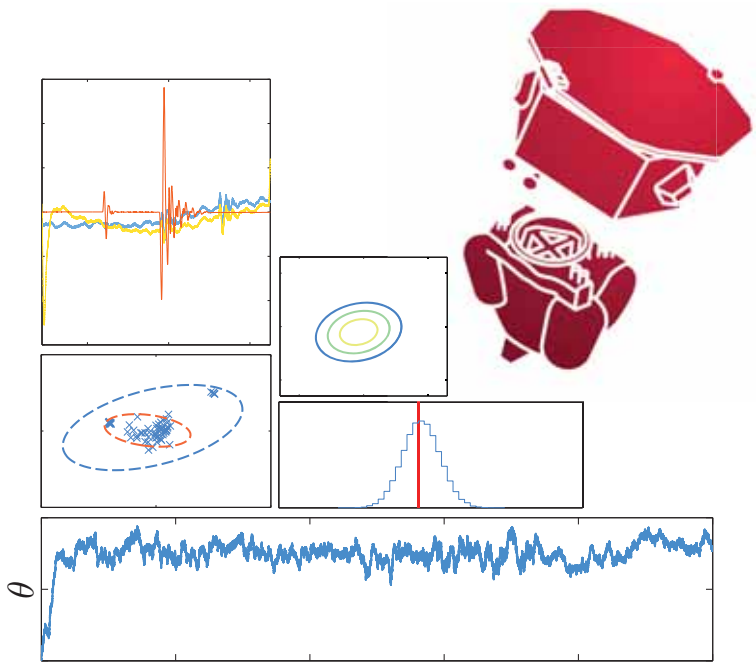
BAYESIAN DATA ANALYSIS FOR LISA PATHFINDER

NIKOLAOS KARNESIS



Techniques applied to system identification experiments

Departament de Física
Universitat Autònoma de Barcelona

Advisors: Dr. Miquel Nofrarias Serra & Dr. Carlos Fernández Sopuerta

October, 2014

IEEC CSIC UAB

Nikolaos Karnesis : *Bayesian Data Analysis for LISA Pathfinder , Techniques applied to system identification experiments* , October, 2014

Memòria presentada per
Nikolaos Karnesis per optar al grau de Doctor en Fìsica.

DIRECTORS:

Nofrarias S. M.

Sopuerta C. F.

CANDIDAT:

Karnesis N.

Barcelona, October 30, 2014

Universitat Autònoma de Barcelona.
Departament de Fìsica.
Programa de doctorat
Tutor: Diego Pavón Coloma.

Nikolaos Karnesis : *Bayesian Data Analysis for LISA Pathfinder , Techniques applied to system identification experiments* , October, 2014

Zero probability events happen every day.

— Josè Alberto Lobo Gutièrrez

To my family.

ABSTRACT

The eLISA concept design consists of a constellation of three space-crafts forming a triangle in the sky. While in a sun centered orbit, it will constantly monitor the distance oscillations between the test bodies enclosed in the different space-crafts. Its principal goal, is to detect oscillations that are caused by passing Gravitational-Waves. The technical complexity of this design was the reason for ESA and NASA to approve the LISA Pathfinder mission (LPF) which aims at testing all the key technologies for future Gravitational-Wave space observatories.

The LISA Technology Package (LTP) instrument onboard the LPF satellite, can be considered as one eLISA arm, squeezed from 10^6 km to $\simeq 30$ cm, and it aims to measure the differential acceleration between two test-bodies with unparalleled precision via a Mach-Zehnder interferometer. Among its objectives we have: The estimation of the acceleration noise models, the derivation of an accurate dynamical model of the system in all degrees-of-freedom, and the estimation of the systems' parameters. In this thesis, we focus on a Bayesian analysis framework to set-up analysis strategies to process the planned system identification experiments.

We first model the system using different approximations, and then we develop and apply Markov Chain Monte Carlo (MCMC) algorithms to simulated data-sets. We report the accuracy on the parameters over the planned system identification experiments, that can be divided in two categories; the x-axis system identification experiments, that are performed over the sensitive axis defined by the line joining the two test masses; and the so-called cross-talk experiments, where different degrees of freedom of the test bodies of the system are excited. The various cross-coupling physical effects that produce signal leakage on the sensitive differential interferometer channel, are then identified and estimated. In addition, the pipeline analysis designed for on-line data analysis during operations is presented.

Finally, we also investigate the possible model selection problems in LPF data analysis, and we apply the reversible jump MCMC algorithm to simulated data sets. Different applications to the x-axis and the cross-talk experiments are considered, where the efficiency of the developed tools is demonstrated. We also show the association of the model selection results to the design of the experiment itself. The above work is integrated to the LTP data analysis dedicated toolbox, the LTPDA.

In God we trust, all others bring data.

— W. Edwards Deming, *The Elements of Statistical Learning*, Springer New York Inc, 2001

ACKNOWLEDGEMENTS

Those four years have been quite a journey... Those four years were spent in studying and working, with the beautiful background of the city of Barcelona. In my eyes there is nothing more I could ask for my post-graduate studies. It all started with an awkward e-mail sent to Carlos, after Kostas showing me the webpage that listed Carlos' professional interests. It was Carlos, that first saw potential in me, even through this poorly written letter, and soon helped me apply for a grant provided by the Generalitat. Without him and his constant guidance, probably I wouldn't be writing these lines now. Then, I met the rest of the Group... A group of wonderful and different people, all of them under the guidance of Alberto. A great professor, scientist, and founder of the group, that always promoted the idea that the core of the scientific research resides in the passion and skills of people, and their ability to work as a team. I consider it an honour that I was able to work alongside him...

It was Victor, Ferran, Miquel, Nacho, Lluis, Ivan, Priscilla, Daniel and Marc, the rest of the wonderful GWART team, a handful of people working with passion and dedication, but also delightful to spend time together. Ferran has a special place in my heart, as he was my companion during the bumpy road trip of the doctoral research. I shall enjoy his defence soon! And I arrive to Miquel, the person who converted me to Bayesian statistics and guided and motivated me during my research. He has been always there for me and my –silly sometimes I admit– questions. But, besides the fruitful discussions around physics, I consider myself extremely lucky that we spent time together both inside and outside the institutes' hallways. I feel grateful for his patience with me, and I certainly hope to keep up with our collaboration in the future.

At this point I shall thank the Generalitat de Catalunya, that funded my research for three years, and also supported me financially to visit Hannover for two months. Also the superb secretariat team of IEEC, Eva and Anna who patiently endured my catalan.

A couple of months after my arrival in Barcelona, I was introduced to the LTPDA team. The core team dedicated to data analysis of the LISA Pathfinder mission; P. W. McNamara, M. Hueller, D. Vetrugno, S. Vitale, L. Ferraioli, E. Plagnol, H. Inchauspe, M. Armano, G. Congedo, B. Weber, J. I. Thorpe, P. Wass, H. Audley and N. Korsakova. It was one of my greatest experiences, to discuss and collaborate with them, and present common line work in conferences and workshops. I have to show my special gratitude to Dr. M. Hewitson, with whom we had an excellent collaboration, and even from start he showed his trust on me.

Life in Barcelona was substantially improved when we founded the International Bonobo Project. It was during a concert of Mishima[1] when Kai told me that he is playing the bass guitar, and was looking for people to play some music. We soon got lucky enough to meet Hugo, who completed the trio, and finally produce some of our own tracks. Thank you guys for your weekly dose of musical creativity.

Last but not least, my parents, who always supported my choices and dreams financially and most importantly, emotionally. Finally, I specially feel indebted to my beloved –since recently– wife, who beared with me, encouraged me, and psychologically supported me, even from distance. And of course, my dear friends who did the same, each one in his own way...

Τα τελευταία χρόνια ήταν για μένα ένα υπέροχο ταξίδι ... Ένα ταξίδι γεμάτο πρωτόγνωρες εμπειρίες με φόντο την πόλη της Βαρκελώνης... Ένα ταξίδι που ξεκίνησε πριν απο τέσσερα χρονια, όταν αναζητώντας μεταπτυχιακά προγράμματα, ο Κώστας μού πρότεινε την ιστοσελίδα του Carlos. Ήταν αυτος, που μετά από ένα κακώς γραμμένο e-mail, είδε πρώτος κάποιες δυνατότητες σε μένα και με βοήθησε να υποβάλλω αίτηση για υποτροφία στην Κυβέρνηση της Καταλονίας. Στην συνέχεια, γνώρισα και την υπόλοιπη ομάδα ... Μια ομάδα από υπέροχους και διαφορετικούς ανθρώπους, που συνεργάζονταν κάτω από την καθοδήγηση του Alberto. Ένας εξαιρετικός καθηγητής, επιστήμονας και ο ιδρυτής της ομαδας, ο οποίος πάντα θεωρούσε ότι ο πυρήνας της έρευνας βρίσκεται στο πάθος και την ικανότητα των ανθρώπων να δουλεύουν σαν μονάδα. Θεωρώ τον εαυτό μου τυχερό που συνεργάστηκα μαζί του, έστω και για μικρό χρονικό διάστημα...

Και φυσικά οι Victor, Ferran, Miquel, Nacho, Lluis, Ivan, Priscilla, Daniel και Marc, το υπέροχο αυτό γκρούπ, με τους οποίους μοιραστήκαμε πλούσιες εμπειρίες. Και φτάνω στον Miquel, ο οποίος μου έδειξε το δρόμο της Μπαεζιανής ανάλυσης και με καθοδήγησε δίνοντάς μου τα σωστά κίνητρα κατά τη διάρκεια της έρευνάς μου. Νιώθω πολύ τυχερός που συνεργαστήκα μαζί του και ελπίζω η συνεργασία μας να συνεχιστεί και στο μέλλον.

Επίσης, δεν θα μπορούσα να παραλείψω τη βασική ομάδα της ανάλυσης δεδομένων του LISA Pathfinder, τους P. W. McNamara, M. Hueller, D. Vetrugno, S. Vitale, L. Ferraioli, E. Plagnol, H. Inchauspe, M. Armano, G. Congedo, B. Weber, J. I. Thorpe, P. Wass, H. Audley και N. Korsakova και ιδιαίτερα τον M. Hewitson, με τον οποίο είχαμε μια εξαιρετική συνεργασία.

Η ζωή μου στη Βαρκελώνη θα είχε λιγότερο ενδιαφέρον, χωρίς το Κάι και τον Ούγκο, που εμπλούτιζαν την καθημερινότητά μου με μουσική. Τέλος, ευχαριστώ την οικογένειά μου, που υποστηρίζει πάντοτε τις επιλογές και τα όνειρά μου. Αισθάνομαι ιδιαίτερα ευγνώμων στην αγαπητή μου σύζυγό, Γλυκερία, η οποία ακόμα και από απόσταση με ενθάρρυνε και με στήριξε ψυχολογικά, αλλά και στους φίλους μου οι οποίοι έκαναν το ίδιο, ο καθένας με τον τρόπο του.

1 http://www.mishima.cat/

Nikolaos Karnesis : *Bayesian Data Analysis for LISA Pathfinder , Techniques applied to system identification experiments* , October, 2014

CONTENTS

LIST OF FIGURES

AO Analysis Object . . . 32
BF Bayes Factor . . . 124
CDF Cumulative Distribution Function . . . 63
CMA Caging Mechanism Assembly . . . 24
CMB Cosmic Microwave Background . . . 6
CPSD Cross Power Spectral Density . . . 51
CRB Cramèr-Rao Bound . . . 52
DA Data-Analysis . . . 72
DDS Data and Diagnostics Subsystem . . . 27
DFACS Drag-Free Attitude Control System . . . 23
DFT Discrete Fourier Transform . . . 93
DM2 Drift Mode 2 . . . 85
DMU Data Management Unit . . . 28
DOF Degrees-of-Freedom . . . 30
DRS Disturbance Reduction System . . . 13
DWS Differential Wavefront Sensing . . . 27
ECDF Empirical Cumulative Distribution Function . . . 63
EGO European Gravitational Observatory . . . 11
EH Electrode Housing . . . 24
ESA European Space Agency . . . 12
ESAC European Space Astronomy Centre . . . 85
eLISA evolved LISA . . . 12
EPB Experimental Performance Budget . . . 113
ET Einstein Telescope . . . 11
EMRI Extreme Mass Ratio Inspiral . . . 7
FEEP Field Emission Electric Propulsion . . . 13
FEE Frond End Electronics . . . 25
FIM Fisher Information Matrix . . . 52
GW Gravitational-Waves . . . 3
GR General Relativity . . . 3
GRS Gravitational Reference Sensor . . . 13
IPTA International Pulsar Timing Array . . . 6

IS Inertial Sensor Subsystem . . . 23
KS Kolmogorov-Smirnov . . . 63
LF Laplace-Fisher . . . 126
LIGO Laser Interferometer Gravitational-Wave Observatory . . . 11
LM Laplace-Metropolis . . . 126
LISA Laser Interferometer Space Antenna . . . 3
LPF LISA Pathfinder . . . 21
LSA Linearized-Signal Approximation . . . 53
LTP LISA Technology Package . . . 21
LTPDA LISA Technology Package Data Analysis
MC Monte Carlo . . . 56
MCMC Markov Chain Monte Carlo . . . 54
MCD Minimum Covariance Determinant . . . 126
MH Metropolis-Hastings . . . 54
MLE Maximum Likelihood Estimation . . . 46
MLDC Mock LISA Data Challenges . . . 14
MVE Minimum Volume Ellipsoid . . . 126
NASA National Aeronautics and Space Administration
OBC On-Board Computer . . . 29
OB Optical Bench . . . 25
OMS Optical Metrology System . . . 25
OPD Optical Pathlength Difference actuator . . . 27
OOL Out-of-Loop . . . 70
OSE Offline Simulation Environment
OSTT On-Station Thermal Test . . . 33
PDF Probability Density function . . . 49
PSD Power Spectral Density
PSRF Potential Scale Reduction Factor . . . 62
RJMCMC Reversible Jump Markov Chain Monte Carlo . . . 126
RLU Reference Laser Unit . . . 25
RMS Root Mean Square
SBIC Schwarz-Bayes Information Criterion . . . 126
SC Space-Craft . . . 22
SISO Single-Input-Single-Output . . . 30
SKA Square Kilometre Array . . . 6
SN Super Novae . . . 9

SNR Signal-to-Noise-Ratio

SMBH Super Massive Black Hole 6

STOC Science and Technology Operations Centre 32

SSM State-Space Model 33

TM Test-Mass 22

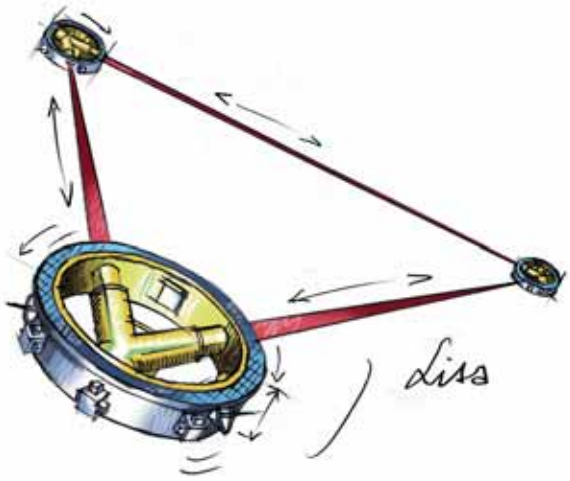
TOQM Thermo-Optical Qualification Model

ULU UV Light Unit 25

WP Work Package 37

Part I

INTRODUCTION



Credit: http://www.esa.int/Our_Activities/Space_Science/LISA

1 GRAVITATIONAL WAVES

1.1 INTRODUCTION

The search for Gravitational-Waves (GW) was greatly accelerated with the discovery discovery of a milisecond pulsar in a binary system [1], where the loss of its energy could be completely explained by gravitational-wave emission according to the theory of General Relativity (GR) [1]. From the discovery of the binary until the present day, there have been numerous attempts and many detectors have been developed devoted to directly measuring the effects of passing GWs. Along the way, the concept of a space-born GW detector was born, the Laser Interferometer Space Antenna (LISA). The goal of a space antenna like LISA is to directly detect gravitational-wave sources and study the violent events that causes them, like binary super-massive black holes.

One of the challenges of this mission is the data analysis part, where the knowledge of the system has to be combined with the signal in the data stream. The local dynamics of the system, together with the movement of the test bodies in various degrees of freedom, may cause undesired signal leakage to the sensitive channel, the one that measures the test masses' oscillations caused by passing GWs. Hence, although lot of work has been done in the field of parameter estimation of gravitational-wave sources, it has to be combined with the dynamics of the local system, that one of the space-craft and the test-bodies enclosed in it.

This is the role of the LISA Pathfinder mission. The LISA Pathfinder mission is a precursor mission to space detectors following the concept of LISA, and its principal goal is to test key technologies, to be directly inherited, and derive a noise and dynamical model of the local system. The LISA Pathfinder mission is going to perform a series of system identification experiments, where the parameters of the system have to be derived and then to estimate the differential acceleration noise between two free-floating test-masses[2]. For the estimation of the dynamical parameters of the system, the data analysis team has developed a set of tools and algorithms, that are going to be discussed extensively in this thesis. The modelling of the system is also discussed, as well as the analysis strategies to be followed during flight operations.

In this first introductory chapter, we explore the concept of propagating GWs, and how they are derived from GR. We discuss the possible sources of GWs, and summarise the attempts of detecting them via ground-based detectors. Finally, the concept of space-born GW telescopes is presented, focusing mostly in the LISA mission concept. The part of the GW spectrum that will be visible to a telescope like Laser Interferometer Space Antenna (LISA) is also discussed. The next chapter 2 enters into details of

1 R.A. Hulse and J.H. Taylor were awarded with the Nobel prize in 1993 for this discovery.
2 Free-floating means that the test masses follow a nominally geodesic motion.

the design of the LISA Pathfinder experiment, while chapter 3 introduces the background theory of the Bayesian statistics that we will use in the analysis of simulated data-streams. Then, in chapters 4 and 5, we discuss the modelling of the dynamics of the system, as well as the parameter estimation results from the aforementioned simulations. Finally, chapter 6 is dedicated to model selection methods implemented to discover the best model that describes the observations for the case of the LISA Pathfinder experiments.

1.2 AN OVERVIEW OF THE CONCEPT OF GRAVITATIONAL WAVES

The concept of Gravitational Waves (henceforth GW) was first introduced with Einstein's Theory of GR [2] and they can be seen as the consequence that no information, in particular the gravitational interaction, can travel faster than the speed of light from any given source to an observer. Just like in electromagnetism accelerated charges produce electromagnetic waves, the GWs are emitted with mass distribution changing over on the geometry of the space-time; they can be interpreted as small ripples on the space-time that periodically stretch and compress the the proper distance between two objects.

As a direct result, the proper distance L separating two test-masses will oscillate under the influence of a GW. The measurable effect of the GWs is produced by the equivalent of tidal forces, that is, gradients of the gravitational potential that affect the space-time curvature. This allows the scientific community to built potential detectors that operate with the same principle, i.e. measuring tiny differential disturbances that may have been caused by GWs. Since this chapter serves as an introduction, it would be beneficial to discuss the basic characteristics of the GW radiation and how they are derived from GR [3, 4, 5, 6] as small perturbations of space-time.

The starting point is to approximate a weak gravitational field by a "nearly" flat space-time, which is valid for the regions away from the sources of GWs. Now, the spacetime can be described as the manifold on which coordinates exist in which the metric has components

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{1.1}$$

If $\eta_{\mu\nu}$ is the Minkowski metric and we assume $|h_{\mu\nu}| \ll 1$, we have made a first approximation of a gravitational field as a small perturbation to the flat space-time. Then, Einstein's equations in the absence of matter ($T^{\mu\nu} = 0$), and far away from the source field, can be written as

$$\Box \bar{h}^{\mu\nu} = \left(-\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \bar{h}^{\mu\nu} = 0 \tag{1.2}$$

where we have used the *transverse-traceless* gauge conditions (TT). Eq. (1.2) is the three-dimensional wave equation and has a solution of the form

$$\bar{h}^{\mu\nu} = A^{\mu\nu} e^{ik_\mu x^\mu} \tag{1.3}$$

where k_μ are the (real) constant components of some one-form (or the wave-vector) and $A^{\mu\nu}$ the so-called wave amplitude tensor. To go a little further, one can restrict

the number of independent components of $\bar{h}^{\mu\nu}$ and reduce it to two. It can be shown [6], that

(a) k_μ is a *null* vector, and if the gauge condition $\bar{h}^{\mu\nu}_{\nu} = 0$ is imposed, we find that $A^{\mu\nu}$ is orthogonal to k_μ.

(b) since there always exists some ξ_μ which will transform from an arbitrary $h_{\mu\nu}$ to the Lorenz gauge, we can choose one that can further reduce the components of $\bar{h}_{\mu\nu}$ as shown in the following eq. (1.4).

$$\bar{h}^{TT}_{\mu\nu} = (A^{TT}_{\mu\nu})\cos\left[\omega(t-z)\right] = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_{xx} & A_{xy} & 0 \\ 0 & A_{xy} & -A_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos\left[\omega(t-z)\right]. \tag{1.4}$$

To reach that point we have assumed that $\vec{k} \to (\omega, 0, 0, \omega)$, to adapt the coordinate system so that the wave propagation direction coincides with the z axis. Here, the ω is a given frequency of the wave. For the given example where the GW is propagating along the z-axis, we can derive the independent states of polarization of the wave as shown in Fig. 1.1.

