\tau_o = 0.81\,r_o/v$, *(Colavita et al., 1987)*

[70] A quadrant detector splits a focused stellar image into four parts or quadrants, with the light in each quadrant being separately detected. The image position is by definition centered on the detector when these four signals are equal in intensity, while an imbalance implies a centering error. The number of photon events registered in each quadrant is counted and latched by electronics, and read every cycle. From these the image position can be calculated and the appropriate signals sent to adaptive mirrors to re-center the image. Quadrant detectors are often used to study the wavefront tilt correction servo as well as higher order adaptive optics systems.

[71] In principle we could calculate the servo bandwidth for which the sum of the visibility loss due to the finite bandwidth and to the detector is minimum. Using the expressions of visibility loss from the

$SNR = 10$ was reached at a brighter visual magnitude.

In conclusion the breakdown of the Strehl ratio for the tilt error left after the servo system is given in tab. IV.4, assuming $v = 14\ m\,s^{-1}$, $f_m = 100\,Hz$ and $SNR = 10$; the servo system currently uses the visible portion of the wavelength range also when the observation is made in the infrared band, but an infrared star tracker is presently under construction.

tab. IV.4 - Servo system: Strehl ratio breakdown

S_{servo}	S_{tilt}	S_{qua}
0.986	0.991	0.995

equations which give η_{tilt} and η_{qua}, it is found

$$f_m = 0.505 \cdot \left(\frac{D}{r_o}\right)^{5/11} \cdot f_o^{8/11} \cdot S(m)^{3/11}$$

where D is the aperture and $S(m)$ is the number of photons detected in 1 s for a star of magnitude m. The CCD detector used for IOTA samples the entire wavelength band from 450 to 950 nm, and is reached by 50% of the flux from each telescope. Assuming $r_o = 10\,cm$ and $f_o = 20\,Hz$ it is found

$$\begin{aligned} m_V &= 10 \quad f \sim 306\,Hz \\ m_V &= 12 \quad f \sim 180\,Hz \\ m_V &= 14 \quad f \sim 110\,Hz \end{aligned}$$

These values are consistent with those presented in *Tango and Twiss (1980)* which were calculated for $D = r_o = 10\ cm$ wind speed of 5 $m\,s^{-1}$ and for an optical bandwidth of 100 nm centered at 500 nm.

6. Conclusion

In tab. IV.5 is presented the Strehl ratio breakdown for the visible band. These values are to be considered as indicators of the expected visibility reduction for IOTA, not as exact values. They were obtained on the basis of sometimes arbitrary assumptions or rough approximations. A similar breakdown for the infrared is presented in tab. IV.6. However, this is an even more approximate result, due to the lack of information about the specifications of the beamsplitter.

tab. IV.5 - Strehl ratio breakdown: visible wavelength range

S_{array}	S_{opt}	S_{OPD}	S_{diff}	S_{servo}
0.528	0.614	0.909	0.960	0.986

tab. IV.6 - Strehl ratio breakdown: infrared wavelength range

S_{array}	S_{opt}	S_{OPD}	S_{diff}	S_{servo}
0.683	0.704	0.993	0.991	0.986

The IOTA detection system in the visible was designed to operate with a 4-element lenslet array so that each of these elements selects an aperture corresponding to r_o. Because of this we can assume that both in the infrared and in the visible the ratio $\frac{D}{r_o}$ is ≈ 1. According to the expression that gives the coherence transfer factor for the atmosphere, with tilt-compensation (*Noll, 1976*)

$$\eta_{atm} = exp\left[-0.134 \cdot \left(\frac{D}{r_o}\right)^{5/3}\right] \tag{149}$$

for IOTA it is $\eta_{atm} = 0.875$, which is only an indicative value that will vary on a night by night basis, according to actual atmospheric conditions.

Given the estimation of the visibility loss due to the instrument and the indicative value for the coherence transfer factor due to the atmosphere, we expect a measured visibility of the order of 0.46. Eventually, for general estimation of the instrument performance, we should consider a decrement in visibility, with respect to the actual visibility of the object, of 50% .

For example, let us consider the magnitude limit that we could reach with the new CCD detector that is going to substitute the *PAPA* for observations in the visible band.

According to the calculations of the signal-to-noise ratio expected for the CCD (Millan-Gabet et al., 1995), we got that the potential magnitude limit for 10 ms of integration is of the order of 10 m_V with $SNR \simeq 14$. A reduction in the module of the fringe visibility

implies a reduction in SNR and hence in the magnitude limit. Assuming for IOTA a loss in visibility (for both atmospheric and instrumental causes) of 50%, the magnitude limit is of the order of 9 m_V with $SNR \simeq 12$. tab. IV.7 presents some of these results.

tab. IV.7 Magnitude limit and SNR obtainable with the 512X512 CCD used for IOTA detection in the visible, assuming a point-like source, for the ideal case $V = 1$, and the expected measured visibility $V = 0.46$

$V = 1$		$V = 0.46$	
m_V	SNR	m_V	SNR
2	687	2	343
4	273	4	136
6	108	6	54
8	42	8	21
10	14	9	12
11	7	10	7
12	3	11	3

Baseline orientation: OPD variation and u-v coverage

0. Introduction

Le potenzialitá osservative di un array interferometrico sono determinate dalle caratteristiche di ogni sua singola baseline: la lunghezza, da cui dipende la massima risoluzione raggiungible con data baseline, e la sua orientazione.

In particolare in questa sezione mi occuperó di valutare l'effetto dell'orientazione della baseline su alcuni dei parametri che definiscono le prestazioni di un interferometro a due elementi posto sulla superficie lunare. La rotazione della Luna fa variare la differenza di cammino ottico (OPD) percorso dalla radiazione raccolta dai due telescopi. Per mantenere le condizioni di coerenza tra i fasci interferenti tale variazione deve essere compensata. Cosí come la rotazione comporta una variazione in OPD, essa fa anche variare il modulo e l'orientazione della proiezione della baseline sul piano u-v : é proprio questo fatto che viene sfruttato per campionare punti diversi del piano con un'unica *baseline* ed elementi fissi. I risultati dello studio della variazione della OPD e della campionatura del piano u-v, in funzione dell' orientazione della *baseline*, sono alla base dello studio che conduce alla determinazione delle caratteristiche dell'*array* interferometrico, fra le quali: numero di *baselines* (e quindi di telescopi), orientazione e lunghezza di ogni singola *baseline*, latitudine del sito dell'*array*.

Nelle sezioni che seguono mi occuperó del caso di un interferometro a due elementi con *baseline* di 1 km. Dopo aver ricavato l'espressione generale per la variazione della OPD e della velocitá con cui tale variazione si produce, in funzione della rotazione lunare, presenteró i risultati del calcolo di queste quantitá per due casi particolari: orientazione N-S e orientazione E-W.

Nella seconda parte del capitolo mi occuperó invece dello studio della copertura del piano u-v ottenibile con diverse orientazioni della baseline. Le conclusioni di queste analisi saranno prese in considerazione nel capitolo *"Interazione fra sottosistemi: un esempio per l'interferometro lunare"* insieme ai risultati relativi allo studio della configurazione del telescopio e alle valutazioni emerse dell'analisi termica.

1. Expressions for the OPD and OPD rate

With reference to fig. B.1, let us define a selenocentric reference system whose reference plane (x, y) is the equatorial plane of the Moon, which is perpendicular to the Moon rotation axis, z, and intersects the surface of the Moon along the equator. With respect to this reference let (R, φ_1, l_1) and (R, φ_2, l_2) be the selenocentric coordinates of two telescopes, T_1 and T_2, on the Moon surface, and h_1 and h_2 their heights above the reference sphere.

fig. B.1 Selenocentric coordinates of the telescopes T_1 and T_2: φ is the latitude, l is the longitude, h is the height with respect to the mean radius, and R is the mean radius of the Moon. The unit vector $\hat{s}$ is the direction of a celestial object with coordinates l_0 and δ

The cartesian coordinates, $x_i,\ y_i,\ z_i$, for each telescope are given by

$$\begin{aligned} x_i &= (h_i + R)\, cos\varphi_i\, cosl_i \\ y_i &= (h_i + R)\, cos\varphi_i\, sinl_i \\ z_i &= (h_i + R)\, sin\varphi_i \end{aligned} \tag{1}$$

where R is the radius of the Moon, and $i = 1, 2$.

The two telescope vectors define a baseline vector, $\mathbf{B}$, whose cartesian components are given by

$$\begin{aligned} B_x &= (h_2 + R)\, cos\varphi_2\, cosl_2 - (h_1 + R)\, cos\varphi_1\, cosl_1 \\ B_y &= (h_2 + R)\, cos\varphi_2\, sinl_2 - (h_1 + R)\, cos\varphi_1\, sinl_1 \\ B_z &= (h_2 + R)\, sin\varphi_2 - (h_1 + R)\, sin\varphi_1 \end{aligned} \tag{2}$$

and whose modulus is given by

$$|\mathbf{B}| = \sqrt{B_x^2 + B_y^2 + B_z^2}. \tag{3}$$

Finally, let $\hat{\mathbf{s}}$ be the unit vector in the direction of a celestial object whose components are given by

$$
\begin{aligned}
s_x &= cos\delta\, cosl_0 \\
s_y &= cos\delta\, sinl_0 \\
s_z &= sin\delta
\end{aligned} \tag{4}
$$

where δ and l_0 are the declination and longitude of the object in the Moon centered reference system.
The difference in the optical path travelled by the light reaching the two telescopes is given by

$$
\begin{aligned}
OPD &= \mathbf{B} \cdot \hat{\mathbf{s}} \\
&= B_x\, s_x + B_y\, s_y + B_z\, s_z
\end{aligned} \tag{5}
$$

and in general OPD is a quantity that varies with time, since the telescope longitudes are changing, due to the Moon rotation, according to

$$
l_j(t) = l_r + l_{j,i} + \omega\,(t - t_i) \tag{6}
$$

where l_r is the longitude of a reference meridian circle, $l_{j,i}$ is the longitude of telescope T_j with respect to the reference circle at the initial instant, t_i, and ω is the lunar angular velocity.

In the two following sections I give the expressions for the baseline modulus, $B = |\mathbf{B}|$, the OPD, and $\Delta OPD/\Delta t$ for a generic orientation of the baseline, as a function of the latitude and longitude of the telescopes. With little loss of generality, I assume $h_1 = h_2 = 0$ and write the angular coordinates of the telescopes as

$$
\begin{aligned}
\varphi &= \varphi_1 \\
\varphi_2 &= \varphi + \Delta\varphi \\
l &= l_1 \\
l_2 &= l + \Delta l
\end{aligned}
$$

where $\Delta\varphi = \varphi_2 - \varphi_1$ and $\Delta l = l_2 - l_1$, and hereafter I will refer to the coordinates of telescope T_1 as to the coordinates of the array.

1.0.1 OPD

According to eqq.(2) and (5) the OPD for baseline coordinates (R, φ, l) and object unit vector $\hat{\mathbf{s}}$, is

$$
\begin{aligned}
OPD &= R\,\{[cos(\varphi + \Delta\varphi)\, cos(l + \Delta l) - cos\varphi\, cosl] \cdot cos\delta\, cosl_0\} \\
&\quad + R\,\{[cos(\varphi + \Delta\varphi)\, sin(l + \Delta l) - cos\varphi\, sinl] \cdot cos\delta\, sinl_0\} \\
&\quad + R\,\{[sin(\varphi + \Delta\varphi) - sin\varphi] \cdot sin\delta\}
\end{aligned} \tag{7}
$$

that, by using some fundamental trigonometric relations, and taking into account the time dependence of the longitude of the telescope, can be written as

$$\begin{aligned} OPD(t) &= R\{cos\delta\, cos(\varphi+\Delta\varphi)\, cos(l(t)+\Delta l-l_0) \\ &\quad -cos\delta\, cos\varphi\, cos(l(t)-l_0) \\ &\quad +[sin(\varphi+\Delta\varphi)-sin\varphi]\cdot sin\delta\} \end{aligned} \tag{8}$$

1.0.2 Baseline modulus

From eq.(3) and eq.(2) we can write

$$\begin{aligned} \frac{B^2}{R^2} &= [cos(\varphi+\Delta\varphi)\, cos(l+\Delta l)-cos\varphi\, cosl]^2 \\ &\quad +[cos(\varphi+\Delta\varphi)\, sin(l+\Delta l)-cos\varphi\, sinl]^2 \\ &\quad +[sin(\varphi+\Delta\varphi)-sin\varphi]^2 \end{aligned}$$

and, by using some fundamental trigonometric relations, we get

$$\frac{B^2}{R^2} = 2\cdot[1-cos(\varphi+\Delta\varphi)\, cos\varphi\, cos\Delta l-sin(\varphi+\Delta\varphi)\, sin\varphi]. \tag{9}$$

The quantity $\Delta\varphi$ has its maximum value when the baseline is oriented N-S, and is independent of the baseline latitude. By solving the previous equation with $\Delta l = 0$ we get

$$\Delta\varphi_{max} = arccos\left(1-\frac{B^2}{2R^2}\right) \tag{10}$$

For any orientation of the baseline at α degrees from the N-S direction, $\Delta\varphi$ will be given by

$$\Delta\varphi \simeq \frac{B\, cos\alpha}{R} \tag{11}$$

where I take the segment B as the arc $\widehat{B}$ since it is always $R \gg B$.
The quantity Δl reaches its maximum value when the baseline is oriented E-W. In this case, however, this value depends on the baseline latitude and increases as the latitude increases as it is shown by solving eq.(9) with $\Delta\varphi = 0$:

$$\Delta l = arccos\left(1-\frac{B^2}{2R^2\, cos^2\varphi}\right). \tag{12}$$

For a baseline at α degrees from the N-S direction, we have first to solve for $\Delta\varphi$ and then we can calculate Δl according to:

$$\Delta l = arccos\left[\frac{1-\frac{B^2}{2R}}{cos(\varphi+\Delta\varphi)}-tan(\varphi+\Delta\varphi)\, tan\varphi\right]. \tag{13}$$

1.0.3 Rate of the OPD variation

In order to integrate the signal from the source for a long period of time it is necessary to move the optical delay lines of the interferometer during the observations, to compensate for the variation in OPD due to the Moon rotation. Hence, another quantity of interest is the rate of the OPD variation, which is obtained by differentiating eq.(8) with respect to time, and is given by

$$\frac{dOPD}{dt} = \omega\, R\, cos\delta[cos\varphi\, sin(l(t) - l_0) - cos(\varphi + \Delta\varphi)\, sin(l(t) + \Delta l - l_0)]. \qquad (14)$$

For E-W baselines, $\Delta\varphi = 0$, and the rate of variation has a minimum value, never null, for Δl minimum, which means when the baseline is at the equator. For N-S baselines, $\Delta l = 0$, and the rate of the OPD variation depends, besides $l(t)$, on the latitude. However at the equator $\frac{dOPD}{dt} = 0$ since $OPD = const.$[1].

1.1 N-S baseline

A N-S baseline is defined by having both telescopes at the same longitude, so that $\Delta l = 0$. By making this substitution in eq.(8), and (12), we get:

$$\begin{aligned} OPD_{NS}(t) &= R \cdot \{cos\delta\, cos(l(t) - l_0) \cdot [cos(\varphi + \Delta\varphi) - cos\varphi] \\ &\quad + sin\delta \cdot [sin(\varphi + \Delta\varphi) - sin\varphi]\} \qquad (15) \\ \frac{dOPD_{NS}}{dt} &= \omega R\, cos\delta \cdot [cos\varphi - cos(\varphi + \Delta\varphi)] \cdot sin(l(t) - l_0) \qquad (16) \\ \Delta\varphi &= arccos\left(1 - \frac{B^2}{2R^2}\right). \qquad (17) \end{aligned}$$

Fig. B.2a and B.2b show the results of some computations of the OPD_{NS} variation and its rate of variation calculated for baselines at different latitudes and for various object's declinations. The plots show the variation of the OPD length and the rate of the OPD variation during a lunar night, as a function of time expressed in hours (the lunar night lasts about 328 hours).

Equator

Any N-S baseline at the Moon equator has a constant OPD whose value increases from

[1] This is exactly true when $\varphi = 0°$ is the position of the central hub, equidistant from the two telescopes. This means that in eq.(8) and (12) φ has to be replaced by $-\frac{\Delta\varphi}{2}$ so that when $\Delta l = 0$ the terms which depend on time cancel each other, and the constant OPD value is given by the third, time independent, term.

However we can in general assume that OPD is constant at the equator, when considering baselines of few km, even keeping the assumption that φ is the latitude position of one of the telescopes. In fact, $\Delta\varphi$ is small enough, for the quantity that multiplies the term which is a function of time, $[cos\Delta\varphi - 1]$, to be always $\simeq 0$. For example, for a 1-Km baseline, $\Delta\varphi = 0°.033$ and the cosine difference, at the equator, is of the order of $1.6 \cdot 10^{-7}$ which implies a maximum OPD variation of 28 cm in about 6.5 days, for an object at $\delta = 0°$. For an object at $\delta = 80°$ this variation reduces to 5 cm in 6.5 days. This variation in OPD can be compensated by moving the short delay lines instead of the long ones.

0 to about the baseline length when the object's declination increases from 0° to 80°.

Other baseline latitudes

For any other baseline's latitude the OPD is not constant, and its variation with respect to the time elapsed from the moment of culmination of the object, depends both on the position of the baseline and the object declination. However, for any baseline's latitude, the variation is larger for objects closer to the equator than to the pole.

As far as the OPD rate variation is concerned, this means that the maximum rate is found when $\delta = 0°$. For a baseline at 80°, the maximum latitude considered here, the highest rate is of the order of 10 meters per hour, about 0.28 $cm\ sec^{-1}$. However this value is reached many days either before or after the culmination of the object. For a time interval of 100 hours centered on the culmination time, the speed of OPD variation is never larger than 5 meters per hour, about 0.14 $cm\ sec^{-1}$.

fig. B.2a OPD variation (left) and rate of OPD variation (right) for a N-S baseline at 0° and 30° of latitude and different object's declinations

fig. B.2b OPD variation (left) and rate of OPD variation (right) for a N-S baseline at 45° and 80° of latitude and different object's declinations

1.2 E-W baseline

An E-W baseline is defined by having the two telescopes at the same latitude, so that $\Delta\varphi = 0$ and by substitution in eqq.(8), and (12) we get:

$$OPD_{EW}(t) = R \cdot \{cos\delta\, cos\varphi \cdot [cos(l(t) + \Delta l - l_0) - cos(l(t) - l_0)]\} \quad (18)$$

$$\frac{dOPD_{EW}}{dt} = \omega R\, cos\delta\, cos\varphi \cdot [sin(l(t) - l_0) - sin(l(t) + \Delta l - l_0)] \quad (19)$$

$$\Delta l = arccos\left(1 - \frac{B^2}{2R^2\, cos^2\varphi}\right). \quad (20)$$

From eq.(19) we see that the rate of OPD variation is maximum when the interferometer reaches the object meridian[2]. Fig. B.3a and B.3b show the results of some computations of the OPD_{EW} variation and its rate of variation during the period corresponding to about one lunar night.

Equator

An E-W baseline located at the Moon equator has the maximum range of variation in OPD attainable for any baseline orientation. The maximum value of OPD variation is obtained when also the object is equatorial. In proximity of the culmination point[3], where any OPD is null, at any latitude for an E-W baseline, the rate of the variation is maximum, with the largest value being of the order of 10 meters per hour ($0.28\ cm\ s^{-1}$).

Other baseline latitudes

For any other baseline's latitude the OPD variation is smaller and decreases as the latitude increases. The minimum variation is found for a baseline at 80° of latitude and an object at the same latitude. The rate of variation in consequently smaller than at equatorial latitudes.

The interesting difference with the N-S baseline is that the rate of variation now reaches its largest values when approaching the culmination point. In particular, the maximum value ($0.28\ cm\ s^{-1}$) is obtained for a baseline at the equator and object declination $\delta = 0°$.

[2] In eq.(19):

$$sin(l(t) - l_0) - sin(l(t) + \Delta l - l_0) = -2\, sin\left(\frac{\Delta l}{2}\right)\, cos\left(l(t) - l_0 + \frac{\Delta l}{2}\right)$$

and this quantity reaches its maximum value when $l(t) = l_0$.

[3] According to the choice adopted to express the baseline coordinates, this happens when $l(t) - l_0 = \frac{\Delta l}{2}$

fig. B.3a OPD variation (left) and rate of OPD variation (right) for a E-W baseline at 0° and 30° of latitude and different object's declinations

fig. B.3b OPD variation (left) and rate of OPD variation (right) for a E-W baseline at 45° and 80° of latitude and different object's declinations

1.3 A particular case: N-S baseline parallel to the lunar spin axis

In this paragraph I want to briefly address the requirement, that sometimes appears in the literature, of having no OPD variation to compensate for during the whole integration period. As I showed before, this is possible only when considering an equatorial baseline with orientation N-S. For this requirement to be satisfied at any latitude, the two time dependent terms in eq.(8) must be null. This is equivalent to saying that the baseline has to be oriented N-S and be parallel to the Moon spin axis[4]. To satisfy this last requirement we have to assume that one of the two telescopes (at least) is elevated with respect to the surface by a height h.

By using eq.(2) and requiring that $B_x = B_y = 0$, and $B_z = |B|$, and putting $l = 0$, we get

$$\begin{aligned} (h+R)cos(\varphi+\Delta\varphi) &= R\,cos\varphi \\ (h+R)sin(\varphi+\Delta\varphi) &= R\,sin\varphi + B \end{aligned}$$

so that, by solving for the height h, we find that h varies with the latitude according to

$$h(\varphi) = \sqrt{R^2 + 2BR\,sin\varphi + B^2} - R. \tag{21}$$

In fig. B.4 the quantity h is plotted as a function of the latitude of the baseline, φ. In this case one of the telescopes is assumed to be at the surface level and only the other one is elevated.

From the plots we see that the amount by which the telescope has to be elevated is quite large, even at small latitudes. For example, we notice that at a latitude of the order of $6°$, the telescope should be elevated by about 100 m in order to keep the baseline parallel to the spin axis.

[4] In equation (5) the two terms which are time dependent are $B_x\,s_x$ and $B_y\,s_y$. Putting them equal to zero leads to

$$OPD = B_z\,s_z = R\,[sin(\varphi+\Delta\varphi) - sin\varphi]\cdot sin\delta$$

which, for any given object, is constant.
However this also means that the baseline is oriented parallel to the *z-axis* which, according to the coordinate reference system used, is parallel to the spin axis of the Moon.

fig. B.4 Elevation, h, with respect to the lunar surface of one of the telescopes as a function of the latitude of the baseline, φ. The left plot shows the amount of elevation on the range 0° to 20° North latitude

For a normal N-S baseline at 6° of latitude the maximum (i.e. when the object is at 0° of declination) absolute variation in one night is of the order of 100 m, but in a period of 100 hours centered on the culmination point the maximum variation is of 45 m only, and for a period of 10 hours is less than 5 cm. As far as the speed of OPD variation is concerned, the maximum value through the whole night is less than 0.3 $mm\,s^{-1}$.

According to these results we may conclude that for latitudes close to the equator (up to 30°, the maximum OPD variation, in a time interval of 100 h centered on the culmination point, is about 55 m (in 10 h it is about 5 m), and for integration periods of several hours, the OPD variation could be easily compensated by means of a short delay line. Of course, this implies that a long coarse delay line is present too. The only advantage obtained by elevating the telescope is that there is no need for a variable long delay line. This however implies a tight restriction on the scientific capability of the interferometer: such an instrument could be used just for astrometric measurements and only for objects in the declination range which implies ΔOPD that are covered by the short delay line. Besides, we should take into account the large amount of structural material which is necessary to deliver on the Moon in order to build the telescope support.

In conclusion, although on the basis of rough consideration, I would completely neglect the solution of the telescope elevation in order to keep the OPD constant, and rather concentrate on the analysis of a classical N-S baseline configuration.

