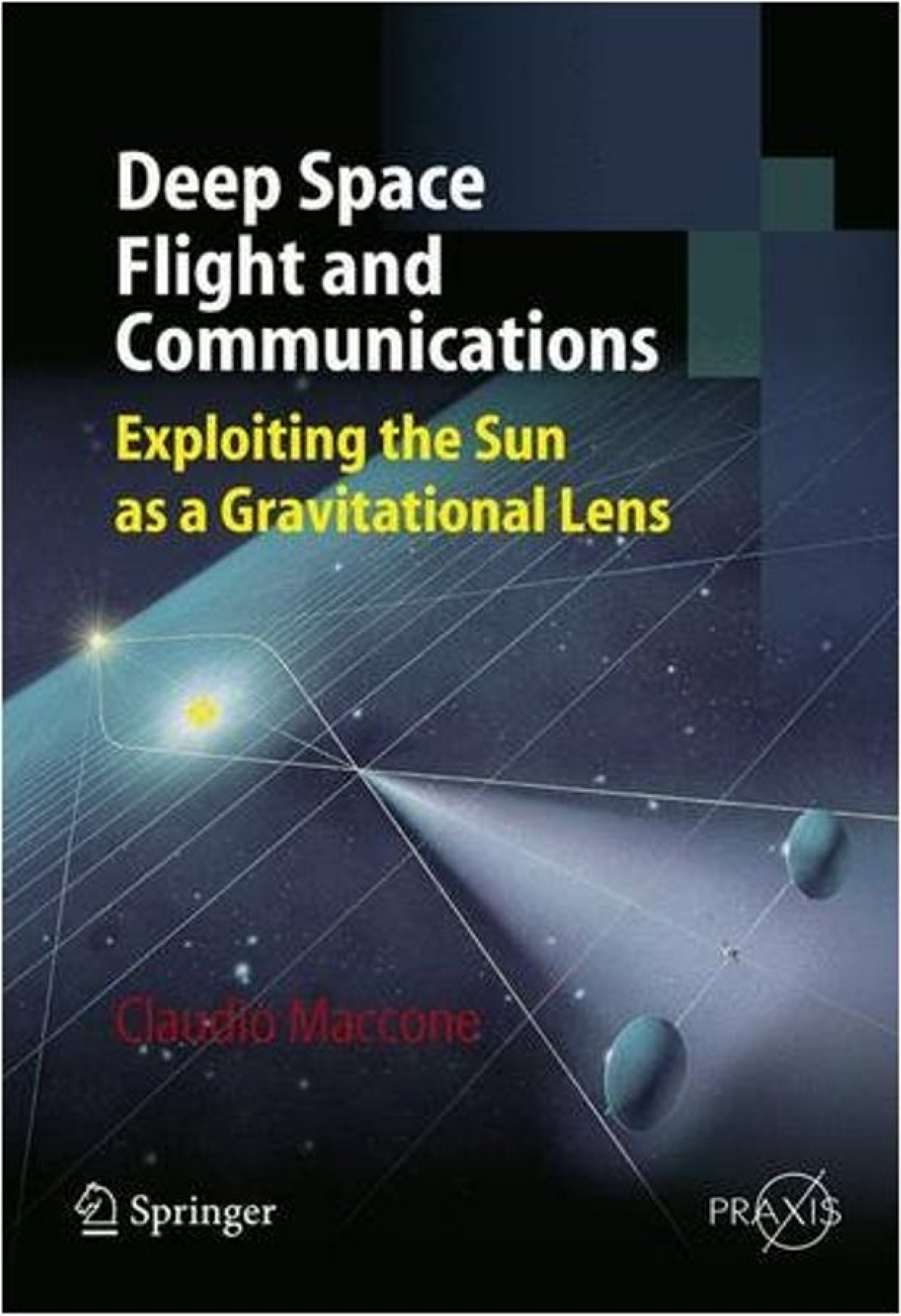
Deep Space Flight and Communications

Exploiting the Sun as a Gravitational Lens



Claudio Maccone

Springer

PRAXIS

Deep Space Flight and Communications

Exploiting the Sun as a Gravitational Lens

Claudio Maccone

Deep Space Flight and Communications

Exploiting the Sun as a Gravitational Lens

Springer

Published in association with
Praxis Publishing
Chichester, UK

PRAXIS

Dr Claudio Maccone
International Academy of Astronautics & Istituto Nazionale di Astrofisica
Via Martorelli 43 - 10155 Torino (Turin)
Italy
Email: clmaccon@libero.it

SPRINGER–PRAXIS BOOKS IN ASTRONAUTICAL ENGINEERING
SUBJECT *ADVISORY EDITOR*: John Mason B.Sc., M.Sc., Ph.D.

ISBN 978-3-540-72942-6 Springer Berlin Heidelberg New York

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

Library of Congress Control Number: 2007939976

Cover design: Jim Wilkie
Project management: Originator Publishing Services Ltd, Gt Yarmouth, Norfolk, UK

Printed in Germany on acid-free paper

Contents

Preface

This is a book about unmanned space missions to the edge of the Solar System. And possibly beyond.

Readers may wish to read first the "Brief Overview" of the scientific and technical problems discussed in this book that is found on pp. xxxi–xxxiv hereafter.

Also, readers might wish to request the DVD of the Lecture that the author gave at NASA–JPL on August 18th, 1999, about "The Sun's Gravity Lens and Its Use for Interstellar Exploration" (running time: about 80 minutes).

This DVD may be requested by email to the author:

Dr. Claudio Maccone, email: *clmaccon@libero.it*

or, in the author's absence, to either of the following co-workers of his:

(1) Dr. Luca Derosa, email: *spacecraft@libero.it*
(2) Dr. Nicolo' Antonietti, email: *n.antonietti@libero.it*.

The present book is the result of merging two previously published, smaller books by the author. The first, *The Sun as a Gravitational Lens: Proposed Space Missions*, corresponds to Part I, and the second, *Telecommunications, KLT and Relativity*, corresponds to Part II of this revised and updated book.

If NASA and ESA decide to fund the first "precursor interstellar mission" to 550 AU or even beyond to 1000 AU in the decades to come, the goal of this book will have been reached.

Claudio Maccone
Torino (Turin), Italy, January 23rd, 2009

Preface to earlier works

I first met Frank Drake in 1987 at Balatonfüred, Hungary, at the *Second International Conference on Bioastronomy* (officially called the 99th Colloquium of the International Astronomical Union, held June 22–27, 1987). Bioastronomy is the newly born branch of science trying to assess whether life exists elsewhere than on Earth. As far as we know, life can come in a variety of forms, either less or more biologically advanced than humans are at present. Next to the very primitive forms of life that biologists are currently trying to detect (e.g., on Mars), it is quite reasonable to admit that civilizations more advanced than ours possibly exist in the Galaxy also. If this is the case, one may try to detect them by using the several large radiotelescopes now available on Earth to pick up the possible "leakage" of radio signals emitted from planets orbiting other nearby stars, just as the Earth is constantly emitting a large quantity of radio waves that have been outflowing into space since humans discovered the technology of radio transmission about 1900 AD.

Frank Drake was the first scientist to try detecting, back in 1960, whether "intelligent" radio signals were being emitted by planets around two nearby Sun-like stars (ε *Eridani* and τ *Ceti*). This was *Project Ozma*, and it marked the beginning of *SETI*, the (radio) *Search for Extra-Terrestrial Intelligence* that has been continuing ever since in some of the most technologically advanced countries, like the U.S.A., Russia, France, Italy, Japan, Australia, and Argentina.