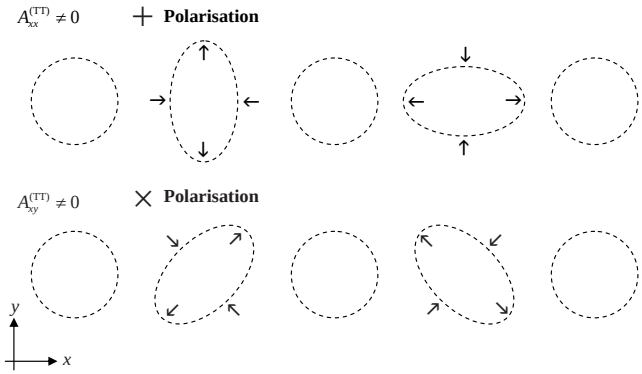


Figure 1.1: The effect of GW propagating along the z-axis, when passing through a "ring" of equally spaced free particles. The particles are initially at rest. The first row of images shows the distortions of the circle produced by a wave with the '+' polarization while the second shows the effect of a 'x' polarised wave. Notice, that contrary to the states of a electromagnetic wave, the two states of a GW are rotated by 45°.

1.2.1 *Gravitational-Wave Sources*

GWs are produced with rapid changes in the distribution of mass. But the principal question is about the possible astrophysical and cosmological sources that emit GW radiation that can be traced in our detectors. It can be proven [7] that far away from

the source and when the motion is small compared to c, the dominant contribution to the strain h is

$$h \simeq \frac{2G\ddot{I}_{\mu\nu}}{rc^4}, \tag{1.5}$$

where $G = 6.67 \times 10^{-11}\ m^3\ kg^{-1}\ s^{-2}$, r the distance to the source, and $\ddot{I}_{\mu\nu}$ is the *reduced quadrupole moment*[3]. From the above, it can be derived that metric perturbations which are spherically symmetric do not produce gravitational radiation. For example, a spherically symmetric collapse of a massive star will generate no GWs.

On the other hand, an optimistic estimation for h produced by a binary Neutron Star[4] located at the Virgo cluster, with both masses equal to the Chandrasekhar mass ($\sim 1.4 M_\odot$), will produce a strain of $|h| \leqslant 6 \times 10^{-21}$. In turn, an asymmetric supernovae originating from a star with $\sim 1 M_\odot$ core located again at the Virgo cluster, will produce $|h| \leqslant 10^{-21}$. This is almost one order of magnitude below the sensitivity of most of the developed *bar detectors* and at the limits of the initial interferometer detectors (see table 1.1 in the following section). Thus, to look for GW sources, we have to search for massive, rapidly moving asymmetric systems.

Another key point to be mentioned, is the event rates for sources within the reach of the developed detectors. For instance, the supernova event rate in our galaxy is calculated to a number of one to three events per century. Therefore, it is of crucial importance to increase the angular sensitivity of the detectors to reach bigger portion of the sky and increase the potential number of GW sources.

Below we will describe the potential GW sources in more detail, classified in order of their characteristic wave frequency, paving the way to later discuss about the detection schemes. A graphical representation of the GW spectrum can be seen in figure 1.2.

- *Extremely-low frequency band, ($10^{-18} - 10^{-15}$ Hz):* It was predicted that during the cosmic inflation, the produced cosmic GW background would imprint its signature in the polarization of the Cosmic Microwave Background (CMB). The BICEP2 experiment on the South pole, claimed to have confirmed recently this hypothesis by detecting this signature (the so-called "B-mode" polarisation of the CMB) [8, 9, 10]. At the moment, BICEP2 is at the level of upgrading the instrumentation for more sensitive measurements, while other space missions like Planck are expected to solidify this result with better statistical accuracy.

- *Very-low and low frequency band, ($10^{-9} - 1$ Hz):* The detection of any source emitting in the Very-low frequency band ($10^{-9} - 10^{-5}$), is extremely demanding in terms of detector requirements. This bandwidth is populated mostly with binary Super Massive Black Holes (SMBHs). The developed strategy for detection, is through measuring any simultaneous deviation of know pulsar periods that may be caused by GWs passing through the Earth [12]. There are detector arrays already devoted to this task, like the International Pulsar Timing Array (IPTA)[6], and the future Square Kilometre Array (SKA) [13], that are expected to be the

3 The reduced quadrupole moment is defined as $I_{\mu\nu} = \int \rho(\vec{r}) \left(x_\mu x_\nu - \frac{1}{3}\delta_{\mu\nu} r^2 \right) dV$, where ρ is the matter density. It is also worth mentioning that this is valid in GR, but not in other alternative theories of gravity.

4 Here we have assumed the signal frequency to be $f = 1$ kHz and a circular orbit of coordinate radius $R = 20$ km.

5 http://science.gsfc.nasa.gov/663/research/

6 http://www.ipta4gw.org/

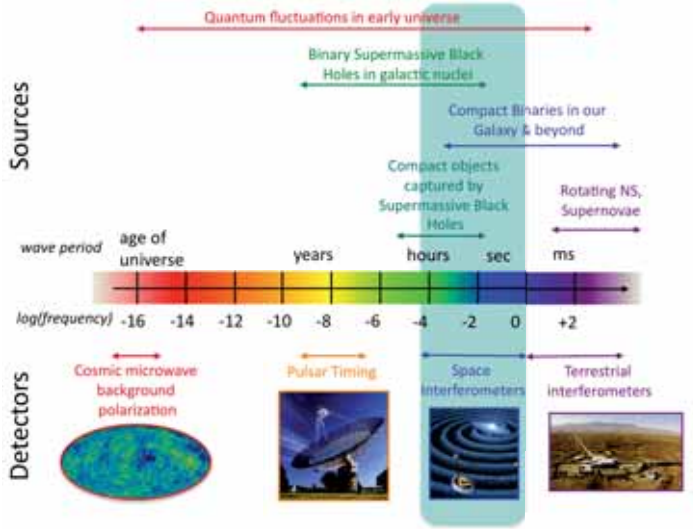


Figure 1.2: The GW spectrum and the detection frequency range for the different observatories. Credit: NASA[5]. An more intuitive and interactive plot, together with detectors sensitivity curves, can be found in [11].

first to claim evidence of GWs.

For the case of the low frequency band ($10^{-5} - 1$ Hz), it is estimated that there is a number of sources that can be used for GW astronomy. They can only be detected by space-borne observatories like the eLISA mission (see section 1.3.2 and figure 1.7). In particular the potential sources are coalescences of SMBHs of masses of $10^5 - 10^9$ $M_\odot$, white dwarf binaries, Extreme Mass Ratio Inspirals (EMRIs) of $10^5 - 10^7$ $M_\odot$, and ultra-compact binaries inside our galaxy [14, 15].

- *High frequency band, ($1 - 10^4$ Hz):* In this region, mainly GWs produced by systems in their last moments of their inspiral, or very rapid bursts caused by supernovae explosions are expected [16]. The large ground interferometer detectors are focused in this bandwidth. More specifically, one of the possible sources are NS-BH[7] mergers in their last few minutes of their inspiral, out to a cosmological redshift $z = 0.4$. The enhanced detectors will be also able to see tidal disruptions of a NS by its BH companion that can be cross-validated with short gamma-ray bursts detection. The final stages of BH-BH merger and ringdown with larger masses than $\sim 10M_\odot/10M_\odot$ at $z = 1$ are also expected, as well as, deformed spinning NS, known fast pulsars and asymmetric stellar core collapses that cause supernovae.

7 of masses $M_{NS} \sim 1.4M_\odot$ and $M_{BH} \sim 10M_\odot$.

1.3 GRAVITATIONAL WAVE DETECTION

Until recently, GWs were not even considered as tools for astronomical observations because of the obvious difficulties in their detection. Astronomers are mostly relying in the observation of the electromagnetic spectrum, thus limiting their measurements to a small fraction of the available information. First because only a 4% of the universe is observable with optical means [6], and secondly, because electromagnetic waves, from the moment of their production, they are dispersed, scattered and absorbed multiple times before they reach our telescopes.

GWs on the other hand, since they interact very weakly with matter, they can provide with useful information directly from the interior of the source. For example, the estimated cross-section for the absorption of GW energy by matter is [17]

$$\sigma = 10^{-24} S, \tag{1.6}$$

where S is the transversal area of the target body. While this can be beneficial, GW astronomy requires extremely sensitive detectors. But since a few decades now, many ideas for GW detectors have been implemented and they are continuously being upgraded to their limits. There are two major categories for ground-based GW detectors. The *resonant mass* or simply *bar* detectors and the *interferometric* detectors [18, 19].

1.3.1 *Ground-Based Detection*

The first ever GW detector was built by J. Weber in the 1960s, a pioneer in the field, who supported the idea of *bar* detectors. He was the first to set-up an array of such detector antennas, operating in room temperature. It is now known, that his claims for GW detection [20], couldn't be reproduced by other groups. But in the end the scientific heritage from -the now surpassed- bar detectors is widely acknowledged, in terms of instrumentation and data analysis. The scientific community now is more focused on the interferometric detectors, that are estimated to detect the first GW source in the high frequency band, in the nearest future.

However, the ground-based GW detectors show an important drawback that affects their sensitivity curve in its lower frequencies, at around ~ 10 Hz. They are limited by the so-called *seismic wall*, the Earths seismic noise, and the gravity gradient noise, and these two can be the principal arguments to build a space-borne GW detector.

1.3.1.1 *Resonant Mass Detectors*

The principle of the bar detectors os based on the idea (as the name suggests) of the resonant amplification of a signal. Consider two test masses that are linked together with a spring of length ℓ_0, as in left panel of figure 1.3. A passing GW will drive oscillations around their resting position, and the measurement will be the mechanical amplification at the spring characteristic frequency Ω.

In reality, a bar detector consists of a solid body in the shape of a cylinder (although many shapes have been proposed and/or developed [17, 21, 22]), while several materials have been considered like niobium, sapphire and aluminium. More details can be found

8 http://www.auriga.lnl.infn.it/auriga/detector/run1/sensitivity.html

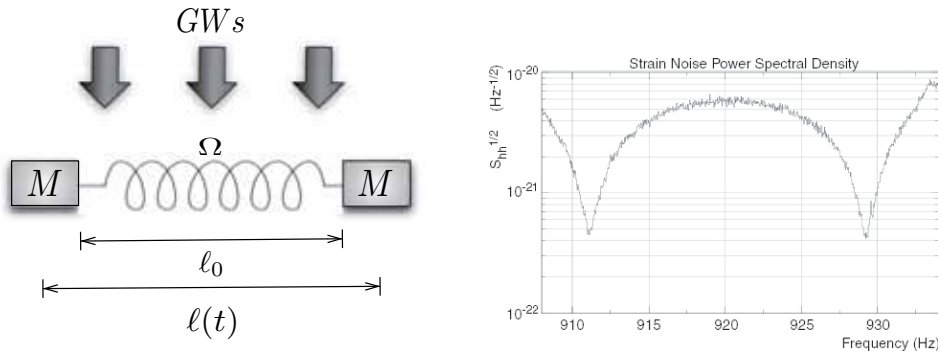


Figure 1.3: *Left:* Principle of operation of a bar detector. *Right:* Typical sensitivity curve of a bar detector, in this case the linear spectral density $\sqrt{S_h}$ of the strain noise versus frequency for the first run of the AURIGA detector[8].

in table 1.1. The primary noise sources for such a design are namely the *thermal noise*, *sensor noise* and the *quantum limit*. In 1971, several institutions agreed for the development of a network of such detectors working in coincidence, in order to do correlation analysis to identify the direction and velocity of incoming waves [18].

After continuous improvements (achieving impresively low temperatures for the masses), in 1990s the pioneering groups of the field created the International Gravitational Event Collaboration. Later in 2006, after almost two decades of operations, no coincidence events were recorded. Nevertheless, they published the upper limit for gravitational wave events at around 1 kHz. The present sensitivity of the detectors could only detect GW bursts of a duration of a few ms, from sources like class II Super Novae (SN)e located inside our galaxy (corresponds to a strain sensitivity of $h \sim 4 \times 10^{-19}$). Since the rate of such events is about 0.004 SNe/year, AURIGA and NAUTILUS antennas continued to be functional for a small period of years, as complementary to interferometer GW detectors. Now they are still being operative but probably are going to be dismantled soon, specially in countries where the focus is shifted to laser interferometers.

Specs	AURIGA	NAUTILUS	EXPLORER	ALLEGRO	NIOBE
Material	Al5056	Al5056	Al5056	Al5056	Nb
M (Kg)	2230	2260	2270	2296	1500
length (m)	2.9	3.0	3.0	3.0	2.75
min freq. (Hz)	912	908	905	895	694
max freq. (Hz)	930	924	921	920	713
$Q_{\pm}$ (10^6)	3	0.5	1.5	2	20
Temp. (K)	0.25	0.1	2.6	4.2	5
Misalignment (*)	5●	2●	3●	6●	16●
$\bar{h}/\sqrt{Hz}$	3×10^{-21}	2.5×10^{-19}	10^{-20}	10^{-20}	10^{-20}
Source: http://igec.lnl.infn.it/					

Table 1.1: List of resonant mass detectors and their characteristics [21]. (*) Angle between bar axis and the perpendicular to the Earth great circle closer to the five detectors.

1.3.1.2 *Laser Interferometry*

Given the narrowband nature of bar detectors and their limited potential sources, the interest shifted to *laser interferometer* type of detectors. The main advantages of these set-ups, is firstly the considerable improvement in the sensitivity and the broader spectrum of sources. The idea behind the interferometer detectors is shown in figure 1.4, and is based on the fact that a passing GW will cause changes $\delta\ell$ to the proper distance ℓ_0 between two test masses. If the orientation of the test-masses and the incoming GW is optimal, the $\delta\ell$ difference will produce a phase shift

$$\delta\phi = \frac{4\pi}{\lambda}\delta\ell. \tag{1.7}$$

A Michelson interferometer is a great instrument to measure this phase shift. Unlike the bar detectors, in this case the obvious solution to increase the the size of the signal in the data-stream, is to increase the optical arms of the detector.

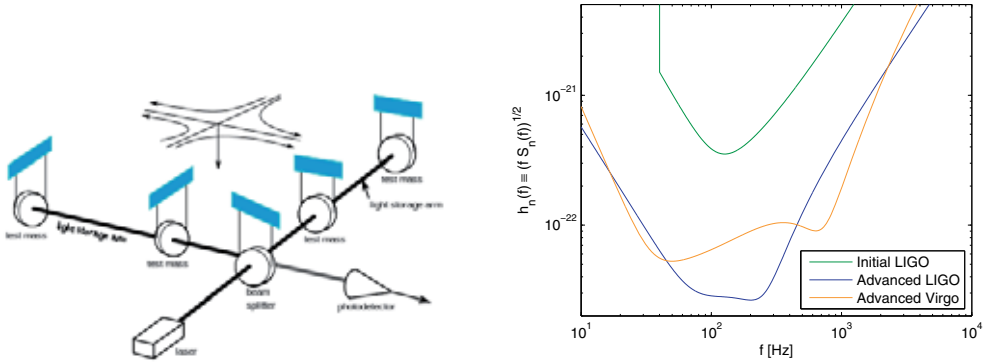


Figure 1.4: *Left:* Interferometer detectors operation principle: A laser beam is firstly being split (by a beam splitter mirror) and then transmitted to the test-masses (mirrors) at the end of the 4 km arms. There, as he light bounces, a Fabry-Perrot cavity is formed. The final measurement happens in the photodetector where the light from both arms interferes. A passing gravitational wave will cause disturbances in the resting distances of the test-masses, and in turn the detectable phase shift $\delta\phi$ as in eq. (1.7). *Right:* Effective GW noise $h(f) \equiv \sqrt{fS_n(f)}$ of three GW ground-based detectors. Credit: [23].

The main noise sources of the ground-based interferometer detectors are:

- The *seismic* noise: the main limit below 10 Hz, often called as the *seismic wall*. The large mechanical vibrations are filtered out with sophisticated pendulum suspensions and various vibration isolation systems (piezo-electric actuators, material layers).
- The *thermal* noise: it is caused mostly by brownian motion of the mirrors and it is limiting at 50 and 250 Hz.
- The *shot* noise: limiting at frequencies above 250 Hz, the random distribution of photons recorded in the photodiode produce fluctuations of the light intensity. Can be mitigated with a powerful enough laser (to the cost of *radiation pressure noise*).

- The *gravity gradient* noise: is caused by the local changes of the newtonian field (atmospheric pressure, seismic density waves) and is limiting at frequencies below 1 Hz.

- The *quantum* noise: the shot noise and effects like zero-point vibrations.

Similar to the case of bar detectors, a network of interferometer detectors has been established (the *Laser Interferometer Gravitational-Wave Observatory (LIGO) scientific collaboration* involves the LIGO antennas, the GEO600 and has signed an agreement of data access with the European Gravitational Observatory (EGO) that includes VIRGO). This network enhances the angular resolution of the detector with techniques based on time-delay between signal detection and the Doppler effect due to relative motion of the antennas.

During the writing of this document, the LIGO collaboration was in the status of upgrading the detectors to the so-called *advanced LIGO/VIRGO*, and they are scheduled to perform their first scientific runs starting from late 2016. There are also developments for new detectors, such as KAGRA in Japan (formerly known as the Large Cryogenic Gravitational-wave Telescope, LCGT), and studies for new concepts, like the Einstein Telescope (ET). The ET is a project for a new generation underground GW detector that will be better insulated from seismic noise, and it will use cryogenic technology for thermal stability. It is still a very broad concept and there is no any final design nor a date to start the construction. The characteristics of the ground-based detectors are summarised in table 1.2.

Detector	Country	Arm length (km)	Approximate date	Generation
GEO600	Germany & UK	0.6	2001-present	1st
TAMA300	Japan	0.3	1995-present	1st
iLIGO	US	4	2004-2010	1st
iVIRGO	Italy & France	3	2007-2010	1st
aLIGO	US	4	*est* 2016	2nd
KAGRA	Japan	3	*est* 2018	2nd
aVIRGO	Italy & France	3	*est* 2017	2nd
ET	European	10	*est* 2030s	3rd

Table 1.2: List of ground-based GW detectors and their basic characteristics.

1.3.2 *Space-Based Detection: The concept of the Laser Interferometer Space Antenna*

Due to the limitations of the terrestial detectors at low frequencies of the GW spectrum, but mostly because of the interesting science in the lower frequency bands, the LISA concept was born. It can be seen as an interferometer, similar to the ones on ground, but gliding into orbit. The first mission concept studies can be tracked back at around 1980s [24, 25] as a purely NASA mission that consisted of six space-crafts in a heliocentric orbit. Later, between the period of the early 1990s to 2005, the space-borne observatory underwent many changes in its design, but it was always widely acknowledged by the scientific community [26] as a telescope of great scientific potential.

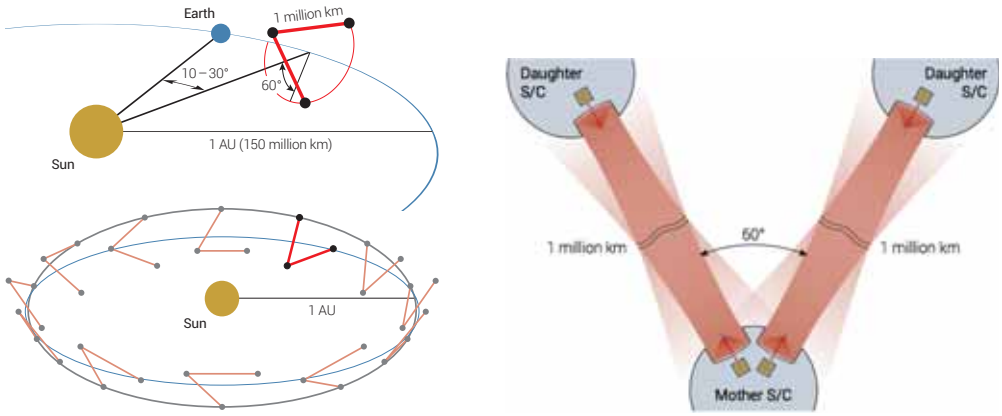


Figure 1.5: *Left:* Cartoon of the eLISA space-crafts constellation and orbit. *Right:* Details for the eLISA space-crafts configuration. The proposed design consists of one "mother" and two "daughters" space-crafts. See text for details. Credit: [27].

But around 2012, NASA dropped from the mission due to budget cuts, and forced the European counterpart to continue with a reduced financial plan and an evolved design. Notwithstanding the adversities, in October 2013 the *Gravitational Universe* [27], a science theme addressed to the European Space Agency (ESA) by the collaboration, was chosen for the L3 mission to be flown in the early 2030s. And in this document the new evolved LISA (eLISA) was introduced.

The eLISA concept is fairly similar to the established LISA design [24, 28], but with two arms instead of three. The distances of the space-crafts has been also reduced from 5×10^6 to 10^6 km, still flying in a heliocentric orbit as in left panel of figure 1.5. The main goals for a space-borne GW observatory like eLISA are [27, 26]:

1. to study the nature of black holes (understand the formation, trace their history and growth rate) and map the space-time around them.

2. to explore the stellar populations and dynamics and structure of the galactic nuclei.