2. U-V coverage

For the definition of the u-v coordinates I refer to the section *Principles of stellar interferometry*. Here I just recall that the u-v plane coverage is determined by the projection in the direction perpendicular to $\hat{\mathbf{s}}$ of a baseline between two telescopes of an array, and that, especially when imaging is required, the possibility of meeting specific scientific goals is determined by the number of u-v points sampled by means of the chosen telescope array configuration[5].

The baseline components, with respect to a right-handed coordinate system, are usually indicated by $(u,\ v,\ w)$ and are measured in units of the central wavelength of the observing band. u and v are measured in a plane normal to the direction to the center of reference of the source, v toward the north as defined by the plane through the origin, the source, and the pole, and u towards the east. The component w is measured in the direction of the center of reference of the source (*Thompson et al. 1986)*.

According to the selenocentric reference system previously chosen for the description of the variation of the orientation of the baseline vector, we can explicitly calculate the $(u,\ v,\ w)$ baseline coordinates.

Let w be in the direction of the point in the sky of declination δ and latitude l_0. v is 90° from w in the plane of the pole and the origin. u, which has to be orthogonal to both w and v, will lie in the equatorial plane rotated by 90° with respect the intersection of the $w-v$ plane with the equatorial plane (see fig. B.5).

[5]The u-v coverage improves by increasing the number of telescopes. However, the cost and complexity of the array increases with the number of telescopes, too, so that a trade-off must be made, at some point, between the most complete u-v coverage and the cost of implementation of the array.

fig. B.5 Relationship between the $(x,\ y,\ z)$ and the $(u,\ v,\ w)$ coordinate system

If B_x, B_y, B_z, are the baseline components as defined in *section 1.*, the transformation to get the components of the baseline in the new coordinate system is given by

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} -sinl_0 & cosl_0 & 0 \\ -cosl_0\, sin\delta & -sinl_0\, sin\delta & cos\delta \\ cosl_0\, cos\delta & sinl_0\, cos\delta & sin\delta \end{pmatrix} \cdot \frac{1}{\lambda} \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix}. \tag{22}$$

By taking the equation for u and v, squaring and adding them, we find the locus of the projected baseline components u and v, which is described by the equation of an ellipse where the object's relative longitude (i.e. the hour angle) acts as the free parameter. The equation for the ellipse is

$$u^2 + \left(\frac{v - \frac{B_z}{\lambda}\, cos\delta}{sin\delta} \right)^2 = \frac{B_x^2 + B_y^2}{\lambda^2}. \tag{23}$$

The arc of ellipse that is traced out during any observation depends on the selenocentric coordinates of the baseline, the declination of the source, and the range of hour angle, or degrees from the zenith, covered.

Some examples of u-v coverage are shown in fig. B.6a, B.6b, and B.6c for different baseline orientations (N-S, E-W and NW-SE), baseline latitudes (0°, 45°, 80°) and source declinations (0°, 30°, 60°). The u and v values are calculated according to

$$u = -\frac{B_x}{\lambda}\, sinl_0 + \frac{B_y}{\lambda} cosl_0 \tag{24}$$

$$v = -\frac{B_x}{\lambda} cosl_0\, sin\delta - \frac{B_y}{\lambda} sinl_0\, sin\delta + \frac{B_z}{\lambda} cos\delta \quad (25)$$

from which we see that for any object with $\delta = 0°$ the ellipse degenerates in a straight line, because B_z depends only on the latitude of the baseline and is not affected by the rotation of the Moon.

Each of the figs. B.6a, B.6b, and B.6c presents a set of 9 u-v coverage plots for, from left to right, a NS, NW-SE, and E-W orientation, and for baseline latitudes of 0°, 45°, and 80° going from bottom to top. Each diagram presents the result for ten baselines, whose length varies from 100 m to 1000 m. The points covered with a 1 km baseline are always the outermost in the plot, as indicated by the label 1000. The hour angle range is 180°, i.e. it was assumed that an integration could last for the entire lunar night. The u and v coordinates are given in unit of giga-rad (10^9 rad) with $\lambda = 0.5$ μm.

Fig. B.6a presents the u-v coverage for a source at 0° declination. As I explained above, for any orientation of the baseline the ellipse traced by the projection of the baseline degenerates in a straight line. In the case of the E-W baseline, since the B_z component is null, the v coordinate is always zero, so that, notwithstanding the latitude of the baseline, the u-v coverage that can be obtained with this set of baselines is always the same. The NW-SE baseline is the one that has the best coverage for objects at 0° declination, and by increasing the latitude, the coverage gets more compact than for the N-S baseline.

Fig. B.6b presents the u-v coverage for a source at 30° declination. The coverage is improved with respect to the the 0° declination, for any orientation. The E-W orientation allows, at any latitude, coverage of a complete ellipse on the u-v surface: at 0° of latitude this is the orientation that gives the best coverage. At higher latitudes, also the NW-SE orientation allows a good coverage, even if the points are spread on a large area.

Fig B6c presents the u-v coverage for a source at 60° declination. This is the case for which each baseline orientation presents the wider, and for the N-S and NW-SE baselines also the more complete, coverage. The E-W orientation offers the best coverage, as far as the number of points covered with each baseline is concerned, at any latitude, however the NW-SE is competitive when baselines up to 300 - 400 m are considered.

The case for a N-S baseline at 0° latitude is the same, notwithstanding the declination of the object. The u-v coverage is always very poor because, as we have seen in the previous sections, the projection of the baseline on the plane of the sky is constant.

Finally a comment about the number of baselines for which the u-v coverage has been calculated. For the prototype of the lunar interferometer I have been considering in this study, I assumed two elements separated by a baseline of 1 km. In this case only the outermost lines in each diagram should be considered for the evaluation of the u-v coverage. The coverage shown in the plots could be obtained by adding three fixed elements on the same line, at different spacings (actually, with 5 elements, whose spacing is a multiple of 100 m, as it is assumed in the plots, we would have only 9 non-redundant

baselines, so that any one of the curves should be missing. However, in general with 5 elements it is possible to find a linear configuration which is non-redundant and hence to have 10 different baselines). However, any additional collecting element should be rather exploited to obtain different baseline orientations. It is evident that none of the solutions presented in figs. B.6a, B.6b, and B.6c, because of the gaps in the u-v coverage, is optimal for imaging[6]. The linear configuration is still adequate for stellar diameter and limb-darkening measurements and for studies of orbits of binaries. Apart from this consideration, the main reason for choosing two-dimensional arrays is for imaging more complex objects than binary stars, a capability that can be greatly enhanced by the extensive u-v coverage obtainable by appropriate configuration of a two-dimensional array.

Hence, as last step in this process, the configuration of the array has to be addressed. For a given number of telescopes, the u-v plane coverage has to be optimized by a number of trials with different configurations (often considered are a "Y" shape array and a more general "cobweb" array). Finally, the u-v coverage results should be evaluated according to the science objectives that we want to accomplish, and to the cost-benefit ratio calculated by increasing the number of collectors. However, this additional study is beyond the scope of my present work. In the section *Interaction among subsystems: an example for the lunar interferometer* I will consider the case for a two-element array with a 1 km baseline, assuming that this will be the first stage of a multi-element array to be built on the lunar surface.

[6]The minimum number of telescopes that can provide imaging capability is three, because it provides one closure phase which is necessary to obtain any information on the phase of the complex visibility (see section *Principles of stellar interferometry*).

fig. B.6a u-v coverage for a NS, NW-SE, E-W orientation (going from left to right), and for baseline latitudes of 0°, 45°, and 80° (going from bottom to top). Results are presented for 10 different baselines, from 100 to 1000 m: the curve for 1000 m is always the outermost. Source at 0° declination

fig. B.6b u-v coverage for a NS, NW-SE, E-W orientation (going from left to right), and for baseline latitudes of 0°, 45°, and 80° (going from bottom to top). Results are presented for 10 different baselines, from 100 to 1000 m: the curve for 1000 m is always the outermost. Source at 30° declination

fig. B.6c u-v coverage for a NS, NW-SE, E-W orientation (going from left to right), and for baseline latitudes of 0°, 45°, and 80° (going from bottom to top). Results are presented for 10 different baselines, from 100 to 1000 m: the curve for 1000 m is always the outermost. Source at 60° declination

The siderostat plus beam-compressor telescope solution for a lunar interferometer

0. Introduction

In this section I analyze some of the aspects which have to be taken into account when choosing the telescope configuration for the elements of a lunar interferometer. In particular I analyze the case of a telescope consisting of the combination of a siderostat (the flat mirror that collects the light from the source) followed by an afocal beam-compressor. I show that by defining two angular parameters (maximum and minimum incidence angle on the siderostat) we can derive a set of requirements and consequent trade-offs which involve: declination and hour angle ranges, dimension of the siderostat, compactness of the telescope support structure.

The results of this analysis will used later, in relation with those of the analysis of the thermal environment and those of the study on the baseline orientation, in the section *Interaction among subsystems: an example for the lunar interferometer*.

In *section 1.* I will briefly address the motivation for the choice of a siderostat plus beam-compressor solution for a lunar interferometer, while the analysis of the telescope parameters is presented in *section 2.*. The last section deals with other issues related to the telescope configuration, which are not main drivers for the choice of the configuration parameters of the telescope.

1. Siderostat plus beam-compressor

For any stellar interferometer it is important to choose an instrument configuration that keeps the number of reflections to a minimum, in order to optimize the transmission of the radiation. This is important when the number of telescopes is large: each beam of a N-element array has to be split (N-1) times in order to interfere with all the other beams and, in addition, the intensity of the beam is decreased by a factor that depends on the number of reflections in the beam combination scheme. Eventually the minimum number of reflections requirement has to be addressed at the level of two subsystems: the telescope assembly and the relay+beam combination system.

As far as the telescope assembly is concerned, the two solutions which imply the minimum number of reflections are a three-mirror Coudé telescope, and a siderostat plus a fixed Cassegrain telescope (which acts as a compressor).

In the coudé solution, independent drive motors are used to drive the elevation and azimuth axes, to move the large tube assembly of the primary and secondary mirror, and the final mirror at the coudé focus is driven by another independent mechanism. With respect to a conventional telescope design like this, the advantage of a siderostat is that the mechanical structure, which must be moved during tracking, can be made considerably more compact. It has to move only a single optical flat rather than the whole telescope assembly. This helps in designing an optical system which has to be movable but also

has to be interferometrically stable. The main advantage comes from keeping fixed the beam-compressor and hence maintaining, at least from a mechanical point of view, the primary-secondary mirror alignment. The siderostat solution allows also an easier access to the pivot on which it turns. This is particular useful in order to simplify the metrology system which has to constantly control the position of the baseline[7] .

In conclusion, these are some of the advantages according to which the siderostat plus beam-compressor solution can be considered a good candidate for a lunar interferometer:

- smaller size and weight of the element that needs to be moved;
- removal of the distortion of the telescope and, consequently, of its emerging wavefront, due to a changing of the gravity vector orientation with respect to the telescope[8];
- freedom from increased mechanical complexity (this may have positive implications in the power budget too);
- easy access to the pivot of the siderostat (this helps in simplifying the metrology system);
- possibility for the beam-compressor optics to be completely shielded by dust contamination, micrometeroids impacts, and to be permanently thermally isolated .

The main drawback is probably the fact that the siderostat is an additional large optical element that increases the payload mass. Also, its large surface area is directly exposed to any environmental contamination.

As far as the telescope mounting is concerned, we can considered both an altitude-azimuth (alt-az) and equatorial mounting[9]. The alt-az mount may be preferred for its mechanical simplicity, with respect to the equatorial mount, but this is at the expense of additional drive circuitry and software, since for the alt-az the required drive rates vary continuously over a wide range of speeds. On the contrary the equatorial mount requires a constant drive rate for rotation around the polar axis, that in the case of the Moon is quite low, about 0.54° *per hour*. However, the issue of the telescope mounting should be addressed at a later stage of the study, when constraints on other parameters, such as the orientation of the telescope and the site latitude, are established.

[7]For each collector pair, the effective baseline is defined as the vector joining the intersection of the rotation axes of the individual collectors.

[8]This may be a minor issue if considered with respect to the effects on a ground-based instrument (the modulus of the gravity vector on the Moon is about 1/6 of the gravity value on Earth), but not with respect to a orbiting instrument which would operate in a *0-g* environment. Also, in this way, the related wavefront error is constant and this reduces the variables in the evaluation of the error budget of the instrument.

[9]In the equatorial mounting one of the axes is directed towards the celestial pole. It is called the *polar axis or hour axis*. The other one, the *declination axis*, is perpendicular to it. Since the polar axis is parallel to the rotation axis of the Earth (or equivalent direction on the Moon), the apparent rotation of the sky can be compensated for by turning the telescope around this axis at a constant rate.
In the alt-az mount one of the axes is in the direction of the local zenith, the other is horizontal. In order to follow the rotation of the sky, the telescope must be turned around both of the axes with changing velocities.

2. Configuration analysis

For the analysis of the configuration with a moveable siderostat and a fixed beam-compressor, I choose to describe the characteristics of the system as a function of two parameters: the incidence angle on the siderostat and the "obscuration" angle due to the telescope support structure, which is the angle, counted from the telescope axis, that is intercepted by the structure of the telescope. In particular I define the quantities:

a) α as the maximum incidence angle on siderostat. Limits on α are due to polarization effects (the phase shift between p and s polarizations begins to change rapidly for incidence angles $\geq 45° - 50°$ (*Born and Wolf, 1980; Traub, 1988*)) and scaling of the size of the siderostat with respect to the diameter of the primary mirror of the beam-compressor.

b) β as the minimum incidence angle on the siderostat, which is equal to one half of the obscuration angle. β depends on the compactness of the telescope, that means on the separation between beam-compressor and siderostat: the larger is the separation the smaller is β, but also the larger is the overall structure.

Both α and β are measured with respect to the normal to the siderostat.

2.1 Relation between α and β and the dimension of the telescope components

There is a direct relation between the choice of the angular parameters α and β and the dimensions of the siderostat (which is related to that of the primary mirror) and of the compactness of the telescope. In both cases these choices directly affect the mass budget of the instrument, which, for an instrument that has to be delivered on the Moon, is quite an important issue. In what follows, I present these relations while some conclusions about the trade-off between mass budget and maximum angular access are presented in *Interaction among subsystems: an example for the lunar interferometer*.

Since the definition of the relation between angular parameters and telescope dimensions is independent of the telescope orientation, I consider the horizontal configuration as shown in fig. TC.1. Let us consider the two limit situations, when the angles of incidence with respect to the horizontal axis are the minimum and the maximum allowed.

From fig. TC.1, we see that all beams with incident directions between 2β and 2α are reflected by the siderostat into the telescope, in a direction parallel to the telescope axis. This also implies that the siderostat has to have a rotation range of $(\alpha - \beta)$ with respect to the telescope-siderostat axis.

As far as the two linear dimensions of the siderostat are concerned, we can see that the one which is parallel to the rotation axis of the siderostat has to be of the same order as the dimension of the primary mirror diameter, whereas the transverse dimension, i.e. that which is perpendicular to the rotation axis, scales as the inverse of the cosine of the maximum incident angle. This requirement is due to the fact that we want to exploit the maximum light collection area also at large incident angles.

fig. TC.1 Minimum (β) and maximum (α) incidence angles counted from the normal to the siderostat: their relation with the linear dimension of the siderostat

For a given primary mirror diameter, d, the relation between α and D, the transverse dimension of the siderostat, is given by:

$$D = \frac{d}{cos\alpha} \tag{26}$$

Tab. TC.1 presents some values of D for different α, as a function of the parameter d.

tab. TC.1 Siderostat transverse dimension, D, for different values of the maximum incidence angle, α (deg), as a function of the primary mirror diameter, d

α	0.0°	30°	45°	50°	55°	60°	65°	70°	75°	80°
D	1.0 d	1.15 d	1.41 d	1.55 d	1.74 d	2.0 d	2.37 d	2.92 d	3.86 d	5.75 d

For incident angles α less than 60° the overall area of the siderostat is less than twice the primary mirror area. This is probably the largest scale factor we should considered since for incident angles larger than $45° - 50°$, the polarization effects begin to change rapidly[10]. From eq.(1) we see that the choice on the size for the collecting area, d, (which

[10]In general, an interferometer can exhibit degraded fringe visibility, arising from polarization-dependent phase shifts at mirror surfaces, even when the light source itself is unpolarized. The degradation

also is related to the magnitude limit reachable by the instrument) eventually determines the fan of angles accessible to the observation, since there is a maximum limit in the dimension D of the siderostat imposed by the mass budget and its relation with the overall transportation costs.

It is useful to formulate a relation which links the diameter of the primary mirror, d, and the ratio between the siderostat dimension D and d (i.e. the coefficient $\delta = \frac{D}{d}$).
For the siderostat I consider an octagonal shape, whose maximum width on the two orthogonal directions are D and d: see fig. TC.2 for definition of the linear dimension of the siderostat.

fig. TC.2 Linear dimension of the siderostat: d is the size of the primary mirror diameter

Then the area of the siderostat (calculated by subtracting from the rectangular area $D \times d$ the areas of the 4 triangles at the corner of the rectangle) is given by

$$A_{sid} = 0.85 \cdot \delta \, d^2 \tag{27}$$

For a given areal density (mass per unit area) ρ_A, the mass difference, Δm, for two different values of δ is given by

$$\Delta m = 0.85 \cdot \rho_A \, d^2 \, \Delta\delta \tag{28}$$

factor ranges from 1 to 0. In principle, if these effects are the same for each telescope, then the net effect on the combined beams is zero. To take into account for likely (small) differences in the two systems, different design approaches can be considered to reduce the effect of polarization degradation (see for example *Traub, 1988*), which, however, will take advantage of small incidence angles to keep the phase shift at each reflection as small as possible.

and the mass difference assuming the same value of δ but different primary mirror diameters ($d_2 > d_1$) is given by

$$\Delta m = 0.85 \cdot \rho_A \, \delta \, d_1^2 \left(\frac{d_2^2}{d_1^2} - 1 \right) \tag{29}$$

Fig. TC.3a shows the linear relation between Δm and $\Delta\delta$ (which is equivalent to the relation between Δm and maximum incidence angle) and fig. TC.3b shows the quadratic relation between Δm and the ratio $\frac{d_2}{d_1}$.

fig. TC.3a, left Relation between Δm and $\Delta\delta$ for a given primary mirror diameter d. Note that the quantity plotted in abscissa is actually $\delta = \Delta\delta + 1$ and the range of δ values considered corresponds to a range for the incidence angle from 0° to 80°, as in tab. TC.1
fig. TC.3b, right Relation between Δm and the ratio $\frac{d_2}{d_1}$ when the primary mirror diameter is increased from d_1 to d_2; the maximum incidence angle does not change

With reference to fig. TC.4, let us now consider the parameter β.

fig. TC.4 Relation between the angular parameter β and the separation between the siderostat and the beam-compressor

β is related to the distance, l, between the boundary of the telescope support, which is responsible for the obscuration cone on the siderostat, and siderostat by :

$$\beta = \frac{1}{2} atan\left(\frac{d}{l}\right) \tag{30}$$

where the separation l is measured taking into account the possibility of shaping the telescope tube, by "cutting" the tube in the area around the secondary mirror, so that β is the minimum incidence angle obtainable. We have also considered that the whole siderostat area has to be available to collect light coming from the direction β.

Tab. TC.2 shows some values of β and the corresponding separation between siderostat and beam compressor as a function of the parameter d.

tab. TC.2 Angular parameter β, in degrees, and corresponding separation between the siderostat and the beam-compressor, as a function of the primary mirror diameter, d

β	5°	10°	15°	20°	22.5°
l	5.76 d	2.75 d	1.73 d	1.20 d	1.0 d

For a given diameter d, to reduce the minimum incidence angle requires increasing the separation between the siderostat and the beam compressor. Viceversa, if we want to reduce β and keep the telescope as much as compact is possible, we need to decrease the

diameter of the primary mirror.

Eventually the choice of the size the diameter of the primary mirror is going to set the limit on the angular range accessible to our optical system, and to affect the mass budget of the instrument, not only for its own contribution, but also because it determines the size and hence the mass of the siderostat and of the supporting structure. For the latter case we are not able, at this early phase of the study, to define a scaling relation between primary mirror size and mass of the telescope support. This would require an analysis of the telescope metering and support subsystems that should be undertaken in a following stage of the design.

According to the performances of presently available and of next generation launchers, a preliminary assessment for the mass of the scientific payload delivered on the Moon ranges between 700 kg and 1400 kg, depending on the propulsion system and on the Earth-orbit to Moon-orbit trajectory chosen (Matra Marconi Space estimations based on an Arianne 5 class launcher performances). According to this assessment, the mass budget allocation to the various elements of the interferometer have to be iteratively generated during the design of the instrument and eventually it will be consolidated in a system level mass budget which ensures compatibility with the launcher mass capability.

As far as the telescope is concerned, we can assume a mass allocation of the order of 15% of the total payload mass, which has to be suballocated among a number of telescope subsystems, including: 1) mirrors, 2) structure, 3) electronics, 4) thermal control, 5) alignment sensors. The largest portion (of the order of the 75-80%) of the telescope mass is allocated to the mirrors and structure subsystems[11], which means that for them together we can allocate from a minimum mass of about 84 kg to a maximum mass of about 168 kg.

2.2 Orientation of the telescope

For the combination siderostat plus beam-compressor, the solution with the telescope positioned vertically, i.e. with the optical axis directed toward the nadir, is not optimal, because the obscuration due to the structure prevents observing around the zenith. Besides, it is the position that requires the largest number of reflections (three) to relay the beam from the end of the telescope to the plane where the combining area is, that hereafter I assume to be the plane of the horizontal axis.

If the telescope is oriented horizontally, there is no need for any relay mirror, since the beam comes out parallel to the horizontal axis. In this case, in order to observe towards the zenith, α should be of 45° and probably we would not go much further beyond the zenith direction because of polarizations effects. The effects of the thermal radiation from the sun and the Moon surface should also be taken into account. The vertical solution is

[11] The estimation on the mass budget allocation are based on a comparative evaluation of data presented in the design project of LUTE (*Lunar Ultraviolet Telescope Experiment, NASA, 1994)* and the preliminary design study by SAIC: *A lunar-based interferometer design for early detection of extra-solar planets (1995).*

the worst, because the mirror area exposed to both fluxes is maximum. The horizontal solution would be the best one, because the mirror area exposed to both the solar and lunar heat flux is minimum.

In order to find a compromise between minimum heat flux exposure and "best sky view", I decided to determine the condition on the telescope orientation according to the limitation imposed by our system on the access to objects close to the zenith direction. In particular I set the requirement that the $2(\alpha - \beta)$ fan of accessible angles is centered on the zenith.

With reference to fig. TC.5, we see that, in order to meet this requirement, it is necessary to tip the telescope with respect to the horizontal axis by

$$\gamma = 90^\circ - [(\alpha - \beta) + 2\beta] = 90^\circ - (\alpha + \beta) \tag{31}$$

fig. TC.5 Orientation of the telescope

Besides this constraint, which defines the position of the telescope with respect to the horizontal, in order to determine the orientation we have to take into account the limitation on the accessible sky. This is set by the angular rotation range of the siderostat with respect to the telescope-siderostat axis, and hence depends on the orientation of the telescope axis: the limitation in angle coverage is in the plane containing the telescope axis, while in the direction perpendicular to the telescope axis the range of accessible angles is, in principle, unlimited.

fig. TC.6a, left Limitation on the range of accessible declination angle when the telescope axis is oriented N-S
fig. TC.6b, right Limitation on the range the accessible hour angle when the telescope axis is oriented E-W

a) If the axis is oriented N-S, the limitation on the rotation range translates into a limitation of the angle range accessible in the plane of the telescope axis and the zenith direction. With reference to fig. TC.6a, the range of accessible declination angles is given by

$$\varphi - (\alpha - \beta) < \delta < \varphi + (\alpha - \beta) \tag{32}$$

where φ is the site latitude.

b) If the axis is oriented E-W, the limitation on the rotation range translates into a limitation of hour angle. With reference to fig. TC.6b, the range of accessible hour angles is given by

$$-(\alpha - \beta) < HA < (\alpha - \beta). \tag{33}$$

c) For any different orientation of the axis the limitation on the rotation angle implies a limitation on both δ and HA.