At the Balatonfüred 1987 conference Frank Drake gave a talk titled *Stars as Gravitational Lenses* that greatly impressed me. He described the huge magnification (i.e., the very intense focusing) that the gravitational field of the Sun would have on the radio waves (or light rays) originating, for instance, at the Galactic center, and then traveling unaltered through the vast Galactic distances until they graze the Sun's surface. These rays are then deflected by the Sun's gravity and made to focus along a line starting at 550 AU from the Sun. If one could construct a spacecraft capable of traveling 550 AU (and beyond, perhaps to 1,000 AU), the spacecraft could transmit back to Earth the greatly magnified radio picture of whatever lies at great distances in the opposite direction to the Sun. I have dubbed both the spacecraft and the entire space mission *FOCAL*.

In the years 1987–1992 I made a personal preliminary study of the FOCAL space mission which made clear to me the five basic points listed below. Meanwhile, my discussions with scientists and engineers, experts in areas of importance to the FOCAL space mission, led me to organize the first one-day international conference on FOCAL. This was the *Space Missions and Astrodynamics I* conference held at Politecnico di Torino (the Engineering School in my home city of Turin, Italy) on June 18, 1992. I later edited the proceedings of this conference and published them in two issues of the *Journal of the British Interplanetary Society* (February and November 1994). By that time interest in FOCAL by various scientists and engineers had grown to such an extent that the time was ripe to submit a formal proposal for the mission to one leading space agency. Opportunity was taken of the *Call for New Mission Ideas* (M3) issued by *ESA*, the *European Space Agency*, early in 1993. So on behalf of a large group of scientists and engineers from both Europe and the U.S.A., on May 20, 1993, I submitted the 50-page FOCAL proposal to ESA, who later included it as Proposal #24 among the Responses to the Call for Mission Concepts for the *Horizon 2000 Plus* plan (see ESA SP-1180, August 1995, p. 115, #24). It was then wittily remarked by the French professor Roger Bonnet, Director of Scientific Programs of ESA, that had FOCAL been approved by the Agency, it would have provided work not just for the present generation of ESA employees, but also for the generations of their sons and grandsons.

This remark by Bonnet obviously pointed out the tremendous amount of work necessary to put up a very deep space mission like FOCAL. However, it could hardly deter far-sighted scientists and engineers from thinking about such a deep space mission. I proceeded to promote the FOCAL mission when I ran the *Space Missions and Astrodynamics II and III Conferences* at the Politecnico in Turin in October 1994 and June 1995, respectively. At the *International Astronautical Federation Congress* in Oslo, October of 1995, the *International Academy of Astronautics* (*IAA*) agreed to hold the *First IAA Symposium on Missions to the Outer Solar System and Beyond* at the Politecnico in Turin on June 25–27, 1996. Some 50 experts from NASA-JPL, Russia, and European countries gathered to discuss the perspectives of future exploration of the outer solar system, and there FOCAL gained ground as the "must" mission before any attempt to go beyond 1,000 AU from the Sun was even conceived.

Now the five basic points summarizing the importance of the FOCAL space mission are as follows:

(1) FOCAL would necessarily be the *first precursor interstellar mission of humankind*. In fact, the minimal distance of 550 AU is about 14 times the distance from the Sun to Pluto, and so FOCAL would, by far, surpass any other ongoing "deep space" mission (such as *Voyagers 1* and *2*, *Pioneer 10* and *11*) or planned (the NASA-JPL *Pluto Express*). Put in better terms still, one might say that any future interstellar mission of humankind will necessarily be a FOCAL mission also, since beyond 550 AU the Sun will *always* be serving as a lens for *some* direction.

(2) Reaching 550 AU takes a long time. How long this flight time might be depends on the propulsion system adopted (possibly nuclear-electric propulsion, solar sailing, magnetic sailing, or a combination of all of them). One can currently imagine this flight taking between 10 and 50 years. It is true that one does *not* have to stop FOCAL at just 550 AU, because *every* point along the straight-line trajectory beyond 550 AU *still is* a focal point. This paves the way to an even more ambitious mission, up to 800 or 1,000 AU, requiring yet more time. One such mission studied at JPL in the 1980s has the name *TAU* (*Thousand Astronomical Units*). It would yield such a wealth of scientific results through the study of parallaxes of stars, the heliosphere[1] and the heliopause,[2] the interstellar medium, and the possible detection of gravitational waves, that the FOCAL (or TAU) mission would be justified independent of the gravitational focusing effect.

(3) Which members of the scientific and technological community would be interested in the results provided by FOCAL? First, astrophysicists would enjoy getting a detailed radio picture of the Galactic center, where a massive black hole is suspected to exist and stars are so close that unexpected physical phenomena could be revealed. This high-resolution radio picture can be obtained only by using the Sun as a gravitational lens jointly with a spacecraft capable of observing on the hydrogen line (1,420 MHz) and/or similar frequencies (1.6 GHz for the OH maser; 22 GHz for the water maser, and so forth). Second, a spacecraft crossing the Kuiper Belt would provide planetary scientists with a wonderful opportunity to investigate the lesser bodies recently discovered to orbit the Sun roughly between 40 AU and 100 AU. Other important fields of investigations, such as plasma physics of the heliosphere and heliopause, determination of parallaxes of stars, perturbation of orbits leading to the possible discovery of new bodies, have already been mentioned. So let me just add that, last but by no means least, space engineers and technologists would have to overcome challenges like the selection of the best propulsion system to get there, how to keep track of the spacecraft at such unprecedented distances, and how to optimally compress information to enable the huge data flow from the FOCAL spacecraft back to the Earth. Advanced technology corporations would support the approval of FOCAL by space agencies as a proving ground for improving technology already in existence.

(4) SETI deserves a separate discussion. As Frank Drake said in 1987, only by exploiting the gravitational lens of the Sun can we expect to detect signals that are extremely weak because they come from so far away in the Galaxy. Consider the former NASA *SETI Targeted Search* (ended abruptly in October 1993 by the U.S. Congress on the ground of "necessary budget cuts"). The goal of this project was to observe 773 Sun-like stars with the highest possible sensitivity

[1] The *heliosphere* is the region surrounding the solar system where the solar wind dominates the interstellar plasma. Despite the word "sphere" as part of the name, the region is not spherical, but extends roughly 150 AU in front, and much farther behind.

[2] The *heliopause* is the imaginary surface bounding the heliosphere.

provided by the collecting area of the largest radiotelescopes on Earth (i.e., the Arecibo telescope located in Puerto Rico, the Goldstone 70-meter Deep Space Network antenna in the Mojave desert of California, the Nançay radiotelescope in France, the Parkes antenna in Australia, etc.). Because of the limited collecting area of these telescopes, the distance of the target stars could not exceed 100 light years. But this distance is very small compared with the size of the Galaxy (100,000 lt-yr in diameter), so even if the NASA-SETI project had been funded, the part of the Galaxy explored for extraterrestrial life would have been very small. In other words, the problem is that the collecting area of radiotelescopes on Earth cannot exceed the current values by much, and hence one cannot detect even weaker signals. One might say that the generation of SETI scientists of the school of Frank Drake in the U.S.A., of Nikolai Kardashev in Russia, and of Jean Heidmann in France, have already done the best they can do on the suface of Earth. It is now up to the space scientists to take the lead in SETI by putting up the first FOCAL space mission. FOCAL could detect signals 2 to 3 orders of magnitude weaker than signals detectable on Earth, and could thus enable the detection of civilizations located much farther out, around the Galactic center 32,000 lt-yr away, or even farther, thus increasing dramatically the probability of humankind's first contact with ET.