3. to test Einstein's theory of General Relativity in the strong field regime. That is, to answer fundamental questions about gravity, for instance, what is the GW propagation speed and if graviton has mass or not.

4. to probe new physics and cosmology using the GWs as a tool and search for unforeseen sources of GWs.

5. to explore energy scales and epochs as close as 10^{-18} to 10^{-10} seconds after the Big Bang.

Since the early days of LISA, it was clear that such a mission would be technologically challenging. As already mentioned, the experiment consists of three drag-free space-crafts in a triangle formation, where inside each one, test-masses in free fall conditions (undisturbed by any force other than gravitation) are placed (right panel of figure 1.5). The importance of a very quiet environment for the test-masses is immediately

recognised, because any local disturbance may be confused with displacements caused by GWs. One important detail about the eLISA space-crafts is that they do not require any kind of formation-flying [25]. The constellation (distance and angles) is evolving freely under the influence of gravitation alone, while the only actuation is torques applied to the space-crafts, firstly to follow the test masses, and secondly to keep the constellation in the filed of view.

At this point, it would be beneficial to describe briefly the payload of the experiment, first to put LISA Pathfinder in context in the following chapter, but also to underline the technological developments of the last decade. In particular, the main instrument blocks on-board are:

- the Optical System. It is responsible for the main scientific measurements, like the noise and the distance between the test-masses. Each optical system unit consists of the *telescope*, the *optical bench* with its subsystems, and the Gravitational Reference Sensor (GRS), which will be described below. The "Mother" space-craft is equipped with two of these units, while each "daughter" with one. The design of the telescope is greatly affected by the planned size of the constellation of the space-crafts, thus the light power required. The current design has a total 60 cm mechanical length, 20 cm diameter, and a field of view of ±7 μrad out-of-plane and ±4 μrad in-plane. The Optical Bench brings support to the various electronic and optical components (such as mirrors, splitters, photodiodes and CCD sensors) in order to direct and interfere the laser beams. The set-up is planned to be inherited directly from LISA Pathfinder Technology Package on-board LISA Pathfinder. A very critical instrument for the mission is the *phasemeter*, which cannot be inherited from LISA Pathfinder due to the higher heterodyne frequency required [29].

- the Disturbance Reduction System (DRS) is the component that constrains the differential residual acceleration error of the test-masses. Basically it consists of two sub-systems, the GRS and the controllers (Drag-Free Attitude and Control System). The DRS controls the space-craft to follow its hosting test-masses by measuring their distance between them, while keeping the the constellation in shape. The sub-systems of GRS are namely the test-masses and housing, the capacitive sensing, the test-mass locking system, and the charge control system . The position of the test-masses is monitored via capacitive readout system, while the degrees of freedom other than the sensitive axis are being controlled by applying voltage to the electrodes. The GRS for eLISA is similar to the GRS of LISA Pathfinder.

- the Micropropulsion System. The LPF mission, originally was planned to use the so-called Field Emission Electric Propulsion (FEEP) thruster, to be abandoned in favour of the Cold Gas thrusters. The Cold Gas μNewton thrusters, that have been selected for the GAIA mission, were verified and declared suitable for LISA Pathfinder.

See [25] and its references for a more detailed description.

Concerning the Data Analysis section of the experiment, great progress has been demonstrated during the last decade; algorithms and statistical tool have been developed in order to extract and resolve the GW signal from the data stream. In addition,

the Mock LISA Data Challenges (MLDC) took place starting at 2006, in order to demonstrate the technical readiness of the data analysis team to disentangle multiple sources present in the data series [30, 31, 32], as in figure 1.6. The data analysis techniques developed for eLISA are very relevant for this thesis, because of the related nature of the instrumentation and the similar statistical challenges. An example of GW sources in the frequency band of eLISA can be seen in figure 1.7.

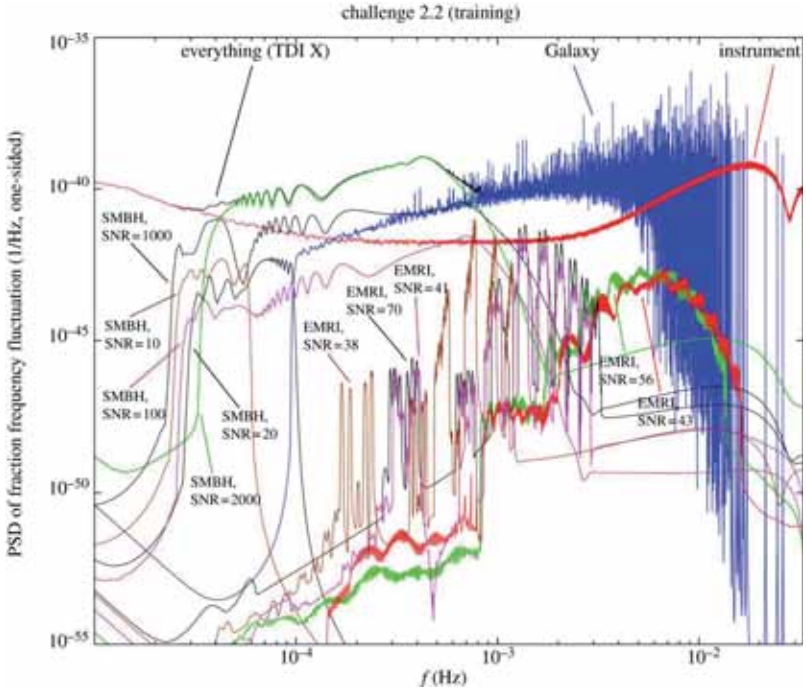


Figure 1.6: A simulation of the strain spectral density recorded by LISA broken out into the individual contributions from galactic sources (blue), instrument noise (red), massive black holes and EMRIs. The black line denotes the spectrum of combined detector output. Credit: [31].

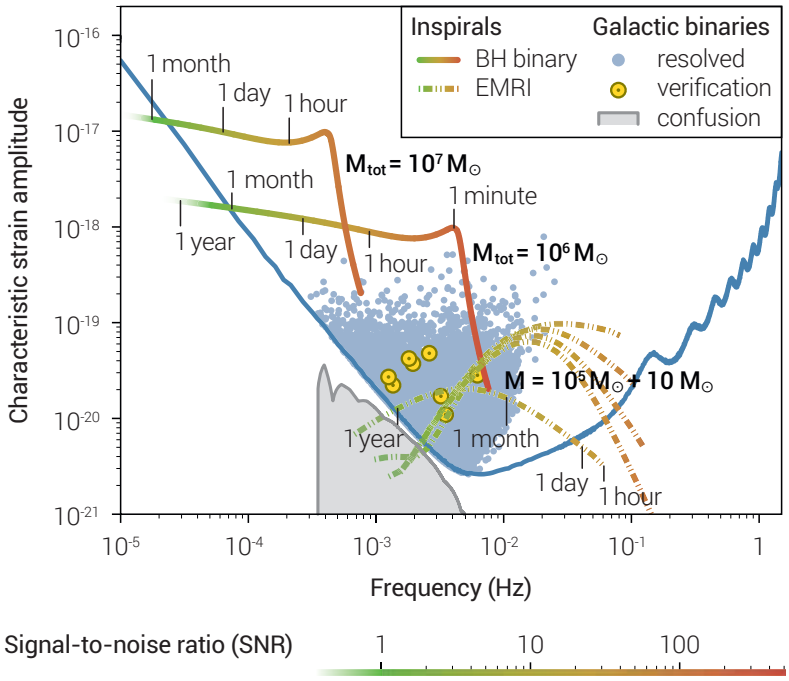


Figure 1.7: Examples of gravitational wave astrophysical sources in the frequency range of eLISA, where the data is plotted in terms of "characteristic strain amplitude". An equivalent plot is shown for an EMRI source at 200 Mpc, with 5 harmonic frequencies evolving simultaneously. Several thousand galactic binaries, with SNRs above 7, will be resolved after one year of observation. Millions of other binaries result in a "confusion noise" that varies over the year. The average level is represented as grey shaded area. Credit: [27].

[1] R.A. Hulse and J.H. Taylor. Discovery of a pulsar in a binary system. *Astrophys.J.*, 195:L51–L53, 1975. doi:10.1086/181708. 3

[2] Albert Einstein. The Foundation of the General Theory of Relativity. *Annalen Phys.*, 49:769–822, 1916. doi:10.1002/andp.200590044. 4

[3] J. Weber. *General Relativity and Gravitational Waves*. Dover Books on Physics Series. Dover Publications, 1961. URL: http://books.google.gr/books?id=t6_AXrEnVZYC. 4

[4] M. Maggiore. *Gravitational Waves: Volume 1: Theory and Experiments*. Gravitational Waves. OUP Oxford, 2008. URL: http://books.google.gr/books?id=AqVpQgAACAAJ. 4

[5] Indrajit Chakrabarty. Gravitational waves: An Introduction. *Reson.Sci.Ed.*, 1999. arXiv:physics/9908041. 4

[6] Bernard Schutz. *A First Course in General Relativity*. Cambridge University Press, 2nd edition, June 2009. URL: http://www.amazon.com/A-First-Course-General-Relativity/dp/0521887054. 4, 5, 8

[7] Bernard F. Schutz. Gravitational Waves on the back of an envelope. *Am.J.Phys.*, 52:412–419, 1984. doi:10.1119/1.13627. 5

[8] The BICEP2 Collaboration. BICEP2 2014 Results Release, 2014. URL: http://bicepkeck.org/. 6

[9] P.A.R. Ade et al. BICEP2 I: Detection Of B-mode Polarization at Degree Angular Scales. 2014. arXiv:1403.3985. 6

[10] A.R̃. Ade, P. W. Aikin, R. D. Barkats, et al. Detection of b-mode polarization at degree angular scales by bicep2. *Phys. Rev. Lett.*, 112:241101, Jun 2014. URL: http://link.aps.org/doi/10.1103/PhysRevLett.112.241101, doi:10.1103/PhysRevLett.112.241101. 6

[11] C. Moore, R. Cole, and C. Berry. Gravitational Wave Sources, and Detector sensitivity curves, 2014. URL: http://www.ast.cam.ac.uk/~rhc26/sources/. 7

[12] G Hobbs, A Archibald, Z Arzoumanian, et al. The international pulsar timing array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27(8):084013, 2010. URL: http://stacks.iop.org/0264-9381/27/i=8/a=084013. 6

[13] R.N. Manchester. Pulsar searching and timing. *Int.J.Mod.Phys.*, D22:1341007, 2013. doi:10.1142/S0218271813410071. 6

[14] Pau Amaro-Seoane, Sofiane Aoudia, Stanislav Babak, et al. Low-frequency gravitational-wave science with elisa/ngo. *Classical and Quantum Gravity*, 29(12):124016, 2012. URL: http://stacks.iop.org/0264-9381/29/i=12/a=124016. 7

[15] Pau Amaro-Seoane, Sofiane Aoudia, Stanislav Babak, et al. eLISA/NGO: Astrophysics and cosmology in the gravitational-wave millihertz regime. *GW Notes*, 6:4–110, 2013. arXiv:1201.3621. 7

[16] Curt Cutler and Kip S. Thorne. An Overview of gravitational wave sources. pages 72–111, 2013. arXiv:gr-qc/0204090, doi:10.1142/9789812776556_0004. 7

[17] E. Coccia, J.A. Lobo, and J.A. Ortega. Gravitational wave observatory based on solid elastic spheres. *Phys.Rev.*, D52:3735–3738, 1995. doi:10.1103/PhysRevD.52.3735. 8

[18] Piero Rapagnani. Gravitational wave detectors on the earth. *Class.Quant.Grav.*, 27:194001, 2010. doi:10.1088/0264-9381/27/19/194001. 8, 9

[19] J. Alberto Lobo. The Detection of gravitational waves. *Lect.Notes Phys.*, 617:210–241, 2003. arXiv:gr-qc/0202063. 8

[20] J. Weber. Evidence for discovery of gravitational radiation. *Phys.Rev.Lett.*, 22:1320–1324, 1969. doi:10.1103/PhysRevLett.22.1320. 8

[21] Peter Aufmuth and Karsten Danzmann. Gravitational wave detectors. *New Journal of Physics*, 7(1):202, 2005. URL: http://stacks.iop.org/1367-2630/7/i=1/a=202. 8, 9

[22] J. Alberto Lobo. Lisa. 2004. URL: http://arxiv.org/abs/gr-qc/0404079, arXiv:gr-qc/0404079. 8

[23] Thibault Damour, Alessandro Nagar, and Miquel Trias. Accuracy and effectualness of closed-form, frequency-domain waveforms for nonspinning black hole binaries. *Phys. Rev. D*, 83:024006, Jan 2011. URL: http://link.aps.org/doi/10.1103/PhysRevD.83.024006, doi:10.1103/PhysRevD.83.024006. 10

[24] LISA International Science Team. The LISA Assessment Study Report, YellowBook ESA/SRE(2011)3, February 2011. URL: http://sci.esa.int/lisa/48364-lisa-assessment-study-report-yellow-book/. 11, 12

[25] Oliver Jennrich et al. NGO Assessment Study Report, YellowBook ESA/SRE(2011)19, January 2012. URL: http://sci.esa.int/ngo/49839-ngo-assessment-study-report-yellow-book/. 11, 13

[26] Committee for a Decadal Survey of Astronomy and Astrophysics; National Research Council. *New Worlds, New Horizons in Astronomy and Astrophysics*. The National Academies Press, 2010. URL: http://www.nap.edu/openbook.php?record_id=12951. 11, 12

[27] Pau Amaro Seoane et al. The Gravitational Universe. 2013. URL: http://arxiv.org/abs/1305.5720, arXiv:1305.5720. 12, 15

[28] the NASA/ESA LISA Project Team. Laser Interferometer Space Antenna (LISA). a response to the Astro2010 RFI for the particle astrophysics and gravitation panel. Technical report, April 2009. URL: http://lisa.nasa.gov/documentation.html. 12

[29] D. Shaddock, B. Ware, P. G. Halverson, R. E. Spero, and B. Klipstein. Overview of the lisa phasemeter. *AIP Conference Proceedings*, 873(1):654–660, 2006. URL: http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.2405113, doi:http://dx.doi.org/10.1063/1.2405113. 13

[30] K.A. Arnaud, G. Auger, S. Babak, et al. Report on the first round of the Mock LISA data challenges. *Class.Quant.Grav.*, 24:S529–S540, 2007. arXiv:gr-qc/0701139, doi:10.1088/0264-9381/24/19/S16. 14

[31] K.A. Arnaud, S. Babak, J. Baker, et al. An Overview of the second round of the Mock LISA Data Challenges. *Class.Quant.Grav.*, 24:S551–S564, 2007. arXiv:gr-qc/0701170, doi:10.1088/0264-9381/24/19/S18. 14

[32] Stanislav Babak et al. The Mock LISA Data Challenges: From Challenge 3 to Challenge 4. *Class.Quant.Grav.*, 27:084009, 2010. arXiv:0912.0548, doi:10.1088/0264-9381/27/8/084009. 14

2 LISA PATHFINDER

So far, we have made a short introduction to GWs physics and detection, more focused to the space-based experiment part, the eLISA. In this chapter, we will describe the LISA Pathfinder mission in more details, since it's more relevant for the purpose of this thesis. This chapter is organised as follows: In the first section, the main idea, purposes and goals of the mission are going to be described. In the second section, the payload and main instrumentation are going to be briefly discussed, and finally, we will summarise the current status of the satellite.

At the time around 2000, the ESA and NASA realised the technological challenges in the design of LISA. In addition, a great part of the flight hardware could not be fully tested on ground. For these reasons, a mission called SMART-2 (Small Missions for Advanced Research in Technology) and later re-named to LISA Pathfinder (LPF) [1, 2, 3], was selected as a precursor mission to a space-borne GW observatory similar to the concept of eLISA. The LPF was selected by the ESA Science Programme Committee (SPC) in November 2000. It was further reconfirmed by the same body and by the ESA Council in May 2002, as part of ESA's new 'Cosmic Vision' Scientific Programme.

During the first phases of the design of the mission, the two main components of the payload of the satellite were the LISA Technology Package (LTP), from the ESA part, and the DRS coming from the NASA counterpart. The LTP is basically the principal element of LPF, while from the DRS module only the thrusters and controllers from the initial design are included. In this thesis we will focus on the European contribution of the LPF.

The LPF is going to be launched with the ESA VEGA rocket, and put in orbit around the first Sun-Earth Lagrange point (L1). At the moment, it is at its final stages before the scheduled launch in half of 2015.

2.1 THE MISSION

The scientific objective of the LPF mission consists of the first in-flight test of low frequency gravitational wave detection metrology [4, 5]. In particular, while the purpose of eLISA is to directly measure the effect of changes in the spacetime geometry produced by passing GWs, the LPFs' *main aim is to demonstrate that free-falling bodies follow geodesics in space-time by more than two orders of magnitude better than any past, present, or planned mission* (with the exception of eLISA itself) [6].

The scheme for this test is to "squeeze" an eLISA optical arm from 10^6 km down to 30 cm, while keeping the main principle of measurement. The main scientific measurement of LPF is the distance between two test masses (as in an eLISA arm) by means of

laser interferometry. In the end, the relaxed in comparison to eLISA differential acceleration noise requirement is

$$S_{\Delta\alpha}^{1/2}(f) \leqslant 3\times 10^{-14}\left[1+\left(\frac{f}{3\ \mathrm{mHz}}\right)^2\right]\ \mathrm{m\ s^{-2}/\sqrt{Hz}}, \tag{2.1}$$

in the frequency range of $1\ \mathrm{mHz} \leqslant f \leqslant 30\ \mathrm{mHz}$. The last equation means, that the total noise contribution for each subsystem in the differential acceleration measurement $\Delta\alpha$, does not exceed the value of $S_{\Delta\alpha}^{1/2}(f)$ for the given frequencies. And this comes directly from the requirements of eLISA/LISA experiment, because the ability to measure space-time curvature requires free-falling test-mass pairs with very low relative acceleration of non-gravitational origin, and light beam tracking of the test-masses with the lowest possible instrumental noise. Of course together with the previous requirements, there are secondary demands that follow. To measure with this sensitivity, we need

- Very low thrust level ($\sim 10\ \mu\mathrm{N}$) and very quiet in terms of noise to command the Space-Craft (SC) to follow the Test-Masss (TMs).
- 18-degrees of freedom dynamical control laws, and
- A gravitationally flat and stable SC to host the test bodies.

Having described the principal aim of LPF, there is a series of individual goals to be achieved during the lifespan of the mission. In particular, the LPF will [6]:

1. Test and develop key technologies to be inherited to a future GW observatory, like the eLISA mission concept..
2. Characterise and model all noise sources of an instrument operating following the working principles of LISA, dedicated for low frequency Gravitational-Wave detection.
3. Demonstrate that electromagnetic and locally generated gravitational fields can be suppressed to a level that allows geodesic motion with a precision better than $\Delta\alpha \simeq 7\times 10^{-15}\ \mathrm{ms^{-2}/\sqrt{Hz}}$ at $1\ \mathrm{mHz}$.
4. Demonstrate the residual acceleration of the spacecraft relative to a local inertial frame with $\alpha_{SC} \simeq 2\times 10^{-13}\ \mathrm{ms^{-2}/\sqrt{Hz}}$ at $1\ \mathrm{mHz}$.
5. Operate as a differential dynamometer, and it will measure forces at $1\ \mathrm{mHz}$ with amplitude uncertainty of $\Delta\alpha/\sqrt{T} \simeq 2\times 10^{-17}$ during a period of $T \simeq 5$ days.
6. operate as a gradiometer with a resolution of $\simeq 1.5\times 10^{-14}\ \mathrm{s^{-2}/\sqrt{Hz}}$. This together with the gravitational compensation abilities ($\simeq 10^{-6}\mathrm{s^{-2}}$) of LPF, opens the way to a new class of high-resolution geodesy missions.

In the following section we will describe the main instrument on-board, the LTP, which performs this sensitive measurement, together with its subsystems.

2.2 LISA TECHNOLOGY PACKAGE

Figure 2.1: Artistic representation of the LISA Technology Package. The test-masses inside the vacuum enclosure can be seen, as well as the optical elements of the experiment.

The LTP encloses all the elements for the sensitive differential measurement between the TMs. It is also responsible for the *drag-free* control of the space-craft. Drag-free control is just one of the many tools used to achieve test-mass geodesic motion. The main difference between drag-free and geodesic motion, is that geodesic motion is the lack of relative acceleration between free test-masses other than the one due to spacetime curvature, while drag-free motion is the lack of acceleration of the spacecraft relative to a local inertial frame. This is achieved (unlike other space missions), by controlling the attitude of the LPF space-craft with the payload, by commanding forces to the μN thrusters.

In the following, we will break-down and describe the LTP and its various subsystems as well as a few other important elements of the experiment; the Inertial Sensor Subsystem, the Optical Metrology System, the Drag-Free Attitude Control System (DFACS), and the Data and Diagnostics Subsystem.

- **Inertial Sensor Subsystem (IS)**: one of the major components, it consists of the test-masses and their surroundings; namely the vacuum enclosure, the electrode housing, the frond-end electronics, the charge management unit, the caging mechanism, and the vacuum enclosure (see figure 2.2). The IS is responsible for the capacitive sensing and applying forces to control the motion of the TMs. In addition, it needs to guarantee the mechanical accuracy to allow TM alignment under the requirements [7].