From the point of view of the integration time, one in general would try to increase it

by starting observing the object as soon as it appears on the horizon[12]. However, an instrument placed on the Moon takes advantage of the low rotation rate of the planet to increase the integration time, so that we can reserve our, in principle, unlimited angular range for the direction perpendicular to the telescope axis, in order to reach the largest declination coverage. In this case the range of angular rotation of the siderostat determines how far from the zenith we can start tracking the object, which in turns determines how much of the baseline foreshortening (OPD variation) due to the Moon rotation we can exploit.

Let us then consider the telescope axis oriented E-W. As far as the declination range is concerned, we have no restriction, as it only depends on the latitude where the interferometer is located[13].

To determine the maximum range of time, T_{max}, during which we can track the same object, we make use of the moon rotation rate, which is of about 0.54° *per hour*. If the hour angle maximum range is of $2(\alpha - \beta)$, then T_{max} is given by

$$T_{max} = \frac{2(\alpha - \beta)}{0.54}\ hours \tag{34}$$

Tab. TC.3 presents some values of T_{max}.

tab. TC.3 E-W telescope orientation: maximum period of time (in hours) during which an object can be tracked as a function of the maximum angular range centered on the zenith direction (in degrees)

$2(\alpha - \beta)$	15°	30°	45°	60°	80°
T_{max} (h)	28	55	83	111	148

As an example of the advantage of the long integration times reachable on the Moon, let us consider for the interferometer baseline an N-S orientation at the equator and the siderostat-telescope axis oriented E-W. This configuration would allow $2(\alpha - \beta)\ /13^\circ$ days of integration with the same baseline, that could be of particular interest for high precision astrometric measurements of very faint objects (see section *1.2* in *Interaction among subsystems: an example for the lunar interferometer* for some numerical evaluations).

3. Other aspects related to the telescope configuration

Site latitude

In *section 1.2* we saw that in order to have a symmetric sweep about the zenith the

[12]This is strictly true for a space instrument; for a ground-based one this is not possible because of the strong absorption and refraction of the atmosphere, when the object is very low on the horizon.

[13]For the North hemisphere $\delta \geq \varphi - 90^\circ$, so the whole declination range is accessible for an instrument located at the equator.

telescope has to be tipped by γ. If γ could be chosen close, or equal, to the site latitude the telescope mounting might actually be equatorial (certainly when the telescope axis is oriented N-S), and this would be an advantage as far as the drive mechanical simplicity of the telescope drive system is concerned.
However, for a lunar interferometer the latitude selection depends on other factors, such as the orientation of the baseline, thermal consideration, and constraints on the landing site, which are drivers more important for this choice.

Thermal effects

As far as the orientation of the telescope is concerned, from the results presented in section *Thermal analysis for a telescope of the lunar interferometer* we see that a telescope at 25° with respect to the horizontal, and protected by a cylindrical shade, has the best performance with respect to those obtained with either a protected or unprotected solution for the telescope. This is due, mainly, to the fact that the telescope is not exposed perpendicularly to the solar vector, and to the lunar flux as well. However, when the night time is considered, the differences among the several solution vanish.

This is an effect comparable to that we could expect locating the telescope at higher latitudes. The maximum temperature during the day decreases going toward higher latitude because of the reduction of the effect of the solar flux, and by tilting the telescope with respect to the the horizontal we would expect to reduce also the effect due to the lunar flux. However, when we consider the night thermal behavior, no significative advantage is offered by moving the telescope to higher latitudes (see section *Thermal analysis for a telescope of the lunar interferometer*).

Thermal analysis for a telescope of the lunar interferometer

0. Introduction

The lunar thermal environment is characterized by large temperature swings and long period of time at the temperature extremes. The lunar surface temperature can vary from 93 K at sunrise to as hot as 395 K at local noon[14]. The surface temperature rises rapidly after sunrise, remaining above 300 K for a period of about 240 hours, then falls rapidly at sunset and it is near 100 K for about 290 hours. This environment presents a challenge for any equipment which must operate on the lunar surface, and it may create especially severe demands on the components of an optical interferometer. Without adequate thermal control these extreme thermal conditions can cause permanent misalignment of the optical components, introduce structural deformations, and damage electronic and mechanical device.

The thermal environment for an astronomical instrument operating on the lunar surface is more severe than that encountered by orbiting space instruments, because the lunar surface acts as a heat source during the lunar day and also complicates heat rejection by limiting the view to deep space. For an orbital instrument, the typical dark period is much lesser than the dark period of the Moon which results in much smaller temperature swings. An enlightening example is presented in tab. MT.1 which shows the maximum and minimum temperatures at radiative equilibrium for a 1-m^2 flat plate located on the lunar surface, in Low Earth Orbit (LEO) and in High Earth Orbit (HEO). For all cases, the direct solar radiation absorbed by the plate is the same, whereas the planetary radiation is quite different, due to the large difference in the distance from the closest planet. This is mainly responsible for the large difference in the variation between the radiation equilibrium temperature of the plate when considering the lunar and LEO cases.

In the following sections I will present a preliminary thermal analysis for a telescope of the lunar interferometer. Since thermal analysis supporting the design of any lunar equipment must include the influence of the lunar surface, in the first section I will present some of the relevant thermal characteristics of the lunar surface. The results for the thermal analysis which I present are due to Sherry Walker, from NASA Marshall Space Flight Center, who specifically performed it for a 16-m telescope to be located at the lunar equator. Further conversations and discussions with engineer Walker, allowed me to adapt these results to the case of a 1-m class telescope for a lunar interferometer[15].

[14] These are data calculated on the basis of the thermal properties of samples from the Apollo 11 mission, which landed at the Moon equator. Surface temperatures were also monitored at the Apollo 15 (26° N latitude) and Apollo 17 sites (20° N latitude) by means of thermocouples in cables placed several centimeters above the lunar surface. At the Apollo 15 site the maximum temperature was 374 K and the minimum 92 K. Temperatures at the Apollo 17 site were about 10 K higher (*Langseth and Keihm, 1977*).
In the following sections I will present results for thermal analysis calculation for an instrument located at the lunar equator. This is why I refer to the Apollo 11 data as representative of the maximum range of temperature variation during a lunar cycle.

[15] The interest in a 1-m class telescope derives mainly from mass budget considerations. For some evalu-

The conclusions derived in this study will be used as driver constraints for the definition of the parameters of the telescope configuration and for the discussion of the baseline orientation, as it is shown in the section *Interaction among subsystems: an example for the lunar interferometer.*

tab. MT.1 Comparison of the effects of different space environments on the thermal behaviour of a 1 m^2 flat plate, located on the lunar surface, in Low Earth Orbit (LEO), and in High Earth Orbit (HEO). It is assumed that the plate has emissivity and solar absorptivity of 0.8 on both sides. *From Walker and Alexander, 1993*

	Lunar	LEO	HEO
Altitude (m)	3	$5 \cdot 10^5$	$10 \cdot 10^8$
Absorbed heat (W)			
Solar	1082	1082	1082
Planetary	1080	475	5
Radiation equilibrium temperature			
Maximum (K)	442	399	313
Minimum (K)	98	231	64

1. Thermal characteristic of the lunar surface

The major role that the Moon has in influencing the thermal behaviour of a body placed on its surface, is due to the extreme characteristics of the thermal parameters of the lunar surface: high solar absorptivity, high emissivity, and low thermal conductivity. As a result of these characteristics the lunar surface heats rapidly after sunrise, radiates thermal energy very efficiently, and cools rapidly after sunset. Radiation and heat transfer models used in the thermal analysis should include radiation interchange with the lunar surface and the models should be adjusted until predicted surface temperatures approximate measured temperatures (when available). To accomplish this task typically requires modeling of the lunar subsoil as well as of the lunar surface layer, which in turns requires knowing the values of the thermophysical properties of the lunar soil.

Because of lack of extensive direct measurements of temperature fields and heat fluxes, the knowledge of the thermal behaviour of the lunar surface is based on the results of laboratory experiments to determine the thermophysical properties of samples of lunar soil collected during the Apollo missions. These experimentally derived data are then introduced in the transient heat equation for the lunar surface layers which is then solved

ations see section *2.1* in *The siderostat plus beam-compressor telescope solution for a lunar interferometer* and section *1.1* in *Interaction among subsystems: an example for the lunar interferometer.*

for the temperature as a function of time. In what follows I refer specifically to the results obtained by *Cremers et. al. (1972)* who examined samples collected during the Apollo 12 mission. They consist of rock particles of size which varies from (occasionally) about 1 cm down to below 0.1 μm. Since it is apparent that the lunar surface layer consists mainly of those kinds of particles which are randomly distributed, the lunar surface thermal behaviour study is usually approached as that of the thermal transfer problem of an evacuated porous medium.

The thermal energy transport problem for a porous evacuated medium is one involving both conduction and radiation heat transfer. These can be combined so that the total heat flux may be considered as an entity and it is possible to establish a model for the heat transfer in the lunar surface layer. Because of the excellent insulating properties of evacuated rocks powders, the depth to which there will be significant daily temperature variation is on the order of a meter whereas the Moon's diameter is of the order of $3 \cdot 10^6\ m$. Therefore the heat flow problem can be treated as one dimensional, i.e. we only consider the heat that is exchanged perpendicularly to a plane area on the lunar surface. The heat flow equation is then expressed by

$$\rho\, c(T) \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(k(T) \frac{\partial T}{\partial x} \right) \tag{35}$$

where ρ is the density, c the specific heat, and k the conductivity of the regolith[16].

The condition at the surface is expressed by the heat balance between the incoming solar radiation and the energy which is emitted to space plus that which is conducted into the surface layer (it is assumed that the heat flux from the Moon interior is zero). These conditions are written as:

$$F(t) = \epsilon \sigma T_0^4 - k(T) \frac{\partial T}{\partial x} \quad at\ x\ =\ 0 \tag{36}$$

where ϵ is the emissivity, which is a function of temperature of the regolith, T_0 is the temperature at the lunar surface, and F, the solar flux, is a function of the latitude and longitude of the lunar site in question.

The thermophysical properties of the lunar regolith which are needed to solve the two previous equations have been measured in the laboratory. However, we have to remind ourselves that, as far as the sample is concerned, there is no record of how far below the surface it came from or how representative of the overall regolith it is. Results for thermal conductivity and specific heat as a function of the temperature are presented in fig. MT.1.

[16]Regolith is a terrestrial term also used for the Moon. It has been defined as a general term for the layer or mantle of fragmental and unconsolidated rock material, whether residual or transported and of highly varied character, that nearly everywhere forms the surface of the land and overlies or covers bedrock. It includes rock debris of all kind, including volcanic ash (*Lunar sourcebook, 1991*).

fig. MT.1 Thermal conductivity (*data from Cremers et al., 1972*) and specific heat values (*data from Jones et al., 1975*) of the lunar regolith as a function of temperature

The maximum temperature, calculated by means of eq.(1)[17], occurs at lunar noon and is 389 K. The minimum temperature occurs at lunar sunrise and is 86.1 K. Fig. MT.2, from *Cremers et al. (1972)*, shows the results of the calculation of the temperature variation during a whole lunar day, for the lunar surface and three different depths.

The temperature distribution presented by *Cremers et al. (1972)* should be a good representation of the true surface temperatures at the Apollo 12 site (unless there are significant variations of density or composition within the first meter of depth). The results should apply to other mare regions[18] on the lunar equator as well. However, when projecting the results to uplands regions there could be differences, apart from those due to a different orientation with respect to the solar flux vector: the presence of a larger number of rocks, which have higher conductivities (by about a factor of 100) would make these results less applicable.

[17] For this computation made by *Cremers et al. (1972)* the available data on the specific heat were those derived by *Robie et al. (1970)* on the basis of analyses of samples collected at the Apollo 11 site.

[18] "Mare" is a dark-colored flat area on the Moon surface formed by large-volume eruptions of low-viscosity basaltic lavas. In current usage, mare basin designates a circular, multi-ring impact structure and "maria" the dark-colored lavas which fills such basins.

fig. MT.2 Temperature variation of lunar surface layers, during a lunar day (about 27.3 days), and at three different depths: 20, 50 and 100 mm. The plot starts at ends at noon of the lunar day

2. Thermal investigation of a telescope located on the lunar surface

The operation of a lunar interferometer requires stable thermal environment, whose local variation has to be controlled by means of thermal control systems and compensation devices (athermalization). Because of the extreme temperature variations during the whole lunar cycle, I am interested in investigating the thermal behaviour of a telescope located on the Moon surface to derive which period during the lunar cycle is most favorable to the operation of an interferometer. This result is to be related to the requirements imposed on the telescope configuration and the baseline orientation, and eventually, when considering the design for a lunar interferometer, the estimation of the thermal range in which the interferometer is expected to operate would enable us to set requirements on the thermal control system of several instrument subsystems.

2.1 Thermal analysis: the theoretical approach

Let us consider an elemental system consisting of a flat plate of area A located on the Moon surface; it is supported by means of a pedestal of section area A' and height Δh. For the thermal analysis of this system we have to take into account the heat flux from the sun as well as the radiative exchange between the plate and the deep space and the plate and the lunar surface. Also the conduction from or to the lunar surface through the pedestal has to be considered (the heat flow through the first layers of the lunar surface is already considered in the calculation for the estimate of the temperature of the surface itself).

With reference to the sketch in fig. MT.3, we can write the equation for the thermal balance of the system:

$$\alpha F_S A + \alpha \epsilon_M A \sigma T_M^4 + \frac{kA'}{\Delta h} (T_M - T) = 2 \cdot \epsilon A \sigma T^4 \tag{37}$$

where α is the absorptivity of the plate, F_S is the solar constant (1358 Wm^{-2} at 1 AU), ϵ_M is the emissivity of the Moon surface layer, which varies with temperature (maximum value is about 0.98 at 80 K, minimum value is 0.92 at 440K, (*Cremers et al., 1972*)), σ is the Stefan-Boltzmann constant ($5.67 \cdot 10^{-8}$ $Wm^{-2}K^{-4}$), k is the thermal conductivity of the pedestal. T_M is the temperature of the lunar surface layer as it given by analytical models like that by Cremers et al. that I presented in the previous section. The value of T_M already takes into account the heat conduction between internal and superficial lunar layers, which is characterized by a very low value of thermal conductivity. However, we have to be aware that the very installation of our apparatus (with a pedestal which is anchored in depth within the lunar soil, in a more realistic situation than that presented in the sketch of fig. MT.3) may modify the conductivity of the lunar surface area which directly interacts with it.

fig. MT.3 Thermal interchanges for a flat plate supported by one pedestal on the lunar surface

Eq.(3) is eventually solved for T as a function of time, which means as a function of the change in solar illumination and in the lunar surface temperature.

Let now turn our attention to the lunar night. The term which represents the solar flux vanishes, and the function describing the lunar surface temperature tends to a minimum value. Let us also consider different emissivity for the two sides of the plate. The equation

of the thermal balance becomes

$$\alpha\epsilon_M\sigma T_M^4 + \frac{k\gamma}{\Delta h}\,(T_M - T) = \sigma\epsilon_u T^4 + \sigma\epsilon_d T^4 \tag{38}$$

where $\gamma = \frac{A'}{A}$ and ϵ_u and ϵ_d are the emissivity of the plate towards the deep space and the lunar surface respectively.

If now we assume that the parameters in the conduction term can be adjusted so that we obtain a good insulation of the plate from the surface, and if we neglect the conduction contribution, then by solving for the equilibrium temperature T we get

$$T = \left(\frac{\epsilon_d \cdot \epsilon_M}{\epsilon_u + \epsilon_d}\right)^{0.25} \cdot T_M. \tag{39}$$

From this equation we can see that an appropriate choice of the emissivity of the material which covers the face of the plate towards the lunar surface may contribute to decreasing the plate equilibrium temperature (in principle to reach a value which is lower than the equilibrium value of the lunar surface. In general, due to the dimension of the Moon, we would consider it as a heat sink and expect that the temperature of equilibrium reached by the plate is the same as that of the lunar surface).

2.2 Numerical solution

What I presented in the previous section can be considered the generalized mathematical model of the thermal balance problem for an elemental system. However, we have usually to deal with complex systems and want to obtain numerical solutions for the equation of the thermal balance applied to them. In order to do this, thermal systems are often simulated by means of what is called the "nodal approach". This consists in dividing the structure into nodes, each defined as a subvolume of the structure in thermal equilibrium, and then calculating node temperatures and heat flux interchange among the nodes. Eventually, to apply this method to complex systems requires that the mathematical model be expressed in a form that allows for the processing of information at discrete points (nodes) by a computer. If the number of points is large enough an accurate picture of the temperature distribution can be obtained.

Nowadays the most common way to proceed in the thermal analysis for space systems is to makes use of the combination of two computer programs. The first one is the *radiation* program, which in input requires the system geometry, orientation, altitude, and surface radiation property values (emissivity and solar absorpitivity). The output of the radiation program defines the absorbed energy from the Sun, and from other heat sources (in our case the heat flux from the Moon). The radiation program is also used to define all the radiation interactions between the different surfaces of the system (by defining relative radiation exchanges and view factors). After the radiation aspects of the overall problem are established, a heat transfer model of the system is assembled. This model accounts for all the thermal characteristics of the system and allows us to predict the temperatures of the system as a function of time. In general, elements of this model include the system

geometry, electrical power dissipation, radiation heat transfer, conduction heat transfer, and thermal characteristics of the thermal control hardware. On the basis of this model, a *thermal analyzer* computer program performs actual heat transfer calculations that yield either a transient or a steady state solution. The outputs are typically temperatures (versus time if a transient solution) and heat flow through particular conductors.

3. Thermal analysis for a lunar telescope

In this section I present a brief model description and the results of the thermal analysis performed for the 16-m primary mirror of a telescope located at the lunar equator. The results of this analysis will be further discussed in order to be applied to the case of the 1-m class telescopes considered for the lunar interferometer.

3.1 Model description

The temperature profile and physical properties of the lunar surface layer are not constant but vary depending on location and local terrain. For this study flat terrain near the telescope has been assumed. The temperature dependent values of soil thermal conductivity and soil specific heat used for the model of the lunar surface are those presented in fig. MT.1.

The mirror was modelled as a segment of a sphere, divided into four nodes. Silicon carbide (SiC) was assumed for the mirror material, with a thermal conductivity of 125 $W\,m^{-1}\,K^{-1}$ and a specific heat of 0.21 $W\,hr\,kg^{-1}\,K^{-1}$. For the computation of the effects of shades and enclosures a 1.25 cm thick, multi-layer insulation (MLI) was used, with effective conductivity[19] of 0.000208 $W\,m^{-1}\,K^{-1}$.
Absorbed heat flux and radiation interchange were computed using the Thermal Radiation Analysis System (TRASYS) which gave the input for a System Improved Numerical Differencing Analyzer (SINDA) model used for heat transfer computations to determine actual nodal temperatures as a function of time. Geometry, surface optical properties, location, and local time are input to the TRASYS program, which considers direct solar radiation as well as radiation interchange between model surfaces, including the lunar surface. TRASYS generates a network of radiation conductor between all model surfaces, along with arrays containing time dependent solar heat flux values for all nodes for a period of one lunar cycle.

This information is input to the SINDA model, along with nodal heat capacitance values and conductors representing linear conduction through the telescope components. No convective heat transfer was included in the SINDA model. Output from the SINDA program is the temperature at each node at time intervals throughout the lunar day/night cycle.

[19]The thermal performance of a multilayer insulation is usually described in terms of effective conductivity, i.e. the resulting conductivity across the whole thickness, from the innermost to the outermost surfaces of the insulation.

3.2 Results

Thermal analyses were performed to provide an insight into the effects on the temperature of the mirror due to a ground shade, different enclosure concepts, site latitude, and local terrain. Unless otherwise indicated, the telescope is pointed directly overhead for all analyses.

Ground shade

A telescope placed on bare regolith was compared to one placed on a ground shade composed of 1.25 cm thick MLI. The ground shade reduces the mirror temperature by about 15 K at noon and by about 7 K for the most of the lunar night. These results indicate that a ground shade should not be considered as an optimum device to reduce the overall telescope temperature. Some other mechanism to reduce the radiation interchange between telescope and lunar surface needs to be studied.

Enclosures

The study examined two basic enclosure concepts. One is a dome that covers the telescope during the day and is removed completely at night. The other concept is a cylindrical sunshade attached to the mirror support structure. This concept was analyzed in two orientations: pointing directly overhead and pointing 75° to the north to prevent direct sunlight from entering the telescope.

All four concepts show that the mirror can be maintained below 100 K for at least 170 hours per lunar cycle, with a maximum of 200 hours predicted for the dome enclosure. In addition for all four cases the minimum temperature is in the 80 to 85 K range. The main discriminator between the four concepts is the maximum mirror temperature which occurs at about local noon. For the telescope with the cylindrical sunshade and no enclosure at all, this temperature ranges from 325 to 450 K. However, the 450 K temperature has little relevance because it represents the unlikely case of the telescope pointed directly at the sun. The maximum temperature reached by the mirror with a dome is of only 190 K. This concept is clearly preferred when low temperatures and (important for an interferometer) temperature stability are desired. Results for the effects of different enclosure are presented in fig. MT.4. The diagrams also shows the temperature variation of the lunar surface.

fig. MT.4 Enclosure concepts and resulting temperatures for the mirror *(Walker, 1992)*

Site latitude

The effect of the site latitude is mainly of interest for telescopes that are designed for both day and night viewing, since a location considerably north or south the equator can simplify the design of the sunshade of the telescope. The study considered equatorial site and sites at 15°, 30°, and 45° north latitude and did not incorporate a ground shade.
The results show that latitude has essentially no effect on the minimum temperature of the mirror, but does affect the maximum mirror temperature. Moving the telescope from the equator to 15° north decreases the maximum mirror temperature by less than 5 K, but moving to 45° north results in a 39 K decrease in the maximum mirror temperature. Fig. MT.5 presents the effects on the maximum temperature of the mirror due to different site latitudes.

fig. MT.5 Effects of site latitude on the maximum temperature of the mirror *(Walker, 1992)*

Local terrain

Local features such as craters and hills might be expected to affect the surface temperature at a particular site. The previous analyses all assumed that the telescope was located on a flat area. The study considered also the cases of the telescope in the center of a crater 50 m in diameter and 12 m deep, and on the top of a hill 50 m high. No ground shade was included for these locations and all sites are on the lunar equator. The results from these calculations show that differences in the maximum and minimum temperatures are small, but either a hill or a crater site slightly reduces the maximum mirror temperature. A hill site also reduces the minimum mirror temperature, while a crater location increases the minimum temperature by about 4 K. This is due to the fact that a hill location offers the telescope a better view to deep space than either a flat or crater location. Results for higher latitude would probably show larger temperature differences, especially for a crater site which might be in shadow for much of the day.

4. Thermal performance of a 1-m class telescope for the lunar interferometer

For the evaluation of the thermal performance of a 1-m class telescope for the lunar interferometer, I will adapt the results obtained for the 16 meters telescope presented above.

4.1 Assumption

I assume that the 1-m telescope has insolation characteristics and exterior optical properties (emissivity and absorpitivity) similar to those of the 16-m telescope, and that the thermal capacitance (mass time specific heat) per squared meter of two telescopes is of the same order. The last assumption is set in order to properly estimate the time required to reach the thermal equilibrium. Since this depends on the thermal capacitance, whenever

the thermal capacitance per squared meter is of the same order, the two telescopes cool down in about the same time, independently of the area of the two mirrors (if the thermal capacitance per squared meter is smaller for one of the two telescopes, this telescope will cool in less time.) Clearly this approximation is applicable only if the external optical properties (above all emissivity) are the same in the two cases (*Walker, private communication).*

Actually these assumptions are quite general and applicable to any lunar telescope at the beginning of the design study. In the specific case of my work, what I am interested in is a realistic estimation of the thermal behaviour of the telescope that allows me to set the range of variability for other parameters related to the telescope configuration and the baseline orientation. Once this is accomplished and both configuration and material for the telescope components are selected, a thermal analysis properly tailored for the telescopes of the interferometer should be performed.