(5) Politicians of the highest caliber might also be interested in supporting an epoch-making space mission like FOCAL for their own personal prestige. In fact, the *costs* of the first precursor interstellar mission of humankind could hardly be supported by a single national space agency, and international cooperation would be necessary. The U.S. and NASA could, at a minimum, provide the JPL and Deep Space Network expertise, the Russians could provide the launcher, the Europeans could be in charge of the scientific payload, as could the Japanese, etc. Finally, the members of the United Nations could give their patronage to a space mission of unsurpassed "grandeur" as FOCAL, symbol of the first expansion of humankind outside the solar system.

Claudio Maccone
Secretary of the Interstellar Space Exploration Committee and
Member of the SETI Committee,
International Academy of Astronautics

Acknowledgments

As a member of two committees of the International Academy of Astronautics—the SETI Committee and the Interstellar Space Exploration Committee—I am greatly indebted to my fellow members for many helpful discussions as well as for encouragement to publish this book.

Within the SETI international community I had the honor and privilege of discussing the gravitational lens of the Sun with the late, incomparable Carl Sagan as well as with the initiator of experimental SETI, Frank Drake. In Russia, Nikolai Kardashev gave me the opportunity to present my views on the FOCAL space mission at the Sternberg Astronomical Institute in Moscow in October 1996. To these three outstanding scientists I would like to express my deepest feelings of gratitude. SETI Institute scientists and engineers also taught me a great deal: Jill Tarter, Kent Cullers, Seth Shostak, Tom Pierson, and many others served as my teachers of experimental SETI, while Laurance Doyle and others did the same on the subject of habitable zones and extrasolar planet searching. John Billingham, former SETI Chief at NASA-Ames and Coordinator of the IAA SETI Committee, has always encouraged my SETI-related research, from FOCAL to the intended Moon farside radiotelescope, and to the "crazy" topic of "SETI via Wormholes". In Europe, inspiring personalities have been Jean Heidmann in France and Ivan Almar in Hungary, both of whom also attended and provided outstanding contributions at the conferences I ran in Turin. May I also add the names of François Biraud in France and Robert S. Dixon at the Ohio State University, who independently foresaw back in 1983, as I did, that SETI could be greatly improved by the replacement of the Fast Fourier Transform by the *Karhunen–Loève Transform* (KLT). Finally, I am happy to acknowledge the contribution of the outstanding SETI work of Stuart Bowyer and Dan Werthimer at UC Berkeley, of Paul Shuch in New York, of the only master of experimental SETI in Italy, Stelio Montebugnoli of the CNR Radio Astronomy Institute in Bologna, and of Cristiano Cosmovici of the CNR in Rome.

The other IAA Committee of which I am part is the Interstellar Space Exploration Committee, created back in 1984 by the leading British physicist Leslie Shepherd. To him, as well as to the present coordinator, Giovanni Vulpetti of Telespazio in Rome, I am indebted for many inspiring conversations about interstellar flight, including relativistic flights. In particular, I must here mention Giovanni's outstanding discovery (1992–1995) of a Sun flyby by virtue of a solar sail that would literally throw the solar sail out of the solar system at a much higher speed than previously believed. If the FOCAL spacecraft was to be a solar sail, it might reach 550 AU in 25 to 30 years. Solar sails also are the leading field of expertise of one more Committee member, Gregory Matloff of NYU, to whom I am very grateful for suggestions, friendship, and hospitality. Another member of the Committee to whom I am very grateful for his great cooperation is Macgregor S. Reid, Technical Executive Assistant to the Director of JPL, who hosted me at JPL several times and put me in touch with many outstanding experts working there. One of these is Rob Stahle, the *Pluto Express* program manager, who obviously faces problems that are very much akin to those of the FOCAL mission.

The organization of the above-mentioned conferences in my home town of Turin would not have been possible without the decisive support of Giancarlo Genta, former director of the Mechanical Department at the Politecnico and now president of the G. Colombo Center for Astrodynamics. Similarly, the organization of a *SETI-Day* on June 28, 1996, at the Academy of Sciences of Turin, was greatly helped by another friend, Danilo Noventa of the Peace Watch Committee in Turin. Special thanks are due to Salvatore Santoli of the Rome Branch of INT (International Nanobiological Testbed), editor of the Exobiology Section of the *Journal of the British Interplanetary Society* and leading nanotechnology expert, for his constant, friendly, and wonderful cooperation. Also my former student, Michele Piantà, shared with me interest in the study of the FOCAL space mission and greatly helped me with the preparation of the tables and drawings that are contained in this book.

Turning now to the publication of the book, I would like to stress my particular gratitude to Prof. Richard A. Blade of the University of Colorado at Colorado Springs for his hospitality, for understanding the importance of the FOCAL space mission, and for editing and supporting the publication of this book by Springer-Praxis.

The author wishes to thank his Publisher, Clive Horwood, and Dr. John Mason, Chief Subject Advisory Editor of Praxis, for agreeing to publish this book which combines his two previous, smaller books (corresponding to Part I and Part II of this upgraded book). The author also wishes to thank Neil Shuttlewood of OPS Ltd. for re-setting the equations, certainly not an easy task.

And, of course, I am indebted to Alenia Aerospazio (Divisione Spazio), the space company where I have been employed since 1985, for the opportunity to do research.

Finally, I want to thank my parents, whose loving efforts, despite their modest financial means, encouraged and supported my studies over many years.

Claudio Maccone
Torino (Turin), Italy, August 24, 1997

Foreword

Wherever in space there are intelligent creatures like us, they will be driven to explore and understand our Universe, just as we do. We and they wish to glimpse the farthest depths of space with the greatest clarity allowed by the laws of nature. To this end, we build, at great expense, ever more powerful telescopes of all kinds on Earth, and now in space. As each civilization becomes more knowledgeable they will recognize, as we now have recognized, that they have been given a single great gift: a lens of such power that no reasonable technology could ever duplicate or surpass its power. This lens is civilization's star: in our case, the Sun. The gravity of each such star acts to bend space, and thus the paths of any wave or particle, in the end creating an image, just as familiar lenses do.

This lens can produce images which would take perhaps thousands of conventional telescopes to produce. It can produce images of the finest detail of distant stars and galaxies. Every civilization will discover this eventually, and surely will make the exploitation of such a lens a very high priority enterprise. One wonders how many such lenses are being used at this moment in time to scan the Universe, capturing a flood of information about both the physical and biological realities of our time.

We are just beginning to appreciate the power of such a lens, and to contemplate its exploitation. In this book is written the theory and potential performance of such a lens, not only for light and radio waves, but even for gravitational waves and neutrinos. But such a lens can only be utilized if a major challenge is met. This challenge derives from the fact that the magnificent images created by the lens for any electromagnetic waves, including light and radio, are formed at a distance of at least 550 AU from the Sun. Thus, at this very moment images of fantastic clarity and brightness are being created far out in space. The challenge to us is to send adequate detectors to these great distances to capture those images.

In this book Claudio Maccone describes this technological challenge, and how it might be met by the FOCAL mission. He points out that there are many scientific bonanzas in addition to the gravitational lens which will accrue from such a mission.

Here can be found the detailed technical requirements of the mission, as well as firm and accurate quantitative values for the imaging abilities of the lens.

This is a technical primer for what, in the long run, may be the most important space mission we will conduct. Readers of this book should wonder, as they read, how many times this same text has been created over the eons on the planets of other stars, and how many stars already are serving as the super-powerful eyes of the creatures of distant planets.