  The TMs, are designed to reduce as much as possible the noisy external forces. They are identical, 1.96 kg Gold-Platinum alloy cubes. The choice of the material (73% gold and 27% platinum) is due to its high density and its magnetic properties (very low magnetic susceptibility $\chi_m \simeq 10^{-5}$). The high density of the alloy, together with a greater area of the TM, allows better sensing and more accurate commanded forces with the electrodes.

The Caging Mechanism Assembly (CMA) is the mechanism responsible for the release of the TMs marking the beginning of the scientific operations. One of the major requirements of the CMA is the accuracy of the TMs injection into geodesic orbit, and their stability during launch and separation of the space-craft modules. This instrument must release the TM within an error margin of $200\ \mu m$ and a velocity $u \leqslant 5 \times 10^{-6}\ ms^{-1}$. During launch a hydraulic actuator applies $\simeq 2 \times 10^{3}\ N$ on each TM. Ground tests have already shown that the above requirements are met for the CMA.

The TMs are enclosed inside the Electrode Housing (EH) which monitors their position, in all six degrees of freedom (per TM), via electrostatic sensing. The EH was required to be mechanically robust and accurate, and provide a stable thermal environment for the TMs. For that reason the material for the construction was shifted to molybdenum ($_{42}Mo$), because of its high thermal conductivity value. The sensing per TM is performed by twelve gold coated sapphire electrodes with $\simeq 4\ mm$ gap between each TM face and the electrodes.

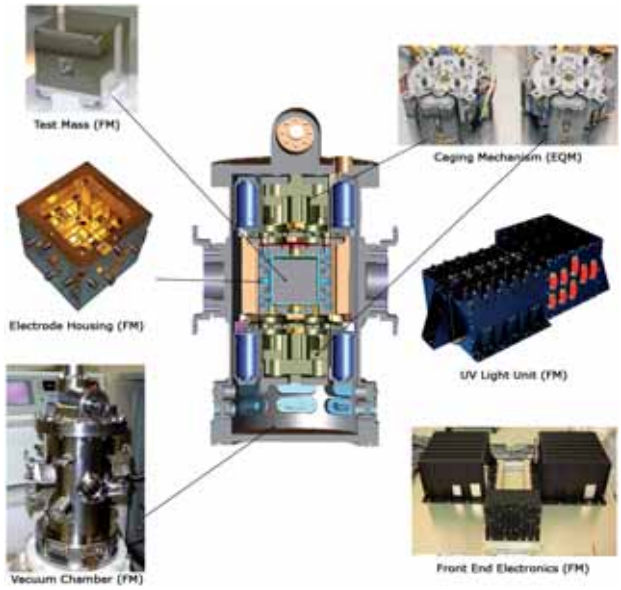


Figure 2.2: The main components of the Inertial Sensor. Credit [8].

The principle of the capacitive measurement is based on computing the induced voltage of the electrodes; any possible displacement of the TM from its central initial position will cause capacitance changes between the electrode pair, from which the read-out signal for the TM positioning will be calculated. The dominant sources of noise inside the EH are the displacement noise and noise originating from environmental and thermal gradient effects. The displacement noise for one channel at 100 kHz is estimated to be at the level of

$$S_{x_{th}}^{1/2} \simeq 1.1\ nm\ Hz^{-1/2}, \tag{2.2}$$

where each channel is comprised by a pair of opposite electrodes as shown in figure 2.3. Figure 2.3 shows a cartoon of the electrodes, as well as the Frond End Electronics (FEE), and their basic principle of operation.

TM charging due to cosmic rays is another notable noise source. Any non-zero charge on the TM will induce a bias on the read-out of the system, since it is based on capacitive sensing [9]. For that reason, two main strategy have been employed. First, a series of experiments have been planned to measure the charge of the TMs for the duration of the in-orbit operations. If necessary, DC bias voltages are going to be applied through the electrodes to compensate stray potentials on the TMs.

Secondly, a mechanism has been developed to dismiss the accumulated charge from both TMs [10, 11]. This mechanism is based on illuminating the TM and housing with UV light. With similar photoelectric properties of the two surfaces, a net flow of photoelectrons will be produced travelling from the TM to the housing. The module designed for this job is the UV Light Unit (ULU).

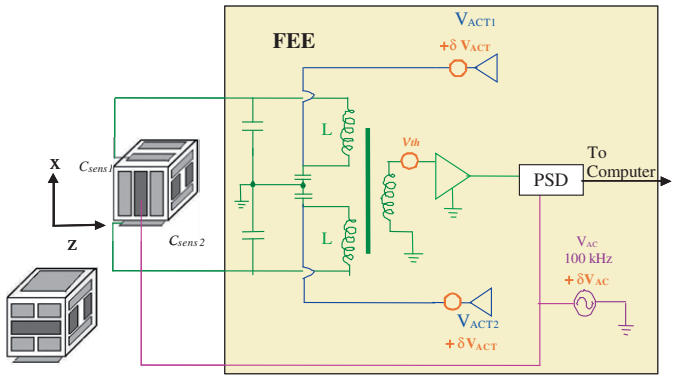


Figure 2.3: Schematic of the TM electrode configuration, simplified to a single channel (electrode pair along x-axis) connected to its FEE (including actuation). The dark coloured electrodes inject a 100 kHz ac bias, while the lighter coloured ones are those used for sensing and actuation. The TM motion along the x-axis causes an unbalance of the bridge, which is detected by the pre-amplifier. The TM is biased to a voltage $V_M \simeq 0.6$ V with a signal that is also the reference for the phase sensitive detector (abbreviated as PSD here) at the output of the pre-amp. Signals are A/D converted and read by an on board computer. The actuation force signals, calculated numerically according to control laws, act through a modulation of a numerically synthesized ac drive signal V_{ACT} (at audio frequencies). The force signal is finally D/A converted and applied to the electrodes. In the lower left we see the alternative electrode configuration (two electrodes pre face of the TM). The quantities denoted with δ reflect dominant noise sources. Credit [7].

- **Optical Metrology System (OMS)**: responsible for the most sensitive scientific measurements, the OMS includes the Optical Bench (OB), the Reference Laser Unit (RLU), the laser modulator and the phasemeter. The measurement scheme goes as follows:

  The laser source (RLU) is a $Nd:YAG$, 35 mW transmitting at 1064 nm, non-planar ring oscillator [12] directed via optical fibres to the laser modulator. The

modulator generates two beams with 1.2 kHz difference to be fed into the OB[1] interferometer. The four interferometers developed for the LTP [13, 14, 15, 16] are designed to perform –in combination– the two sensitive measurements of displacement.

For the LTP, the non-polarising Mach-Zehnder interferometer was chosen to be the most suitable [13, 17]. This decision was reached after a study of different types of interferometry, many of them using polarising components. It was shown that the polarising components are susceptible to temperature fluctuations, and results in the laboratory were not easily reproducible. For a Mach-Zehnder interferometer set-up, the measurement is the photocurrent produced by the beat note signal at the frequency difference between two interfering beams

$$I(t) = A(1 - c\cos(2\pi f_{het} t + \phi_{int}(t))), \tag{2.3}$$

where f_{het} the heterodyne frequency, I is the measured photocurrent, A the average photocurrent of the heterodyne signal, c the interferometric contrast and ϕ_{int} the interferometric phase given by

$$\phi_{int} = \frac{2\pi}{\lambda}(\ell_1 - \ell_2). \tag{2.4}$$

Here, λ is the laser wavelength and ℓ_1 and ℓ_2 are the length of the interferometer arms. Any displacement $\delta\ell$ of the optical components (the TMs in the case of the LTP) will cause a change in the arm-length and thus in the interferometric phase ϕ_{int}, as in eq. (1.7) of the first chapter. A mrad change in the phase of the 1 kHz heterodyne signal corresponds to a position change of the TM in the sub-nanometre range.

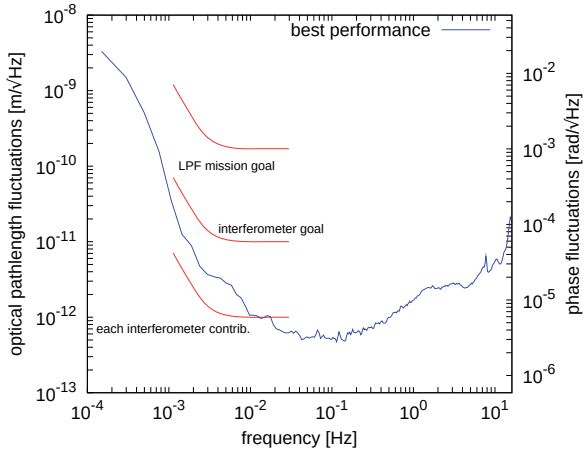


Figure 2.4: Sensitivity of the differential interferometer measurement. Credit: [18].

1 The OB is made mainly by Zerodur ceramic glass of dimensions $200 \times 212 \times 22.5$ mm. The primary material of the OB elements, like the mirrors and beam splitters is fused silica.

The OMS includes in total, four interferometers. The principal scientific measurement is performed by the $x_2 - x_1$ (or simply x_{12}) interferometer, which monitors the distance between the two TMs with a desired sensitivity of the order of ~ 0.1 nm (see figure 2.4). The second interferometer x_1 measures the distance of the first TM and the OB, and is more noisy than the x_{12} due to the pollution of the signal originating from the noisy thrusters. The third *Reference* interferometer is a rigid equal arm interferometer which provides the system noise floor, and is used to stabilise the fibre pathlengths via the Optical Pathlength Difference actuator (OPD). Finally, the *Frequency* Mach-Zehnder unequal-arm interferometer is used for corrections of laser frequency fluctuations.

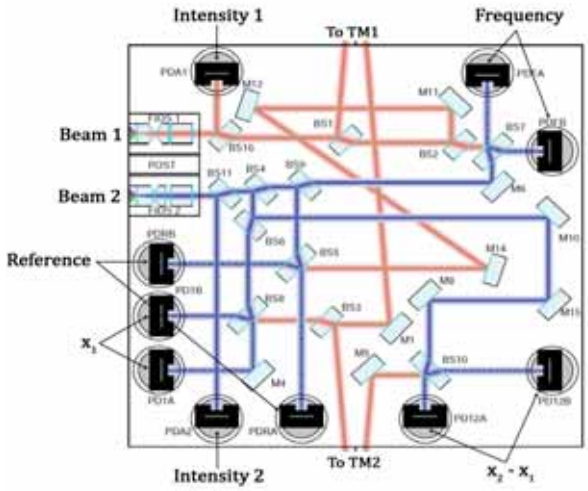


Figure 2.5: Schematics of the Optical Bench. See text for details. Credit: [4].

For each interferometer read-out, two quadrant photodiodes are utilised. The quadrant photodiodes can also be used to determine the angular rotation of the TMs. This output is transmitted to the *Phasemeter* [14]. The phasemeter receives the current measured from each photodiode, converts to voltage, and digitises it to 0.1 kHz. For each channel, the single bin Fourier transform at heterodyne frequency is applied. The resulting complex data vectors and DC values are sent to the on-board computer (see DMU below) for downsampling again to 10 Hz and further processing. There, all the measurements are combined to produce the final longitudinal measurement, Ψ, and two angular ones, ϕ and η, for the TMs. The calculation of the angular rotation along ϕ and η is twofold; (1) via Differential Wavefront Sensing (DWS), where the DC signals from each quadrant is combined to estimate the position of the beam on the photodiode, and (2) via the phase difference with respect the heterodyne frequency for each quadrant of the photodiode. For the latter, the position error of the beam is smaller, but the range is reduced in comparison to the first method.

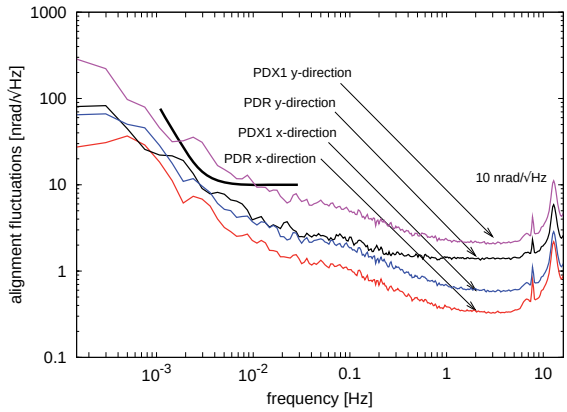


Figure 2.6: Sensitivity of the interferometer TM rotation from laboratory tests. Credit: [18].

- **Data and Diagnostics Subsystem (DDS)**: is a collection of sensors and actuators (Diagnostics) focusing into monitoring various disturbances inside the LTP. The sub-system completes the Data Management Unit (DMU), the on-board computer. In particular, the subsystems of the DDS are [19]:

  1. The *Magnetometers*: There are four *fluxgate* magnetometers placed in the LTP, surrounding the TMs. They constantly monitor the magnetic field originating from different known and unknown sources inside the satellite. The planned analysis of the data-set acquired, includes the extrapolation of the magnetic field to the position of the TMs, and the estimation of the magnetic force noise to the total noise budget. In combination with the magnetic coils, the magnetic properties of each TM are going to be estimated. More details can be found in [20, 21, 22].
  2. The *Magnetic Coils*: Two magnetic coils are positioned just outside the vacuum enclosure along the x-axis (defined by the line joining the two TMs). As already stated above, during the magnetic dedicated experiments, the coils will apply oscillating magnetic field along the x-axis, with the aim of estimating the magnetic moment $\vec{m} \equiv \{m_x, m_y, m_z\}$ and magnetic susceptibility χ_m of the TMs.
  3. The *Temperature Sensors*: Temperature monitoring is crucial for the outcome of the mission since thermal gradients may induce excessive force noise on the TMs. Consequently, 23 sensors have been placed around the LTP [23, 19].
  4. The *Heaters*: There are 14 heaters placed on the EH, the Optical Window and the suspension struts of the LTP. Their functionality is to heat the environment in proper constant levels of temperature to later estimate the thermal noise contribution to the overall differential acceleration noise. Secondly, the data analysis will be focused on the disentanglement of the various thermal effects inside the IS; namely the *radiation pressure*, the *asymmetric outgassing*, and the *radiometer effect*.
  5. The *Radiation Monitor*: It is essentially a particle counter to track the rate of charged particles bombarding the space-craft. Although the LTP is shielded

from cosmic events, some energetic particles ($E \succeq 100\ MeV$) will penetrate the outer material layers and add charge to the TMs. The radiation monitor is mainly comprised by two scintillator plates forming a telescope. Coincidence events (particles that pass through both surfaces) are recorded, together with the energy deposited. For more details see [24, 25].

6. The *Data Management Unit* (DMU): It operates as a module of the LTP On-Board Computer (OBC) and is designed to control the DDS and handle the interferometer data streams. The OBC, in-turn interfaces with the SC and LTP sub-systems. More information can be found in [25, 26].

- μ-*Newton Trusters*: During the first phases of the LPF implementation, the FEEP technology for the thrusters was considered [27]. The idea behind FEEP-type thrusters is based on the acceleration of ions by means of electrostatic force in high vacuum. But due to delays in the qualification schedule, the concept was abandoned in favour of the *Cold Gas* thrusters design [2], that have been developed by Thales Alenia Space - Italy (TAS-I). The Cold Gas thrusters concept is based on a very fine control of the gas flow emitted (N_2), and controlled by a closed loop system. Measuring the mass flow rate is also possible, with piezo-electric actuation used to move the thruster valve.

  The selection and integration process for the thrusters was relatively short, since they are already verified for the GAIA mission [28], that is already being launched and taking scientific measurements.

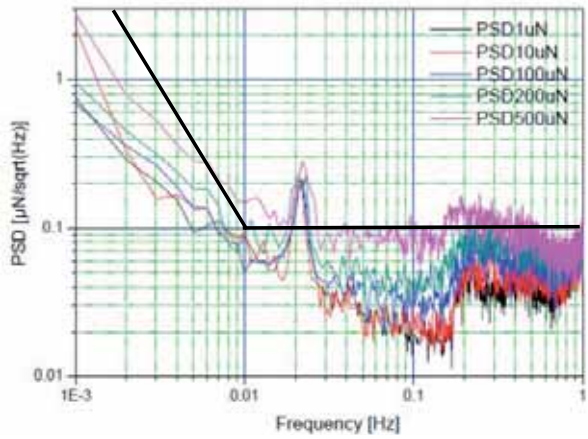


Figure 2.7: Noise performance of the Cold Gas thrusters, in different levels of applied force.

- *Drag-Free Attitude Control System* (DFACS): Again, one of the most important components of the experiment, the DFACS is the computer calculating and commanding forces and torques to the three bodies of the system (TMs and SC) via control laws. The controller and its modes of operation is discussed shortly in the following section.

2.3 OPERATION MODES

As already mentioned above, the TMs position is measured by the OMS and the IS and the attitude of the SC (three more Degrees-of-Freedom (DOF)) is sensed by the star-tracker. In overall, the DFACS have to control a sum of 15 DOF to perform the following tasks:

1. to achieve robust control of the LPF SC.
2. to stabilise its dynamics and give the best performance in assisted free-fall.
3. to keep the spacecraft panels pointed to the sun, and the communication antenna pointed to Earth.

To perform successfully the above, a great effort has been put in the design of the DFACS to meet the noise requirements. In fact, the controllers maintain a detailed balance between the selected drag-free DOF and their measuring bandwidth. The design itself is based on a feedback interconnection that yields a system of decoupled simple Single-Input-Single-Output (SISO) systems [29, 30]. The controllers have a number of operation modes, most of them switched on during station-keeping, TM release, or for smoother transition between modes. Below we list the most important scientific modes, and their basic operation principles.

- *Science Mode 1*: Operating in this mode (often abbreviated as M3), the TM2 is controlled along the x-axis to follow TM1. Then, the sensitive differential measurement is calculated as [31, 32, 33]

$$\begin{aligned} x_{12} &= S_{sus} G \Big[-K_{sus,x} n_{x_{12}} + \tfrac{(F_{TM1} - F_{TM2})}{m_{TM}} \\ &+ (\omega_1^2 - \omega_2^2) x_1 + \omega_2^2 (\delta x_2 - \delta x_1) \Big], \end{aligned} \tag{2.5}$$

  where $(F_{TM1} - F_{TM2})/m_{TM}$ is the differential force per unit mass, $(\omega_1^2 - \omega_2^2)x_1$ is the acceleration due to differential stiffness, S_{sus}, G and K_{sus} denote transfer functions of the system and $\omega_2^2(\delta x_2 - \delta x_1)$ represents the coupling effect due to elastic distortion along the sensitive axis. The last effect, can be neglected at a first approximation. The ω terms denote the parasitic spring-like stiffnesses to the TMs. By examining eq. (2.5), and if we discard cross-talk effects, we can distinguish the main contributions to the overall residual noise. This may become more clear if we re-write a variation of eq. (2.5) in the acceleration domain. That way, terms representing transfer functions vanish and the contributions to the dynamics become more apparent. A simple realisation of the differential acceleration between the two TMs can be written as

$$\alpha_{12} = \left[\frac{d^2}{dt} + \omega_2^2 \right] x_{12} + (\omega_1^2 - \omega_2^2) x_1 - G_2 \frac{F_{TM2}}{m_{TM}} + G_1 \frac{F_{TM1}}{m_{TM}}, \tag{2.6}$$

  where G_1 and G_2 are the actuator gains and here we assume

$$\begin{aligned} x_1 &= o_1(t - \tau), \\ x_{12} &= o_{12}(t - \tau) + \delta_{12} o_1(t - \tau), \end{aligned} \tag{2.7}$$

where as previously stated, the o_1 and o_{12} are the two interferometer readings respectively, the τ is a system delay and $\delta_{12} o_1(t-\tau)$ some interferometric cross-coupling. The dynamics and modelling of the three bodies of the system will be revisited in chapter 4. From eq. (2.6), one of the important contributions to the total residual acceleration is the total differential force acting on the TMs. The term $(\omega_1^2 - \omega_2^2)x_1$ shows "pollution" of the first interferometer channel into the differential one, an effect that depends on the material of the TMs and the configuration of the IS.

The Science Mode 1 is the mode where most of the experiments analysed and studied in this thesis are performed. All planned parameter estimation experiments are going to be performed during this system configuration and its variations. More details in Chapters 4, where the system identification along the sensitive axis is studied, 5, where the coupling effects between different DOF is investigated, and 6, where we perform model selection on the aforementioned experiments. A summary of Science mode 1 and the measurement channels are presented in table 2.1

Table 2.1: Summary of Science mode 1 and its sub-modes. The 1.1 case is a transition sub-mode before reaching the 1.2-optical. The 1.2-backup mode will be used in case of malfunction or high unexpected noise for the optical read-outs. The different DOF controlled can be seen in figure B.1.

Mode		**Measurements**	
		IS	**OMS**
Science Mode 1.1		$x1$, $y1$, $z1$, $\theta1$, $\eta1$, $\phi1$, $x2-x1$, $y2$, $z2$, $\theta2$, $\eta2$, $\phi12$	*none*
Science Mode 1.2	▷ All optical	$y1$, $z1$, $\theta1$, $y2$, $z2$, $\theta2$	$x1$, $\eta1$, $\phi1$, x_{12}, $\eta2$, $\phi2$
	▷ Backup	$y1$, $z1$, $\theta1$, $y2$, $z2$, $\theta2$, $x1$, $\eta1$, $\phi1$, $\eta2$, $\phi2$	x_{12}

Name of Loop	**DOF**
Drag-Free	$x1$, $y1$, $z1$, $\theta1$, $y2$, $z2$
Suspension	$x2$, $\eta1$, $\phi1$, $\eta2$, $\phi2$, $\theta2$
Attitude	η_{SC}, θ_{SC}, ϕ_{SC}

- *Normal modes*: They are used as transition modes to the science phase of the experiment. During the Normal Mode (and its sub-modes) the measurements are performed only from the IS readings.
- *Accelerometer modes*: The Acceleration Mode and its multiple variations are designed to bring the SC smoothly into full operation. It initially monitors TM1 and TM2 after their release with the IS and performs the first measurements/tests with the OMS.