In conclusion, the mirror is assumed to be SiC, with absorpitivity $\alpha = 0.09$ and emissivity $\epsilon = 0.03$; all multilayer insulations (MLI) considered are 1.25 cm thick and have $\alpha = 0.22$ and $\epsilon = 0.82$. The total thermal capacitance for the 16-m telescope is of 11240 kJK^{-1}, hence the 1-m class telescope has a thermal capacitance per squared meter of about 56 $kJK^{-1}m^{-2}$.

4.2 Comparison with results obtained for LUTE

I also took into consideration the results of the thermal analysis performed for LUTE[20], a 1-m telescope for UV observation to be located at 40° North on the lunar surface.

[20]Lunar Ultraviolet Telescope Experiment.

fig. MT.6 Mirror temperature variation during one day/night lunar cycle, calculated for the LUTE baseline design (*LUTE Phase A Final Report: NASA TM 4594, 1994*).

In this case the thermal analysis was much more detailed and took into account the effects of properly shaped sunshade, optimized insulation and of the site latitude.

Fig. MT6 shows the results of this analysis. As it could be expected the main difference with the 16-m telescope consists of the temperatures reached during the day. Because of the higher latitude, of the sunshade and optimized insulation, the maximum temperature is lower than in all cases examined for the 16-m, but for the case with a dome shade.

However, as far as the nighty behaviour is concerned, there are not evident differences. The minimum temperature is about 70 K and, since for the thermal models adopted in both thermal analyses, errors of the order of 5-10 K are not uncommon, the results for the minimum temperature from the two analyses can be considered equivalent.
As far as the transient time after sunset is concerned, in the initial period (from about 336 to 440 hours) the 16-m telescope performance is equivalent (when a tilted cylindrical shade is considered) or even better, with the dome shade solution, than the LUTE performance. After the 18th day the performance is the same, the telescope follows the lunar surface behaviour: in particular, in both cases, the telescope is at the equilibrium temperature for a period of about 5 days.

In conclusion, the main difference between the two cases is due to the use of a sunshade for LUTE. This is the main feature responsible for the lower maximum temperature of the LUTE mirror, and allows the instrument to be operated during the day. However, as far as the night is concerned, an open telescope will cool more quickly after sunset, which is an advantage when considering night time operations.

4.3 Results for a 1-m class telescope of the lunar interferometer

With the assumptions presented at the end of section *4.1* (which allow me to apply the results for the 16-m telescope also to a 1-m class telescope) and on the basis of the results presented in the previous section, I am able now to present what can be considered a preliminary assessment of the thermal performance of a 1-m class telescope of the lunar interferometer. These are only results to start with for a preliminary definition of the parameters of other subsystems. As I said above, once these parameters are set, a detailed thermal analysis based on the specifications for this telescope should be performed.

I first consider the telescope located at the Moon equator, for the three cases: telescope unprotected (only ground-shade), telescope protected by a cylindrical shade and tilted by 75° towards north, telescope protected by a dome shade during the day time and then unprotected during the night time. Tab. MT.2 shows the values for the temperature at sunrise, noon, sunset. Note that the temperature at sunrise is actually the equilibrium temperature that the telescope reaches during the last part of the night.

tab. MT.2 Estimated temperature at sunrise, noon, and sunset for a 1-m class telescope of the lunar interferometer located at the Moon equator

	sunrise	noon	sunset
unprotected	80 K	395 K	250 K
tilted shade	80 K	325 K	220 K
dome shade	80 K	190 K	175 K

The following two tables present the estimated temperature excursions and mean temperature variations per hour for different periods during both the night and the day time. The night equilibrium temperature is assumed to be 80 K in all of the examined cases. For the night time I have considered three periods: from sunset to the 18th day (the main part of the transient after sunset), from the 18th to the 22nd day, and from the 22nd day to sunrise, when the telescope is practically in thermal equilibrium[21]. For the day time I have considered only two periods: from sunrise to noon and from noon to sunset.

[21] The lunar cycle lasts for about 27.3 days. I assume that sunset occurs about 8 hours before the 14th begins, i.e. 327.6 hours after sunrise.

tab. MT.3 Estimated temperature excursions (ΔT) and temperature variations per hour ($\dot{T} = \frac{\Delta T}{\Delta t}$) during the lunar day for a 1-m class telescope on the Moon equator

Time period	sunrise to noon $\sim 164\ h$	noon to sunset $\sim 164\ h$
ΔT unprotected	$315\ K$	$145\ K$
ΔT tilted shade	$245\ K$	$105\ K$
ΔT dome shade	$110\ K$	$15\ K$
$\dot{T}$ unprotected	$1.92\ Kh^{-1}$	$0.88\ Kh^{-1}$
$\dot{T}$ tilted shade	$1.50\ Kh^{-1}$	$0.64\ Kh^{-1}$
$\dot{T}$ dome shade	$0.67\ Kh^{-1}$	$0.09\ Kh^{-1}$

tab. MT.4 Estimated temperature excursions (ΔT) and temperature variations per hour ($\dot{T} = \frac{\Delta T}{\Delta t}$) during the lunar night for a 1-m class telescope on the Moon equator. The equilibrium temperature is assumed for all cases $\sim 80\ K$

Time period	sunset to 18th day $\sim 104\ h$	18th to 22nd day $96\ h$	22nd to dawn $\sim 127\ h$
ΔT unprotected	$130\ K$	$15\ K$	5-10 K
ΔT tilted shade	$105\ K$	$15\ K$	$5\ K$
ΔT dome shade	$75\ K$	$10\ K$	$< 5\ K$
$\dot{T}$ unprotected	$1.25\ Kh^{-1}$	$0.16\ Kh^{-1}$	$\sim 0.08\ Kh^{-1}$
$\dot{T}$ tilted shade	$1.00\ Kh^{-1}$	$0.16\ Kh^{-1}$	$\sim 0\ Kh^{-1}$
$\dot{T}$ dome shade	$0.72\ Kh^{-1}$	$0.10\ Kh^{-1}$	$\sim 0\ Kh^{-1}$

We saw in *section 3.2* that the site latitude has a non-negligible effects on the maximum temperature of the mirror, as the plot in fig. MT.5 shows. The temperature variation shown in the plot, which refers to an unprotected mirror without ground shade, is actually due to the different inclination of the solar vector at different latitudes. It is immediately verified that the temperature approximately scales as the 0.25 power of the cosine of the latitude[22]. This is the principal effect due to a change in site latitude that we should expect

[22] $T \propto (F_{sun} \cdot cos\phi)^{1/4}$, where ϕ is the site latitude and F_{sun} is the solar flux.

for any of the surfaces which are exposed to the sun. Hence, as a first approximation we can scale by the same factor the results for the maximum temperature of tab. MT.2.

As far as the night time is concerned, results for minimum temperature at latitudes up to 45° show that the variation is always less than 1 K (*Walker, private communication*). Then, because of the variation in the maximum temperature, we would expect that the time to cool down after sunset changes with the latitude, too. However, the relevant parameter which determine the time to reach the equilibrium is temperature at sunset rather than maximum temperature. Eventually, at least for latitudes up to 45°, the latitude is not expected to change the time to reach the thermal equilibrium by more than 10 to 20 hours (*Walker, private communication*).

Interaction among subsystems: an example for the lunar interferometer

0. Introduction

The purpose of this section is to show an example of the relations which link parameters belonging to different subsystems, and how their determination is influenced by environmental conditions.

In particular, in this section I show how the results of the analyses presented in the sections *Thermal analysis for a telescope of the lunar interferometer*, *The siderostat plus-beam compressor solution for a lunar interferometer*, and *Baseline orientation: OPD variation and u-v coverage* can be combined together to define a set of parameters for the preliminary design of a two-element lunar interferometer.

For the following analysis I consider a two-element array, with 1 km baseline, which will be operated during the lunar night. Taking into account time and costs for developing a multi-dimensional array on the lunar surface (which however should be the final goal), it is very likely that the first configuration available will consist of two collectors. The baseline length should be of the order of the minimum required to reach interferometric performances superior to those reachable from the Earth (see section *Interferometry from the Earth and from the Moon*). Finally, the requirement of operating the instrument only during the night is due to the extreme thermal conditions to which the interferometer is exposed. As far as the mechanical and optical performances, the main concern to be addressed is the large variation in temperature, rather than the absolute values (both maximum and minimum), that the instrument undergoes (see section *Thermal analysis for a telescope of the lunar interferometer*). However, the absolute temperature values are important quantities to be considered if we want to operate in the infrared, where the background radiation from the instrument will limit the performance of a detector. In order to meet the requirements imposed by the wavefront error budget on the optics and the telescope structures, it would be preferable for the instrument to be operated at lower temperatures, but above all to keep the temperature variation undergone by the different components of the instrument in the range of about $100\,K$ (see *Appendix L*). These conditions could be maintained by means of thermal control systems like either cryogenic fluids or active systems as heaters, refrigerators, heat pumps. However this will result in a significative increase in the mass and/or power required for the thermal control system[23]. But also, these conditions can be met by protecting the telescopes of the interferometer by means of a dome-shade during the day period, and letting them passively cool down during the night period. This solution will require active thermal control systems only for those electronic components whose temperature must be maintained within ranges of few degree kelvin.

[23]Precise evaluation can be done only when the instrument configuration is established. For general evaluation I refer to *Walker, Alexander, and Tucker (1995).*

In what follows I will take into consideration the results of the thermal analysis obtained for a telescope provided with a dome-shade. I first apply them to determine the maximum hour angle (HA) range on the basis of the duration of the period of thermal equilibrium for the telescope. Then the other parameters related to the telescope configuration will be derived. Finally, I will calculate the u-v coverage obtainable with the two-element interferometer, for both a 0° and 30° baseline latitude, for the maximum integration time allowed by the thermal equilibrium condition.

1. Two-element lunar interferometer: 0° latitude, dome-shade

Let us refer to tab. MT.4 in section *Thermal analysis for a telescope of the lunar interferometer*, in particular to the dome-shade case. The night has been divided into three time periods, in order to separate the transient period from the period during which the telescope is in thermal equilibrium. During the whole thermal equilibrium period the maximum temperature variation is less than $5\,K$, and probably most of it occurs in the early hours of the period. Athermalization systems for different parts of the telescope can be optimized for this range of temperature variation, so that we can consider the telescope thermally stable and assume that the wavefront is not affected by any error induced by transient thermal distortions. We then require that we can follow (track) one object for the whole period. This requirement first determines the minimum hour angle coverage that is necessary for such a long observation.

Eventually, according to tab. MT.4, the last 130 h of each lunar cycle could be allocated to either astrometric or visibility observations of very faint objects (thanks to the long time of integration available) or to observations which last less time but require high precision in the measurement. From sunset to the 18th day and from the 18th day to the 22nd, observations will have shorter integration time, but still quite longer than what is achievable on Earth. Even by means of adaptive optics, the maximum integration time is still given by the ratio of the phased beam diameter (which is equal to the mirror diameter in this case) to the wind velocity, and for a 1 m telescope it is of the order of 0.1 s. As far as the thermal stability is concerned, we could consider integrations which are one hour long between the 18th and 22nd day ($\Delta T/\Delta t \sim 0.10\,K h^{-1}$), and either 1000 s long ($\Delta T \sim 0.20\,K\ in\ 1000\ s$) or snap-shot observations, during the first period of the lunar night[24].

[24] These values of temperature variation during an integration time are acceptable when considering the limits imposed by the baseline attitude error. Let us perform a rough calculation of this limit. An uncertainty $\Delta\mathbf{B}$ in the measurement of the baseline vector $\mathbf{B}$ causes, in an angle measurement, a corresponding uncertainty of magnitude $\frac{\Delta B}{B}$. Since the maximum resolution achievable by the instrument is given by λ/B, this implies the generally adopted requirement that ΔB is known to the order of $\lambda/2$ or better. Let us assume that, for the lunar interferometer, ΔB has to be of the order of $10^{-7}\,\mu m$ (that means $\lambda/50$ at 0.5 μm). Let us now assume that the ΔB error is due to failure in (or absence of) the thermal compensation of the longest linear dimension in the structure of the telescope, i.e. $\Delta B = \Delta L$ where L is the length of the uncompensated part. Let L be of the order of the separation between siderostat and beam-compressor. According to tab. TT.2 in section *The siderostat plus beam-compressor telescope solution for a lunar interferometer*, we can assume that L varies from a minimum of 2 m to a maximum of 10 m, when the primary mirror diameter is of the order of 1 m. On the basis of these assumptions we

1.1 Telescope parameters

The thermal stability condition sets the requirement for the minimum range in HA to be covered by the telescope. In principle we could orient the telescope in direction N-S and expand this range up to 180°. However, since we will not be able to exploit the increase in HA range to increase the integration time, it is certainly preferable to choose an E-W orientation, which allows us to observe in the whole declination range available. Then we will require that the limitation in the HA range implied by the E-W orientation matches the minimum HA range set by the thermal stability condition.

By adopting this solution, to track the same object for 130 h requires that the rotation range of the siderostat, given by $2(\alpha - \beta)$, is about 70°. For comparison, it is interesting to notice that, for IOTA the rotation range of the siderostat is of the order of 80° which in principle would allow a maximum integration time of about 5 h. Actually, for IOTA the rotation range of the siderostat is exploited to point objects which are up to about 2 h and 30 min far from the zenith.

1.1.1 Inclination with respect to the horizontal and trade-off for the α and β parameters

In order to have the HA fan equally distributed on both side of the zenith, the telescope must be tilted by $\gamma = 90° - (\alpha + \beta)$. Different tilt angles may be suggested, according to the choice of the values of α and β.

β is determined by the structural constraints. However, because of the dome-shade solution, we do not have external envelopes surrounding the primary-mirror structure (for example a cylindrical shade) to consider. Hence, the limitation will be set directly by the metering structure which connects the primary and the secondary mirrors.

Let us consider possible β values in a range from 5° to 15°. I calculated the related values of α and γ by taking into account that (from the thermal stability condition) $\alpha - \beta = 35°$. Tab. SI.1 presents the results of the calculation of the values of α, γ, D (the siderostat maximum linear dimension), and l, the compactness parameter (the separation between the siderostat and the boundary of the telescope support. For the definition of all these quantities see section *The siderostat plus beam-compressor solution for a lunar interferometer*).

---

have

$$\frac{\Delta L}{L} = CTE \cdot \Delta T \; = 10^{-8} to \, 5 \cdot 10^{-8}$$

where CTE is the coefficient of thermal expansion of the material. According to the state-of-the-art for composite material structures, CTE values in the axial direction below $10^{-7}\,C^{-1}$ are achieved (*Synott et al., 1991*), and an improvement of a factor 10 is strongly pursued. For this calculation I assume a conservative $CTE = 10^{-7}\,C^{-1}$, so that the maximum temperature variation we can tolerate during an integration time is given by

$$\Delta T = \frac{\Delta L}{L} \cdot CTE^{-1} \; = \; 0.1 \, to \, 0.5 \, K.$$

tab. SI.1 Telescope parameters: maximum incidence angle, α, inclination with respect to the horizontal, γ, siderostat maximum diameter, D, and the compactness parameter, l, as a function of β, the minimum incidence angle due to the obscuration cone. d is the primary mirror diameter, the HA range is to be covered in 130 h

β	α	γ	D	l
5°	40°	45°	$1.30\,d$	$5.76\,d$
7.5°	42.5°	40°	$1.36\,d$	$3.73\,d$
10°	45°	35°	$1.41\,d$	$2.75\,d$
12.5°	47.5°	30°	$1.48\,d$	$2.14\,d$
15°	50°	25°	$1.55\,d$	$1.73\,d$

We notice that a small inclination with respect to the horizontal implies a large incidence angle on the siderostat: for $\gamma = 0^\circ$, i.e. telescope parallel to the horizontal, α would be larger than 60°, a too large value if we consider the polarization-dependent phase shifts arisen at the mirror surface[25]. Thus, even if smaller γ value would imply the advantage of having a more compact system, values of γ less than 25° are not considered. As I said before, the thermal stability condition sets a lower limit for the rotation range of the siderostat, but it does not prevent it from having a larger rotation range. Nevertheless, the results in tab. SI.1 show that this limit allows us to span large enough ranges of feasible values for the telescope parameters, and to form and trade-off a large number of combinations.

For any β, we have one free parameter, the primary mirror diameter, d, which determines the dimensions of both D and l and eventually the mass of the telescope. Let us then consider how the mass of the siderostat increases by increasing the primary mirror diameter and how this influences the range of combinations available for trade-off.

For the values of β considered in tab. SI.1, the maximum variation of the ratio $\delta = (D/d)$ is $\Delta\delta = 0.25$. From eq.(2) in *The siderostat plus beam-compressor telescope solution for a lunar interferometer*, we derive $\Delta m = 0.2125\,\rho_A \cdot d^2$. Assuming both siderostat and primary mirror made of silicon carbide ($\rho_A = 28\,kg\,m^{-2}$), $\Delta m = 5.95\,kg$ when $d = 1\,m$ and $\Delta m = 23.8\,kg$ when $d = 2\,m$. It is likely that the increment associated with a 2-m primary mirror diameter is too large to allocate in the mass budget of the telescope, so that we would have to limit the trade-off among the combinations which imply small β values. Because of the smaller Δm involved, the case with 1 m diameter keeps all the combinations available for the trade-off, and drives us to select primary mirror diameters of the order of 1 m. However, this is not the principal driver towards the choice of a 1-m class telescope. The main factor to be considered is certainly the mass budget allocation for the payload of the launcher[26].

[25] See section 2.1 in *The siderostat plus beam-compressor telescope solution for a lunar interferometer.*

[26] For a given δ, to double the size of d (in this case from 1 to 2 m) implies to increment the mass of

Let us then concentrate on a 1-m primary mirror and consider the trade-off between the maximum incidence angle and the compactness factor. The range of variation of α is relatively small (10°), whereas l increases by 4 m. A good compromise can be reached by choosing a maximum incidence angle of 45° (acceptable as far as polarizations effects are concerned) to which corresponds a separation between siderostat and beam-compressor of 2.75 m. Assuming that a β value of 10° is the minimum achievable due to the obscuration cone of the telescope, the inclination of the telescope with respect to the horizontal has to be 35°.

In conclusion, the parameter values which characterize the preliminary design of a telescope for the lunar interferometer are those given in tab. SI.2.

tab. SI.2 Parameters for the preliminary design for a telescope of the lunar interferometer located at 0° latitude

latitude	$\varphi = 0°$
telescope orientation	E-W
inclination with respect to the horizontal	$\gamma = 35°$
maximum angle of incidence	$\alpha = 45°$
minimum angle of incidence	$\beta = 10°$
primary mirror diameter	$d = 1\,m$
siderostat maximum linear dimension	$D = 1.41\,m$
separation between siderostat and beam-compressor	$l = 2.75\,m$
primary mirror mass	$M_P = 0.785\,\rho_A$
siderostat mass	$M_{sid} = 1.2\,\rho_A$

where ρ_A is the aerial density of the mirrors. Values for ρ_A between 25 and 30 $kg\,m^{-2}$ imply a total mass for the mirrors between about 50 and 60 kg. In order to deliver both of them on the Moon by one launch, these values should be compatible with the payload mass of the order of 100 kg (conservative) which actual launchers can provide.

---

the siderostat by $\Delta m = 3 \cdot M_{sid}$ (where M_{sid} is the initial mass of the siderostat). For SiC mirrors this is equivalent to a minimum $\Delta m \simeq 71.5\,kg$ for the siderostat, to which we should add the increase in mass for the primary mirror itself, which is of the order of $66\,kg$. According to the mass budget allocation based on available launchers, the maximum mass allocable to the telescope is of the order of $168\,kg$ (see section *2.1* in *The siderostat plus beam-compressor telescope solution for a lunar interferometer*). This means that two separated launches would be necessary to deliver on the Moon surface both the primary mirror and the siderostat. In this case the launch cost would dramatically increase because of, in addition to the cost of a second launcher, the cost of the operations to assemble the spacecraft elements will have to be taken into account.

1.2 Baseline orientation and u-v coverage

For an equatorial baseline both the E-W orientation and the N-S orientation present interesting aspects.

For any object's declination, the E-W orientation gives the maximum range of OPD variation available with a given baseline. In this case the maximum u-v coverage is determined by the maximum integration time available according to the thermal equilibrium condition. As an example, in fig. SI.1 I show the u-v coverage obtained with a two-element E-W baseline in 130 h of integration, for two different object's declinations.

fig. SI.1 u-v coverage with an E-W equatorial baseline (maximum length 1 km): object at 30° (right) and 60° (left) north declination.

In order to exploit the maximum range of variation in OPD, the delay line of the interferometer has to have a "round-trip" length of the order of 600 m, which has to be covered during one observation (the maximum OPD variation is actually reached only for objects at 0° of declination). This can be realized by means of a coarse delay line, consisting of a number of stations among which the instrument switches as far as the OPD changes, complemented by a fine delay line that covers the OPD variation values which are between those covered by two adjacent stations of the coarse line. By looking at the diagram of the OPD variation for an equatorial E-W baseline (fig. B.3a in *Baseline orientation: OPD variation and u-v coverage*), we can set the length of the fine delay line. For example, in order to have 10 hours of integration centered on the culmination point, with no need of changing the coarse delay line, the fine delay line has to be 50 m long (100 m round-trip). In this case, in order to cover the maximum delay obtainable

in 130 h of continuous integration, the coarse delay line would have 5 stations in a fixed position (the solution with one movable station, that travels the whole 300-m distance, can be considered too).

The maximum variation speed, required for the the fine delay line, in proximity to the culmination point, is of the order of 1.5 $mm\,s^{-1}$ (the OPD rate variation is of the order of 3 $mm\,s^{-1}$).

For an equatorial baseline the N-S orientation may be taken into consideration if we are interested in astrometric measurements. Since, for a given object's declination, the OPD does not change during the whole integration period, the long period of thermal equilibrium allows, in principle, for high precision measurements on very faint objects. In tab. SI.3 I present the visible limiting magnitudes obtainable as a function of the signal-to-noise ratio, or equivalently of the photon-noise-limited accuracy in the trigonometric parallax, assuming maximum angular resolution of 0.1 mas[27]. The data in tab. SI.3 are calculated assuming V =1, 1-m diameter for the collecting optics, overall (optics and detector) efficiency of 0.4, a bandwidth of 1000 $\mathring{A}$, and 100 h of integration time.

tab. SI.3 Limiting magnitudes, in the visible, reachable in 100 h integration as a function of the signal-to-noise ratio (SNR), or equivalently the accuracy of the trigonometric parallax ($\Delta\alpha$) in μas

SNR	σ_α (μas)	m_V
10	1.6	$\sim$ 33
100	0.16	$\sim$ 28
1000	0.01	$\sim$ 23

Nevertheless, to be able to observe on the whole declination range, it is necessary to have a delay line system which can cover up to about 1000 m of delay. This can be accomplished, as in the case of the E-W baseline, by means of a coarse delay line consisting of a number of fixed stations, plus a fine delay line which covers the gaps between two stations of the coarse line. With a 50-m long fine delay line, 9 fixed stations would be necessary. The difference with the case of the E-W baseline is that, after the delay line is positioned to compensate for the delay, it does not need to be moved for the whole integration. In case, the fine delay can compensate for any minor variations in the OPD due to unexpected events.

[27] For an interferometer, the photon-noise-limited accuracy in the measurement of the trigonometric parallax, $\Delta\alpha$, is related to the angular resolution α and the signal-to-noise ratio, SNR, by

$$\sigma_\alpha = \frac{\alpha}{2\pi \cdot SNR \cdot V}.$$

2. Two-element lunar interferometer: 30° of latitude, dome-shade

Results from the thermal analysis show that, for an instrument that operates only during the lunar night, a change in latitude should not make an appreciable difference as far as the equilibrium temperature is concerned. However, the time required to reach the equilibrium could be something like 10-20 h shorter, with a gain of the same order in the duration of the equilibrium period.