Frank D. Drake
Professor of Astronomy and Astrophysics
and Former Dean, Division of Natural Sciences,
University of California, Santa Cruz

Figures

Tables

Abbreviations and acronyms

AC	Aurora Collaboration
AOCS	Attitude and Orbit Control System
ASS	All-Sky Survey
AXAF	Advanced X-ray Astrophysics Facility
BAM	Bordered Autocorrelation Method
BETA 2	Software program
BPPP	Breakthrough Propulsion Physics Program
CBR	Cosmic Background Radiation
CMB	Cosmic Microwave Background
CMR	Cosmic Microwave Radiation
CNR	Consiglio Nazionale delle Ricerche
COBE	COsmic Background Explorer
DAstCom	Data on Asteroids and Comets
DBS	Direct Broadcast Satellite
EHOF	Extended Heliocentric Orbital Frame
ESA	European Space Agency
ET	ExtraTerrestrial
FFT	Fast Fourier Transform
FOCAL	Space mission
FT	Fourier Transform
GRO	Black holes within the Milky Way
HAE	High-Amplification Event
HIF	Heliocentric Inertial Frame
HPBW	Half-Power Beam Width
HR	Hertzsprung–Russell diagram
HST	Hubble Space Telescope
IAA	International Academy of Astronautics
INT	International Nanobiological Testbed

IR	InfraRed
IRAS	InfraRed Astronomical Satellite
ISP	InterStellar Probe
ISP STDT	ISP Science and Technology Definition Team
JBIS	*Journal of the British Interplanetary Society*
JPL	Jet Populsion Laboratory
KL	Karhunen–Loève
KLT	Karhunen–Loève Transform
LDR	LIGO Data Replicator
LMC	Large Magellanic Cloud
MAnE	Mission Analysis Environment
META	Project
MHD	MagnetoHydroDynamic
NASA	National Aeronautics and Space Administration
Nd:YAG	Neodymium-doped Yttrium Aluminum Garnet
NYU	New York University
PCA	Principal Components Analysis
PL	Period–Luminosity
PLL	Phase Locked Loop
RA	Right Ascension
SETI	Search for ExtraTerrestrial Intelligence
SETISAIL	Solar Sail space mission for SETI
SETV	Search for ExtraTerrestrial Visitation
SGF	Solar Gravitational Focus
SIM	Space Interferometry Mission
SIRTF	Space InfraRed Telescope Facility (Spitzer Space Telescope)
SMC	Small Magellanic Cloud
STAIF	Space Technology and Applications International Forum
SVD	Singular Value Decomposition
TAU	Thousand Astronomical Units (space mission)
TF	Tully–Fisher (relation)
UV	UltraViolet
VLBI	Very Large Baseline Interferometry
ZAMS	Zero Age Main Sequence

A brief overview of the Sun as a gravitational lens

Two foci of the gravitational lens of the Sun are predicted to exist by the general theory of relativity (see Figure 1).

(1) *A focus for electromagnetic waves*, located along a line starting at a distance of 550 AU (Astronomical Units)—that is, 3.17 light days, or 14 times the distance from the Sun to Pluto. It will be proved that any point beyond this minimal distance is a focus also. Thus, any spacecraft that can fly to 550 AU and beyond can take full advantage of the huge radio magnifications of any astronomical object lying on the other side of the Sun with respect to the spacecraft position.

(2) *A focus for gravitational waves and neutrinos*, located within the solar system at distances 22.45 AU and 29.59 AU (roughly between the distances of the orbits of Uranus and Neptune). The physical justification for the existence of this focus is that

   (a) a gravitational wave can *penetrate* through the Sun because such waves scatter significantly only in the presence of significant mass density rather than the charge on electrons which scatter electromagnetic waves; and

   (b) the bulk of the Sun's mass is more highly concentrated within its inner layers than within its outer layers (i.e., the Sun's radial density is maximum at the center and zero at the surface).

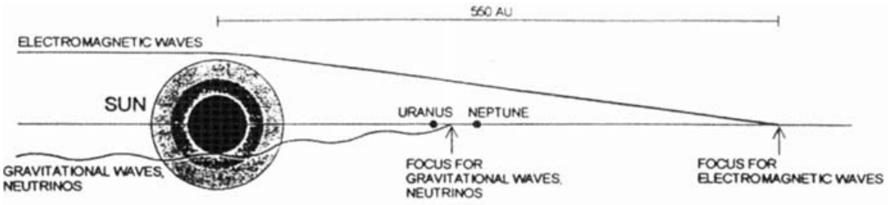


Figure 1. Comparison between the paths of electromagnetic waves and gravitational waves being focused by the gravitational field of the Sun.

THE PROPOSED FOCAL SPACE MISSION TO 550 AU AND BEYOND TO 1,000 AU

Part I of this book is devoted to studying the *FOCAL space mission*. By this we mean to let a spacecraft travel to 550 AU and beyond until it can detect the greatly magnified radio pictures of any radio source located on the other side of the Sun with respect to the spacecraft position.

In this book no study will be made of space missions intended to detect the gravitational waves focused by the Sun, since these missions were studied by David Sonnabend of JPL in 1979 (see [1]).

Part I opens with a review of the gravitational lens of the Sun based on Einstein's general relativity (Chapter 1). This is the so-called "naked" lens, namely the purely convergent lens due to gravity only (i.e., no Coronal effects are taken into account). It is then shown that

(1) The naked Sun's focus lies on a focal sphere of about 550 AU in radius.
(2) The gain (= magnification) of such a gravitational lens is huge: about 57 dB at the hydrogen line frequency of 1,420 MHz, and similarly for other frequencies of radioastronomical interest.
(3) The FOCAL spacecraft position must be very tightly ($\sim$100 km) aligned with the source of electromagnetic waves and the Sun center.

Chapter 2 summarizes the cruise science that could be profitably done by the FOCAL spacecraft while *en route* to 550 AU and beyond. The implications for cosmology (re-calibration of the size of the Universe), for nuclear processes associated with cosmogony, and for stellar parallaxes computation would be profound.

The FOCAL direction of exit out of the solar system is determined by which nearby star we wish to see magnified. In Chapter 3, we compute such a direction, as well as the relevant Sun flyby characteristics, for the nearest 50 stars. And in Chapter 4 we show that a time-optimized design of the spacecraft trajectory to reach a distance of 550 AU is vital for success. This is no simple task, however, since it must be achieved by either conventional chemical engines, and/or by nuclear-electric propulsion and/or by solar sailing (see [2]). Each propulsion system has its own advantages.

NEW TOPICS: GL-SETI, THE SOLAR CORONA EFFECTS "PUSHING" THE FOCUS FARTHER OUT, AND FOCUSED POWER PROPULSION

Part I of this book continues with five further chapters summarized hereafter.

(1) The newly re-written Chapter 6 deals with the use of the gravitational lenses of nearby stars to get magnified radio pictures of objects emitting electromagnetic waves from much beyond the nearby lensing stars. This effect could be used for either astrophysical investigations and for SETI. When used for SETI, it originates a new kind of SETI search, called "GL-SETI" (an acronym for

"Gravitational Lensing SETI"), in addition to the two traditional approaches of the SETI "targeted search" and "all-sky survey", as described by the SETI League President, Richard Factor, in Sections 6.2 and 6.3. His introduction is just qualitative, however. A more profound, mathematical investigation of GL-SETI reveals the diffficulties of probing this new research field, as shown in Section 6.4 by the new equation found by the author to relate the magnification of the lensing star, the distance of the ET transmitter, and the power of the ET transmitter. Finally, in Section 6.5 another application of GL-SETI is provided in case SETV, the Search for Extra-Terrestrial Visitation (within the solar system) is proved in the future to be real science, rather than pure science fiction.

(2) In Chapter 7 the first investigation (to the best of our knowledge) is presented about the gravitational lenses of the four nearest stars to the Sun in the Galaxy: Alpha Centauri A, B, C (Proxima), and Barnard's star. This mathematical description is intended to lead to a future, full description of the "radio bridge" between each one of these stars and the Sun, obtained by exploiting the gravitational lenses of all of them. These results show the enormous gains and energy savings that would affect the telecommunication link between the Sun and each one of these stars, were both gravitational lenses used at the same time. And this might indeed be used in the future in case human probes were able to reach, say, Alpha Centauri, and needed to keep their radio link with the Earth.