2.4 LTPDA TOOLBOX

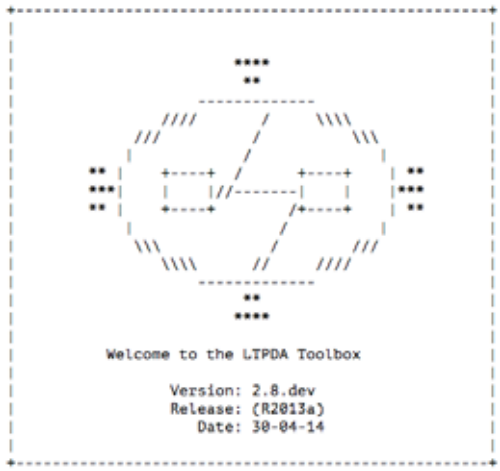


Figure 2.8: The LTPDA logo. The LTPDA toolbox is an Open License product, and the latest release can be downloaded from http://www.lisa.aei-hannover.de/ltpda/index.html.

It is quite evident that the data analysis for the LPF mission poses some particularities; The analysis has to be performed on-line, meaning that the days results must be finalised by the end of the day after, to allow possible re-planning of experiments and investigations. The decision making will take place in the Science and Technology Operations Centre (STOC), which is the same place where the frond-line analysis will be performed. Also, all derived results from the LPF mission must have a long shelf-life, up to the commissioning of the eLISA mission [34, 35].

All the above, led to the creation of the LPF data analysis software, the Matlab™ [36] based LTPDA toolbox. The analysis supported with the LTPDA toolbox is object oriented, which implies that all the processing chain, together with the data, is stored in the so-called *Analysis Objects (AOs)*. It is specialised mostly in all kinds of spectral and time-series analysis, while it contains a long list of LTP models in various formats, parameter estimation algorithms, plotting tools and repository related machinery. The set of requirements for the LTPDA toolbox are listed as follows.

1. The toolbox must provide a flexible and robust data analysis environment for all possible needs of the LTP planned experiments.
2. All methods and functions must come together with the appropriate help and usage information.
3. A graphical user interface must be present, for the non-programming experts.
4. The history of the processing chain must be automatically captured. This ensures that a given result can be reproducible by any member of the collaboration with a single line command.

The LPF system can be modelled in various ways in the LTPDA environment but all can be arranged into two categories; the analytical, and the state-space formats. The

latter, uses the very detailed State-Space Models (SSMs), where at the same time they can be used as a LPF simulator [37, 38].

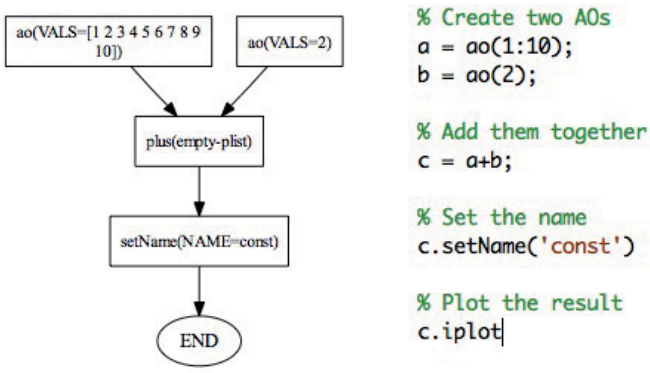


Figure 2.9: Example of the usage of the LTPDA language. Credit: http://www.lisa.aei-hannover.de/ltpda/

The SSM environment allows to combine different versions of the various LTP submodules to generate a "custom-made" simulator. In addition, the different noise sources of each submodule can be manipulated and updated to the latest laboratory results. Other modelling approximations are used when there is a need for a more direct approach to the dynamics of the three bodies of the system. More information in the system is available in Chapter 4.

One important feature, is that the toolbox contains all the mission data analysis pipelines. A pipeline is a form of "one-button-analysis" specially designed to be used during operations. The LTPDA pipeline structure begins with the downloading of the data from repositories, it continues with a pre-designed step-by-step analysis, to end with the submission of the results to dedicated repositories again.

It is worth to mention, that the entire work presented in this thesis, has been developed to be a part of the LTPDA toolbox, coming together with the necessary HTML help. The Markov Chain Monte Carlo class (Appendix A) and the designed system identification pipeline (Appendix D) have already been fully tested and put into stress during the recent LPF training exercises.

2.5 STATUS

Being almost one year prior to the official launch date, the different LPF modules are already integrated and tested excessively in various test campaigns. The more important one was the On-Station Thermal Test (OSTT) campaign in late 2011, where most of the sub-systems where already integrated to the spacecraft and where put into a vacuum tank simulating space conditions [39, 40]. The main purpose of this end-to-end test was to verify the satellite thermal stability and thermal properties, but for LPF, it provided the opportunity to further test the various sub-systems when integrated.

The campaign took place at the Industrieanlagen-Betriebsgesellschaft (IABG) mbH space simulator, in Ottobrun (Germany), (figure 2.10). Inside the vacuum tank, the

Figure 2.10: *Top*: IABG mbH space vacuum chamber with the LPF during the OSTT campaign. Credit: Astrium UK. *Bottom left*: The assembly of the TOQM thermal model. Credit: [40]. *Bottom right*: The LPF SC on vibration tests.

nominal pressure value was below 10^{-4} Pa and an array of high power lamps were resembling the suns' radiation in L1, where the LPF will operate in science mode. Since the EH modules were not installed during the tests, an optical-thermal system replica was used in its place, with two movable piezo-electric driven mirrors mimicking the TMs dynamics. Together with the OMS flight model, this set-up was forming the so-called TOQM. The thermal noise performance of this test can be seen in figure 2.11.

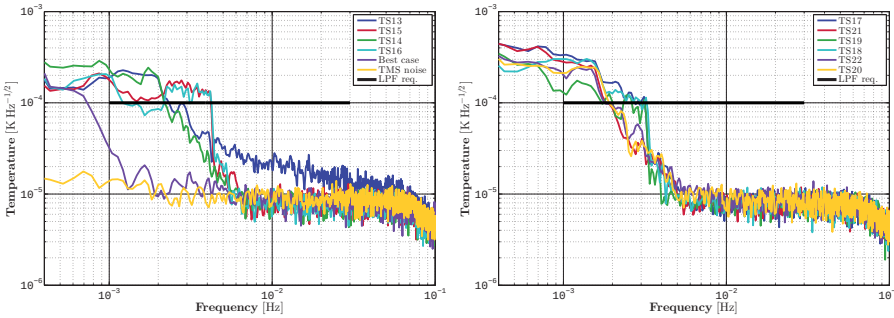


Figure 2.11: (This caption will change) *Left*: PSD for the optical bench temperature sensors. *Right*: PSD of the different Strut temperature sensors. Credit and more details: [39].

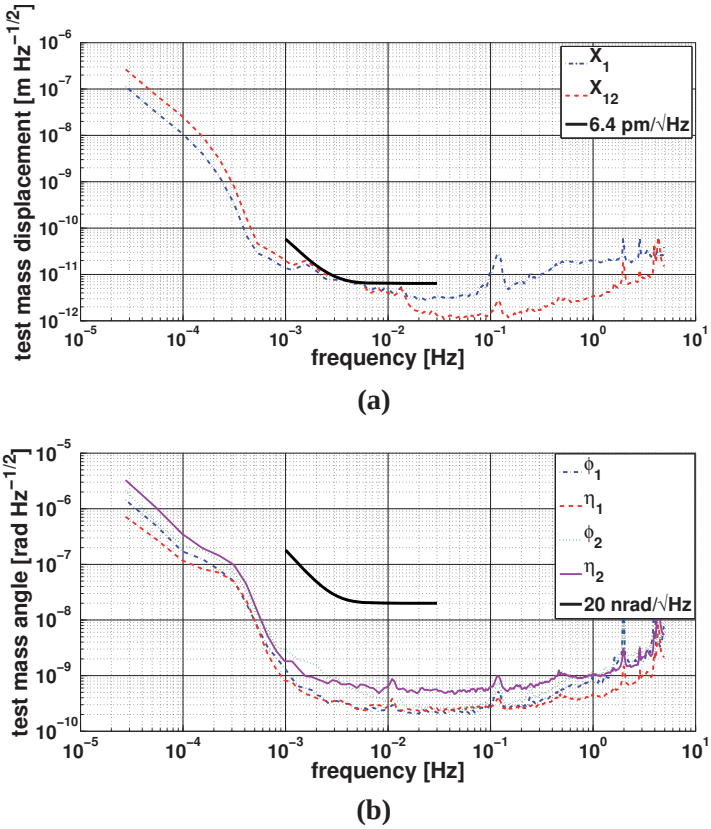


Figure 2.12: Performance measurement of the OMS during the OSTT campaign. *(a)* longitudinal measurement, *(b)* angular measurement via DWS. Credit and more details: [40].

Up to now, there have been a series of tests on the flight hardware and software, that put the readiness of the mission to stress. The most important of them being

Day	N	N+1	N+2	N+3	N+4	N+5	N+6	N+7	N+8	N+9
Team 1	Online	Offline	Online	Offline	Travel	Off Duty	Off Duty	Travel	Online	Offline
Team 2		Online	Offline	Online	Offline	Travel	Off Duty	Off Duty	Travel	Online
Team 3					Online	Offline	Online	Offline	Travel	Off Duty
Team 4						Online	Offline	Online	Offline	Travel

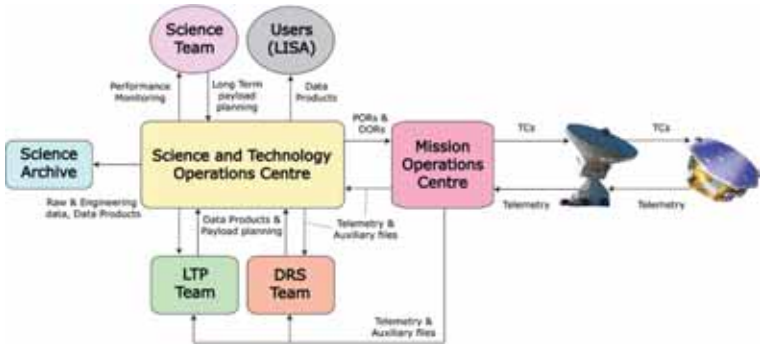


Figure 2.13: *Top:* Example schedule table during operations. Each team comprised by a handful of scientists and engineers will be online every other day, with time provided to finalise any pending analysis from their on-duty day. Credit: LISA Symposium X invited talk by M. Hewitson. *Bottom:* Schematics of the Science Operations Scheme and data flow of the ground segment. Credit: [2].

the laser assembly test campaign and the OMS system tests [41]. The various relevant subsystems, like the RLU, the DMU, the Optical Bench and the Phasemeter, were all delivered and assembled. The next system level test was performed again in 2011 (the so-called sine test), where the SC was undergone through a series of vibrations, to evaluate its overall stability.

In overall, the LPF SC was proven to be in excellent state, and the various sub-systems where verified to be operating nominally. The overall integration process of the SC is essentially completed, and the LPF is in the final stages before being transported to French Guiana. Some of the results from the hardware test campaigns can be seen in figures 2.10, and 2.12.

Being so close to launch and operations, the LPF scientists and engineers team has already started to organise the data acquisition and analysis scheme. The ninety days of LPFs lifetime is packed with experiments spanning from several hours to a few days. A plan for personnel shifts has already been proposed and tested in the so-called STOC exercises [42, 43, 44]. The schematics of the daily schedules and data flow can be seen in figure 2.13.

In this study, we will use the data-sets generated for these simulations. In particular, we focus on the data from STOC simulations 2, 3, and 4. Finally, The LPF collaboration

members have been already divided into the so-called Work Packages. Each Work Package (WP) is dedicated to a specific group of experiments, and their members are chosen due to their specialisation in the particular field.

[1] M Armano, M Benedetti, J Bogenstahl, et al. Lisa pathfinder: the experiment and the route to lisa. *Classical and Quantum Gravity*, 26(9):094001, 2009. URL: http://stacks.iop.org/0264-9381/26/i=9/a=094001. 21

[2] Paul W. McNamara. The LISA Pathfinder Mission. *Int.J.Mod.Phys.*, D22:1341001, 2013. doi:10.1142/S0218271813410010. 21, 29, 36

[3] S. Vitale, P. Bender, A. Brillet, et al. LISA and its in-flight test precursor SMART-2. *Nucl.Phys.Proc.Suppl.*, 110:209–216, 2002. 21

[4] Paul McNamara and Giuseppe Racca. Introduction to LISA Pathfinder, LISA-LPF-RP-0002, January 2012. URL: http://sci.esa.int/lisa-pathfinder/45819-introduction-to-lisa-pathfinder/. 21, 27

[5] Stefano Vitale. Space-borne Gravitational Wave Observatories. *Gen.Rel.Grav.*, 46:1730, 2014. arXiv:1404.3136, doi:10.1007/s10714-014-1730-2. 21

[6] The LISA Pathfinder Science Working Team. The science case for LISA Pathfinder, ESA-SCI(2007)1, January 2007. URL: http://sci.esa.int/lisa-pathfinder/40572-the-science-case-for-lisa-pathfinder/. 21, 22

[7] R Dolesi, D Bortoluzzi, P Bosetti, et al. Gravitational sensor for lisa and its technology demonstration mission. *Classical and Quantum Gravity*, 20(10):S99, 2003. URL: http://stacks.iop.org/0264-9381/20/i=10/a=312. 23, 25

[8] Giuseppe D. Racca and Paul W. McNamara. The LISA Pathfinder mission: Tracing Einstein's geodesics in space. *Space Sci.Rev.*, 151:159–181, 2010. doi:10.1007/s11214-009-9602-x. 24

[9] P J Wass, H M AraG'Ijo, D N A Shaul, and T J Sumner. Test-mass charging simulations for the lisa pathfinder mission. *Classical and Quantum Gravity*, 22(10):S311, 2005. URL: http://stacks.iop.org/0264-9381/22/i=10/a=023. 25

[10] P. J. Wass, L. Carbone, A. Cavalleri, et al. Testing of the UV discharge system for LISA Pathfinder. In S. M. Merkovitz and J. C. Livas, editors, *Laser Interferometer Space Antenna: 6th International LISA Symposium*, volume 873 of *American Institute of Physics Conference Series*, pages 220–224, November 2006. doi:10.1063/1.2405047. 25

[11] M O Schulte, D N A Shaul, D Hollington, et al. Inertial sensor surface properties for lisa pathfinder and their effect on test mass discharging. *Classical and Quantum Gravity*, 26(9):094008, 2009. URL: http://stacks.iop.org/0264-9381/26/i=9/a=094008. 25

[12] R. Lange and B. Smutny. Bpsk laser communication terminals to be verified in space. In *Military Communications Conference, 2004. MILCOM 2004. 2004 IEEE*, volume 1, pages 441–444 Vol. 1, Oct 2004. doi:10.1109/MILCOM.2004.1493308. 25

[13] G. Heinzel, C. Braxmaier, R. Schilling, et al. Interferometry for the LISA technology package (LTP) aboard SMART-2. *Class.Quant.Grav.*, 20:S153–S161, 2003. doi:10.1088/0264-9381/20/10/318. 26

[14] G. Heinzel, V. Wand, A. Garcia, et al. The LTP interferometer and phasemeter. *Class.Quant.Grav.*, 21:S581–S587, 2004. doi:10.1088/0264-9381/21/5/029. 26, 27

[15] Frank Steier, Felipe Guzman Cervantes, Antonio F. Garcia Marin, et al. The end-to-end testbed of the optical metrology system on-board LISA Pathfinder. *Class.Quant.Grav.*, 26:094010, 2009. arXiv:1203.4842, doi:10.1088/0264-9381/26/9/094010. 26

[16] H Audley, K Danzmann, A Carcía Marín, et al. The lisa pathfinder interferometry-hardware and system testing. *Classical and Quantum Gravity*, 28(9):094003, 2011. URL: http://stacks.iop.org/0264-9381/28/i=9/a=094003. 26

[17] Vinzenz Wand. *Interferometry at low frequencies: Optical phase measurement for LISA and LISA Pathfinder*. PhD thesis, Gottfried Wilhelm Leibniz Universität Hannover, 2007. 26

[18] A F Carcía Marín, G Heinzel, and K Danzmann. Interferometry for lisa and lisa pathfinder. *Journal of Physics: Conference Series*, 66(1):012052, 2007. URL: http://stacks.iop.org/1742-6596/66/i=1/a=012052. 26, 28

[19] Miquel Nofrarias. *Thermal Diagnostics in the LISA Technology Package Experiment*. PhD thesis, Universitat de Barcelona, 2007. 28

[20] J. Sanjuan, A. Lobo, P. Cañizares, et al. The Lisa Technology Package (LTP) Diagnostics In Lisa Pathfinder. In *Bulletin of the American Astronomical Society*, volume 42 of *Bulletin of the American Astronomical Society*, pages 576–+, jan 2010. 28

[21] M. Diaz-Aguiló, E. García-Berro, A. Lobo, N. Mateos, and J. Sanjuan. The magnetic diagnostics subsystem of the lisa technology package. *Journal of Physics Conference Series*, 228(1):012038–+, may 2010. doi:10.1088/1742-6596/228/1/012038. 28

[22] Marc Diaz-Aguiló, Enrique García-Berro, and Alberto Lobo. Inflight magnetic characterization of the test masses onboard lisa pathfinder. *Phys. Rev. D*, 85:042004, Feb 2012. URL: http://link.aps.org/doi/10.1103/PhysRevD.85.042004, doi:10.1103/PhysRevD.85.042004. 28

[23] Josep Sanjuan Munoz. *Development and validation of the thermal diagnostics instrumentation in LISA Pathfinder*. PhD thesis, Universitat Politecnica de Catalunya, 2009. 28

[24] I Mateos, M Diaz-Aguiló, F Gibert, et al. LISA PathFinder radiation monitor proton irradiation test results. *Journal of Physics: Conference Series*, 363(1):012050, 2012. URL: http://stacks.iop.org/1742-6596/363/i=1/a=012050. 29

[25] P. Cañizares, M. Chmeissani, A. Conchillo, et al. The LISA Pathfinder DMU and Radiation Monitor. *Classical and Quantum Gravity*, 28(9):094004–+, may 2011. arXiv:1009.5651, doi:10.1088/0264-9381/28/9/094004. 29

[26] L. Gesa, V. Martin, A. Conchillo, A. Lobo, and I. Lloro. The LISA PathFinder DMU Software, a Global Overview. In G. Auger, P. Binétruy, and E. Plagnol, editors, *Astronomical Society of the Pacific Conference Series*, volume 467 of *Astronomical Society of the Pacific Conference Series*, page 277, jan 2013. 29

[27] S. Rocca, C. Menon, and D. Nicolini. Feep micro-thrust balance characterization and testing. *Meas. Sci. Technol*, 17:711–718, 2006. URL: http://www.iop.org/EJ/abstract/0957-0233/17/4/016. 29

[28] L. Eyer, B. Holl, D. Pourbaix, et al. The Gaia mission. 2013. arXiv:1303.0303. 29

[29] Walter Fichter, Alexander Schleicher, and Stefano Vitale. Drag-free control design with cubic test masses. In Hansjorg Dittus, Claus Lammerzahl, and SlavaG. Turyshev, editors, *Lasers, Clocks and Drag-Free Control*, volume 349 of *Astrophysics and Space Science Library*, pages 361–378. Springer Berlin Heidelberg, 2008. URL: http://dx.doi.org/10.1007/978-3-540-34377-6_17, doi:10.1007/978-3-540-34377-6_17. 30

[30] Walter Fichter, Peter Gath, Stefano Vitale, and Daniele Bortoluzzi. Lisa pathfinder drag-free control and system implications. *Classical and Quantum Gravity*, 22(10):S139, 2005. URL: http://stacks.iop.org/0264-9381/22/i=10/a=002. 30

[31] Stefano Vitale et al. The LISA Technology Package on board SMART-2, The LTP Definition Document, Unitn-Int 10-2002/Rel. 1.3, October 2002. URL: http://www.rssd.esa.int/SP/LISAPATHFINDER/docs/Top_level_documents/LTP_DD.pdf. 30

[32] N. Brandt et al. Experiment performance budget, M3 Optical, FEEP. Technical Report S2-ASD-RP-3036, ASD, April 2010. 30

[33] A. Schleicher and R. Saage. DFACS General Design. Technical Report S2-ASD-TN-2001, ASD, August 2013. 30

[34] F Antonucci, M Armano, H Audley, et al. LISA Pathfinder data analysis. *Classical and Quantum Gravity*, 28(9):094006, May 2011. URL: http://stacks.iop.org/0264-9381/28/i=9/a=094006?key=crossref.41fc94b94ce025596175f5fe463e5a86, doi:10.1088/0264-9381/28/9/094006. 32

[35] Martin Hewitson, M. Armano, M. Benedetti, et al. Data analysis for the LISA Technology Package. *Class.Quant.Grav.*, 26:094003, 2009. doi:10.1088/0264-9381/26/9/094003. 32

[36] MATLAB. *version 8.30 (R2014a)*. The MathWorks Inc., Natick, Massachusetts, 2014. 32

[37] M. Nofrarias, F. Antonucci, M. Armano, et al. State space modelling and data analysis exercises in lisa pathfinder. In G. Auger, P. Binétruy, and E. Plagnol, editors, *Astronomical Society of the Pacific Conference Series*, volume 467 of *Astronomical Society of the Pacific Conference Series*, page 161, jan 2013. 33

[38] M Hewitson, M Diaz-Aguiló, and A Grynagier. A linear MIMO model of LPF implemented in LTPDA. Technical Report S2-AEI-TN-3069, 2011. 33

[39] Ferran Gibert, Miquel Nofrarias, Nikolaos Karnesis, et al. Thermo-elastic induced phase noise in the LISA Pathfinder spacecraft. 2014. arXiv:1405.5442. 33, 35

[40] F. G. Cervantes, R. Flatscher, D. Gerardi, et al. Lisa technology package flight hardware test campaign. In G. Auger, P. Binetruy, and E. Plagnol, editors, *Astronomical Society of the Pacific Conference Series*, volume 467 of *Astronomical Society of the Pacific Conference Series*, page 141, jan 2013. 33, 34, 35

[41] F Antonucci, M Armano, H Audley, et al. Lisa pathfinder: mission and status. *Classical and Quantum Gravity*, 28(9):094001, 2011. URL: http://stacks.iop.org/0264-9381/28/i=9/a=094001. 36

[42] M Armano. Plan for the STOC-LTP Simulation 4. Technical Report S2-ESAC-PL-5024, 2012. 36

[43] M Armano. Plan for the STOC-LTP Simulation 2. Technical Report S2-ESAC-PL-5025, 2012. 36

[44] M Armano. Report on the STOC-LTP Simulation 4. Technical Report S2-ESAC-RP-5020, 2012. 36

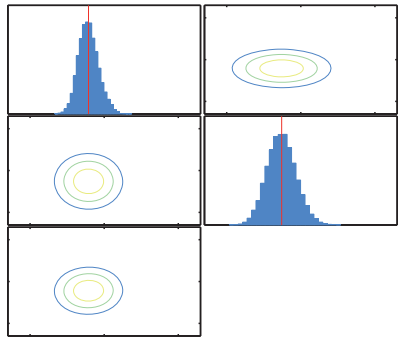
Part II

LISA PATHFINDER DATA ANALYSIS

"Bayesians address the question everyone is interested in, by using assumptions no-one believes",

and

"Frequentists use impeccable logic to deal with an issue of no interest to anyone".