2.1 Telescope parameters

Let assume the best case, i.e. at 30° of latitude the equilibrium period increases by 20 h with respect to the equilibrium period at 0° of latitude. For 150 h of integration the corresponding minimum HA range is of about 80°. Assuming, as in the previous case, that β varies from 5° to 15°, we can derive the telescope parameters given in tab. SI.4.

tab. SI.4 Telescope parameters: maximum incidence angle, α, inclination with respect to the horizontal, γ, siderostat maximum diameter, D, and the compactness parameter l, as a function of β, the minimum incidence angle due to the obscuration cone. d is the primary mirror diameter, the HA range is to be covered in 150 h

β	α	γ	D	l
5°	45°	40°	$1.41\,d$	$5.76\,d$
7.5°	47.5°	35°	$1.48\,d$	$3.73\,d$
10°	50°	30°	$1.55\,d$	$2.75\,d$
12.5°	52.5°	25°	$1.64\,d$	$2.14\,d$
15°	55°	20°	$1.74\,d$	$1.73\,d$

We notice that the compactness parameter is unchanged with respect to the previous case, because it depends only on β, whereas, for each value of β, both α and γ scales by 5°. If we want to keep α within 45° (for which polarization effects are not important), the only combination available is that with $\beta = 5°$, which though implies $l = 5.76\ m$. The siderostat mass would be the same as in the previous case, but the mass of structure by which siderostat and beam-compressor are connected would increase by about[28] a factor 9, which may be difficult to allocate in the mass budget of the telescope. Also, for a preliminary design, we may prefer to have a larger margin on the obscuration angle value, and consequently we would choose a larger β. For example, by taking $\beta = 10°$, α would be of 50° (still within the limit value for polarization effects), $\gamma = 30°$ and the mass of the siderostat would increase by a quantity which is a factor 0.14 the mass given in tab. SI.2 (for a SiC mirror with $\rho_A = 28\ kg\,m^{-2}$, Δm would be of about 3.3 kg), and the mass

[28] This is based on a rough evaluation utilizing the scaling factor R between the 2 linear dimensions: $R = L'/L$. If $M = \rho \cdot L^3$ then M' will be given by $M' = R^3 \cdot M$. In this case $L'/L = 5.76/2.75$ which implies $M' \sim 9\,M$.

of the structure would be unchanged.

Let us finally go through the exercise of keeping γ fixed at 30° and investigate different combination of α and β which are compatible with this value of γ. Tab. SI.5 shows the results for some of these combinations.

tab. SI.5 Combinations of α and β values, related HA range, and corresponding integration time in h, obtained assuming $\gamma = 30°$

α	β	ΔHA	$\Delta t\ (h)$
40°	20°	40°	74
42.5°	17.5°	50°	92
45°	15°	60°	111
47.5°	12.5°	70°	130
50°	10°	80°	150

It is interesting to notice that a variation of 10° in the maximum incidence angle allows us to make the integration time twice longer. On the other hand, the combination with $\alpha = 47.5°$ and $\beta = 12.5°$ is equivalent, as far as the integration time is concerned, to the case with $\alpha = 45°$ and $\beta = 10°$ but the telescope at 0° of latitude.

In conclusion, the preliminary set of parameter values for the telescope of the lunar interferometer that I suggest in order to exploit the maximum integration time available at 30° of latitude, are those presented in tab. SI.6.

tab. SI.6 Parameters for the preliminary design for a telescope of the lunar interferometer located at 30° latitude

latitude	$\varphi = 30°$
telescope orientation	E-W
inclination with respect to the horizontal	$\gamma = 30°$
maximum angle of incidence	$\alpha = 50°$
minimum angle of incidence	$\beta = 10°$
primary mirror diameter	$d = 1\,m$
siderostat maximum linear dimension	$D = 1.55\,m$
separation between siderostat and beam-compressor	$l = 2.75\,m$
primary mirror mass	$M_P = 0.785\,\rho_A$
siderostat mass	$M_{sid} = 1.32\,\rho_A$

2.2 Baseline orientation and u-v coverage

For any object's declination, the OPD varies as a function of the Moon rotation rate, for both baseline orientations. However, with an E-W orientation the OPD variation is larger, over the maximum integration time allowable according to the thermal equilibrium condition. From fig. B.2a and B.3a in *Baseline orientation: OPD variation and u-v coverage*, we see that the OPD maximum variation over 150 h of integration centered on the culmination point, is about 600 m for the E-W baseline and 150 m for the N-S baseline.

Actually, for the E-W baseline, the situation is quite similar to that presented for the 0° latitude case. In fact, the increment in HA range makes the maximum values of OPD that have to be compensated almost equal. The solution with a coarse delay line consisting of 5 fixed stations, plus a 50-m long fine delay line is still valid. The maximum rate of variation for the fine delay line is of the order of 1.2 $mm\,s^{-1}$ (2.4 $mm\,s^{-1}$ for the OPD variation rate), reached with objects at 0° of declination.

fig. SI.2 u-v coverage with an N-S baseline at 30° (maximum length 1 km): object at 30° (right) and 60° (left) north declination, 150 h of integration

In the case of the N-S orientation, the maximum variation in OPD during one integration is smaller than for the E-W baseline, nevertheless the maximum absolute delay (about 900 m for an object at 80° of declination) that has to be compensated is greater, and to accomplish this it is still necessary to have 8-9 fixed stations for the coarse delay line plus a 50-m long fine delay line. However, during one integration time, it is necessary to change the coarse delay only one time, while the fine delay line will constantly move to provide for most of the compensation. The maximum rate of variation for the fine

delay line is of about 0.5 $mm\,s^{-1}$ (1 $mm\,s^{-1}$ for the OPD variation rate), reached at the beginning of the observation with objects at 0° of declination.

As far as the u-v coverage is concerned, it is significantly improved in the case of the N-S orientation, while the coverage with the E-W orientation mainly takes advantage from the longer integration time available. As an example in fig. SI.2 and SI.3 I show the u-v coverage obtained with a N-S baseline and an E-W baseline in 150 h of integration and for two different object's declinations.

fig. SI.3 u-v coverage with an E-W baseline at 30° (maximum length 1 km): object at 30° (right) and 60° (left) north declination, 150 h of integration

3. Conclusions

3.1 Telescope configuration

Since the HA range depends on both α and β, there is a limit in the maximum HA coverage that can be obtained with this telescope configuration. If the maximum α is 45° and the minimum β is 5°, then the maximum value for $2(\alpha - \beta)$ is 80° which is equivalent to about 150 h of integration. Moving to higher latitudes in order to have longer thermal equilibrium periods, does not offer any advantage as far as the maximum integration period is concerned, because this configuration would not allow us to track an object for the entire period. If a siderostat plus beam-compressor telescope solution is chosen for the lunar interferometer and, on the basis of the results from the thermal analysis, the instrument is going to operate only during the night of the lunar cycle, then an equatorial location seems to be the most appropriate.

3.2 Baseline orientation

One of the advantages of locating the interferometer on the lunar surface, is to have the possibility to integrate the signal from the source for very long periods. However, to do this requires compensating for the large OPD variation occurring during the period of the observation. However, the variation of the OPD in itself is not a problem, as far as there are means to precisely compensate for that. Also, when we address this problem, we should take into account the different requirements that imaging and astrometric measurements may set.

With a 1-km baseline the maximum resolution, in the visible range, is of the order of 0.1 mas. To satisfy the requirement of accuracies of the order of 1% on astrometric measurements, the OPD has to be stable within 5 μm. This requirement is to be allocated to the fine delay line, and clearly a configuration which involves constant or slowly variable OPD would simplify the design of the metrology system that controls the OPD fluctuations. In this case a N-S baseline at equatorial or low latitude position should be chosen.

For imaging, on the contrary, it is important to have the largest u-v plane coverage is possible, which translates in desiring the largest variation in OPD[29]. For our simple two-element interferometer, the best u-v coverage attainable, exploiting the long periods of integration, is obtained with an E-W baseline, independently of the baseline latitude.

However, for any kind of visibility measurement, the OPD variation has to be controlled to a high degree of precision to avoid deterioration of the fringe visibility. In order to keep the visibility loss within 1%, at visible wavelengths, the maximum rms error on the OPD compensation has to be of the order of 9 nm.

The speed at which the fine delay line has to move is very low, for both orientations, with respect to those at which ground instruments work[30]. However, a problem may arise from the necessity to run the delay line smoothly enough to keep the rms error in the delay line position within $5 - 10$ nm, which is a requirement valid for both imaging and astrometric measurements. Besides, this is a problem that has to be addressed independently of the baseline orientation.

3.3 Interferometer preliminary design

On the basis of the evaluations performed above, which of course deal only with some of the aspects to be considered for selecting the configuration of the array and of the telescope, I suggest some of the characteristics of the preliminary interferometer design.

[29]The OPD variation is directly related to the variation of the projection of the physical baseline on the plane of the sky perpendicular to the line of sight, i.e a plane tangent to the celestial sphere. Since the u and v coordinates are defined as the components of the projected baseline, this means that the range of OPD variation is also related to the percentage of u-v coverage attainable with a given baseline.

[30]The maximum speed required for the E-W orientation is of the same order of that required for a ground instrument with a 35 m baseline (the ratio between the amount of HA covered in one hour on the Moon and on the Earth is about 0.036). Also, the fine delay line of IOTA can travel at constant speed of the order of $4\,\mu m\,s^{-1}$.

The starting point consists of a two-element array, but it will have to be incremented as soon as possible with other collecting elements: imaging requires at least three telescopes to obtain one closure phase, and astrometric measurements require to be made with respect two orthogonal baselines to get full information on the angular position of an object.

The instrument is operated only during the night of the lunar cycle, and each telescope is provided with a dome-shade that keeps the thermal variation between day and night below $100\,K$. Because of the night time operation there is no need to locate the interferometer at high latitudes: this can be an advantage as far as payload delivery on the Moon surface is concerned.

On the Moon there are two "privileged" categories of landing sites: equatorial sites and polar sites. They are privileged because an equatorial site is overflown at each orbit by a Moon satellite in equatorial orbit, while a polar site is also overflown at each orbit by a satellite in polar orbit. The sites which are not either near the pole or near the equator are not overflown by each orbit. It is necessary to wait until the movement of the orbit ascending node allows the satellite to fly within reach of the designated landing site , i.e. until when the closest distance between the satellite and the landing site is smaller than the cross range that can be covered by the lander. This cross range is directly proportional to the mass of propellant on board the lander.

To overcome this issue, there are two ways which must be traded or even combined:
- to increase the quantity of consumable on board the spacecraft to stay in Moon orbit for a longer time to wait for a favorable position with respect to the landing site;
- to increase the quantity of propellant on board the lander in order to increase its cross range and thus to reduce the waiting time.

Both solutions induce a mass penalty, and eventually the trade depends on the nature of the spacecraft (i.e. what are its needs in Lunar orbit)[31].

For an equatorial site and an E-W baseline, we will be able to perform visibility measurements exploiting the maximum variation in OPD available for a two-element instrument. In order to compensate for the OPD variation, the instrument is provided with a delay line system consisting of a coarse delay line with 5 fixed stations, and a short delay line 50-m long, which allows tracking the same object for a period of time up to 130 h.

The siderostat plus beam-compressor assembly has a primary mirror diameter of 1 m, is oriented E-W, and tilted with respect to the horizontal by 35°.

Starting from this preliminary design, the next step in the development of the project for a lunar interferometer will involve the detailed evaluation of the requirements on the components of the instrument and the definition of a single set of parameter values which properly fit with each other in view of the fulfillment of the mission goals.

[31] A third solution (to be verified) would be to select the launch date in such a way that the targeted Moon landing site is quickly overflown, but the feasibility of this solution and the resulting constraint on the launch window (from Earth) are to be investigated.

L'ambiente lunare

This section consists in a review of the information available about the lunar environment aspects which are of main interest for the operation of a lunar interferometer. In particular I describe the effects induced by the lunar dust and the lunar seismic activity. Other aspects taken into consideration are the flux of micrometeroids, the characterization of the lunar atmosphere, and the cosmic and solar radiation at the lunar surface. I do not deal with the thermal environment, because it is already presented in the chapter *Thermal analysis for a telescope of the lunar interferometer*. More information on these arguments may be found in the references within the following paragraphs.

0. Introduzione

Lo scopo di questa sezione é di presentare alcuni dei fattori che contribuiscono a descrivere l'ambiente lunare, con particolare attenzione a quegli aspetti che sono di interesse per la realizzazione e l'operazione di uno strumento astronomico sulla superficie della Luna.

La descrizione che segue si basa essenzialmente sui risultati ricavati dagli esperimenti condotti durante le missioni Surveyor 3 e varie missioni Apollo. Un maggiore spazio é dato alla descrizione degli effetti prodotti dalla "polvere" lunare e alla descrizione dei fenomeni sismici lunari, in vista della relazione che questi fattori possono avere con l'operazione di un interferometro lunare. Gli altri aspetti presi in considerazione sono il flusso di micrometeoroidi, la caratterizzazione dell'atmosfera lunare e la radiazione cosmica e solare presente al livello della superficie lunare. Non si prende qui in considerazione l'ambiente termico che é dettagliatamente presentato nella sezione *Analisi termica per un telescopio dell'interferometro lunare*.

La descrizione che si fornisce, di carattere piú che altro qualitativo, sará di referimento nell'ambito della valutazione della massima visibilitá raggiungibile da un interferometro lunare nella sezione *Uno "strawman budget" per la visibilitá dell'interferometro lunare*

1. Polvere lunare

La superficie lunare é coperta da piccoli frammenti prodotti da impatti di meteoroidi[32] sul sottostante strato roccioso. Le dimensioni dei frammenti che compongono quello che viene indicato genericamente col nome di regolith[33], variano da 40 a 280 μm. Nella maggior parte dei campioni prelevati durante le missioni Apollo e Luna, il 25% delle componenti del regolith sono di dimensioni minori di 20 μm, e circa il 10% ha dimensioni minori di 10 μm. La distribuzione caotica di queste componenti, associato con l'assenza di umiditá fanno della superficie lunare un ambiente tipicamente "polveroso", particolarmente deleterio per ogni tipo di componente ottica.

[32] Per una definizione di meteoroide si veda sezione *3*.

[33] Regolith é il termine generale per indicare lo strato o il mantello di materiale roccioso frammentario e non consolidato che costituisce la superficie del suolo (sia terrestre che lunare) e copre gli strati rocciosi sottostanti.

1.1 Effetto della polvere su componenti esposte all'ambiente lunare

Alcune delle componenti del Surveyor 3, esposte per un periodo di 31 mesi (32 giorni lunari) all'ambiente lunare, hanno evidenziato significative alterazioni nelle superfici esposte. In particolare sono stati riscontrati: deposito di particelle di regolith, *sputtering*[34] e degrado delle prestazioni nelle coperture per il controllo termico.

La polvere depositata su superfici di alluminio levigate ha causato un notevole aumento del rapporto fra radiazione assorbita e radiazione incidente (l'assorbanza varia da un valore di 0.26 per superfici relativamente pulite, fino a 0.75 per alte concentrazioni di polvere). É stata riscontrata anche una diminuizione in riflettanza (rapporto fra radiazione riflessa e radiazione incidente) dovuta sia ad effetti indotti dall'esposizione alla radiazione solare che al deposito di polvere. In particolare questa diminuizione é piú marcata per piccole lunghezze d'onda (0.6 - 1.0 μm) rispetto a lunghe lunghezze d'onda (2.0 - 2.4 μm).

Per quanto riguarda le coperture usate per il controllo termico, é stato riscontrato che un deposito di polvere di 10^{-5} - 10^{-4} g per centimetro quadrato, puó essere responsabile dell'aumento di un fattore 2 o 3 nella radiazione termica assorbita *(Carrol and Blair, 1972)*.

Strati di polvere molto fine depositati sui filtri della fotocamera del Surveyor 3, sono ritenuti responsabli di una diminuizione del coefficiente di trasmissione del 25%. Anche lo specchio della camera é risultato contaminato dalla polvere, causando una evidente perdita nella qualitá del contrasto nelle immagini prodotte. Non risulta invece che la polvere abbia degradato le prestazioni dei retroriflettori usati per misurazioni di distanze per mezzo di segnali laser inviati dalla Terra. Questo comunque non é un dato particolarmente significativo, visto che le prestazioni richieste alle ottiche di un telescopio per osservazioni astronomiche sono molto piú elevate.

1.2 Meccanismi per la distribuzione della polvere

Il basso grado di gravitá e la quasi totale assenza di atmosfera, fanno sí che la polvere possa essere facilmente sollevata e che, una volta sollevata dalla superficie lunare, ogni singola particelle possa viaggiare indisturbata per distanze anche molto grandi (dipende dalla velocitá iniziale). Ci sono due categorie di eventi che creano sollevamento di polvere: quelli legati ad attivitá umane e quelli legati a fenomeni naturali.

Nella prima categoria rientrano: atterraggio e decollo dei veicoli di trasporto, operazioni da parte di astronauti o moduli automatici sul suolo lunare; nella seconda categoria rientrano: impatti di micrometeoroidi e carica elettrostatica indotta sulle particelle da irradiazione ultravioletta.

Per quanto riguarda le attivitá umane, é stato valutato (*Johnson and Dietz, 1991*) che

[34]Lo *sputtering* é un fenomeno che occorre quando particelle energetiche colpiscono la superficie di un solido o di un liquido, causando l'emissione di particelle e l'erosione della superficie del solido. Le particelle "schizzate" via dall'oggetto bersaglio possono essere costituite da atomi e molecole neutri o ionizzati, ma anche da parti di dimensioni maggiori.

la maggior parte di esse sono in grado di eiettare particelle di polvere ad una distanza massima di 1 km dal luogo in cui sono state sollevate. Invece, la polvere sollevata dai gas di scarico dei veicoli in approccio o allontanamento dalla Luna puó coprire una distanza pari all'intera circonferenza lunare, che peró viene percorsa in un tempo relativamente breve (dell'ordine di un'ora).

Collocando gli strumenti ottici ad un'adeguata distanza (maggiore di 1 km) da un'eventuale base permanente sulla Luna, e proteggendo le ottiche per un periodo di tempo sufficientemente lungo (dell'ordine al massimo di qualche ora) durante le operazioni di atteraggio e decollo di veicoli di trasporto, si dovrebbe poter far fronte al problema della contaminazione prodotta da attivitá umane.

Per quanto riguarda il sollevamento indotto da cause naturali, l'impatto di micrometeoroidi sul regolith produce uno sciame secondario di particelle di polvere di entitá proporzionale all'energia di impatto della micrometeoroide. Queste particelle secondarie, anche se piú lente delle primarie, possono potenzialmente danneggiare tanto le componenti ottiche che quelle struttuturali dello strumento. A questo proposito peró, ulteriori studi sono necessari per ottenere valutazioni quantitative sia dell'effettivo grado di danneggiamento che delle possibilitá di mitigazione del danno stesso (*Johnson and Dietz, 1991*).

Particolarmente interessante é il fenomeno del sollevamento di polvere prodotto da fenomeni di fotoconduttivitá indotta.
Il risultato di misure fatte sulla conduttivitá elettrica di materiale prelevato dal suolo lunare, mostra che rispetto a quando il materiale é in zona d'ombra, irradiazione infrarossa causa un aumento di conduttivitá pari a circa un fattore 10, mentre irradiazione ultravioletta causa un aumento pari ad un fattore 10^6 (*Alvarez, 1977*). Questi grossi cambiamenti in conduttivitá in seguito ad irradiazione possono produrre ampi spostamenti di particelle cariche attraverso il terminatore solare[35] sulla superficie lunare. Il numero di particelle cariche sollevate al passaggio dal giorno alla notte, potrebbe essere tale da produrre vere e proprie nubi di polvere che si spostano sulla superficie lunare. Evidenza di questo fenomeno é stata registrata da esperimenti condotti durante la missione Apollo 17, che hanno registrato un incremento nel conteggio di particelle cariche ogni volta che il rivelatore dello strumento era attraversato dal terminatore solare. In particolare questo incremento inizia circa 40 ore prima e termina circa 30 ore dopo il sorgere del sole. *McDonnel (1979)* ha calcolato che la levitazione di tali particelle potrebbe giungere fino ad altezze di 10 m sulla superficie lunare.

Le componenti di queste nuvole di particelle cariche in movimento sulla superficie lunare possono facilmente aderire su ogni superficie che incontrano, formando un rivestimento elettricamente carico. Poiché la polvere sollevata durante la fase notturna ha la massima capacitá di carica (perché la conducibilitá del materiale durante la notte é minima), essa verrebbe prontamente attratta dalle superfici cariche creando su di esse degli spessi depositi. Questi strati di polvere verrebbero poi solo in parte rimossi al passaggio attraverso

[35]Terminatore é la linea che separa zone di luce e zone d'ombra della Luna o di un qualsiasi pianeta.

il successivo terminatore solare. Le implicazioni che questo possibile meccanismo di sollevamento e deposito di polvere comporta, sono notevoli per l'operazione di uno strumento ottico, soprattutto se deve operare durante la fase notturna. Va comunque sottolineato che questa é un'ipotesi che richiede verifica sperimentale in sito, soprattutto per quanto riguarda l'altezza a cui le particelle cariche possono essere sollevate.

2. Attivitá sismica

I dati relativi all'attivitá sismica della Luna sono stati acquisiti fra il 1969 e il 1977 durante il periodo operativo della rete di stazioni sismiche collocate in sito in occasione di diverse missioni Apollo.

In base ai risultati dell'analisi di questi dati, sono stati identificati quattro tipi principali di fenomeni sismici naturali: lunamoti profondi, lunamoti superficiali, lunamoti termici e lunamoti prodotti da impatti di meteoroidi. Essi sono espressione dell'attuale stato dinamico dell'interno del pianeta e dell'ambiente interplanetario che circonda la Luna (*Nakamura et al., 1982*).

2.1 Lunamoti profondi

Questo tipo di eventi, il piú abbondante, é caratterizzato da piccole magnitudini, minori del 3° nella scala Richter[36] e si originano a profonditá comprese fra 800 e 1000 km dalla superficie. Il loro manifestarsi é fortemente correlato con gli effetti di marea prodotti sulla Luna dai moti relativi della Terra e del Sole, e, secondo le ipotesi piú ricorrenti, ne sono ritenuti la causa.

Le forze di marea sollevano ed abbassano la superficie lunare di circa 10-20 cm ogni giorno. Poiché peró l'orbita lunare é leggermente ellissoidale, le forze di marea indotte variano periodicamente, a seconda della separazione relativa fra Luna e Terra. Questo causa un accumularsi e successivo rilascio di energia (dell'ordine di 0.2 bar), che se concentrata in zone di discontinuitá presenti in profonditá all'interno della Luna puó innescare un moto tellurico. Il meccanismo con cui si producono questi eventi profondi non é peró ancora stato definito e ulteriori studi sarebbero necessari per far luce su questo problema. In generale, non si ritiene comunque che questo tipo di eventi possano avere effetti di disturbo per possibili attivitá umane sulla superficie lunare.

2.2 Lunamoti superficiali

I lunamoti superficiali sono gli eventi sismici piú energetici che sono stati registrati sulla Luna (il piú intenso di quelli registrati durante gli 8 anni delle osservazioni Apollo, era di magnitudine maggiore di 5), sebbene essi siano rari in confronto con tutti gli altri tipi di eventi sismici. La loro occorrenza non é correlata con fenomeni mareali e pertanto si

[36]La massima energia rilasciata da questi eventi é dell'ordine di 10^6 J. La magnitudine dell'evento si ricava in base alla formula

$$m = \frac{logE + 1.2}{2.4}$$

dove E é l'energia rilasciata. (*Richter, 1958).*

ritiene che siano di origine tettonica. É probabile che essi siano legati al rilascio di energia termoelastica che si é accumulata nelle zone piú superficiali del pianeta, a causa di un globale raffreddamento e della conseguente contrazione della Luna.