(3) In Chapter 8 is given the rather difficult mathematical theory of the Solar Corona effects on radio waves grazing the surface of the Sun. We start from the well-known Baumbach–Allen model of the Corona and then go over to finding the actual minimal focal distances by taking into account the frequency of the grazing waves. The result is that the action of the Corona counterbalances and even wins over the action of gravity by "pushing out" the actual minimal focal distance beyond 550 AU. This is "unfortunate" from the point of view of planning space missions intended to reach the actual focuses of the gravitational lens of the Sun, but is much more realistic than naively hoping for a focus at 550 AU, just as if there was no Corona around the Sun!

(4) The mathematical theory of the Corona given in Chapter 8 is then applied in Chapter 9 to the very special case of the Cosmic Microwave Background (CMB), whose Planck spectrum has its peak frequency at 160.378 GHz. For this frequency, we show that the Sun's minimal focal distance is 763 AU because of the Corona effects, rather than just 550 AU. We also point out that the NASA Interstellar Probe (ISP), currently under study at JPL and expected to be launched in 2010 in the direction of the incoming interstellar wind, would reach 763 AU around 2057. ISP could thus prove for the first time the physical existence of the Sun's gravity lens if suitably equipped with a photometer or a bolometer tuned to 160 GHz.

(5) Finally, an entirely new possibility, using the Sun's gravitational lens for propelling a spacecraft over to interstellar distances, is explored in Appendix E. Although we had to confine our description to the basic ideas only (i.e., without equations), this possibility could one day prove vital to help human expansion

into space up to the nearest stars and perhaps beyond. This exploitation of the Sun's gravity lens might possibly be achieved in either of two ways.

(a) exploiting the gravitational focusing of other stars on synthetic solar sails; or
(b) placing a solar power station on the opposite side of the Sun at distances higher than 550 AU and then moving it slowly toward the Sun. In this case, the minimal focal distance of the Sun's lens would be pushed farther and farther out, and so would be the sail located at its minimal focal distance. This second technique has one great advantage over the first one: it would work in all directions out from the Sun center, thus enabling a really free selection of the destination stars.

In conclusion, this book offers an even more comprehensive vision of the phenomenon of the Sun's gravity lens, now extended to the gravitational lenses of the nearest four stars in the Galaxy.

The author is convinced that the new, grand goal of humankind in the new millennium will be the exploration of our Galactic neighborhood and its human settlement. Hopefully, this book will be of use to future generations of space scientists and engineers willing to exploit the Sun's gravity lens and its unusual properties for applications to interstellar flight and to the scientific exploration of the Galaxy.

REFERENCES

[1] D. Sonnabend, "To the Solar Foci," Ph.D. thesis published by Caltech, *JPL Publication 79-18*, Pasadena, CA, June 1, 1979.

[2] G. Vulpetti, "3D High-Speed Escape Heliocentric Trajectories by All-Metallic-Sail Low-Mass-Sailcraft," paper IAA-L-0505 presented at the *Second IAA International Conference on Low-Cost Planetary Missions*, held April 16–19, 1996, at the Johns Hopkins University Applied Physics Laboratory, W. Farquhar (coordinator), and then printed in *Acta Astronautica*, **39**(1–4) (1996), 161–170.

Part I

Space missions to the Sun's gravity focus (550 to 1,000 AU)

1

So much gain at 550 AU

1.1 INTRODUCTION

The gravitational focusing effect of the Sun is one of the most amazing discoveries produced by the general theory of relativity. The first paper in this field was published by Albert Einstein in 1936 [1], but his work was virtually forgotten until 1964, when Sydney Liebes of Stanford University [2] gave the mathematical theory of gravitational focusing by a galaxy located between the Earth and a very distant cosmological object, such as a quasar.

In 1978 the first "twin quasar" image, caused by the gravitational field of an intermediate galaxy, was spotted by the British astronomer Dennis Walsh and his colleagues. Subsequent discoveries of several more examples of gravitational lenses eliminated all doubts about gravitational focusing predicted by general relativity.

Von Eshleman of Stanford University then went on to apply the theory to the case of the Sun in 1979 [3]. His paper for the first time suggested the possibility of sending a spacecraft to 550 AU from the Sun to exploit the enormous magnifications provided by the gravitational lens of the Sun, particularly at microwave frequencies, such as the hydrogen line at 1,420 MHz (21 cm wavelength). This is the frequency that all SETI radioastronomers regard as "magic" for interstellar communications, and thus the tremendous potential of the gravitational lens of the Sun for getting in touch with alien civilizations became obvious.

The first experimental SETI radioastronomer in history, Frank Drake (Project Ozma, 1960), presented a paper on the advantages of using the gravitational lens of the Sun for SETI at the *Second International Bioastronomy Conference* held in Hungary in 1987 [4], as did Nathan "Chip" Cohen of Boston University [5]. Non-technical descriptions of the topic were also given by them in their popular books [6, 7].

However, the possibility of planning and funding a space mission to 550 AU to exploit the gravitational lens of the Sun immediately proved a difficult task. Space scientists and engineers first turned their attention to this goal at the June 18, 1992, *Conference on Space Missions and Astrodynamics* organized in Turin, Italy, led by the author (see Figure 1.1). The relevant *Proceedings* were published in 1994 in the *Journal of the British Interplanetary Society* [8]. Meanwhile, on May 20, 1993, the author also submitted a formal Proposal to the European Space Agency (ESA) to fund the space mission design [9]. The optimal direction of space to launch the FOCAL spacecraft was also discussed by Jean Heidmann of Paris Meudon Observatory and the author [10], but it seemed clear that a demanding space mission like this should not be devoted entirely to SETI. Things like the computation of the parallaxes of many distant stars in the Galaxy, the detection of gravitational waves by virtue of the very long baseline between the spacecraft and the Earth, plus a host of other experiments would complement the SETI utilization of this space mission to 550 AU and beyond. The mission was dubbed "SETISAIL" in earlier papers [11], and "FOCAL" in the proposal submitted to ESA in 1993.

1.2 THE MINIMAL FOCAL DISTANCE OF 550 AU FOR ELECTROMAGNETIC WAVES

The well-known Schwarzschild solution to the Einstein field equations is the mathematical foundation upon which the theory of the gravitational lens of the Sun rests. From it a long string of formulas can be developed. Since those formulas are derived in standard textbooks, we shall simply rewrite without proofs the basic equations needed to explain the advantages provided by the gravitational lens of the Sun, suggesting the interested reader consult the basic references [8, 12, and 13] for the relevant mathematical demonstrations.