3 BAYESIAN STATISTICS

This chapter serves as an introduction to basic Bayesian statistics principles, as we are going to use throughout the rest of the document. The basic concepts and tools for Bayesian Inference are going to be explained in a brief manner, while more can be found in classic textbooks or lecture notes [1, 2, 3, 4, 5, 6]. First, the fundamental ideas of Bayesian (and Frequentists, for the sake of completeness) approach is going to be summarised. Then, Bayesian statistical methods like the Markov Chain Monte Carlo, and the Fisher Information Matrix formalism, will be shown in more detail. These techniques and their advancements are relevant for this thesis and for LPF pipeline analysis.

But at this point, it would be convenient to introduce the notation used in this chapter. From the Frequentist point of view, the probability is is seen as the relative frequency of experiments conducted under identical conditions, usually denoted as $\Pr(\cdot)$, while the Bayesian case (degree of belief) as $\pi(\cdot)$. To keep things simple, in this document we will follow the Bayesian notation, with additional explanation if required.

As most detection and system identification problems, for LPF we make the assumption that the observed data-sets y can be accurately described by a model $h(\vec{\theta})$ of the system determined by a set of parameters $\vec{\theta}$, plus the noise

$$y = h(\vec{\theta}) + n. \tag{3.1}$$

The noise, at a first level, is assumed to be uncorrelated, stationary with zero-mean. Then, if we also assume Gaussianity, the random noise process determines a natural inner product $(\cdot|\cdot)$ and the associated norm on the vector space of measurements is given by [7, 8]

$$(a|b) = 2\int_0^\infty df\left[\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)\right]/\tilde{S}_n(f)\,, \tag{3.2}$$

where the tilde $(\tilde{})$ denotes the operations in frequency domain, and the $(^*)$ the complex conjugation. S_n is the one-sided PSD of the noise. The following equation

$$\left\langle (n|a)(n|b)\right\rangle = (a|b), \tag{3.3}$$

also holds for any given real vectors a and b. Here $\langle\cdot\rangle$ is the ensamble average over all possible realisations of the noise n. Then, it can be proven [7, 9], that the probability for the measured noise time-series $n(t)$ to take a specific value $n_0(t)$ (or simply n_0), can be expressed as

$$\pi(n = n_0) \propto \exp\left(-\frac{1}{2}(n_0|n_0)\right). \tag{3.4}$$

If we combine eqs. 3.1 and 3.4, we make the realisation that the most probable parameters of the model of the system are lying on the maximum of the $\exp\left[-1/2(y-h|y-h)\right]$

surface. This quantity later will be defined as the likelihood function of the system parameters.

3.1 THEORETICAL BACKGROUND: THE FREQUENTIST APPROACH

The main attribute of the Frequentist approach (often called as "orthodox" or "sampling theory"), is that it treats the variables as *the relative frequency of occurrence of a given event in sequential repetitions of the same experiment*[1]. That means that the probability is regarded as a long-run relative frequency [4]. In particular, the the so-called *sampling distribution of the statistic* that depends on a given parameter set $\vec{\theta}$ is defined, where it can be calculated for each possible random sample. Then, this statistic is written as

$$\pi(y|\vec{\theta}) \tag{3.5}$$

and θ is treated as an unknown set of constants. This distribution measures how the statistic varies over all possible samples, given the unknown $\vec{\theta}$. From the Bayesian point of view, the Frequentist statistics performs inferences in the parameter space (the unobservable), based on a probability distribution in the sample space (the observable). In Bayesian statistics (see also the following Chapter 3.2), quite the opposite holds; the inference is performed in the parameter space based on a probability distribution in the parameter space, leading to a posterior distribution as

$$\pi(\vec{\theta}|y). \tag{3.6}$$

In the Sampling theory framework, most experiments are analysed using the the p-value and confidence intervals quantities. The p-value is one of the statistical significance tests, and in most cases has a threshold assigned to $p = 0.05$. It can be interpreted as the probability, given a null hypothesis for the probability distribution of the data, that the outcome would be as extreme as, or more extreme than, the observed outcome [10]. Additionally, a confidence interval for a parameter θ is the function $L = [\theta_{min}(y), \theta_{max}(y)]$ of the data y. This function calculates the confidence level of the parameter to lie in L. Usually this level is assigned to 95%.

At a first level, the Frequentist Inference is *point estimation*, which is in essence any real-valued statistic or function involving the sample data that is used to approximate the true parameters $\vec{\theta}_{true}$. Various approaches for this estimators are developed, the most important being the Maximum Likelihood Estimation (MLE)[2], where one tries to maximise the the statistic of eq. (3.5). This estimator $\hat{\theta}_{MLE}(y)$ can be simply written as

$$\hat{\theta}_{MLE}(y) = \underset{\vec{\theta}}{\operatorname{argmax}}\{\pi(y|\vec{\theta})\}, \tag{3.7}$$

where

$$\underset{x}{\operatorname{argmax}} f(x) = \{x|\forall y : f(y) \leqslant f(x)\}. \tag{3.8}$$

1 Its theoretical foundation lies on the Strong Law of Large Numbers.
2 The MLE was introduced by Sir R. A. Fisher in 1822

The MLE (or any other estimator in general) is classified as a good estimator if it is *unbiased*, so that $E[\hat{\vec{\theta}}(y)] = \vec{\theta}_{true}$ and efficient $\mathrm{var}[\hat{\vec{\theta}}(y)] = \mathrm{CRB}(\vec{\theta}_{true})$. The CRB denotes the Cramer-Rao Bound that will be discussed later in Section 3.3. The likelihood function, as already mentioned, can be written straightforwardly from eq. 3.1 and 3.4, as

$$\pi(y|\vec{\theta}) = C \times e^{-\frac{1}{2}\left(y - h(\vec{\theta})\middle| y - h(\vec{\theta})\right)} = C \times e^{-\chi^2/2}, \tag{3.9}$$

where

$$\chi^2 = \left(y - h(\vec{\theta})\middle| y - h(\vec{\theta})\right). \tag{3.10}$$

In many cases, instead of maximising $\pi(y|\vec{\theta})$, it is much more convenient to maximise the logarithm of the likelihood $-\frac{1}{2}\chi^2$.

For the error calculation in the Frequentist framework, it is assumed that the parameters of the given system are kept fixed (to the estimated set $\hat{\theta}_i$), while the noise $n(t)$ of the instrument is sampled from a fictitious probability distribution. Then the error is the fluctuations of a given parameter θ_i (we take the error of a single parameter for the sake of simplicity)

$$\mathrm{var}\hat{\theta} = \left\langle \left(\hat{\theta}(y) - \left\langle \hat{\theta}(y) \right\rangle \right)^2 \right\rangle \tag{3.11}$$

In practice, the errors of the system parameters can be retrieved by making use of the Fisher Information Matrix (see Section 3.3).

Another useful quantity is the SNR, defined crudely as the ratio of the signal over the background noise. It can be proven that the SNR is also a point estimator [7], here defined as

$$\frac{S}{N}[h] = \rho(\vec{\theta}) \equiv \frac{(h|y)}{\sqrt{(h|h)}} = \left(\hat{h}\middle|y\right), \tag{3.12}$$

where the hat (ˆ) denotes the normalised signal with respect to the scalar product of eq. (3.2), so that $(\hat{h}|\hat{h}) = 1$. The SNR $\rho(\vec{\theta})$ is a random variable with Gaussian PDF of unit variance, and it becomes optimal when the model h accurately describes the given system h_y. So if $h \simeq h_y$, from eq. (3.12) and (3.1)

$$\rho(\vec{\theta}) = \left(\hat{h}_y\middle|y\right) = \left(\hat{h}_y\middle|h_y + n\right) = \left(\hat{h}_y\middle|h_y\right) + \left(\hat{h}_y\middle|n\right) = \left(h_y\middle|h_y\right)^{1/2}. \tag{3.13}$$

Frequentist analysis is usually much easier to prepare because many things do not need to be specified, such as prior distributions, initial values for numerical approximation, and usually the likelihood function. Most frequentist methods have been standardised to "procedures" where less knowledge and programming are required, and they can be applied fast and straightforwardly in most parameter estimation cases.

3.2 THEORETICAL BACKGROUND: THE BAYESIAN APPROACH

The probability from the Bayesian point of view is seen a little differently than from the frequentist one described in the previous section. We have seen that in the frequentist approach, the probability of a given parameter is translated as the outcome of multiple (infinite) identical experiments. Also, the probabilities calculated are considered before the data acquisition, because they are based on all possible random samples, not the specific random sample we obtained.

On the other hand, in Bayesian statistics, the probability is treated as the assignment of a *degree of belief* for the given event, based on the evidence at hand. In this framework, one can add the level of knowledge for a given system, into the total available information to solve a particular problem. This last feature is very relevant for LPF, as the level of knowledge for the system will increase from day-to-day experiments. Another dissimilarity is that the unknown parameters are treated as random variables with an assigned distribution probability function and that the data-set used is the measured one. Frequentist inference uses both observed data and future data that is hypothetical. In reality, the difference between the two "schools" of statistics has foundation in the definition of probability and data treatment, but in the end, it's the nature of the problem to solve that defines the best strategy.

During mission operations, the measurements from LPF are going to be available in the telemetry as y time-series. If we assume a model $h(\vec{\theta})$, depending on the set of parameters $\vec{\theta}$[3], we can associate the *posterior probability distribution* with the *likelihood* function, the *prior* information and the *evidence* of the model in the prominent Bayes theorem as

$$\pi(\vec{\theta}|y) = \frac{\pi(y|\vec{\theta},\mathcal{M})p(\vec{\theta})}{\pi(y|\mathcal{M})}. \tag{3.14}$$

Here, $\pi(\vec{y}|\vec{\theta},\mathcal{M})$[4] is the likelihood of the parameters $\vec{\theta}$ over the data-set y, and $p(\vec{\theta})$ is the prior distributions of the parameters. The posterior $\pi(\vec{\theta}|\vec{y})$, represents our state of knowledge about the truth of the hypothesis in the light of the data. So, our prior knowledge is modified by the experimental measurements through the likelihood function, to finally yield the posterior probability. In a sense, Bayes' theorem encapsulates the process of learning [11].

The evidence, or *marginal likelihood* of the model $\pi(\vec{y}|\mathcal{M})$ is constant over the parameter space and can be calculated as

$$\pi(\vec{y}|\mathcal{M}) = \int \pi(\vec{\theta},\vec{y}|\mathcal{M})d\vec{\theta}. \tag{3.15}$$

The evidence of a given model, denotes the ability of this model to explain the observations. For parameter estimation problems it serves only as a normalisation constant, and can be omitted from eq. (3.14)

$$\pi(\vec{\theta}|y) \propto \pi(\vec{y}|\vec{\theta},\mathcal{M})p(\vec{\theta}), \tag{3.16}$$

3 Remember that the unknown parameters are treated as random variables. This reflects the lack of knowledge about the system of interest, like for example random instrumental noise.

4 The $\mathcal{M}$, denotes the given model.

but this will be explored in more detail in Chapter **??**. Here, it has to be noted that the terms likelihood and probability are not synonyms. The likelihood computes the dependance of the parameters of the model given the observed data. The prior $p(\vec{\theta})$, as already stated contains any prior belief for the parameters that we may have available beforehand, and the posterior distribution is the term that denotes the joint posterior knowledge of the estimated parameters. The single posterior Probability Density function (PDF) for each of the parameters $\vec{\theta} \equiv \{\theta_1, \theta_2, \cdots \theta_n\}$ is obtained by marginalising the joint posterior probability as

$$\pi(\vec{\theta}|y) = \int d\theta_1 \int d\theta_2 \cdots \int \frac{\pi(\vec{y}|\theta_n, \mathcal{M}) p(\theta_n)}{\pi(\vec{y}|\mathcal{M})} d\theta_n, \tag{3.17}$$

and the posterior mean for the parameters is calculated as [12]

$$\langle \theta_i \rangle_\pi = \frac{\int \theta_i \pi(\theta_i|\vec{y}) d\theta_i}{\int \pi(\vec{\theta}|\vec{y}) d\vec{\theta}}. \tag{3.18}$$

The quadratic moment is then

$$\langle \theta_i \theta_j \rangle_\pi = \frac{\int \theta_i \theta_j \pi(\theta_i, \theta_j|\vec{y}) d\theta_i d\theta_j}{\int \pi(\vec{\theta}|\vec{y}) d\vec{\theta}}, \tag{3.19}$$

with the $\langle \cdot \rangle_\pi$ denoting the integration over the posterior PDF. In the Bayesian case, the error is taken from a single experiment, by measuring the spread of the posterior distribution. So the variance of each parameter θ_i is taken as

$$\sigma_i^2 = \langle (\Delta\theta_i)^2 \rangle = \langle (\theta_i - \langle \theta_i \rangle_\pi)^2 \rangle_\pi = \langle \theta_i^2 \rangle_\pi - \langle \theta_i \rangle_\pi^2. \tag{3.20}$$

One of the advantages of Bayesian statistics, is that it has a straightforward way of dealing with nuisance parameters. They are always marginalised out of the joint posterior distribution, simplifying the given problem. But, Bayesians are criticised that they introduce subjectiveness into the problem by incorporating the prior belief. There has been a long discussion about the level of subjectiveness in combination with non-informative prior probabilities, but this is out of the scope of this document, as it touches more philosophical grounds[5].

In the end, we have chosen to follow the path of Bayesian analysis for LPF because of the following advantages.

- The prior distributions, and the elimination of nuisance parameters.
- Exact inferences (e.g., confidence interval) which do not rely on large sample approximations, are available through Bayesian approach.
- Bayesian answers have simple interpretation than the Frequentists ones, because, as we have already seen, we infer by a process which is closer to the empirical process of learning. An example of this is the confidence intervals in both schools.

5 More information about these discussions can be found in [13, 14, 6, 10].

3.2.1 *Multiple data channels and multiple experiments*

For the LPF case, a series of system identification experiments have been planned. Each one of them is targeting to the estimation of a particular parameter set, and are designed to stimulate the systems' sensitivity to this set. Working in a Bayesian framework we are able to consider the outcome of each experiment as prior information for the next experiment in line. This can be crucial for LTP analysis, since the final aim is to completely model the dynamics of the system (and characterise the noise sources), specially because of the nature of the LTP itself, where cross-talk effects are present. For these reasons, a combination scheme for the planned experiments has been developed [15]. Furthermore, there will be a great number of data channels available, that can be used in combination in the analysis for greater accuracy of the estimation. An example of the read-out channels can be the DWS outputs of the photodiodes, the IS readings or the Temperature Sensors, etc. If we consider

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_{ch}} \end{pmatrix}, \qquad \vec{h} = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_{N_{ch}} \end{pmatrix}, \qquad \vec{n} = \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_{ch}} \end{pmatrix}, \tag{3.21}$$

then, eq. (3.1) can be generalised for N_{ch} number of channels simply as

$$\vec{y} = \vec{h}(\vec{\theta}) + \vec{n}. \tag{3.22}$$

We can then split the scheme in two categories; one where the parameter set to be estimated remains the same through the experiments, and one where different parameter sets are involved in the analysis. For the first case we can assume, without loss of generality, a number of experiments $N_{exp} = 2$, and consequently two data-sets, $\vec{y}_1$ and $\vec{y}_2$. Then, taking into account eq. (3.22), eq. (3.16) becomes [15]

$$\pi(\vec{\theta}|\vec{y}_1, \vec{y}_2) \propto \overbrace{\underbrace{p(\vec{\theta})}_{\text{prior}} \times \overbrace{\pi(\vec{y}_1|\vec{\theta}, \mathcal{M})}^{\text{likelihood}}}^{\text{prior}} \times \underbrace{\pi(\vec{y}_2|\vec{\theta}, \mathcal{M})}_{\text{likelihood}}. \tag{3.23}$$

The interpretation of the last equation is twofold. We can either (safely in most cases) assume independent experiments and write the joint likelihood function as the product of likelihoods of the individual experiments, or regard the product prior × likelihood of the first experiment as the prior for the second experiment in line.

Quite frequently, it is more convenient to maximise the logarithm of the likelihood of eq. (3.23), because the expressions can be further simplified. The logarithm of the posterior is then written as

$$\begin{aligned} \log\left(\pi(\vec{\theta}|\vec{y}_1, \vec{y}_2, \cdots, \vec{y}_{N_{exp}})\right) &\propto \log\left(p(\vec{\theta}) \prod_j^{N_{exp}} \pi(\vec{y}_j|\vec{\theta}, \mathcal{M})\right) \\ &\propto \log\left(p(\vec{\theta})\right) + \Lambda(\vec{\theta}), \end{aligned} \tag{3.24}$$

where the log-likelihood is defined as

$$\Lambda(\vec{\theta}) = \sum_{j}^{N_{exp}} \log\left(\pi(\vec{y}_j|\vec{\theta},\mathcal{M})\right). \tag{3.25}$$

The same way of thinking can be extended to the case where experiments depend on different parameter sets $\Theta = \{\vec{\theta}_1, \vec{\theta}_2, \vec{\theta}_3\}$. Following [15], at first level we can assume again that both $\vec{y}_1$ and $\vec{y}_2$ data-sets and the corresponding model, depend on $\vec{\theta}_1$. On the other hand we can assume that $\vec{y}_1$ and $\vec{y}_2$ depend also on $\vec{\theta}_2$ and $\vec{\theta}_3$ respectively. Then, from eq, (3.23), the joint likelihood is

$$\pi(\vec{y}_1, \vec{y}_2|\Theta) \equiv \pi(\vec{y}_1, \vec{y}_2|\vec{\theta}_1, \vec{\theta}_2, \vec{\theta}_3) = \pi(\vec{y}_1|\vec{\theta}_1, \vec{\theta}_2) \times \pi(\vec{y}_2|\vec{\theta}_1, \vec{\theta}_3), \tag{3.26}$$

where we have omitted the $\mathcal{M}$ for the sake of intelligibility. This can be further simplified if we assume independent prior densities, so that

$$p(\vec{\theta}_1, \vec{\theta}_2, \vec{\theta}_3) = p(\vec{\theta}_1) \times p(\vec{\theta}_2) \times p(\vec{\theta}_3). \tag{3.27}$$

For the majority of the LTP experiments, the prior PDFs are independent (see Chapter 4). From eq. (3.26) and (3.27) we can write the joint marginal posterior distribution for $\vec{\theta}_1$ and $\vec{\theta}_2$ for both $\vec{y}_1$ and $\vec{y}_2$ data-sets as

$$\pi(\vec{\theta}_1, \vec{\theta}_2|\vec{y}_1, \vec{y}_2) = \pi(\vec{\theta}_1, \vec{\theta}_2|\vec{y}_1) \times \frac{\pi(\vec{\theta}_1|\vec{y}_2)}{p(\vec{\theta}_1)}. \tag{3.28}$$

From the above, we can see that in order to take into account the second data-set $\vec{y}_2$, that depends on $\vec{\theta}_1$ and $\vec{\theta}_3$, we need to consider the marginal prior and posterior distributions of the common parameter set and that the prior of $\vec{\theta}_1$ needs to be cancelled out. Otherwise it would enter twice into the resulting posterior. This trick can be generalised to any arbitrary higher dimension case.