Nakamura et al. (1980), hanno mostrato che i lunamoti superficiali hanno molte caratteristiche in comune con i terremoti che si producono all'interno delle placche tettoniche terrestri (detti intraplacca per distinguerli dai terremoti di faglia che si producono lungo i confini delle placche tettoniche). Per esempio, entrambi sembrano originarsi in zone di preesistente debolezza (discontinuitá) della litosfera[37]. Anche l'abbondanda relativa di eventi di grossa e piccola intensitá é simile, il che suggerisce che meccanismi simili siano all'origine di questi eventi su entrambi i pianeti. A questo proposito é da notare che terremoti intraplacca si verificano di norma su larghe scale temporali, ma sono in genere di elevata magnitudine. Non si puó escludere quindi la possibilitá che eventi sismici di magnitudine elevata ($m \simeq 7$) possano verificarsi anche sulla Luna, ma essi non sono stati rivelati nel breve periodo durante cui sono state condotte le osservazioni del programma Apollo.

Esistono peró sostanziali differenze fra eventi sismici superficiali lunari e terrestri dovuti alle diverse caratteristiche dell'ambiente lunare, che influenza il modo di propagazione delle onde sismiche. L'effetto in termini di moto vibrazionale del terreno é alquanto diverso per lunamoti e terremoti di pari magnitudine. Ad esempio, gli spettri di potenza dei lunamoti mostrano che la maggior parte dell'energia é associata a frequenze maggiori rispetto a quelle a cui si distribuisce per un terremoto. Le onde sismiche sono molto meno attenuate sulla Luna che non sulla Terra, col risultato che l'energia rilasciata da un lunamoto sará distribuita su una zona molto piú vasta, intorno all'epicentro, di quella che sarebbe sulla Terra. Infine, i segnali sismici sulla Terra sono relativamente brevi e di carattere impulsivo. Sulla Luna invece i segnali sismici presentano un graduale incremento in intensitá nell'arco di alcuni minuti, seguito da un decadimento di durata estremamente lunga (spesso dell'ordine di alcune ore). La caratterizzazione dell'effetto dei lunamoti superficiali in termini del moto vibrazionale del terreno é un ambito che richiede ulteriori studi, in particolare per la valutazione dei possibili rischi di danneggiamento a strutture poste sulla superficie lunare.

2.3 Lunamoti termici

Questi sono eventi sismici di intensitá molto piccola, causati da variazioni di temperatura che si manifestano alla superficie lunare, o in prossimitá di essa. Essi sono misurabili solo entro distanze di pochi km dal punto in cui si producono, e sono caratterizzati da periodicitá e caratteristiche costanti. In particolare, la periodicitá di questi eventi strettamente correlata con il periodo di lunazione e la tendenza degli eventi di presentarsi durante il giorno lunare, suggerisce che essi siano innescati da tensioni di origine termica causate da espansione (eventi che si presentano nelle prime ore del giorno lunare) e contrazione (eventi che si presentano in prossimitá o subito dopo il tramonto) del materiale lunare.

[37] Lo strato solido piú esterno della superficie di un pianeta. Consiste della crosta e della parte superiore del mantello.

Poiché a profonditá dell'ordine di 50 cm la temperatura é pressoché costante, il materiale interessato da tali tensioni termiche, puó essere sia materiale roccioso superficiale che il regolith. L'analisi delle caratteristiche dei segnali registrati suggerisce comunque che questo tipo di tensione sia in grado di innescare improvvisi slittamenti di regolith deposto lungo superfici in pendenza, piuttosto che la frattura, in prossimitá di punti di discontinuitá, del materiale roccioso. Anche in questo caso peró, le ipotesi avanzate richiedono una conferma sperimentale in sito e soprattutto nuovi dati sono necessari per quantificare l'entitá dello spostamento superficiale che il fenomeno comporta.

2.4 Effetti prodotti da impatti di meteoroidi

La sorgente da cui originano gli impatti di meteoroidi non é, chiaramente, interna alla Luna, quindi a rigori essi non rappresentano dei veri e propri eventi sismici lunari. Gli impatti di meteoroidi comunque sono secondi, in termine di numero di eventi prodotti, solo alle forze mareali, fra le cause di fenomeni sismici sulla Luna.

Eventi sismici dovuti a queste collisioni variano in un ampio intervallo di energia, che corrisponde all'ampio intervallo in massa degli oggetti in collisione, che va da circa 10^{-1} fino a 10^3 kg. Comunque, sebbene i sismografi delle missioni Apollo abbiano registrato segnali associati ad oggetti la cui massa é stimabile fra 500 g e 50 kg, la maggior parte dei segnali sono dovuti ad impatti di masse molto piú piccole. *Oberst e Nakamura (1991)* hanno classificato gli impatti in due categorie: "piccoli" impatti, se il segnale sismico prodotto é stato registrato da una sola delle stazioni di rivelamento; "grossi" impatti, se l'evento é stato registrato da tutti gli elementi della rete. Sebbene, oltre alla massa, altri fattori quali la velocitá, la densitá e l'angolo di impatto, contribuiscono a determinare la tipologia dell'impatto, *Oberst e Nakamura* hanno valutato che in genere oggetti di massa minore di 1 kg producono piccoli impatti, mentre oggetti con masse maggior producono grossi impatti. Durante il periodo di attivitá della rete sismica lunare, sono stati registrati 511 piccoli impatti e 416 grossi impatti, e in ambo i casi si é registrata una tendenza al raggruppamento degli eventi, molto piú marcata per i piccoli impatti.

3. Micrometeoroidi

Il termine meteoroide viene usato per indicare corpi solidi naturali, che viaggiano attraverso lo spazio, di dimensioni troppo piccole per essere identificati come asteroidi o comete. Meteoroidi con diametri minori di 1 mm (e masse minori di 10^{-2} g) sono generalmente classificati come micrometeoroidi[38].

La tenue atmosfera lunare permette anche alle micrometeoriti di piú piccole dimensioni di giungere alla superficie lunare con la massima velocitá, che é dell'ordine dei 10 $km\,s^{-1}$ ma che, in alcuni casi, puó anche superare i 50 $km\,s^{-1}$. Analisi di rocce lunari hanno rivelato la presenza di numerosi microcrateri sulle loro superfici, attribuiti appunto all'azione delle micrometeoriti. Dalla combinazione dei dati sulle misure di flusso di particelle di massa

[38]Il termine meteorite si riferisce invece a meteoroidi che sono precipitati attraverso un'atmosfera e sono stati in seguito recuperati.

minore di 10^{-7} g (il cui impatto é in grado di produrre microcrateri di 10 μm di diametro) ottenute dall'analisi delle superfici esposte di alcune parti dei veicoli delle missioni Apollo, con i dati calcolati per le rocce lunari, si ricava la stima del flusso di particelle, in funzione del diametro del cratere prodotto, presentato in tab. E.1 (da *Johnson et al., 1988*)

tab. E.1 Flusso di micrometeoroidi sulla superficie lunare, in funzione del diametero del cratere prodotto

Diametro del cratere (μm)	$Crateri \cdot m^{-2} \cdot anno^{-1}$
> 0.1	3000000
> 1.0	12000
> 10	3000
> 100	0.6
1000	0.001

Flussi di questo ordine di grandezza sono chiaramente un problema per ogni tipo di superficie ottica ad essi esposta. L'uso di opportune schermature metalliche di 2-3 mm di spessore dovrebbe essere sufficiente per proteggere tali superfici dall'impatto di particelle fino ad alcuni milligrammi in massa (responsabili per crateri di diametro dell'ordine del millimetro). Queste protezioni non sono peró sufficienti per bloccare particelle di massa superiore, dell'ordine del grammo, che sebbene siano eventi rari, sono in grado di produrre crateri di dimensioni dell'ordine del centimetro.

4. Atmosfera

L'atmosfera lunare é estremamente tenue. La concentrazione di gas é pari a circa $2 \cdot 10^5$ mol/cm^3 durante la notte lunare, ed é dell'ordine di 10^4 mol/cm^3 durante il giorno lunare. Questi valori sono circa 14 ordini di grandezza inferiori a quelle riscontrati per l'atmosfera terrestre, un fatto che conduce spesso ad affermare, erroneamente, che la Luna sia priva di atmosfera.

La maggior parte dell'attuale conoscenza della composizione dell'atmosfera lunare é basata su valutazioni teoriche, piú che su misure sperimentali. Questo é dovuto al fatto che le misure previste nell'ambito di diversi esperimenti delle missioni Apollo, sono risultati contaminati dai gas rilasciati durante le missioni stesse (Il gas rilasciato in sei missioni Apollo é pari a sei volte quello della stessa atmosfera naturale della Luna).

4.1 L'atmosfera naturale della Luna

I maggiori costituenti dell'atmosfera lunare sono neon, idrogeno, elio ed argon. La sorgente del neon e dell'idrogeno é cosituita dal vento solare. Anche l'elio deriva in gran parte dal vento solare, ma circa il 10% di esso é probabilmente di origine lunare, prodotto di qualche decadimento radioattivo. La componente di argo é in gran parte costituita da ^{40}Ar che é un prodotto del decadimento radiativo del ^{40}K presente nel materiale lunare. Solo un 10%

circa dell'argo é sottoforma di ^{36}Ar, e proviene dal vento solare (*Lunar Sourcebook, 1991*).

Per la valutazione delle concentrazioni delle varie componenti dell'atmosfera lunare, bisogna tenere conto di alcuni fattori. Innanzitutto il problema della contaminazione, che influenza in modo particolare la determinazione delle concentrazioni durante il giorno (si veda il successivo paragrafo *4.2*). Il secondo fattore da considerare é che la concentrazione delle specie di origine solare é direttamente influenzata dalla variazione nelle abbondanze delle componenti del vento solare, in corrispondenza dei cicli solari e altri fenomeni dell'attivitá solare. Terzo aspetto, é che la differenza in temperatura fra giorno e notte diminuisce andando verso latitudini piú elevate, risultando in minori differenze fra la composizione dell'atmosfera nelle ore del giorno e della notte man mano che ci si avvicina alle regioni polari. Infine occorre tener presente il diverso comportamento delle singole specie in funzione della temperatura. In particolare l'abbondanza dell'argo diminuisce nelle ore notturne, invece che aumentare, in modo molto graduale. Questo é dovuto al fatto che esso é un gas che condensa alle temperature tipiche della notte lunare. Al contrario, neon, elio e idrogeno non condensano, anzi la loro abbondanza aumenta durante la notte, poiché la concentrazione dei gas che non condensano a queste temperature é proporzionale all'inverso della temperatura stessa ($\propto T^{-5/2}$ da *Hodges et al., 1974*).

4.2 L'atmosfera artificiale della Luna

La presenza su larga scala di attivitá umane sulla Luna puó dare inizio ad un processo irreversibile di inquinamento e completa alterazione dell'atmosfera lunare.

Ognuna delle missioni Apollo ha rilasciato una massa di gas equivalente a quella dell'atmosfera naturale della Luna; molti di questi gas sono condensabili, pertanto la loro concentrazione aumenta duarnte il giorno e diminuisce durante la notte. Alle concentrazioni attuali, molti di questi gas sono rimossi dall'atmosfera lunare per mezzo di processi di ionizzazione, in seguito all'interazione con particelle energetiche del vento solare, e grazie alla presenza del campo elettrico intraplanetario che rimuove dall'atmosfera lunare le particelle ionizzate. Il processo di allontamento per mezzo di riscaldamento delle particelle fino a raggiungere la velocitá di fuga, é invece trascurabile. Peró se l'atmosfera lunare raggiunge una massa 100 volte superiore a quella attuale, l'importanza, in termini di efficacia nell'allontanamento delle particelle, di questi due processi si inverte. Infine, per masse dell'ordine di 10^7-10^8 kg questi due processi di perdita raggiungono uno stato stazionario, se non subiscono addirittura un'inversione di tendenza. In questo caso, la Luna acquisirebbe una vera e propria atmosfera permanente (*Vondrak, 1990*).

5. Radiazioni

La radiazione che raggiunge la seperficie della Luna é assai diversa da quella che raggiunge la superficie terrestre, poiché la Luna non possiede né un forte campo magnetico (si valuta che sia compreso fra 10^{-2} e 10^{-4} volte quello della Terra all'equatore) né un'atmosfera sufficientemente densa da bloccare o attenuare la radiazione. Per questo motivo la radiazione che giunge sulla superficie lunare é assai varia, sia per quanto riguarda le componenti che l'energia. Si prendono ora in considerazione i principali tipi di radiazione che si registrano

sulla Luna.

Innanzitutto la radiazione solare: circa il 7% dello spettro solare é concentrato nella zona della radiazione ultravioletta. Poiché la costante solare alla distanza della Luna dal Sole é di 1393 $W\,m^{-2}$, questo significa che il flusso ultravioletto totale che raggiunge la superficie lunare é dell'ordine di 97 $W\,m^{-2}$. In particolare questa radiazione é ritenuta responsabile del degrado delle caratteristiche dei rivestimenti per il controllo termico, come é stato riscontrato in materiali utilizzati nella missione Surveyor 3 e missioni Apollo.

Il secondo tipo di radiazione é dovuta a particelle cariche di provenienza cosmica e solare. Esse possono essere raggruppate in tre gruppi che si differenziano per l'energia e l'entitá del flusso.

1) raggi cosmici galattici di alta energia ($1 - 10\ GeV/nucleone$), caratterizzati da flussi dell'ordine di $1 \cdot cm^{-2}\,s^{-1}$ e profonditá di penetrazione fino ad alcuni metri.
2) particelle prodotte da *flares* (brillamenti) solari, con energie comprese fra 1 e 100 $MeV/nucleone$, caratterizate da flussi che raggiungono valori di $100 \cdot cm^{-2}\,s^{-1}$ e profonditá di penetrazione dell'ordine del centimetro.
3) particelle provenienti dal vento solare, di bassa energia ($\sim 1000\ eV$), caratterizzate da flussi elevati, dell'ordine di $10^8 \cdot cm^{-2}\,s^{-1}$ e piccole profonditá di penetrazione (molto minori del cm) (*Johnson et al., 1988*).

Le profonditá di penetrazione riportate, si riferiscono solo alle particelle primarie. Reazioni fra le particelle di alta energia (raggi cosmici e particelle da *flares* solari) che interagiscono con il materiale lunare causano uno sciame di radiazione che penetra molto piú in profonditá, fino a diversi metri. Di questo si deve tenere conto soprattutto per la schermatura delle componenti elettroniche e di qualsiasi tipo di rivelatore per osservazione astronomica. Per quanto riguarda le particelle di vento solare invece, il problema maggiore deriva dall'alto flusso piuttosto che dall'energia. L'impatto di tali particelle potrebbe favorire il deterioramento di superfici esposte all'ambiente lunare, in particolare le superfici ottiche. Opportune schermature e collocazione delle parti elettroniche a profonditá di alcuni metri nel terreno lunare, dovrebbero diminuire il rischio di danneggiamento associato a queste forme di radiazione.

Strawman visibility budget for the lunar interferometer

0. Introduction

In this section I propose a strawman visibility budget for the lunar interferometer, obtained following the scheme adopted to estimate the instrumental visibility for IOTA. However, in this case the problem is reversed: assuming a given visibility goal for the instrument, I will suggest error allocations (by means of the value of the Strehl ratio, see section *0.1* in *Instrument visibility for IOTA*) to different subsystems in order to match the final goal. These allocations should be used as starting requirements for the selection of materials, design and configuration of the subsystems, but they are likely going to be modified until the design of these subsystems is fixed.

A strawman error[39] budget is not supposed to be a refined (scattered) error allocation to all the possible subsystems. Rather, it has to show error allocation with respect to terms grouped together according to which subsystem they belong to and to the class of error cause considered, so that it should be easier for the designer and the manufacturer to state if an adequate system has been produced, given more degree of freedom for "internal" error parameters to be properly adjusted.

In particular, I will separate, whenever it is possible, contributions to the error budget from *environment* dependent and *non-environment* dependent terms, where the latter definition indicates those conditions which are strictly related to the system as it is built. Even if this kind of classification of error causes will not always be straightforward, it is intended to be a convenient means to assess the benefits of spending resources to improve the system assembly, versus spending them to control the environmental impacts.

As far as the characterization of the error sources is concerned, the main difference for a Moon-based interferometer with respect to an Earth-based interferometer is the absence of the atmosphere. This makes other environmental conditions, sometimes much more extreme, as the thermal ones, become more relevant than what they are for a ground instrument, for which many environmental effects are negligible in comparison with the disturbance due to the atmosphere. The analogy between the role of the atmosphere on Earth and of the "environment" on the Moon, can be used to derive the visibility goal for the lunar interferometer.

Then, visibility allocations to four main subsystems of the interferometer will be made: the total visibility will be expressed as the product of their related Strehl ratios, namely:

$$S_{goal} = S_{OPT} \times S_{OPD} \times S_{DIF} \times S_{SERVO} \tag{40}$$

[39]Because of the relation between *rms* wavefront error and Strehl ratio, $S = exp[-(2\pi\sigma/\lambda)^2]$, in what follows I will use equivalently the expression "visibility budget" and "error budget". They both give the same kind of information, however the first one may be more meaningful for the scientist who is going to utilize the instrument, while the second one may be preferred by the manufacturer and the engineer who are going to build the (components of) instrument.

where S_{OPT} is referred to visibility loss due to any optical component, S_{OPD} is referred to visibility loss due to optical path change during an integration time, S_{DIF} is referred to visibility loss due to diffraction effects in the beam, and S_{SERVO} is referred to visibility loss due to imperfect tilt correction by the servo system[40].

With reference to the preliminary design suggested for the lunar interferometer in section *Interaction among subsystems: an example for the lunar interferometer*, I present, for each subsystem, a visibility budget suballocation to several components. However, even when the suballocation is more detailed, as in the case of the optical subsystem, it should always be conceived with enough flexibility to allow it to be modified during the development of the design: the aim of this error budget is to help perform first order design\manufacturing trades.

1. Determination of the visibility goal

To determine the visibility goal I consider two aspects: the effect of visibility loss on the SNR of the fringe detection, or equivalently on the increase in the magnitude limit for the observation, and the comparison with the best visibility performances we can expect to obtain with ground instruments.

As far as this comparison is concerned, there is a fundamental difference in the magnitude limit that, assuming the same detectors and operation conditions, can be reached on the Earth and on the Moon. Assuming that the ground interferometer is equipped with adaptive optics, so that it can exploit the maximum aperture area of its collectors, the difference is made by the integration time. In fact, adaptive optics cannot work as an efficient system to cophase the telescopes, and the relative phase of the telescopes will still drift with a time constant characteristic of the atmospheric turbulence. An estimate of the relative phase drift is still given by the ratio of the phased beam diameter to the wind velocity in the relevant part of the atmosphere (*Ridgway, 1992*), which we assume to be of the order of 10 ms^{-1}. With adaptive optics we can assume that the diameter of the phased beam is equal to the mirror diameter so that the maximum integration time for the ground interferometer will be given by $\tau = d/v_{wind}$. Since the difference in magnitude limit between the two systems is given by

$$\Delta m = 2.5\, log_{10} \left(\frac{d_E^2 \cdot \tau}{d_M^2 \cdot T} \right) \tag{41}$$

where d_E and d_M are the aperture diameters of the instrument and τ and T are the integration times, on Earth and on the Moon respectively, for a ground interferometer to reach the same magnitude limit of the lunar one, its diameter should be of $10\, d_M$, if $T = 100\ s$, and of the order of $21.5\, d_M$, if $T = 1000\ s$, where for our lunar interferometer d_M

[40] We assume that a servo system is in any case necessary to correct for any tilt in the wavefront that may arise during the operation of the instrument, even if the requirements on this component are not so much demanding as for a ground instrument, for which it has the main task to correct for the wavefront tilt induced by the atmospheric turbulence.

$= 1\ m$ and I used the relation $d_E^2 \cdot \tau \simeq d_E^3/10\, ms^{-1}$.[41]

For the evaluation of the effect of visibility loss on the SNR I considered the same performances and conditions of operations of the 512×512 CCD detector (quantum efficiency in the visible $\simeq 80\%$, read-noise $= 6\ e^-\ rms$, spectral bandwidth $\simeq 46\%$ centered at 0.65 μm) we are using for IOTA, and applied the formula to compute the SNR as a function of the measured visibility and of the source magnitude in the visible range as given by *Millan-Gabet et al., 1995*. Tab. VB.1 presents some results for three values of visibility (1, 0.75, and 0.5) and for three different periods of integration time (1000 s, 1 h, and 100 h). The values in the three columns give the magnitude limit in conditions of low (about 10), medium (about 100) and high (about 1000) SNR.

tab. VB.1 Visual magnitude limit, m_V, in conditions of low, medium, and and high SNR for different integration times ($\tau = 1000\ s$, 1 h, 100 h), as a function of the fringe visibility, V

	$V = 1$					
τ	m_V	SNR	m_V	SNR	m_V	SNR
1000 s	27	10	22	135	17	1360
1 h	28	13	24	121	19	1026
100 h	33	12	29	101	24	1026
	$V = 0.75$					
τ	m_V	SNR	m_V	SNR	m_V	SNR
1000 s	26	14	22	101	17	1019
1 h	28	9.5	23	121	18	1220
100 h	33	9.5	28	121	23	1220
	$V = 0.5$					
τ	m_V	SNR	m_V	SNR	m_V	SNR
1000 s	26	9	21	107	16	1077
1 h	27	11	22	128	17	1290
100 h	32	11	27	128	22	1290

From these results we can conclude that in order to keep the increase in magnitude limit within 1 mag, we can tolerate a maximum visibility loss of 25%, for which this condition is satisfied even in the most demanding case of $SNR \sim 1000$. Eventually the requirement on the sensitivity of the measurements is determined by the scientific objectives[42]. However, the lunar interferometer should be able to cover a wide range of different scientific

[41] Here I am comparing performances between interferometers and not between lunar interferometers and large ground-telescopes (see section *Interferometry from the Earth and from the Moon.*)

[42] In particular, for a better dynamic range (i.e. the ability to see weak structures in the presence of bright objects), more photons are needed per data sample, to determine the fringe parameters accurately.

objectives, so that there will not be a unique sensitivity requirement. To take into account for the large range of sensitivities expected, we suggest for the visibility goal an optimistic maximum loss of 20%. Of course, the sensitivity that will actually be achieved will be determined only in part by physics, for the rest it will be determined by cost trade-offs in the design.

On the other hand it is necessary to check if this performance goal is compatible with the visibility performance expected for ground instruments.

Let us assume that the instrument operates, by means of adaptive optics, always in the conditions $D/r_o = 1$, where D is the maximum aperture diameter and r_o is the coherence length[43]. In this condition, the visibility loss associated with atmospheric effects is of the order of 13%, i.e. $S_{atm} = 0.87$. Assuming, for the ground-based instrument, an instrumental visibility goal of the same order of that imposed by the atmosphere, we would get for the total visibility $S_{tot} \sim 0.75$. This can be a demanding goal to meet[44]: actual instruments can reach $S_{tot} \sim 0.5$, and improvements are certainly to be expected. Likely an instrumental visibility of the order of 0.75 is a more feasible goal to meet (*CHARA, Technical Report No.15, May 1995*), and it would imply $S_{tot} \sim 0.65$.

Finally, we could substitute the visibility loss term due to the atmosphere for a ground instrument, with the visibility loss term due to environmental causes for a lunar one. Also in this case we will require that the instrument visibility is of the same order of the "environmental" visibility, which automatically implies a total visibility $S_{tot} \sim 0.75$

According to all the above evaluations, I eventually set a preliminary visibility loss goal for the lunar interferometer of 20%, i.e. $S_{tot} = 0.8$, to which corresponds a total wavefront rms error of $\lambda/13$.

In the following sections I will indicate several subsystems, and main items within each subsystem, among which this wavefront error has to be distributed. In particular, the ideal wavefront error allocation would follow a form which takes into account environmental and non-environmental causes, as it is also suggested by the analogy with ground instruments where we make a distinction between atmospheric and non-atmospheric related effects. The ideal form for the wavefront error/visibility budget allocation is presented in fig. VB.1.

More photons per sample are also needed for successful reconstruction of complex objects, with more than one non-zero resolution element (pixel) in the image.

[43] See fig. IV.1 in section *0.1* in *Instrument visibility for IOTA*.