The geometry of the Sun gravitational lens is easily described: incoming electromagnetic waves (arriving, for instance, from the center of the Galaxy) pass *outside* the Sun and pass within a certain distance r of its center. Then the basic result following from the Schwarzschild solution shows that the corresponding *deflection angle* $\alpha(r)$ at the distance r from the Sun center is given by

$$\alpha(r) = \frac{4GM_{Sun}}{c^2 r}. \tag{1.1}$$

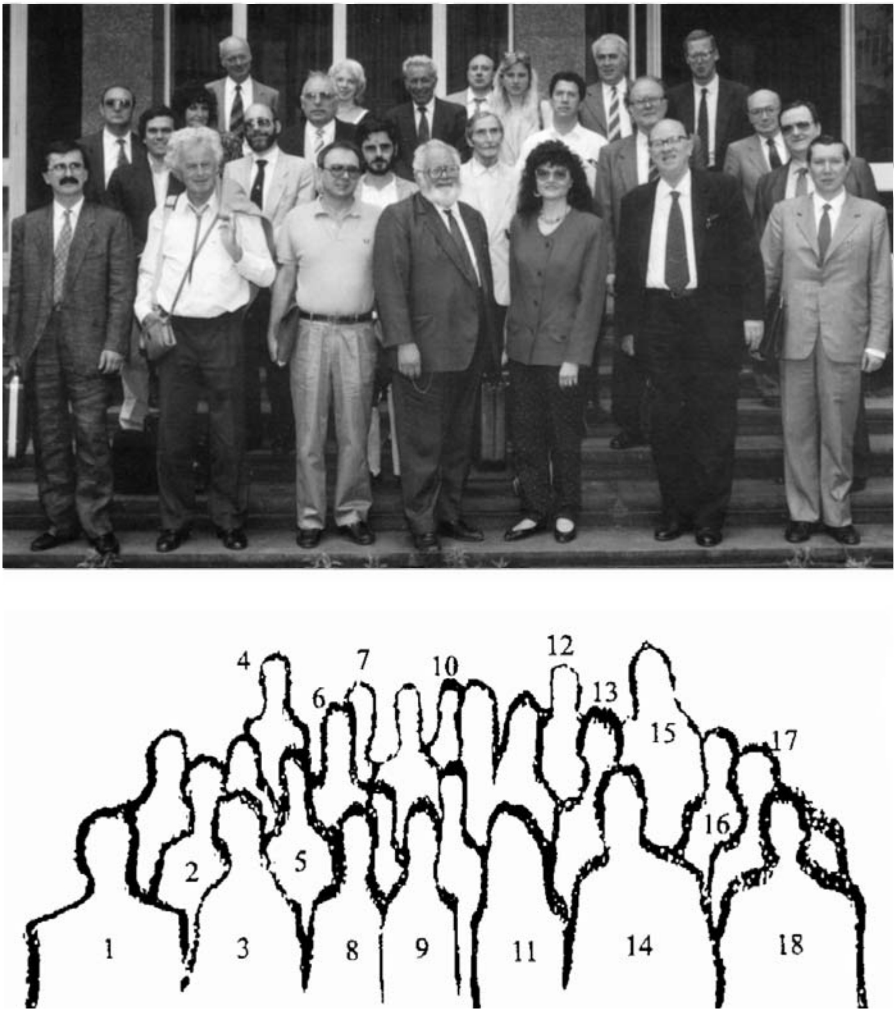
Figure 1.2 depicts the various parameters.

The light rays (i.e., electromagnetic waves) cannot pass through the Sun's interior (whereas gravitational waves and neutrinos can), so the largest deflection angle α occurs for those rays just grazing the Sun surface (i.e., for $r = r_{Sun}$). This yields the inequality

$$\alpha(r_{Sun}) > \alpha(r) \tag{1.2}$$

with

$$\alpha(r_{Sun}) = \frac{4GM_{Sun}}{c^2 r_{Sun}}. \tag{1.3}$$



1. Stelio Montebugnoli
2. Ed Belbruno
3. Jean Heidmann
4. Jorg Strobl
5. Gregory Matloff
6. Ettore Antona
7. Constance Bangs
8. Renato Pannunzio
9. Sigfrido Leschiutta
10. Giovanni Vulpetti
11. Maria Sarasso
12. Rinaldo Bertone
13. Franco Palutan
14. Vittorio Banfi
15. Mario Pasta
16. Federico Bedarida
17. Luciano Santoro
18. Claudio Maccone

Figure 1.1. First conference ever about the FOCAL space mission to 550 AU, held on June 18, 1992, at the Politecnico di Torino (Turin, Italy).

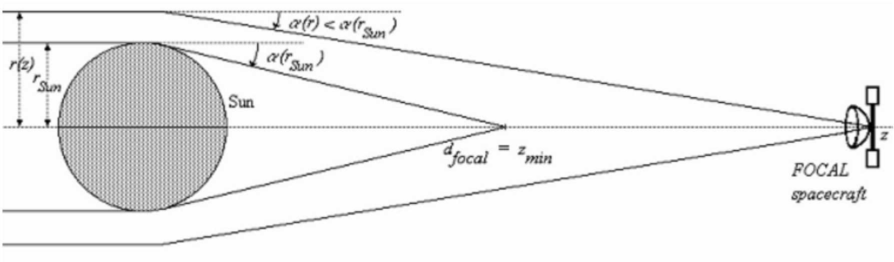


Figure 1.2. Basic geometry of the gravitational lens of the Sun, showing the minimal focal length and the FOCAL spacecraft position.

From the illustration it should be clear that the minimal focal distance d_{focal} is related to the tangent of the maximum deflection angle by the formula

$$\tan(\alpha(r_{Sun})) = \frac{r_{Sun}}{d_{focal}}. \tag{1.4}$$

Moreover, since the angle $\alpha(r_{Sun})$ is very small (its actual value is about 1.75 arcsec), the above expression may be rewritten by replacing the tangent by the small angle itself:

$$\alpha(r_{Sun}) \approx \frac{r_{Sun}}{d_{focal}}. \tag{1.5}$$

Eliminating the angle $\alpha(r_{Sun})$ between equations (1.3) and (1.5), and then solving for the minimal focal distance d_{focal}, one gets

$$d_{focal} \approx \frac{r_{Sun}}{\alpha(r_{Sun})} = \frac{r_{Sun}}{\frac{4GM_{Sun}}{c^2 r_{Sun}}} = \frac{c^2 r_{Sun}^2}{4GM_{Sun}}. \tag{1.6}$$

This basic result may also be rewritten in terms of the *Schwarzschild radius*

$$r_{Schwarzschild} = \frac{2GM_{Sun}}{c^2}, \tag{1.7}$$

yielding

$$d_{focal} \approx \frac{r_{Sun}}{\alpha(r_{Sun})} = \frac{r_{Sun}}{\frac{4GM_{Sun}}{c^2 r_{Sun}}} = \frac{r_{Sun}^2}{2r_{Schwarzschild}}. \tag{1.8}$$

Numerically, one finds

$$d_{focal} \cong 542 \text{ AU} \approx 550 \text{ AU} \approx 3.171 \text{ light days}. \tag{1.9}$$

This is the fundamental formula yielding the minimal focal distance of the gravitational lens of the Sun—that is, the minimal distance from the Sun's center that the FOCAL spacecraft must reach in order to get magnified radio pictures of whatever lies on the other side of the Sun with respect to the spacecraft position.

Furthermore, a simple, but very important consequence of the above discussion is that *all points on the straight line beyond this minimal focal distance are foci too*, because the light rays passing by the Sun farther than the minimum distance have smaller deflection angles and thus come together at an even greater distance from the Sun.

And the very important astronautical consequence of this fact for the FOCAL mission is that *it is not necessary to stop the spacecraft at 550 AU. It can go on to almost any distance beyond and focus as well or better*. In fact, the farther it goes beyond 550 AU the less distorted the collected radio waves by the Sun Corona fluctuations. The important problem of Corona fluctuations and related distortions is studied in Chapter 8 of this book.