The remaining detail in the developed scheme, is to compute accordingly the likelihood function

$$\pi(y|\vec{\theta}) = C \times e^{-\frac{1}{2}\left(\vec{y}-\vec{h}(\vec{\theta})|\vec{y}-\vec{h}(\vec{\theta})\right)} = C \times e^{-\chi^2/2}, \tag{3.9}$$

for the investigations with multiple data channels. The inner product $(\cdot|\cdot)$ of (3.2) contains $S_n(f)$, which for the single channel case is the PSD of the noise (a frequency series vector). Now we have to compute the cross-spectrum matrix of the noise

$$\tilde{S}_n(f) = \begin{pmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,N_{ch}} \\ S_{2,1} & S_{2,2} & \cdots & S_{2,N_{ch}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N_{ch},1} & S_{m,2} & \cdots & S_{N_{ch},N_{ch}} \end{pmatrix}, \tag{3.29}$$

where the non-diagonal terms are the Cross Power Spectral Density (CPSD) of the discrete time series of the output data channels. Of course, if the measurement channels are independent, the matrix of eq. (3.29) is diagonal. Finally, if we define the residuals

$\vec{r}(\vec{\theta}) = \vec{y} - \vec{h}(\vec{\theta})$, we can combine the expressions for multiple data channels and multiple experiments, and modify the log-likelihood function of eq. (3.25) to

$$\Lambda(\vec{\theta}) \propto -\frac{1}{2} \sum_{j}^{N_{exp}} \left(\begin{bmatrix} \tilde{r}_{1,j}(\vec{\theta}) & \tilde{r}_{2,j}(\vec{\theta}) & \cdots & r_{N_{ch},j}(\vec{\theta}) \end{bmatrix}^{\star} \times \tilde{S}_n^{-1} \times \begin{bmatrix} \tilde{r}_{1,j}(\vec{\theta}) \\ \tilde{r}_{2,j}(\vec{\theta}) \\ \vdots \\ r_{N_{ch},j}(\vec{\theta}) \end{bmatrix} \right), \quad (3.30)$$

which is expressed in frequency domain. The last relation is constructed automatically in our LTPDA machinery, with an arbitrary number of channels and experiments. The only limitation is, of course, the available computational power, since that for large data-segments, eq. (3.30) becomes computationally heavy.

3.3 THE FISHER INFORMATION MATRIX

The Fisher matrix methodology has been used extensively when involving any kind of detection schemes, in studies of expected accuracy of the parameters of the sources. For the case of LPF, as we will see below, it is used again for parameter accuracy investigations and experiment design. Here we display the basic concepts following Vallisneri [16], who has gathered and presented the various interpretations of the Fisher Information Matrix (FIM) formalism.

A great advantage of the FIM formalism is that it can be calculated analytically and numerically. For the cases where an analytical model is available, the FIM is quite straitforward to implement. But its different interpretations give rise to possible confusion. The inverse of the FIM F^{-1} is approximated by three different point of views; the Cramèr-Rao Bound, the Frequentist point of view, and the Bayesian point of view. Below, we briefly summarise the individual approximations, concluding to the one used for LPF data analysis.

- **The Cramèr-Rao Bound (CRB) approximation**: The CRB can be derived, simply by bounding the frequentist variance estimator, from Section 3.1

$$\mathrm{var}(\hat{\theta}) = \left\langle \left(\hat{\theta}(y) - \left\langle \hat{\theta}(y) \right\rangle \right)^2 \right\rangle. \quad (3.11)$$

  If the ensemble product

$$\langle u(y), w(y) \rangle = \int u(y) w(y) \pi(y|\vec{\theta}_{true}) dy, \quad (3.31)$$

  is defined, we can set the bounds on the estimator with the Schwartz inequality

$$\mathrm{var}\hat{\theta} \equiv \langle u(y), w(y) \rangle \geqslant \frac{\langle u(y), w(y) \rangle^2}{\langle u(y), w(y) \rangle}, \quad (3.32)$$

which holds true for any function of the data. With (3.32), together with the assumption that $\hat{\theta}$ is an unbiased estimator[6] we can prove [16] that for the multivariate case

$$\text{covar}(\hat{\theta}_i, \hat{\theta}_j) \geqslant F^{-1}_{ij}(\theta_{true}), \qquad \text{(CRB)} \tag{3.33}$$

which is the CRB for unbiased estimators[7]. Here the F^{-1} is the inverse of the FIM, defined if we consider eq. (3.25) as

$$F_{ij} = \left\langle \left(\frac{\partial \Lambda(\vec{\theta})}{\partial \theta_i} \middle| \frac{\partial \Lambda(\vec{\theta})}{\partial \theta_j} \right) \right\rangle \Bigg|_{\theta=\theta_{true}} . \tag{3.34}$$

evaluated at $\theta = \theta_{true}$.

From eq. (3.3), (3.1) and (3.9), where we have assumed Gaussian properties of the noise, we arrive to the more simple and useful relation of

$$F_{ij} = \left(\frac{\partial h(\vec{\theta})}{\partial \theta_i} \middle| \frac{\partial h(\vec{\theta})}{\partial \theta_j} \right) \Bigg|_{\theta=\theta_{true}} . \tag{3.35}$$

One of the criticisms against CRB is that practically, there is no likelihood function against the (3.33). So, in general, and in most cases for any unbiased estimator, the CRB is not reached with an unknown level of underperformance. In addition, the biased case of CRB, cannot be generalised since the result is affected from the particular estimator.

- **The Frequentist high SNR approximation**: The reasoning starts with the true signal/system h_y and the optimal SNR from (3.13)

$$A = \left(h_y \middle| h_y \right)^{1/2} . \tag{3.36}$$

Then, after assuming a very large A[8], the $\theta_{true} - \hat{\theta}^{MLE}$ becomes so small that the estimator is implied to achieve the CRB. The parameter-estimation errors $\Delta\vec{\theta}$ have the Gaussian probability distribution [7, 9]

$$p(\Delta\theta_i) = C' \times e^{-1/2 F_{ij} \Delta\theta_i \Delta\theta_j}, \text{ with } C' = \sqrt{\det(F/2\pi)}, \tag{3.37}$$

and therefore the root-mean square error for the parameter θ_i is given by

$$\sqrt{\langle (\Delta\theta_i)^2 \rangle} = \sqrt{\Sigma^{ii}}, \text{ with } \Sigma \equiv F^{-1}. \tag{3.38}$$

Actually, it can be proven that the bound is not reached in many cases because we are neglecting higher order corrections, and also in order the FIM to be consistent, other measures have to taken into account, like the *likelihood mismatch criterion*.

6 An "unbiased" estimator means that $\left\langle \hat{\theta}(y) \right\rangle = \theta_{true}$ and consequently $\partial_{\theta_{true}} \left\langle \hat{\theta}(y) \right\rangle = 1$. For the biased estimator, $\left\langle \hat{\theta}(y) \right\rangle = \theta_{true} + b(\theta_{true})$, where b is the bias, and $\partial_{\theta_{true}} \left\langle \hat{\theta}(y) \right\rangle = 1 + \partial_{\theta_{true}} b(\theta_{true})$.

7 The full expression, including any bias b can be written as $\text{covar}(\hat{\theta}_i, \hat{\theta}_j) \geqslant (\delta_{im} + \partial_m b_i(\theta_{true})) F^{-1}_{mj} (\delta_{jl} + \partial_j b_l(\theta_{true}))$.

8 The high SNR coincides with the Linearized-Signal Approximation (LSA).

- **The Bayesian high SNR approximation**: In the Bayesian context and in the high SNR regime, the inverse of the FIM is seen as the covariance of the posterior $\pi(\vec{\theta}|\vec{y})$ when $\vec{\theta} = \vec{\theta}_{MAP}$[9], assuming Gaussian noise and uniform priors for the parameters [17]. The introduction of Gaussian prior information centered around $\vec{\theta}_{true}$ in the formulae does not affect greatly the results [16]. At a first level, the result is equivalent to the Frequentist treatment, but the interpretation is quite different (see Sections 3.2 and 3.1).

  An interesting feature, is that in the limit $A \to 0$, the variance of the posterior tends to the value of the variance of the prior PDF, whilst in the Frequentist approach, the variance tends to zero for any non-uniform prior.

For low SNRs, like in some GW detection cases, the FIM can be singular or ill-conditioned, making its inversion problematic. This may also apply to LPF data analysis, but the advantage is that for the planned system identification experiments, the injections are known, and the SNR is large. With this at hand, it is safer to adopt any of the aforementioned interpretations of the FIM. But to be consistent, for most cases we select the Bayesian context, that allows for certain advantages, like the usage of prior densities.

3.4 MARKOV CHAIN MONTE CARLO METHODS

As the name suggests, the idea behind the Markov Chain Monte Carlo (MCMC) is to mix Monte Carlo methods and Markov Chains. A Markov chain $\{\zeta(t)\}$ is a sequence of dependent random variables such that the probability distribution of $\zeta(t)$ given the past variables depends only on $\zeta(t-1)$. This conditional probability distribution is called a *transition kernel* or a *Markov kernel* K. Eventually, the aim is to construct Markov chains that follow the desired *target, stationary* distribution. The existence of a stationary distribution requires that the kernel K allows free moves in all-over the parameter space (or in other words, there is always positive probability reaching any point on the parameter space)[10]. Another prerequisite is the so-called *ergodicity* of the chains. This means that a kernel K that produces ergodic Markov chain with a stationary distribution f, will generate simulations from f. Then if all requirements are met, any integral g can be approximated with the *Ergodic Theorem*

$$\frac{1}{T}\sum_{t=1}^{T} g(\zeta(t)) \to \langle g(\zeta)\rangle\,. \tag{3.39}$$

Yet, the working principe of a MCMC, is to construct a target density f, together with a Markov kernel K with f being its stationary distribution, generate Markov chains, and finally evaluate the integral using eq. (3.39). While there could be frustration in the choice of a suitable Markov kernel, the solution is to choose one that can be arbitrary up to a certain level, like the *Metropolis-Hastings (MH)* algorithm[11]. The trick is to

9 Here, $\vec{\theta}_{MAP}$ is the set of parameter values that maximise the posterior distribution (*Maximum A Posteriori*).

10 The kernel and stationary distribution f, must satisfy $\int_{\mathcal{Z}} K(x,y)f(x)dx = f(y)$.

11 The MH algorithm was first introduced in a 1953 paper by Metropolis *et al* [18], to be later generalised by W. K. Hastings in [19].

link the target f with an arbitrary conditional density q that is easy to simulate in practice. A more detailed description can be seen in Algorithm 1.

Algorithm 1 The Metropolis-Hastings (see figure 3.1)

1: If the current state of chain is $\zeta(t_n) \equiv \zeta$
2: Draw a new state $\zeta(t_{n+1}) \equiv \zeta'$ from a *proposal distribution* $q(\zeta'|\zeta)$.
3: Take

$$\zeta(t_{n+1}) = \begin{cases} \zeta' \text{ with probability } \alpha(\zeta, \zeta') \\ \zeta \text{ with probability } 1 - \alpha(\zeta, \zeta') \end{cases}$$

,
where

$$\alpha(x, y) = 1 \wedge \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)} \right\}. \tag{3.40}$$

4: Repeat until equilibrium is reached.

In principle, the MH algorithm is a MCMC version that is guaranteed to converge no matter the choice of the proposal distribution. The proposal only affects the *rate* of convergence [19]. It should be chosen to maintain a fine balance; to allow exploring a great part of the parameter space, without going into very low probability areas and getting stuck. A very common solution that maintains this balance, is the multivariate Gaussian distribution.

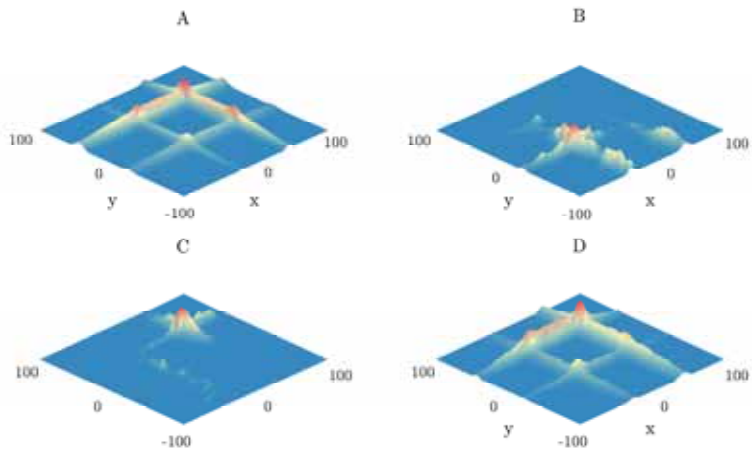


Figure 3.1: Simple example of the performance of the MH algorithm in different configurations. In all cases the initial point is in $[-100, 100]$. In panel A the theoretical, target distribution is shown. In panel B we have used a narrow proposal distribution, assuming no prior information for the parameter set $\vec{\theta} \equiv \{x, y\}$. It is evident that the poor choice of the proposal distribution in combination with the "flat" priors, result to minima trapping or no convergence of the chains. The sampling is immediately improved in panel C, when assuming Gaussian prior PDFs with large $\sigma_x = \sigma_y = \sqrt{3 \times 10^3}$. A wider proposal distribution allows the sampling of a larger part of the parameter space, as shown in panel D.

After the implementation of a MH procedure, the typical question is about the stoping criteria. Or, in other words, "after how many iterations have we reached the goal of the MH?" As in any Monte Carlo (MC) sampling method, the aim of the MH is to simulate independent samples from the target distribution. At a first touch with the algorithm, a rule of thumb has been established that gives a lower bound on the number of iterations [10]. Let us assume that the target distribution is a separable distribution along D directions. If the largest standard deviation of the space of probable states is σ_{max}, a MH method whose proposal distribution generates a random walk with step size ε must be run for at least [10, p. 369]

$$N_{iter} \simeq \left(\frac{\sigma_{max}}{\varepsilon}\right)^2, \tag{3.41}$$

iterations to obtain an independent sample. Actually, the σ_{max} is one of the unknowns to be investigated, so a criterion like (3.41) is not very practical. For that reason, other tools have been implemented to monitor the statistical properties of the MCMC chains, and they are going to be summarised in section 3.4.1.

Now, going back to the LPF intended analysis, we can select our target distribution as the eq. (3.24) that we have seen earlier

$$\log\left(\pi(\vec{\theta}|\vec{y}_1, \vec{y}_2, \cdots, \vec{y}_{N_{exp}})\right) \propto \log\left(p(\vec{\theta})\right) + \Lambda(\vec{\theta}), \tag{3.24}$$

while we neglect the evidence $\pi(\vec{y})$ because it is constant over the parameter space $\vec{\theta}$. For our parameter estimation purposes, we can sample the joint posterior PDF which will assign the posterior PDFs $\vec{\theta}_{est}$ for the parameter set. The $\vec{\theta}_{est}$, correlation and the errors $\vec{\sigma}$ can be extracted directly from the MCMC chains.

The role of prior densities

The choice of prior PDFs depends heavily on the problem to solve, and incorporates the idea of subjective view of probability of the Bayesian school. In some instances, there might not be any prior information available, and we may simply assign uniform PDFs for each parameter. In other cases, there might be more information available and this allows us to introduce a known density centered at the best guess value before the estimation. In general, the types of prior PDFs are categorised in the following families; the *non-informative* and the *informative* priors.

- *Non-informative family:* The first case reflects the desire to remove any subjectiveness from the Bayesian procedures, and perform inference solely with the data at hand. A non-informative prior can be for example a uniform distribution. It is also common to use Gaussian distributions $\theta \sim \mathcal{N}(\theta_{guess}, \sigma)$ with a very large spread, that in the area of interest it can be approximated as flat. But the known composition of the normal distribution yields good properties for numerical approximations. The renowned *Jeffreys* prior falls in this area also, written for a given parameter θ as

$$p(\theta) \propto \sqrt{F(\theta)}, \tag{3.42}$$

where F is the FIM. Jeffreys' prior generalises the location and scale invariance-based non-informative priors to arbitrary (continuous and one-to-one) reparameterizations.

- *Informative family:* Any distribution in this category reflects any kind of partial knowledge about the parameters. Here, we meet *conjugate*[12] priors. Conjugate distributions arise when the likelihood times the prior produces a recognisable posterior kernel. This introduces many advantages for analytical problems, when the resulting posterior belongs to a known family of distribution, making it easier to sample from. For example, any prior that belongs to the family of

$$\mathcal{P} = \{p(\theta) = \psi(\theta)^{\nu} \exp(\xi^{\mathsf{T}}(\theta)\gamma)\}, \tag{3.43}$$

works in an efficient way with the exponential-type of likelihood of (3.9), to produce a posterior of the same family $\mathcal{P}$.

Any prior density that is not normalised over the parameter space, e. g.

$$\int_{\theta} \pi(\theta) d\theta = +\infty, \tag{3.44}$$

is called *improper* prior and it fails to follow the basic law of probability, that the integral of (3.44) must result to unity[13]. But they can still be used together with a well defined evidence to yield a "proper" posterior distribution. In our case, this type of densities will not be used, and further explanation is beneath the intended range of this document. More details, and useful discussions can be found in [13, 14, 20].

The proposal distribution

As shown in figure 3.1, the proposal distribution may affect greatly the convergence rate of the algorithm [21]. This can be explained because the *acceptance rate* $\mathcal{R}$, or the percentage of samples that are accepted over the total proposed, is solely depending on the choice of the proposal distribution q. A high acceptance rate, is generally desired, but it does not necessarily mean that the algorithm is behaving satisfactorily. Quite often, a high value of $\mathcal{R}$ indicates that the chain is moving too slowly on the surface of the posterior. In contrast, if the average $\mathcal{R}$ is low, it implies that the MH explores regions with low probability. Nonetheless, a low acceptance rate not an issue, except for the computing time point of view, because it explicitly indicates that a larger number of simulations are necessary.

The traditional rule of thumb for the optimal acceptance rate urges to "*tune the scale of the proposal distribution, so that the acceptance rate of the algorithm is roughly* 1/4". This was properly formulated by Gelman *et al* [22], and was generalised for D-dimensional multivariate distributions of the form of

$$\pi(\vec{\theta}) = \prod_{i=1}^{D} f(\theta_i), \tag{3.45}$$

12 There are many more categories of prior densities belonging to the informative family. Some of them are the *maximum entropy* or the *minimum description length* priors. But their description is out of the scope of this thesis. More information can be found in [4, 5, 6].

13 In practice they are considered the limits of proper priors.

for any one-dimensional density f. In [22], it was proven that the optimal normal jumping kernel has its scale $\simeq 2.38/\sqrt{D}$ times the scale of the D-dimensional target distribution[14].

Naturally, a good candidate for the proposal distribution is the multivariate Gaussian. Among the available candidate distributions, it is relatively straightforward to implement, since it is symmetric and makes the $q(\cdot|\cdot)$ from (3.40) to vanish[15]. But a significant advantage is that it can be tuned (or scaled) easier because we can get a first estimation of the spread of the target posterior distribution by the FIM technique. Since a normal distribution with

$$\Sigma = F^{-1}\Big|_{\vec{\theta}=\vec{\theta}_{MAP}} \tag{3.46}$$

becomes a good approximation of the posterior around its maximum, the normal PDF

$$\mathcal{N}(\vec{\theta}, 2.38/\sqrt{D} \times \Sigma), \tag{3.47}$$

becomes a suitable proposal density that generates the desired acceptance ratio.

The efficiency of the MH algorithm can be further improved if we consider updating the parameters on the direction of the eigen-directions of the covariance matrix according to [23]

$$\theta^{\mu} \to \theta^{\mu} + \left(\sqrt{\frac{\beta}{D}}\right) \sum_{\nu=1}^{D} \frac{V_{\mu\nu}}{\sqrt{E_{\nu}}} n_{\nu}, \tag{3.48}$$

where $n_{\nu} \sim \mathcal{N}(0,1)$, E_{ν} the eigen-values, $V_{\mu\nu}$ the eigen-vectors, and β is the heat factor (see eq. (3.49) below). Since the covariance matrix can be interpreted also as a "mapping" of the posterior surface, a symmetric proposal distribution that updates as in (3.48), is proven to move quicker to the maximum of the posterior.

In the end, the suitability of the proposal distribution q, is always determined by the given estimation problem. Many alternatives to the standard Gaussian have been developed, like the *adaptive proposal* [24] or more complex multimodal distributions like in [12] where *delayed-rejection* MH schemes are implemented[25].