[44] Above all if, in order to reach the same magnitude limit of a lunar instrument, the telescope diameter has to be of the order of 10 meters or larger.

fig. VB.1 Schematic of wavefront error budget (or equivalently visibility budget) form for the lunar interferometer

2. Visibility budget for the optical subsystem

Following the scheme used in *Instrument visibility for IOTA*, the Strehl ratio for the optical subsystem is given by the product of these terms:

$$S_{OPT} = S_{TEL} \times S_{SID} \times S_{REL} \times S_{BS} \times S_{WIN} \tag{42}$$

where S_{TEL} refers to the telescope, i.e. the beam-compressor which consists of primary plus secondary mirror, S_{SID} refers to the siderostat, S_{REL} refers to all the flat mirrors of the relay system, S_{BS} refers to the beamsplitter, and S_{WIN} is referred to the all the windows the beam meets on the way to the combination area. Among all these components the most complex is certainly the telescope, for which we can indicate many items to which to suballocate the wavefront error.

A schematic of the visibility budget allocation among the components of the optical subsystem is presented in fig. VB.2, and a more detailed explanation of each term is given in the following paragraphs from 2.1 to 2.5.

fig. VB.2 Schematic of the visibility budget allocation among the components of the optical subsystem

2.1 Visibility loss allocation for the telescope

2.1.1 Non-environment related terms

• For the primary mirror:

1) surface optical errors: these depend on the development in the mirror manufacturing and polishing techniques. However, the error budget for this term should be forced to be as small as possible, because of the uncertainty in the error that is related to the following item.

In order to specify the errors of the surface mirror it is appropriate to make use of the structure function of the wavefront, as defined by *Hill (1990)*.[45].

By setting the Strehl ratio goal for the optical surface and by means of the structure function one can find the optical tolerances for the surface on different distance scales. In tab. VB.2 I present some results for three different Strehl ratios. The Strehl ratio of 0.95 is the goal assumed for the optical tolerances of a ground-based instrument whose total instrumental visibility goal is of 0.75. The column named σ_w gives the *rms* wavefront error obtained by solving the simple Strehl ratio equation with a given value of S, without considering the structure function method.

tab. VB.2 *rms* wavefront error at different spacial scale, x in m, for a 1-m mirror, as a function of the Strehl ratio S. σ_w is the *rms* wavefront error derived by solving the simple Strehl ratio equation without considering the structure function method

S	σ_w	$x(m)$				
		0.005	0.2	0.5	0.8	1.0
0.99	$\lambda/62$	$\lambda/126$	$\lambda/53$	$\lambda/33$	$\lambda/35$	$\lambda/57$
0.98	$\lambda/44$	$\lambda/88$	$\lambda/37$	$\lambda/23$	$\lambda/25$	$\lambda/39$
0.95	$\lambda/27$	$\lambda/63$	$\lambda/45$	$\lambda/14$	$\lambda/30$	$\lambda/48$

[45]The application of the structure function to optical tolerance has been described by *Hill (1990)*. The structure function is defined as the function which describes the error in the wavefront as a function of the separation between two points on the wavefront itself. It is given by

$$\delta^2(x) = \left(\frac{\lambda}{2\pi}\right)^2 6.88 \left(\frac{x}{r_o}\right)^{5/3} \left[1.0 - 0.975 \left(\frac{x}{D}\right)^{1/3}\right]$$

where D is the mirror diameter, $\delta(x)$ is the *rms* wavefront difference between two points on the wavefront separated by x, and r_o is defined by the requirement on the Strehl ratio, $S = exp[(-\frac{2\pi}{\lambda}\sigma_w)^2]$ by means of

$$r_o = D \cdot \left(\frac{\sigma^2}{0.134}\right)^{-3/5}.$$

In particular, the results from the structure function allow us to relax the tolerances on the telescope optics at large scales. Otherwise, it could be very difficult to reach very high Strehl ratio values over the full aperture.

For manufacturing considerations we have to recall that the rms mirror surface error has to be 0.5 times the rms wavefront error.

Even if this will be quite demanding for the mirror optical tolerances, in order to meet the final visibility goal we have set, we allocate to this term a Strehl ratio of 0.99.

2) fabrication to Moon changes: this factor is related to the capability of building the mirror in environmental conditions which are as closed as possible to the lunar conditions. It involves both gravity and thermal environment. As far as gravity is concerned, experience (related to large mirrors fabrication) has already been developed for the construction of the Hubble Space Telescope. For this program the manufacturer used a "0-g metrology mount" to simulate the on-orbit 0-g environment. Because the metrology mount is always somewhat imperfect, some residual error will exist. However, for the lunar telescope the scenario is quite different, because it is exposed to a temperature range wider than the about 32 K range over which the Hubble telescope operates. The uncertainty of the metrology mount's performance operating over such a large lunar temperature range (at least of the order of 100 K) could completely mislead the measurements of the effect on the mirror due to the 1-g to 1/6-g variation. Because of this uncertainty in the actual simulation systems, we allocate to this item a large error: according to estimations performed for LUTE, we suggest an allocation of $S = 0.98$.

• For the secondary mirror:

3) surface optical errors: as for the primary mirror, i.e. $S = 0.99$.

• For the primary-secondary mirror system:

4) initial alignment of the two mirrors with respect to the optical axis of the system: lateral displacement and tilt of the mirrors. According to the analysis performed for IOTA we require $S \simeq 0.995$.

5) defocus: the maximum axial relative displacement between primary and secondary mirror left after the alignment operation. In this case the analysis for IOTA gave $S = 0.93$, which is due to the difficulty in detecting a displacement less than 2 μm by analyzing the interferograms. However, this is a too large error for our visibility budget. To meet our visibility final goal we should require, at least, $S = 0.99$.

6) alignment sensors and actuators system: assuming that the alignment of the telescope is performed as for IOTA, i.e. setting the telescope in autocollimation mode, sensors are needed to analyze the interferograms and to send information to the mirror actuators which move the optics toward the optimal alignment.

Error to be allocated.

2.1.2 Environment related terms

• For the mirrors:

7) thermal figure changes in the mirror substrates: this takes into account radius of curvature changes, non-uniformity in the coefficient of thermal expansion, and thermal gradients across the mirror. All of them are more relevant for the primary than for the, smaller, secondary mirror. From evaluations made for LUTE, the Strehl ratio for this item would be 0.98. In order to meet our visibility budget however, we should require $S = 0.99$.

8) deterioration of the optical surfaces due to dust and micrometeroid impacts: the telescope is protected during half of the day by the dome-shade, but during the observing time it has no protection from dust and micrometeroids. Some sort of shielding can be considered, so that this term can be less relevant than the equivalent one for the siderostat.

Error to be allocated.

• For the alignment:

9) lunar day-night changes: this takes into account the non perfect compensation of the metering structure between primary and secondary mirror and non perfect athermalization between the telescope optics and the telescope supporting structure. Also, deterioration in the thermal protection due to dust has to be considered[46], in case this causes a variation in the day-time temperature that exceeds the temperature range for which the telescope athermalization system has been optimized. Due to the several unknowns involved in this item (thermal properties of materials, temperature range and thermal protection deterioration are all fields which deserve deep investigation), we allocate a large margin of visibility loss: $S = 0.98$.

10) thermal variation during an integration time: this effect may be due to the temperature variation in the first part of the night, during the thermal transient, or to non-perfect thermal insulation of electronics, mechanical drivers, or power system, in contact with some parts of the telescope structure. As far as the first effect is concerned, at the beginning of the night the integration time is not longer than 1 h during which the total temperature variation is of 0.72 K. This is not a large variation to be compensated for, but the results of the compensation depends on the temperature range for which the athermalization of each telescope component has been optimized (see *Appendix L*). The

[46] Coatings exposed to the space environment exhibit radiation-induced darkening which increases with time. For surfaces which were exposed to the lunar environment during the *Surveyor* and *Apollo* missions, the overall discoloration of inorganic coatings has been attributed to the result of several effects, from lunar dust, to solar radiation, and organic outgassing from spacecraft parts. Dust and irradiation played the key role in altering the appearance and usefulness of the coatings. In particular a thin coating of dust renders low absorptivity and low emissivity surfaces ineffective. Even 10^{-5} to 10^{-4} grams of lunar fines per square centimeter can increase absorbed solar thermal energy for a reflective-thermal control surface by a factor as large as 2 or 3 *(Carrol and Blair, 1972)*.

second effect sets requirements on the design of any "thermal source" which has to be located in the telescope area. We require that this item is almost negligible in terms of visibility loss: $S \geq 0.99$.

11) misalignment due to ground vibrations: this is likely a small effect, but for it to be correctly evaluated, studies on the real entity of the moon seismic activity, and an analysis of the degree of isolation that the telescope support structure provides, are necessary.

Error to be allocated.

2.2 Visibility loss allocation for the siderostat

2.2.1 Non-environmental related terms

1) mirror optical errors: as for the primary mirror, $S = 0.99$.

2) ground to Moon changes: as for the primary mirror, but because this is a flat mirror, it should be less affected, as far as the changes are uniform on all its parts. By scaling from the allocation for the primary mirror, we require $S \simeq 0.986$.

2.2.2 Environmental related terms

3) deterioration of the optical surface due to dust and micrometeroids impacts.
Although it may be possible to fabricate a very smooth optical surface, lunar dust contamination could severely degrade system performance. Protection from dust, perhaps including sensors, a protective cover, and a means for "in situ" cleaning may be required.

Error to be allocated.

4) thermal figure changes in the mirror substrate: it takes into account non uniformity in the coefficient of thermal expansion and radial thermal gradients across the mirror. With reference to the more stringent requirement on the primary mirror, for the siderostat we require $S \sim 0.998$.

5) change in the gravitational load during an integration time: although the magnitude of the lunar gravity vector is small, we have to consider that during an integration time the siderostat will change angular orientation with respect to the vertical. For the telescope configuration considered in the section *Interaction among subsystems: an example for the lunar interferometer* the change in orientation may by up to 35°, for the longest integration period and consequently the gravitational load applied to the siderostat can have a variation as large as $0.6\rho_A$ (which is of the order of 16 kg for a SiC substrate), over the whole integration time. If the gravity load affects the surface of the mirror, the wavefront error that it produces may change during the integration time and degrade the visibility.

Error to be allocated.

2.3 Visibility loss allocation for the relay system

2.3.1 Non-environmental related terms

1) surface optical errors: the surface specifications for these mirrors should be as those for the main optics of the system. However they will depend also on the number of reflections that are necessary to lead the beam along the relay system. As shown in section *2.5* in *Instrument visibility for IOTA* the instrument visibility is quite sensitive to any additional reflection (for a $\lambda/80$ rms surface error, each additional reflection decreases the visibility by 2%). In tab. VB.3 I present the the wavefront error tolerances for a single mirror as a function of the Strehl ratio allocated to all the flats of the relay system, assuming that the relay path consists of 10 reflections, as it is the case for IOTA.

tab. VB.3 Maximum rms wavefront error due each single mirror, σ_{mir}, in the relay path as a function of the Strehl ratio allocated to all the flats of the relay system, S_{rel}

S_{rel}	σ_{mir}
0.99	$\lambda/200$
0.98	$\lambda/140$
0.97	$\lambda/115$
0.96	$\lambda/100$

Even if it may be quite demanding for the mirror manufacturing, we have to require $S = 0.99$ in order to meet our visibility goal. Note that the structure function method of allocating errors should be applied to these elements as well, which will result in reduced apparent (but more complexly specified) tolerances.

2.3.2 Environmental related terms

2) deterioration of the optical surfaces due to dust and micrometeroids impacts: these mirrors may be exposed to the harsh lunar environment during the whole day (the dome-shade is limited to cover the telescope area).

Error to be allocated.

3) lunar day-night thermal changes: if the mirrors are directly exposed to the thermal environment, we should evaluate the effects of non-uniformity of the thermal properties of the mirrors substrate over the day-night thermal variation range. Assuming that we can allocate to each mirror the same error allocated to the siderostat, 10 reflections would imply $S = 0.98$ (but this is quite optimistic).

2.4 Visibility loss allocation for the beamsplitter

2.4.1 Non-environmental related terms

1) reflective surface optical errors: as for one flat of the relay system, i.e. $S = 0.999$.

2) polarization effects: depends on the optimization of the multilayer coating utilized for

the beamsplitter. Because of the good results already reached in this field (see *Phillips and Hickey, 1995*), we should be able to allocate to this effect a negligible error: $S \sim 0.999$ or better.

2.4.2 Environmental related terms

3) lunar day-night thermal figures changes: this takes into account non-uniformity in the coefficient of thermal expansion and thermal gradient effects affecting the beamsplitter substrate. We require an error of the same order of that due to the substrate of the relay mirrors: $S = 0.998$.

4) thermal changes during an integration time: in this case the OPL of the beam is affected by a variation which is proportional to the temperature change during the integration, the substrate thickness, and a material related coefficient, $D(\lambda)$, which is a function of the wavelength[47]. As far as δOPL is the same across the whole beam and for any λ, there is no effect on the visibility. However, because of the dependence of D on λ, some degradation in the visibility might occur.
The thermal variation has to be controlled accordingly to the variation of D over the wavelength range of observation. In general, for substrates used at room temperature, $\Delta D \sim 10^{-6}\,K^{-1}$ so that the maximum ΔT should be kept of the order of 10^6 times the ratio between the maximum wavefront error and the substrate thickness (of the order of 1-2 cm). Assuming for ΔD the same behaviour also in the Moon environment, and visible wavelengths, we calculate: $\Delta T \simeq 0.88\ K$ if $S = 0.99$, and $\Delta T \simeq 0.3\ K$ if $S = 0.999$. As in the case of the telescope alignment, the maximum thermal variation we should worry about is of $0.72\ K$, occurring in an interval of $1\ h$ at the beginning of the thermal transient. Hence we require for the thermal control system of the beam combining area to be able to keep $\Delta T \simeq 0.3\ K$, so that we can allocate $S \sim 0.999$, or better, to this item.

2.5 Visibility loss allocation for the windows

2.5.1 Non-environmental related terms

1) manufacturing errors: a difference Δd in the thickness of the parallel plates encountered by the beam, introduces a difference in the OPL which is given by $\delta OPL = (n(\lambda)-1)\cdot\Delta d$. When considering the whole range of wavelength observed, each monochromatic beam travels a different OPL and creates an interference fringe slightly shifted with respect to those at other wavelengths. The effect on the resulting interference fringe may consists in some loss of visibility. The wavefront error allocation to this effect depends on the number of windows the beam encounters (likely only one if the beamsplitter and the detector are located in the same environment-protected enclosure). For one window we require $S = 0.999$.

[47] $D = \alpha \cdot (n(\lambda) - 1) + \beta(\lambda)$ where α is the coefficient of thermal expansion, $n(\lambda)$ is the index of refraction, and $\beta(\lambda)$ is the coefficient of temperature variation of the index of refraction. (See section *1.2.1.1* in *Thermal effects* and *3.1* in *Instrument visibility for IOTA*).

2.5.2 Environmental related terms

2) thermal figure changes in the window material: this takes into account non-uniformity in the coefficient of thermal expansion and radial thermal gradient across the window. We require an error of the same order as that due to the beamsplitter substrate: $S = 0.998$.

3) thermal changes during an integration time: as for the beamsplitter substrate, i.e. $S \sim 0.999$ or better.

3. Visibility budget for the OPD control subsystem

3.1 Non-environmental related terms

1) fine delay line motion: smoothness in the motion of the reflector of the delay line is required to avoid the noise on the signal produced by unintended motion. The smoothness requirement is in turn related to the problem of operating the system in vacuum: continuously operated motors can overheat, moving parts sometimes require high vapor pressure lubricants, the delay line components must not release materials that could condense on the optics of the instrument. These are problems that are currently faced for the delay lines of the ground-based interferometer (for observation in the visible). However, new developments are necessary for the delay line of the lunar interferometer: we should allocate to this item a large margin of error.

2) mirror drives: vibrations generated by the mirror drives can be transmitted to the telescope structure. The pivot points of the mirror drives, of both the siderostat and the steering mirror that follows the beam-compressor, have to be kept at the intersection of the optical axis with the mirror surface, to avoid that rotation motion introduces variations in the optical path length of the chief ray of the system.

3) vibrations generated by other devices: for example vibrations produced by the pumps of fluid cooling systems.

For these three items all together we require an improvement with respect to the allocation suggested for the CHARA[48] array of a Strehl ratio of 0.98. We require $S = 0.99$.

3.2 Environmental related terms

1) microseismic activity: vibrational effects on the structure. Effects on the straightness and orientation of the delay lines components.

Error to be allocated.

2) dust: this can affect the fine delay motion mechanisms as well as the siderostat pivot.

Error to be allocated.

[48] The Center for High Resolution Astronomy (CHARA) of Georgia State University will build a facility for optical/infrared multi-telescope interferometry, called the CHARA array. This array will consist of initially five (with a goal of seven) telescopes distributed over an area approximately 350 m across.

4. Visibility budget for the beam-compression system (related to diffraction effects)

4.1 Non-environmental related terms

1) size of the propagated beam: the choice of the compression factor for the beam compressor involves a trade-off between diffraction effects and size of the optics which follow the beam-compressor. To efficiently propagate light to the beam combiner area, the compressed beam must have minimum diffraction, which translates in having the Rayleigh length much larger than the propagation distance[49].

2) wavelength range: the beam size for which diffraction may have a limited effect on the visibility loss depends on the wavelength of observation: it increases for longer wavelength. This means that a system which is optimized for the visible range may have a large loss in visibility at far-infrared wavelength. As an example I present in tab. VB.4 the Strehl ratio obtained for several beam size values as a function of the wavelength (propagation length is 1 km, telescope diameter is 1 m).

tab. VB.4 Strehl ratio S for several beam sizes, a, and corresponding beam compressor factors, m, as a function of wavelength

m	$a(cm)$	$\lambda\ (\mu m)$	0.12	0.55	2.2	10
20	5		0.95	0.91	< 0.92	$\ll 0.92$
10	10		0.98	0.94	0.91	< 0.92
8	12.5		1	0.96	0.93	0.92
5	20		1	0.98	0.95	0.93
2	50		1	1	0.98	0.96

The Strehl ratio values are derived from table 1. in *Tango and Twiss (1974)* where the maximum ratio between Rayleigh length and propagation length considered is equal to 1, to which corresponds a Strehl ratio of 0.92.

From these results we see that, in order to reach good visibilities over the largest wavelength range possible (to exploit its maximum potential the instrument should cover from

[49]The Rayleigh length is defined by

$$D_R = \frac{\pi\, a^2}{4\,\lambda}$$

where a is the size of the beam (*Tango and Twiss, 1974*). See also section *4.* in *Instrument visibility for IOTA*. Even with significative diffraction, while there may be loss in the total signal, there will be no systematic visibility loss if both interfering beams diffract to the same extent, that is, they travel the same distance path. However, in a Michelson interferometer with a long baseline fixed to the ground, the propagation path is necessarily different for the two beams, and this implies a systematic loss in visibility due to diffraction effects.

UV to far infrared wavelengths), the trade-off to be considered is that of minimum compression factor versus dimension of the optics which follow the beam-compressor. A good compromise could be obtained by selecting either $m = 8$ or $m = 5$. However, in both cases a second beam-compressor may have to placed before the detection area to match the dimension of the beam with that of the detector, or before the fine delay line to reduce the dimension of the following optics. In order to meet our final visibility goal we should require $S = 0.98$, which unfortunately implies a beam size of 20 cm. Note also that we can tolerate lower visibility in the infrared band as far as we can compensate with better visibility performances in other parts: for example a high Strehl ratio for the alignment in the visible, implies almost loss free conditions in the infrared.

4.2 Environmental related terms

3) lunar atmosphere and dust: effects due to the presence of rarefied atmosphere and scattered dust on the beam path may introduced diffraction, which should be evaluated.

Error to be allocated.

5. Visibility budget for the tilt correction servo system[50]

Due to the fact that a large fraction of the error budget is likely to be allocated to the optical system (and in part to the OPD control system), the allowable Strehl ratio for the servo control system has to be set very high, as for an almost a perfect system.

5.1 Non-environmental related terms

1) maximum frequency bandwidth of the wobbler mirror: the wavefront rms error associated with this term is proportional to the quantity $(f_o/f_m)^{\frac{8}{3}}$, where f_m is the maximum frequency bandwidth of the mirror, and f_o is the frequency of variation of the wavefront tilt[51]. For a lunar instrument f_o is mainly related to tilt variations which arise in the instrument itself, and should be kept much smaller than the equivalent f_o imposed by the atmosphere for an earth-based instruments. By keeping it (by means of vibration isolation and suppression systems[52] also much smaller than f_m, eventually the wavefront error related to this term would be almost negligible with respect to that due to the quadra detector performance, hence we allocate $S \sim 0.999$ or better.

2) performance of the detector used in the servo system: in general the tilt-correction

[50]The basic function of the tilt correction servo is to keep the interfering beams of the arms of the array parallel. If the difference in beam tilt is too large, losses in signal-to-noise will occur in the visibility measurement of the interferometer.

[51]For earth-based instruments this is directly related to the turbulence of the atmosphere.

[52]When structures vibrate, they do so in preferred shapes called *mode shapes* that are a consequence of the mass and stiffness distribution in the structure. Once the geometry of the structure is selected, there is relatively little that can be done passively to alter these shapes. Precision optical components that are located at positions of large motion in the mode shape will be difficult to stabilize passively. Active structures technology is an attempt to overcome these inherent limitations by using feedback between actuators and sensors embedded within the structure. They sense deviations from desired response and take corrective action, thereby achieving performance better than passive systems *(Fanson et al., 1990)*.

information necessary to guide the wobbler mirror is provided by sending a small part of the beam to a quadrant detector (see section *5.2* in *Instrument visibility for IOTA*). The visibility loss, due to error in the angular position measurement by means of a quadrant detector, is inversely proportional to the square of the SNR of the detector[53]. Eventually, the Strehl ratio allocated to this term will set the requirement for the performances of the detector. According to the values presented in tab. VB.5, the detector should be optimized to reach a $SNR = 15$, so that we can allocate a Strehl ratio of 0.998.

tab. VB.5 Strehl ratio for the quadrant detector, S_{qua}, as a function of SNR

SNR	S_{qua}
5	0.980
10	0.995
15	0.998

5.2 Environmental related terms

3) deterioration of the optical surface of the wobbler mirror due to dust and micrometeroids.

Error to be allocated.

6. Conclusion

The visibility loss allocation performed in the previous sections is summarized in tab. VB.6, with respect to the four main subsystems of the instrument. The terms TBA (to be allocated) take into account, for each subsystem, the amount of Strehl ratio that was not allocate due to lack of sufficient information about the error source and its effects.

tab. VB.6 Preliminary visibility loss budget for the lunar interferometer allocated among the four main subsystem of the instrument. The term TBA is referred to the error induced by identified error sources that was not possible to quantify

S_{tot}	S_{OPT}	S_{OPD}	S_{DIF}	S_{SERVO}
$0.82 \cdot TBA$	$0.84 \cdot TBA$	$0.99 \cdot TBA$	$0.98 \cdot TBA$	$0.998 \cdot TBA$

If now we consider that the visibility goal we have set at the beginning is about 0.8, it is evident that the margin left for all the TBA items implies a high Strehl ratio, 0.97. In

[53]

$$S_{qua} = 1 - 1.8 \left(\frac{0.438}{SNR} \right)^2 .$$

See also section *5.2* in *Instrument visibility for IOTA*.

fact, we note that all, but for the error due to the performance of the alignment control system, of the TBA terms are environmental related and most of them are due to the effect of the lunar dust on optical surfaces and on some mechanism. This is not a small issue and certainly a larger loss should be associated with its effects.

If we relax our visibility goal to 0.75, the Strehl ratio for the TBA items becomes 0.91, which is still quite demanding. However, it is indicative of the level to which we must be able to control the environmental impacts in order to design a lunar-based competitive instrument with respect to the best performing ground-based interferometer we expect to realize in the next future.

Conclusion

There will be a future for stellar interferometry from the Moon only if we are able to demonstrate that it offers a unique scientific potential or at least that to obtain the same performances by ground-based or space-based interferometry would be very difficult and expensive.

As far as astrometric measurements are concerned, it is likely that the lunar interferometer will have to compete with an orbiting instrument, which will be capable of measurements of global astrometry which are not possible for an instrument on the lunar surface.