We would like to add here one more result that is very important because it holds well not just for the Sun, but for all stars in general. This we'll do without demonstration; that can be found on p. 55 of [12]. Consider a spherical star with radius r_{star} and mass M_{star}, which will be called the "focusing star". Suppose also that a light source (i.e., another star or an advanced extraterrestrial civilization) is located at the distance D_{source} from it. Then ask: How far is the minimal focal distance d_{focal} on the opposite side of the source with respect to the focusing star center? The answer is given by the formula

$$d_{focal} = \frac{r_{star}^2}{\dfrac{4GM_{star}}{c^2} - \dfrac{r_{star}^2}{D_{source}}}. \tag{1.10}$$

This is the key to gravitational focusing for a pair of stars, and may well be the key to SETI in finding extraterrestrial civilizations. It could also be considered for the magnification of a certain source by any star that is perfectly aligned with that source and the Earth: the latter would then be in the same situation as the FOCAL spacecraft except, of course, it is located much farther out than 550 AU with respect to the focusing, intermediate star. Finally, notice that Equation (1.10) reduces to Equation (1.6) in the limit $D_{source} \to \infty$; that is, (1.6) is the special case of (1.10) for light rays approaching the focusing star from an infinite distance.

1.3 THE (ANTENNA) GAIN OF THE GRAVITATIONAL LENS OF THE SUN

Having thus determined the minimal distance of 550 AU that the FOCAL spacecraft must reach, one now wonders what's the good of going so far out of the solar system—that is, how much focusing of light rays is caused by the gravitational field of the Sun. The answer to such a question is provided by the technical notion of "antenna gain", which stems from antenna theory.

A standard formula in antenna theory relates the antenna gain, $G_{antenna}$, to the antenna effective area, $A_{effective}$, and to the wavelength λ or the frequency ν by virtue of the equation (see, e.g., [13], in particular p. 6-117, eq. (6-241)):

$$G_{antenna} = \frac{4\pi A_{effective}}{\lambda^2}. \tag{1.11}$$

Now, assume the antenna is circular with radius $r_{antenna}$, and assume also a 50% efficiency. Then, the antenna effective area is obviously given by

$$A_{effective} = \frac{A_{physical}}{2} = \frac{\pi r_{antenna}^2}{2}. \tag{1.12}$$

Substituting this back into (1.11) yields the antenna gain as a function of the antenna radius and of the observed frequency

$$G_{antenna} = \frac{4\pi A_{effective}}{\lambda^2} = \frac{2\pi A_{physical}}{\lambda^2} = \frac{2\pi^2 r_{antenna}^2}{\lambda^2} = \frac{2\pi^2 r_{antenna}^2}{c^2}\nu^2. \tag{1.13}$$

The important point here is that *the antenna gain increases with the square of the frequency*, thus favoring observations on frequencies as high as possible.

Is anything similar happening for the Sun's gravitational lens also? *Yes* is the answer, and the "gain" (one maintains this terminology for convenience) of the gravitational lens of the Sun can be proved to be

$$G_{Sun} = 4\pi^2 \frac{r_{Schwarzschild}}{\lambda} \tag{1.14}$$

or, invoking the expression (1.7) of the Schwarzschild radius

$$G_{Sun} = \frac{8\pi^2 G M_{Sun}}{c^2} \cdot \frac{1}{\lambda} = \frac{8\pi^2 G M_{Sun}}{c^3} \cdot \nu \tag{1.15}$$

The mathematical proof of equation (1.14) is difficult to achieve.

The author, unsatisfied with the treatment of this key topic given in [1, 3, and 13], turned to three engineers of the engineering school in his home town, Renato Orta, Patrizia Savi, and Riccardo Tascone. To his surprise, in a few weeks they provided a full proof of not just the Sun gain formula (1.14), but also of the focal distance for rays originated from a source at a finite distance, Equation (1.10). Their proof is fully described in [12], and is based on the aperture method used to study the propagation of electromagnetic waves, rather than on ray optics.

Using the words of these three authors' own Abstract, they have "computed the radiation pattern of the [spacecraft] Antenna + Sun system, which has an extremely high directivity. It has been observed that the focal region of the lens for an incoming plane wave is a half line parallel to the propagation direction starting at a point [550 AU] whose position is related to the blocking effect of the Sun disk (Figure 1.2). Moreover, a characteristic of this thin lens is that its gain, defined as the magnification factor of the antenna gain, is constant along this half line. In particular, for a wavelength of 21 cm, this lens gain reaches the value of 57.5 dB. Also, a measure of the transversal extent of the focal region has been obtained. The performance of this radiation system has been determined by adopting a thin lens model which introduces a phase factor depending on the logarithm of the impact parameter of the incident

Table 1.1. Table showing the gain of the Sun's lens alone, the gain of a 12-meter spacecraft (S/C) antenna and the combined gain of the Sun + S/C Antenna system at five selected frequencies is important in radioastronomy. (Here, as throughout the book, **bold type** indicates greater importance).

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	**1,420 MHz**	**327 MHz**	**1.6 GHz**	**5 GHz**	**22 GHz**
Wavelength λ	21 cm	92 cm	18 cm	6 cm	1.35 cm
S/C Antenna Beamwidth	1.231 deg	5.348 deg	1.092 deg	0.350 deg	0.080 deg
Sun Gain	57.4 dB	51.0 dB	57.9 dB	62.9 dB	69.3 dB
12-meter Antenna S/C Gain	42.0 dB	29.3 dB	43.1 dB	53.0 dB	65.8 dB
Combined Sun + S/C Gain	**99.5 dB**	**80.3 dB**	**101.0 dB**	**115.9 dB**	**135.1 dB**

rays. Then the antenna is considered to be in transmission mode and the radiated field is computed by asymptotic evaluation of the radiation integral in the Fresnel approximation."

1.4 THE COMBINED, TOTAL GAIN UPON THE FOCAL SPACECRAFT

One is now able to compute the Total Gain of the Antenna + Sun system, which is simply obtained by multiplying Equations (1.13) and (1.15)

$$G_{Total} = G_{Sun} G_{Antenna} = \frac{16\pi^4 G M_{Sun} r_{Antenna}^2}{c^5} \nu^3. \tag{1.16}$$

Since the total gain increases with the cube of the observed frequency, it favors electromagnetic radiation in the microwave region of the spectrum. Table 1.1 shows numerical data provided by Equations (1.15) and (1.13) for five selected frequencies: the hydrogen line at 1,420 MHz and the four frequencies that the Quasat radioastronomy satellite planned to observe, had it been built jointly by ESA and NASA as planned before 1988 [14] (the definition of dB is $N\,\text{dB} = 10\,\text{Log}_{10}(N) = 10\ln(N)/\ln(10)$).

1.5 THE IMAGE SIZE AT THE SPACECRAFT DISTANCE z

The next important notion to understand is the size of the image of an infinitely distant object created by the Sun lens at the current spacecraft distance z from the Sun ($z > 550$ AU). We may define such an image size as the distance from the focal axis

Table 1.2. Table showing image sizes for a 12-meter antenna, located at distances of 550 AU, 800 AU, and 1,000 AU from the Sun, for the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	**1,420 MHz**	**327 MHz**	**1.6 GHz**	**5 GHz**	**22 GHz**
Wavelength λ	21 cm	92 cm	18 cm	6 cm	1.35 cm
Image size (down 6 dB) **at 550 AU**	**2.498 km**	**10.847 km**	**2.217 km**	**0.709 km**	**0.161 km**
Image size (down 6 dB) **at 800 AU**	**3.033 km**	**13.169 km**	**2.691 km**	**0.861 km**	**0.196 km**
Image size (down 6 dB) **at 1,000 AU**	**3.391 km**	**14.724 km**	**3.009 km**	**0.963 km**	**0.219 km**

(i.e., from the spacecraft straight trajectory) at which the gain is down 6 dB. The formula for this (proven in [8]) is

$$r_{6dB} = \frac{\lambda}{\pi^2}\sqrt{\frac{z}{2r_{Schwarzschild}}} = \frac{c}{2\pi^2\sqrt{GM_{Sun}}}\lambda\sqrt{z} = \frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{\sqrt{z}}{\nu}. \tag{1.17}$$

Thus, the image size *increases* with spacecraft distance z from the Sun. Table 1.2 provides a quantitative feeling of how the image size changes with the spacecraft distance from the Sun.