The annealing procedures

For some problems, the sampled posterior distribution appears to be multimodal, and the convergence of MCMC methods becomes slower. The MCMC chains are spending more time in localities of the posterior distribution, and this requires longer runs of the algorithm. There are a few techniques developed, that allow the MCMC chains to sample the parameter space with more ease during the *search* or *burn-in* phase of the simulation. The search phase happens during the first iterations and the proposal distribution can be scaled to propose each time bigger jumps, thus covering a larger part of the parameter pace. At the same time, the target distribution can be modified

14 The acceptance rate of the associated Metropolis algorithm is approximately $\mathcal{R} = 44.1\%$ for $D = 1$ and declines to $\mathcal{R} = 23.4\%$ as $D \to \infty$ for the normal target, normal proposal case [22].

15 By using a symmetric q, we build a so-called *random-walk* MH. A random-walk MH means that in algorithm 1, each $\zeta(t_{n+1})$ is simulated from $\zeta(t_{n+1} = \zeta(t_n) + \varepsilon_t$, where ε here is a random perturbation.

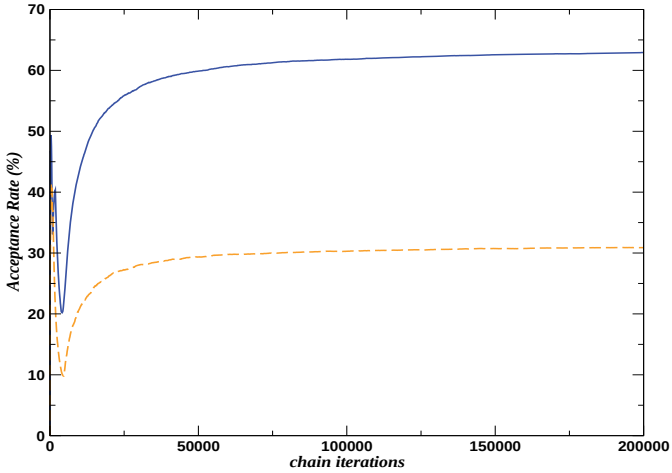


Figure 3.2: Acceptance ratio improvement by updating to eigen-directions. The blue solid line represents the acceptance ratio from a standard multivariate Gaussian proposal, while the dashed orange shows the acceptance ratio for the improved eigen-direction scheme. Credit: [23].

from $\pi(\vec{\theta}|\vec{y})$ to $\pi(\vec{\theta}|\vec{y})^{1/T}$ where $T \geqslant 1$ to smoothen the distribution and increase the acceptance rate. The temperature T follows

$$T = \begin{cases} 10^{\beta[1-(T_h/T_c)]}, & \text{with } 1 \leqslant i \leqslant T_h \\ 10^{\beta[1-(i/T_c)]}, & \text{with } T_h \leqslant i \leqslant T_c \\ 1 & \text{with } i \geqslant T_c \end{cases}, \tag{3.49}$$

where β is a positive constant, i the i-th sample of the chain and $[T_h, T_c]$ mark the "annealing" profile [15, 26] (see figure 3.3).

Another type of annealing is based on the current estimated SNR [26]. In the *thermostated annealing* case, the PSD of the noise in eq. (3.9) is multiplied by

$$T = \begin{cases} 1, & \text{with } 0 \leqslant SNR \leqslant C_{SNR} \\ \left(\frac{SNR}{C_{SNR}}\right)^2, & \text{with } SNR \geqslant C_{SNR} \end{cases}, \tag{3.50}$$

where C_{SNR} is a constant. This ensures that the chains do not get stuck in localities when reaching a high SNR of the template. The same type of "thermostat" can work with the likelihood values instead of the SNR. We have called it "simple thermostat", and a variant of its implementation can be seen in [27]. The temperature T is controlled every k amount of iterations (where k a fixed constant), by calculating the deviation of the likelihood value, according to

$$T = \begin{cases} T+1, & \text{with } \delta\Lambda(\vec{\theta}) \leqslant C_1 \\ T-1, & \text{with } \delta\Lambda(\vec{\theta}) \geqslant C_2 \end{cases}. \tag{3.51}$$

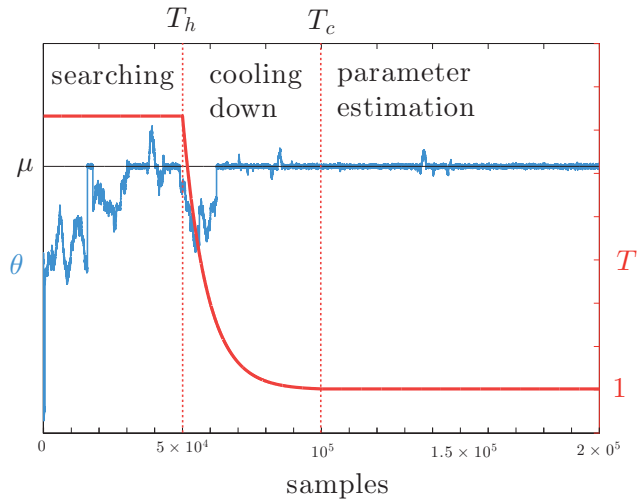


Figure 3.3: Demonstration of the effect of the Simulated Annealing procedure on a parameter chain θ. From the first to the $i = T_h$ sample of the chain, the heating is large thus allowing easier transition in the parameter space. When $T_h \leqslant i \leqslant T_c$, β progressively goes to 1 according to eq. (3.49). This is the "cooling down" phase of the sampling. For $i \geqslant T_c$, $\beta = 1$ and the parameter estimation phase begins. The $i \leqslant T_c$ part of the chain is later discarded, and the $\hat{\mu}$ of the given parameter is estimated. The values of T_h and T_c are of course defined according the complexity of the given problem. Credit: LTPDA toolbox HTML help page.

If this deviation $\delta\Lambda(\vec{\theta})$ is less than a constant C_1, this means that the chains are not mixing in a satisfactory way, and the temperature is rising. The opposite happens when $\delta\Lambda(\vec{\theta}) \geqslant C_2$, and the temperature is lowered.

Another way to increase the efficiency of the algorithm is to provide more information of the surface being sampled. In practice, prior to system identification, only a guess of the parameter values can be used for the calculation of the FIM (and consequently, the Σ of the proposal). Since, the covariance matrix Σ can also be seen as a mapping, or slope, of the posterior surface, it can be updated frequently during the MCMC to increase the acceptance ratio. Typically this can happen during the search or *burn-in* period of the chain, and if the computation of the FIM and its inversion are cost-efficient (not computationally demanding), the Σ is calculated at each MH iteration. Otherwise, the proposal can be updated in a constant period of the MCMC iterations.

3.4.1 *Convergence diagnostics*

When implementing MCMC methods, one of the remaining issues is to construct the necessary diagnostics machinery. For a healthy MH run, we must monitor the *mixing* of the chains, the acceptance ratio, and finally the convergence of the algorithm. The convergence and mixing of the MCMC chains, or in other words, the ability of sufficiently sampling the parameter space, can be monitored in two ways; by "visual" and by statistical means. In the LTPDA implementation, we choose the following strategies to assess the MCMC simulation [28].

- Regular plotting of the traces during the MCMC run. The convergence run can be simply monitored visually, by looking the progress of the parameter traces. This is a superficial diagnostic, but in most cases it can be proven to be very helpful. The same holds for the sampled covariance matrix of the posterior distribution, where a lot of information about the posterior surface can be derived.

- Plotting the auto-correlation of the parameter chains. Since the aim of a MCMC algorithm is to simulate independent samples from a target distribution, the auto-correlation between the draws of the Markov chain may reveal the convergence progress. The auto-correlation ρ_k is the correlation between every draw and its k-th lag

$$\rho_k = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.52}$$

  The expected behaviour would be that the k-th lag autocorrelation is smaller as k increases. If the autocorrelation is still relatively high for higher values of k, this indicates high degree of correlation between our draws and slow mixing (figure 3.4).

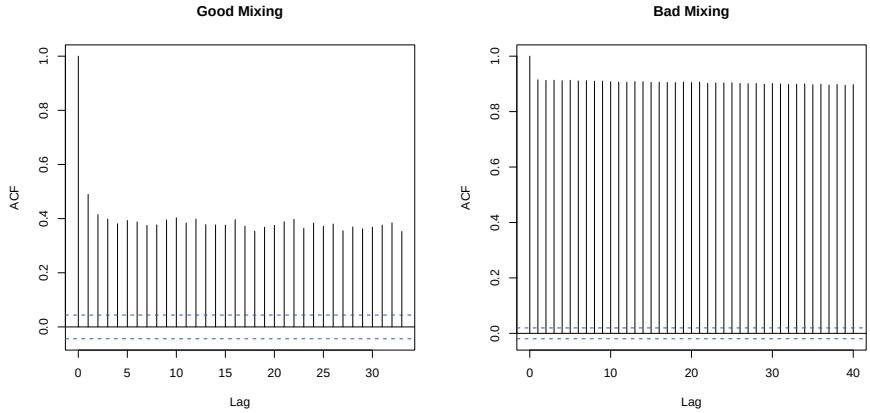


Figure 3.4: Example of good and bad mixing of the chains based on the autocorrelation of the samples.

- Another graphical procedure to characterise the mixing of the chains is the Yu-Mykland cumulative sum path plots, that is applied to every single parameter chain. If the chosen summary statistic is $g(X)$, then the estimate of its mean based on the retained iterates is

$$\hat{\mu} = \frac{1}{N_{total} - T_c} \sum_{j=T_c+1}^{N_{total}} g(X^{(j)}) \tag{3.53}$$

  where N_{total} is the total number of samples, and the cumulative sum is

$$\hat{S}_t = \sum_{j=T_c+1}^{t} \left(g(X^{(j)} - \hat{\mu}) \right). \tag{3.54}$$

If we plot $\hat{S}_t$ versus the samples t we obtain a plot that begins and ends in zero, called the cusum plot. If the mixing of a parameter chain is slow, then the cusum plot will appear to be smoother and fluctuating further from zero. Conversely, a fast mixing chain is accompanied with a "hairier" cusum plot, like the θ_1 parameter in figure 3.5. But, this method was criticised for being rather

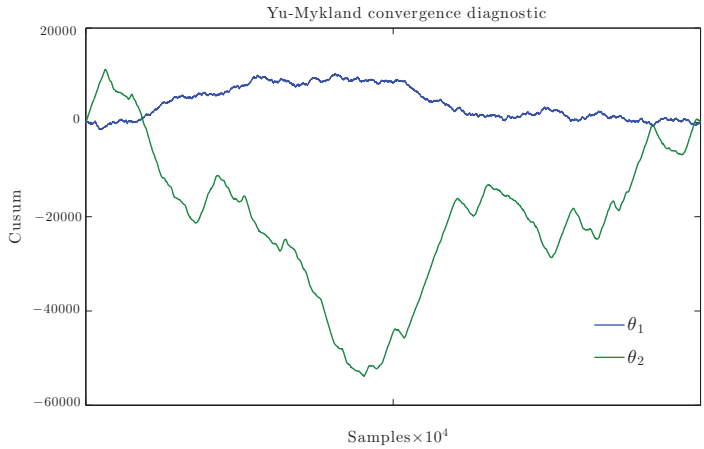


Figure 3.5: Example of the Yu-Mykland convergence diagnostic for testing the mixing of the MCMC chains for the set $\vec{\theta} = [\theta_1, \theta_2]$. From this plot we can derive that parameter θ_1 is mixing faster that θ_2. This figure is produced at the end of each MCMC run of the LTPDA algorithm implementation.

subjective, and Brooks [29, 30] later proposed the "index of hairiness" $\mathcal{I}$ as follows. If we define the d_i as

$$d_i = \begin{cases} 1, & \text{if } S_{i-1} > S_i \text{ and } S_i < S_{i+1} \\ & \text{if } S_{i-1} < S_i \text{ and } S_i > S_{i+1} \\ 0, & \text{otherwise} \end{cases}, \tag{3.55}$$

then

$$\mathcal{I}_t = \frac{1}{t - T_c - 1} \sum_{i=T_c+1}^{t-1} d_i, \quad T_c + 2 \leqslant t \leqslant N_{total} \tag{3.56}$$

In the end, the $\mathcal{I}$ takes values from 1, that indicates a high level of mixing, to 0, that corresponds to smoother cusum plots and slow mixing of the MCMC chains.

- One of the most widely used convergence diagnostics is the Potential Scale Reduction Factor (PSRF) $\hat{R}_c$, introduced by Gelman and Rubin [31]. In essence the PSRF measures the change of the variance of m different MCMC simulations. If the change is small enough then we can assume that the chains are converging to the same target distribution. For each θ, the individual chains are compared

to the mix of mn values from all sequences. The first step is to compute the between-sequence variance

$$B = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j - \bar{\theta})^2, \tag{3.57}$$

and within-sequence variance

$$W = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{t=1}^{n}(\theta_{jt} - \bar{\theta}_j)^2, \tag{3.58}$$

where θ_{jt} is the value after t iterations in chain j. Then, if

$$\hat{V} = \frac{n-1}{n}W + \left(1 + \frac{1}{m}\right)\frac{B}{n}, \tag{3.59}$$

the PSRF is calculated as

$$\hat{R}_c = \frac{(df+3)\hat{V}}{(df+1)W}. \tag{3.60}$$

The $\hat{R}_c$ takes values $\geqslant 1$, with $\leqslant 1.2$ being a threshold to assess the convergence of the chains. The PSRF is mostly used in parallel runs of MCMC simulations, but it can also be calculated for a single run. The chains are "chopped" in different, equal in length segments, treated again as independent runs, where $\hat{R}_c$ is computed.

- The final test is to perform a Kolmogorov-Smirnov (KS) test to the statistics of the MCMC chains. The KS test [32], compares the sampled distribution for each chain with a reference probability distribution. If $f(x)$ is the probability density function associated with a random process $X_n = \{x_1, x_2, \cdots, x_n\}$, then the Cumulative Distribution Function (CDF) is

$$F(x) = Pr[X \leqslant x] = \int_{-\infty}^{x} f(u)du. \tag{3.61}$$

The Empirical Cumulative Distribution Function (ECDF) is defined $F_n = k/n$, with k being the number of observations that are $\leqslant x$. If the null hypothesis is that the two data-sets are generated from the same statistic, it can be proven that the maximum distance between the two ECDFs has a limiting distribution which is independent from the statistical properties of the corresponding random variable. Then the test can be computed as

$$d_k = \max|F_n(x) - F_m(x)|. \tag{3.62}$$

between two data-sets, with $k = (nm)/(n+m)$. Finally, being a Frequentist criterion, a significance level α is defined (usually set to 0.05) as the probability to reject the null hypothesis. The KS test has already been implemented for LPF analysis of the noise in [33].

All the aforementioned convergence diagnostics, are implemented and integrated into the LTPDA version of the MCMC [34]. They are computed automatically at the end of each simulation. A more complete collection of convergence diagnostics methods, and more detailed explanations, can be found in [28].

[1] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995. URL: http://books.google.gr/books?id=TRXrMWY_i2IC. 45

[2] W.S. Kendall, F. Liang, and J.S. Wang. *Markov Chain Monte Carlo: Innovations and Applications*. Lecture notes series. World Scientific, 2005. URL: http://books.google.gr/books?id=cB-ZZX2HcWQC. 45

[3] G. L. Bretthorst. *Bayesian spectrum analysis and parameter estimation*. Number 48 in Lecture notes in statistics. Springer, New York [u.a.], 1988. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+025617311&sourceid=fbw_bibsonomy. 45

[4] William M. Bolstad. *Introduction to Bayesian Statistics*. Wiley, 2007. URL: http://books.google.gr/books?id=gQw6U09uZd8C. 45, 46, 57

[5] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. 45, 57

[6] Jun S. Liu. *Monte Carlo strategies in scientific computing*. Springer, 2001. 45, 49, 57

[7] Curt Cutler and Éanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D*, 49:2658–2697, Mar 1994. URL: http://link.aps.org/doi/10.1103/PhysRevD.49.2658, doi:10.1103/PhysRevD.49.2658. 45, 47, 53

[8] Leor Barack and Curt Cutler. LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy. *Phys.Rev.*, D69:082005, 2004. arXiv:gr-qc/0310125, doi:10.1103/PhysRevD.69.082005. 45

[9] Lee S. Finn. Detection, measurement, and gravitational radiation. *Phys. Rev. D*, 46:5236–5249, Dec 1992. URL: http://link.aps.org/doi/10.1103/PhysRevD.46.5236, doi:10.1103/PhysRevD.46.5236. 45, 53

[10] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. 46, 49, 56

[11] D. S. Sivia. *Data Analysis: A Bayesian Tutorial*. Clarendon (Oxford Univ. Press), Oxford, 1996 (ISBN: 0-19-851762-9 or 0-19-851889-7 in paperback). 48

[12] Miquel Trias. *Gravitational-Wave observation of compact binaries: Detection, parameter estimation and template accuracy*. PhD thesis, Universitat de les Illes Balears, 2010. 49, 58

[13] James O. Berger. *Statistical decision theory and Bayesian analysis*. Springer series in statistics. Springer, New York, NY [u.a.], 2. ed edition, 1985. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+027440176&sourceid=fbw_bibsonomy. 49, 57

[14] Christian P. Robert. *The Bayesian Choice: a decision-theoretic motivation*. Springer, New York, 1994 (ISBN: 0-387-94629-3). 49, 57

[15] M. Nofrarias, C. Röver, M. Hewitson, et al. Bayesian parameter estimation in the second LISA Pathfinder mock data challenge. *Physical Review D*, 82(12):1–14, December 2010. URL: http://link.aps.org/doi/10.1103/PhysRevD.82.122002, doi:10.1103/PhysRevD.82.122002. 50, 51, 59

[16] Michele Vallisneri. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Physical Review D*, 77(4):1–20, February 2008. URL: http://link.aps.org/doi/10.1103/PhysRevD.77.042001, doi:10.1103/PhysRevD.77.042001. 52, 53, 54

[17] David Nicholson and Alberto Vecchio. Bayesian bounds on parameter estimation accuracy for compact coalescing binary gravitational wave signals. *Phys. Rev. D*, 57:4588–4599, Apr 1998. URL: http://link.aps.org/doi/10.1103/PhysRevD.57.4588, doi:10.1103/PhysRevD.57.4588. 54

[18] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J.Chem.Phys.*, 21:1087–1092, 1953. doi:10.1063/1.1699114. 54

[19] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970. doi:10.1093/biomet/57.1.97. 54, 55

[20] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, New York, 1994 (ISBN: 0-471-92416-4). 57

[21] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 02 1997. URL: http://dx.doi.org/10.1214/aoap/1034625254, doi:10.1214/aoap/1034625254. 57

[22] A. Gelman, G.O. Roberts, and W.R. Gilks. Efficient metropolis jumping rules. In J. M. Bernado et al., editors, *Bayesian Statistics*, volume 5, page 599. OUP, 1996. 57, 58

[23] Luigi Ferraioli, Edward K. Porter, Michele Armano, et al. Improving Bayesian analysis for LISA Pathfinder using an efficient Markov Chain Monte Carlo method. *Exper.Astron.*, 37:109–125, 2014. doi:10.1007/s10686-014-9372-7. 58, 59

[24] Heikki Haario, Eero Saksman, and Johanna Tamminen. Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14(3):375–395, 1999. URL: http://dx.doi.org/10.1007/s001800050022, doi:10.1007/s001800050022. 58

[25] Miquel Trias, Alberto Vecchio, and John Veitch. Delayed rejection schemes for efficient Markov-Chain Monte-Carlo sampling of multimodal distributions. *arXiv*, 2009. 58

[26] Neil J. Cornish and Edward K. Porter. Searching for Massive Black Hole Binaries in the first Mock LISA Data Challenge. *Class.Quant.Grav.*, 24:S501–S512, 2007. arXiv:gr-qc/0701167, doi:10.1088/0264-9381/24/19/S13. 59

[27] Andrew F Neuwald and Jun S Liu. Gapped alignment of protein sequence motifs through monte carlo optimization of a hidden markov model. *BMC bioinformatics*, 5, 2004. URL: http://dx.doi.org/10.1186/1471-2105-5-157, doi:10.1186/1471-2105-5-157. 59

[28] Mary Kathryn Cowles and Bradley P. Carlin. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904, 1996. 60, 63

[29] URBAN HJORTH and ANNA VADEBY. Subsample distribution distance and mcmc convergence. *Scandinavian Journal of Statistics*, 32(2):313–326, 2005. URL: http://dx.doi.org/10.1111/j.1467-9469.2005.00424.x, doi:10.1111/j.1467-9469.2005.00424.x. 62

[30] Stephen P. Brooks and Gareth O. Roberts. Convergence assessment techniques for markov chain monte carlo. *Statistics and Computing*, 8(4):319–335, December 1998. URL: http://dx.doi.org/10.1023/A:1008820505350, doi:10.1023/A:1008820505350. 62

[31] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992. http://www.stat.columbia.edu/~gelman/research/published/itsim.pdf. 62

[32] A.N. Kolmogorov. Sulla determinizione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari*, 4:83?91, 1933. 63

[33] Luigi Ferraioli, Giuseppe Congedo, Mauro Hueller, et al. Quantitative analysis of lisa pathfinder test-mass noise. *Phys. Rev. D*, 84:122003, Dec 2011. URL: http://link.aps.org/doi/10.1103/PhysRevD.84.122003, doi:10.1103/PhysRevD.84.122003. 63

[34] N. Karnesis, Nofrarias M., and M. Hewitson. MCMC Parameter Estimation for LPF: Design & Code Review. Technical report. 63