The great potential of a lunar interferometer is mainly in the fields of narrow angle astrometry, of visibility measurements with a two-aperture instrument, and of imaging. Hence these are the fields in which we should investigate for the performances of a lunar interferometer to be better than those of a ground-based interferometer (a part from the possibility of observing in the UV which is a privilege of extraterrestrial sites) and of a interferometric system realized by means of several free-flyers.

In the research work I have presented I analyzed some of the aspects which allow a realistic evaluation of the actual potential of an lunar interferometer. In particular, I showed the need of having a clear view of the whole design architecture, which is a requirement for the management of any space mission. According to this aim I presented an example of the interaction between some of the interferometer subsystems and I defined a set of essential parameters to be considered for the realization of a lunar interferometer. The result of this work is summarized in tab. C.1 where the data from tab. SI.2 and from section *3.3* of the chapter *Interaction among subsystem: an example for the lunar interferometer* are merged.

However, there are other subsystems and interactions to be considered. For example, I did not deal with the power subsystems and the electronic components of the system, whose thermal interactions with different parts of the instrument are certainly to be analyzed.

By means of the visibility strawman budget of the instrument I offered a final comprehensive presentation of the study, and of its practical utilization, developed on the lunar interferometer.

The visibility budget is one of the elements by which the potential of the instrument is defined and hence it represents a means to evaluate the worthwhileness of the realization of a lunar interferometer.

In order to perform a realistic budget it is necessary to know the architecture of the whole design, to single out the main subsystems, and to determine their contribution to the loss in the instrument performances. This also implies to know the architecture of the subsystems, the interactions among subsystems and, expecially for a lunar interferometer, their interactions with the environment.

tab. C.1 Parameters of the strawman design for the lunar interferometer (ρ_A is the aerial mass for the mirrors)

GENERAL CHARACTERISTICS	
number of elements	2 (initial)
baseline	
maximum length	1 km
latitude	equatorial
orientation	E-W
delay line	
coarse	5 stations separated by 100 m
fine	maximum extension: 50 m
thermal protection	dome shade
thermal variation	100 K (maximum)
integration time	130 h (maximum)
TELESCOPE	siderostat plus beam compressor
latitude	$\varphi = 0°$
telescope orientation	E-W
inclination with respect to the horizontal	$\gamma = 35°$
maximum incidence angle	$\alpha = 45°$
minimum incidence angle	$\beta = 10°$
primary mirror diameter	$d = 1\,m$
maximum linear dimension of the siderostat	$D = 1.41\,m$
distance between siderostat and beam-compressor	$l = 2.75\,m$
mass of the primary mirror	$M_P = 0.785\,\rho_A$
mass of the siderostat	$M_{sid} = 1.2\,\rho_A$

To develop this strawman budget the knowledge and experience I acquired working within the IOTA projected turned to be very important. Actually, the visibility budget for the lunar interferometer is modelled on the scheme adopted for the estimation of the instrumental visibility of IOTA.

In conclusion of this work it seems appropriate to me to show together the final results about the performances offered by the two instruments, which are summarized in tab. C.2 (for the instrumental visibility of IOTA) and in tab. C.3 (for the visibility budget of the lunar interferometer). The comparison between these results is indicative of the

evaluation that will have to be performed when comparing the performances of a ground-based interferometer of the next generation with those expected for a lunar interferometer whose feasibility requirements can realistically be met.

tab. C.2 Instrumental visibility of IOTA: allocation of the Strehl ratio among the four main subsystems of the instrument

S_{array}	S_{opt}	S_{OPD}	S_{diff}	S_{servo}
0.528	0.614	0.909	0.960	0.986

tab. C.3 Visibility budget for the lunar interferometer: allocation of the Strehl ratio among the four main subsystems of the instrument (TBA = *to be allocated*)

S_{tot}	S_{OPT}	S_{OPD}	S_{DIF}	S_{SERVO}
$0.82 \cdot TBA$	$0.84 \cdot TBA$	$0.99 \cdot TBA$	$0.98 \cdot TBA$	$0.998 \cdot TBA$

Even if IOTA is not the ground-based instrument with the highest performances, it is evident that the constraints on some of the subsystem components of the lunar interferometer will be very demanding.

This is probably not the case for the wavefront tilt servo control system, because the main disturbance is due to the vibrations within the instrument and not to the atmosphere fluctuations. As far as the vibrations induced on the structure of the instrument are properly controlled, the lunar solution is more favorable than the ground-based one.

The lunar case has to be preferred also when considering the diffraction effects. If we assume that the transmitted beams travel the same path length, the lunar beam is much less distorted because of the absence of the initial distortion introduced by the atmosphere turbulence. However, this aspect will be fundamental in establishing the maximum baseline length (or in requiring a greater degree of complexity and a larger number of transmitting optical components, if we want to exploit a baseline longer than that achievable with only one beam-compressor).

As far as the system for the control of the optical path difference is concerned, we notice that many of the problems which need to be faced to obtain a high degree of smoothness in the operation of the delay lines are, or will be, to be solved also for the ground-based systems. For example, the problems related to the operation of the delay lines in vacuum are to be considered also on the ground, when the system is designed for observations in the visible range. Hence, we can expect that the technological development necessary to

reach the ambitious Strehl ratio value of 0.99 will be a follow up of that developed for ground-based instruments.

Finally, we consider the term related to the optical components, which probably imposes the most demanding constraints because of the large number of reflections necessary for the relay of the beam. In order to satisfy the total budget allocation of 0.84, it is required a surface specification equivalent to a Strehl ratio of 0.99 over the whole surface for each mirror. As a comparison we refer to the CHARA Array for which it is expected that the main optics have a surface quality such to guarantee a Strehl ratio equal to 0.95 on any sub-aperture of 30 cm in diameter. We notice that the improvement required in the mirror surface manufacture is not directly related to the development obtainable for ground-based instruments: on the Earth the surface accuracy limit is determined by the atmosphere disturbance, even if it could be quite reduced if adaptive optics systems are going to be largely used. However, we can expect that the realization of actual and future instruments for space astronomy will demand higher quality in the mirror surface so that the development required will benefit also the needs of lunar interferometry.

Eventually, we can assume that, as far as the instrumental requirements are concerned, the lunar interferometer will be able to meet the constraints set by the visibility budget.

At this point the major unknown is related to the value that has to be allocated to the TBA terms. They represents the contribution due to the environmental impact on the components of the different subsystems, in particular the effects due to dust and to seismic activity. The equivalent unknown term for a ground-based instrument is related to the disturbance caused by the atmosphere. However, in this case we are able to give a good estimation of its contribution and we assume that for normal seeing conditions the corresponding Strehl ratio is of the order of 0.875 (actually this value changes from night to night and from hour to hour during the same night). It is necessary to be able to make a similar estimation also for the lunar interferometer. However, the actual knowledge, which for the effects affecting the operation of a lunar interferometer are mostly qualitative, on the lunar environment is not enough. Before starting the detailed design of a lunar interferometer it is necessary to acquire quantitative information about the thermal characteristics, the seismic activity, the mechanism of the dust distribution, and on the atmosphere of the Moon.

The data acquisition could be realized by means of a, or several, mission(s) designed for the operation of a lunar probe on the Moon surface and a detailed *in situ* monitoring of the lunar environment characteristics.

Among the recently proposed missions which are expected to make environment measurements *in situ* there are *LEDA* and *LUNAR-A*.

LEDA (*Lunar European Demonstration Approach*) is one of the missions candidate for the phase 1 of the *ESA Moon Programme* whose final goal is the *establishment of the first mantended outpost to serve science and the utilization of lunar resources (LEDA, 1995).* The phase 1 of the program will be devoted to the development of a range of technologies

which will be of direct benefit to the later phases of the *Moon Programme*. The main objective of the *LEDA* mission is the demonstration of the soft-land of a spacecraft on the lunar surface and its operation *in situ* (among the capabilities to be demonstrated: locomotion on the lunar surface, power, communication and teleoperation systems). The mission shall also provide a scientific return which consists in different sets of environment measurements data. As far as the fields of main interest for the development of a lunar interferometer are concerned, *LEDA* will be provided by instruments for the measurements of the Moon microseismic activity, of the thickness and cohesion of the regolith at various locations, of the temporal variation of the surface temperature and of the thermal properties of the lunar soil, of the dust behaviour related to natural or induced causes, and of the micrometeroids flux. In case the ESA program eventually starts and *LEDA* is the chosen mission for the phase 1, it might be launched by the year 2002 (*LEDA, 1995*).

LUNAR-A is a mission designed by the Japanese *Institute of Space and Astronautical Science (ISAS)* and its launch is planned for August 1997. The scientific objective of the mission is to explore the lunar interior using seismometry and heat-flow measurements. Three missile-shaped penetrators (82.6 cm long and 12 cm in caliber) will be released by a spacecraft and will impact the lunar surface the lunar surface at a speed of the order of 300 $km\,s^{-1}$. Depending on the hardness of the regolith, each penetrator will penetrate to a depth varying from 1 to 3 m.

Each penetrator carries three seismometers, designed to be approximately 10 times as sensitive as the Apollo instruments for both horizontal and vertical displacement measurements. Besides, since the penetrators will be deployed at three different sites (the first next to the Apollo 12 and 14 sites, the second in an antipodal direction on the farside of the Moon, and the third in a site of high latitude on the near side) they will constitute a seismic network of a much larger span than that of the Apollo seismic network.

The heat-flow measurements will be obtained by means of the combination of measurements of thermal conductivity and vertical temperature gradient. In this case the reliability of the measurements depends on the analysis of the temperature field around the penetrator, which is disturbed by the penetrator itself. Preliminary analyses performed in laboratory indicate that the heat-flow can be estimated within an error of 10% (*Mizutani et. al, 1992*).

As far as the next future is concerned, I wish to witness and, maybe, contribute to the development of the feasibility study for a lunar interferometer, while waiting for having soon the results from the measurements performed by the *LUNAR-A* and, likely, the *LEDA* missions.

Bibliography

AA.VV. *A Lunar Optical-Ultraviolet-Infrared Synthesis Array (LOUISA)*, Workshop Proc., NASA Conf. Publ. 3066, 1992.

AA.VV, *A proposed medium-term strategy for optical interferometry in space*, ESA SP-1135, 1990.

AA.VV. *Lunar sourcebook*, G.H.Heiken, D.T. Vaniman, and H.H. French editors, Cambridge University Press, 1991.

AA.VV. *Mission to the Moon*, ESA SP-1150, 1992.

AA.VV. *International Lunar Farside Observatory and Science Station: ILFOSS*, Project Report 1993, ISU, Huntsville, Al, 1993.

AA.VV. *International Lunar Workshop*, Beatenberg (Interlaken) Switzerland, June 1994, ESA SP-1170, 1994.

AA.VV. *Study of a Moon-based/Free-flyer Optical Interferometer*, ESA *Invitation to Tender* AO/1-2918/94/NL/CC, September 1994.

AA.VV. *Horizon 2000 plus*, ESA SP-1180, 1994.

AA.VV. *Lunar Ultraviolet Telescope Experiment (LUTE)*, Phase A Final Report, NASA Technical Memorandum 4594, 1994.

AA.VV. *A lunar-based interferometer design for early detection of extra-solar planets*, Science Applications International Corporation, Report No. SAIC - 95/1037, 1995.

AA.VV. *LEDA (Lunar European Demonstration Approach) - Assessment study*, LEDA-RP-95-02, 1995.

R. Alvarez, *Photoconductive effects on lunar and terrestrial fines*, Proc. Lunar Sci. Conf. 8th, 1977.

W.G. Bagnuolo, *Simulations for the CHARA/GSU interferometer and binary star speckle photometry*, ESO Conference and Workshop Proceedings No.29, 1988.

T. Baumeister and L. Marks, *Standard Handbook for Mechanical Engineers, 7th edition*, McGraw - Hill book company, 1967.

J.M. Beckers, *Adaptive optics for astronomy: Principles, Performance, and Applications*, Ann. Rev. Astron. Astrophys., **31**, 1993.

R.K. Bhatia, *Telescope alignment: is the zero-coma condition sufficient?* Proc. SPIE **2479**, *Telescope control systems*, 1995.

M. Born and E. Wolf, *Principle of Optics*, *6th edition*, 1980.

J.O. Burns and W.W. Mendell, Eds., *Future astronomical observatories on the Moon*, NASA Conf. Pub. 2489, 1988.

B.F. Burke, *Astronomical interferometry on the Moon*, Lunar bases and space activities of the 21st century, Lunar and Planetary Institute, Houston, 1985.

D.F. Busher, *Getting the most out of the C.O.A.S.T.*, Cambridge University, Cambridge, 1988.

M.M. Colavita, *Prospect for imaging and astrometry on the ground and implication for space inter-*

ferometry, Proc. ESA Colloquium on *Targets for space-based interferometry*, 1992, ESA SP-354.

M.M. Colavita, M. Shao, and D.H. Staelin, *Atmospheric phase measurements with the Mark II stellar interferometer*, Applied Optics, Vol. 26, No. 19, 1987.

M.M. Colavita, M. Shao, B.E. Hines, B.M. Levine, and R. Gershman, *Optical system design for a lunar optical interferometer*, Proc. SPIE **1494**, *Space astronomical telescopes and instruments*, 1991.

M.M. Colavita and D.M. Wolff, *Orbiting Stellar Interferometer*, Proc. SPIE **2477**, *Space Born Interferometry II*, 1995.

N.P. Carleton, *Consideration of thermal and mechanical stability for IOTA*, Proc. SPIE **1237**, *Amplitude and Spatial Interferometry*, 1990.

W.F. Carrol, and P.M. Blair, *Spacecraft changes, Part A, lunar dust and radiation darkening of Surveyor 3 surfaces*, "Analisys of surveyor 3 material", NASA, 1972.

C.J. Cremers, R.C. Birkebak, and J.E. White, *Thermal characteristic of the lunar surface layer*, Int. J. of Heat and Mass Transfer, 1972.

J. Fanson, E. Anderson, and D. Rapp, *Active structures for use in precision control of large optical structures*, Opt. Eng., **29**, 1990.

D.L. Fried, *Statistics of a geometric representation of wavefront distortion*, J. Opt. Soc. Am., **55**, 1965.

C.L. Johnson and K.L. Dietz, *Effects of the lunar environment on optical telescopes and instruments*, Proc. SPIE **1494**, *Space astronomical telescopes and instruments*, 1991.

S.W. Johnson, G.J. Taylor, and J.P. Wetzel, *Environmental effects on lunar astronomical observatories*, NASA Conf. Publication 3166, Vol.1, 1988.

W.P. Jones, J.R. Watkins, and T.A. Calvert, *Temperatures and thermophysical properties of the lunar outermost layer*, The Moon 13, D. Reidel Publishing Company, 1975.

J.M. Hill, *Optical design, error budget and specifications for the Columbus Project Telescope*, Proc. SPIE **1236**, *Advanced technology optical telescopes IV*, 1990.

R.R.Jr. Hodges, J.H. Hoffman, and F.S. Johnson, *The lunar atmosphere*, Icarus, **21**, 1974.

D.J. Hutter, *Siderostat/site metrology system for the USNO Astrometric Interferometer*, Workshop Proceedings: Technologies for Optical Interferometry in Space, JPL D-8541, Vol. 1, 1991.

A. Labeyrie and D. Mourand, *Proposed optical interferometer with automated deployment*, Astrophysics from the Moon, AIP Conf. Proc. 207, 1990.

A. Labeyrie, G. Lemaitre, C. Thom, and F. Vakili, *Steps towards an optical very large array*, ESO Conference and Workshop Proceedings No.29, 1988.

M.G. Langseth and S.J. Keihm, *In-situ measurements of lunar heat flow*, Soviet-American Conference on Geochemistry of the Moon and Planets, NASA SP-370, 1977.

L. Lindegren, *Atmospheric limitations of narrow-field optical astrometry*, Astron. Astroph., **89**, 1980.

V.N. Mahajan, *Strehl ratio for primary aberrations in term of their aberration variance*, J. Opt. Soc. Am., **73**, 1983.

J.A.M. McDonnel, *Lunar surface grain motion: electrostatic charging, supercharging (electret effect) and mechanical bonding*, COSPAR Space Research XIX, 1979.

W.W. Mendell, Ed., *Lunar bases and space activities of the 21st century*, Lunar and Planetary Institute, Houston, 1985.

R. Millan-Gabet, F.P Schloerb, W.A. Traub, H.M. Dick, *Signal-to-noise ratio of interference fringe patterns*, in preparation, 1995.

H. Mizutani, M. Kohno, A. Fujimura, I. Yamada, S. Tanaka, and M. Hayakawa, *Lunar penetrator mission*, AIAA 92-0357, 1992.

M.J. Mumma and H.J. Smith, Eds., *Astrophysics from the Moon*, AIP Conf. Proceedings 207, 1990.

M.A. Murison, *Ray Trace Analysis of Selected POINTS Optical Subsystems*, TM 93-07, 1993.

S. Musikant, *Optical materials*, Marcel Dekker, Inc., New York-Basel, 1985.

J. Oberst and Y. Nakamura, *A search for clustering among the meteoroid impacts detected by the Apollo Lunar Seismic Network*, Icarus **91**, 1991.

J.F. O'Brien, G.W. Neat, J.W. Melody, and R.J. Calvet, *Six-axis vibration insolation technology applied to spaceborne interferometers*, Proc. SPIE **2477**, *Space Born Interferometry II*, 1995.

Y. Nakamura, G.V. Latham, H.J. Dorman, and F.K. Duennerbier, *How we processed Apollo lunar seismic data*, Phys. Earth Planet. Interiors, **21**, 1980.

Y Nakamura, G.V. Latham, and H.J. Dorman, *Apollo Lunar Seismic Experiment - Final Summary*, Proc. Lunar Planet. Sci. Conf. 13th, in Geophys. Res, **87**, 1982.

R.J. Noll, *Zernike polynomials and atmospheric turbulence*, J. Opt. Soc. Am., **66**, 1976.

C. Papaliolios, P. Nisenson, and S. Ebstein, *Speckle imaging with the PAPA detector*, App. Opt., **24**, 1985.

M.R. Pearlman, D. Hogan, K. Goodwin, and D. Kurtenbach, *A meterological report for the Mt. Hopkins observatory: 1968-1971*, SAO Special Report 345, 1972.

D. Poulikakos, *Conduction heat transfer*, Prentice Hall, Englewood Cliffs, New Jersey 06732, 1994.

J.D. Phillips and C. Hickey, *Beamsplitters for astronomical optical inteferometry*, Proc. SPIE **2477**, *Space Born Interferometry II*, 1995.

J.D. Phillips, *Itek beamsplitter: calculated performance*, TM94-13, 1994.

R.J Rark and W.C. Young, *Formulas for stress and strain, 5th edition*, McGraw-Hill Book Company, 1975.

R.D. Reasenberg, R.W. Babcock, M.A. Murison, M.C. Noecker, J.D. Phillips, *POINTS: the instrument and its mission*, Proc. SPIE **2477**, *Space Born Interferometry II*, 1995.

C.F. Richter, *Elementary seismology*, Freeman, San Francisco, 1958.

S.T. Ridgway, *Infrared interferometry from the lunar surface*, in *Astrophysic from the Moon*, Annapolis, Maryland, 1990.

F. Rigaut and E. Gendron, *Laser guide star in adaptive optics: the tilt determination problem*, Astron. Astrophys., **261**, 1992.

R.A. Robie, B.S. Hemingway, and W.H. Wilson, *Specific heat of lunar surface material from 90 to 350 K*, Proceed. of the Apollo 11 Lunar Sci. Conf., Vol. 3, 1970.

G. Rousset, P.Y. Madec, and D. Rabaud, *Adaptive optics partial correction simulation for two telescopes interferometry*, ESO Conf. Proc. on *High resolution Imaging by Interferometry II*, 1991.

F.P Schloerb, *Image reconstruction techniques for the Infrared-Optical Telescope Array*, Proc. SPIE **1237**, *Amplitude and Intensity Spatial Interferometry*, 1990.

F.P Schloerb, *Imaging interferometry: lessons from the ground*, Proc. SPIE **1947**, *Spaceborne interferometry*, 1993.

D.J. Schroeder, *Astronomical Optics*, Academic Press, Inc., 1987.

M. Shao, M.M. Colavita, B.E. Hines, J.L. Hershey, J.A. Hughes, *Wide angle interferometry with the Mark III stellar interferometer*, Astron. J., **100**, 1990.

M. Shao and M.M. Colavita, *Potential for long-baseline infrared inteferometry for narrow angle astrometry*, Astron. Astrophys., **262**, 1992.

M. Shao, M.M. Colavita, B.E. Hines, D.H. Staelin, D.J. Hutter, K.J. Johnston, D. Mozurkewich, R.S. Simon, J.L. Hershey, J.A. Hughes, and G.H. Kaplan, *The Mark III stellar interferometer*, Astron. Astrophys., **193**, 1988.

W.H. Steel, *Interferometry*, Cambridge University Press, *2nd edition*, 1983.

S.P. Synott, R.E. Freeland, P.M. McElroy, P.V. Dardzinski, *Materials technology for space optical interferometry*, Workshop Proceedings: Technologies for Optical Interferometry in Space, JPL D-8541, Vol. 1, 1991.

W.J. Tango *Dispersion in stellar interferometry*, App. Opt., Vol. 29, No. 4, 1990.

W.J. Tango and R.Q. Twiss, *Diffraction effects in long path interferometers*, App. Opt., Vol. 13, No. 8, 1974.

W.J. Tango and R.Q. Twiss, *Michelson stellar interferometry*, Progress in Optics XVII, E. Wolf, ed., 1980.

T.A. ten Brummelaar, W.G. Bagnuolo, and S.T. Ridgway, *Strehl ratio and visibility in long-baseline stellar interferometry*, Opt. Letters, **20**, No. 6, 1995.

A.R. Thompson, J.M. Moran and G.W. Swenson, *Interferometry and Synthesis in Radio Astronomy*, New York: John Wiley & Sons, 1986.

W.A. Traub, *Polarization effects in stellar interferometers*, ESO Conference and Workshop Proceedings No.29, 1988.

W.A. Traub, *Optical path fluctuations*, Internal Memorandum, Feb. 1989.

W.A. Traub, M.G. Lacasse, and N.P. Carleton, *Spectral dispersion and fringe detection in IOTA*, Proc. SPIE **1237**, *Amplitude and Intensity Spatial Interferometry*, 1990.

W.A. Traub, *Pupil geometry and pupil re-imaging in telescope arrays*, Proc. SPIE **1237**, *Amplitude and Intensity Spatial Interferometry*, 1990.

W.A. Traub, *Constant-dispersion grism spectrometer for channeled spectra*, J. Opt. Soc. Am. A,

7, 1990.

W.A. Traub, N.P. Carleton, M.G. Lacasse, P. Nisenson, M.R. Pearlman, R.D. Reasenberg, X. XU, C.M. Coldwell, A. Pnasyuk, H.M. Dick, J.A. Benson, C. Papaliolios, R. Predmore, and F.P. Schloerb, *IOTA: Operational Status and Measurements*, Proc. Space Surveillance Workshop, K.L. Schwan, Ed., STK-235, Lincoln Laboratory, 1995.

W.A Traub, N.P. Carleton, and I.L. Porro, *A search technique for planets in nearby stars using the IOTA-PSI interferometer*, submitted to Journ. of Geophysical Research - Planets, 1995.

G.A. Tyler, and D.L. Fried, *Image position error associated with a quadrant detector*, J. Opt. Soc. Am., **72**, 1982.

R.R. Vondrak, *Lunar base activities and the lunar environment*, Lunar bases and space activities of the 21st century, Vol. II, Ed. W.W. Mendell, 1990.

S.T. Walker, *Thermal investigation of a large lunar telescope*, "Engineering, Construction, and Operations in Space III", Vol. II, 1992.

S.T. Walker, and R. Alexander, *The impact of the lunar thermal environment on the design of telescopes for lunar surface operation*, AIAA 93-4775,1993.

S.T. Walker, R.A. Alexander, and S.P. Tucker, *Thermal control on the lunar surface*, Journal of the British Interplanetary Society, Vol. 48, 1995.