It is clear that these image size values are very small compared with the spacecraft distance from the Earth. This means that if we want to observe a certain point-source in the sky, the alignment between this source, the Sun, and the spacecraft position must be extremely precise. In fact, the spacecraft tracking must exceed by far what we are able to do within the solar system today. However, this is not true if the source we want to observe is the center of the Galaxy, which is a very broad source: slight changes in the spacecraft trajectory (say, in a spreading spiral shape as described in Appendix D) would enable us to gradually see much of the Galactic center at the huge resolution provided by the gravitational lens of the Sun.

1.6 REQUIREMENTS ON THE IMAGE SIZE AND ANTENNA BEAMWIDTH AT THE SPACECRAFT DISTANCE z

There are two "geometrical" requirements that must be fulfilled in order that the combined lens system Sun + FOCAL spacecraft antenna can work at best:

(1) Size requirement: the full antenna dish of the FOCAL spacecraft must fall well inside the cylindrical region centered along the focal axis and having radius equal

to r_{6dB}. That is, the spacecraft feed–dish radius must be considerably smaller than r_{6dB}

$$r_{Antenna} \ll r_{6dB} = \frac{c}{2\pi^2\sqrt{GM_{Sun}}}\lambda\sqrt{z} = \frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{\sqrt{z}}{\nu}. \tag{1.18}$$

(2) Angle requirement: the impact–radius circle around the Sun within which electromagnetic waves are focused towards the FOCAL spacecraft must fall well within the antenna beamwidth of the FOCAL spacecraft. In slightly more technical terms, the Half-Power Beam Width (HPBW, the angular width of the main lobe of the spacecraft antenna at the half-power level) should be considerably greater than the angle subtended at the spacecraft distance by twice the incident ray impact radius at the Sun

$$HPBW \gg 2\alpha(r) = \frac{8GM_{Sun}}{c^2 r}. \tag{1.19}$$

Tables 1.3 and 1.4 show that both these conditions are fulfilled at the three FOCAL distances from the Sun for the five selected frequencies, respectively.

1.7 ANGULAR RESOLUTION AT THE SPACECRAFT DISTANCE *z*

The notion of angular resolution of the Sun lens is relevant to the discussion. Angular resolution is simply defined as the ratio of the image size (at the spacecraft distance z from the Sun) to that distance z. From Equation (1.17),

$$\theta_{resolution}(z) = \frac{r_{6dB}}{z} = \frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{1}{\sqrt{z}\nu}. \tag{1.20}$$

Clearly the angular resolution also depends on the spacecraft distance z from the Sun,

Table 1.3. Table showing image sizes vs. the antenna radius for a 12-meter antenna located at various distances from the Sun for the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	**1,420 MHz**	**327 MHz**	**1.6 GHz**	**5 GHz**	**22 GHz**
Wavelength λ	21 cm	92 cm	18 cm	6 cm	1.35 cm
Image size at 550 AU vs. Antenna Radius	**2.498 km ≫6 m**	**10.85 km ≫6 m**	**2.22 km ≫6 m**	**0.71 km ≫6 m**	**0.16 km ≫6 m**
Image size at 800 AU vs. Antenna Radius	**3.03 km ≫6 m**	**13.17 km ≫6 m**	**2.69 km ≫6 m**	**0.86 km ≫6 m**	**0.20 km ≫6 m**
Image size at 1,000 AU vs. Antenna Radius	**3.39 km ≫6 m**	**14.72 km ≫6 m**	**3.01 km ≫6 m**	**0.96 km ≫6 m**	**0.22 km ≫6 m**

Table 1.4. Table showing HPBW vs. aspect angle of the Sun for a 12-meter antenna located at various distances from the Sun for the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	**1,420 MHz**	**327 MHz**	**1.6 GHz**	**5 GHz**	**22 GHz**
Wavelength λ	21 cm	92 cm	18 cm	6 cm	1.35 cm
HPBW at 550 AU vs. 2α	$\mathbf{1.23154^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{5.34798^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{1.09299^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{0.34976^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{0.07949^\circ \gg 1.5 \times 10^{-7\circ}}$
HPBW at 800 AU vs. 2α	$\mathbf{1.23154^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{5.34798^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{1.09299^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{0.34976^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{0.07949^\circ \gg 1.5 \times 10^{-7\circ}}$
HPBW at 1,000 AU vs. 2α	$\mathbf{1.23154^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{5.34798^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{1.09299^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{0.34976^\circ \gg 1.5 \times 10^{-7\circ}}$	$\mathbf{0.07949^\circ \gg 1.5 \times 10^{-7\circ}}$

and it actually improves (i.e., it gets smaller) as long as the distance increases beyond 550 AU.

Table 1.5 gives angular resolutions for the same three distances at the same five frequencies. Let us take a moment to ponder over these numbers. The best angular resolutions achieved so far, in visible light, were obtained by the European astrometric satellite *Hipparcos*, launched in 1989, and dismissed from service in 1993. Though the apogee kick motor of Hipparcos didn't fire, forcing technicians to take

Table 1.5. Table showing angular resolution for three spacecraft distances (550 AU, 800 AU, and 1,000 AU), at the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	**1,420 MHz**	**327 MHz**	**1.6 GHz**	**5 GHz**	**22 GHz**
Wavelength λ	21 cm	92 cm	18 cm	6 cm	1.35 cm
Angular resolution at 550 AU S/C distance	6.3458×10^{-6} arcsec	2.7557×10^{-5} arcsec	5.6319×10^{-6} arcsec	1.8022×10^{-6} arcsec	4.0959×10^{-7} arcsec
Angular resolution at 800 AU S/C distance	5.2267×10^{-6} arcsec	2.2697×10^{-5} arcsec	4.6387×10^{-6} arcsec	1.4844×10^{-6} arcsec	3.3736×10^{-7} arcsec
Angular resolution at 1000 AU S/C distance	4.6749×10^{-6} arcsec	2.0301×10^{-5} arcsec	4.1490×10^{-6} arcsec	1.3277×10^{-6} arcsec	3.0174×10^{-7} arcsec

Table 1.6. Table showing the spatial resolutions for astronomical objects at selected distances from the Sun for a 12-meter spacecraft antenna.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	**1,420 MHz**	**327 MHz**	**1.6 GHz**	**5 GHz**	**22 GHz**
Wavelength λ	21 cm	92 cm	18 cm	6 cm	1.35 cm
Resolution at **0.5 lt-yr (Oort Cloud)**	145 km	632 km	129 km	41 km	9 km
Resolution at **4.29 lt-yr (α Centauri)**	1,248 km	5,422 km	1,108 km	355 km	81 km
Resolution at **10 pc = 32.6 lt-yr**	9,576 km	41,583 km	8,499 km	2,719 km	618 km
Resolution at **100 pc = 326 lt-yr**	9,575 km	415,833 km	84,986 km	2,719 km	6,180 km
Resolution at **1 kpc = 3,260 lt-yr**	957,58 km = 0.006 AU	**4,158,330** km = 0.028 AU	849,861 km = 0.005 AU	271,955 km = 0.001 AU	61,808 km = 0.0004 AU
Resolution at **10 kpc = 32,600 lt-yr (Galactic Center)**	9,575,870 km = 0.06401 AU	41,583,000 km = 0.27797 AU	8,498,610 km = 0.05681 AU	2,719,550 km = 0.01818 AU	618,082 km = 0.00413 AU
Resolution at **50 kpc = 160,000 lt-yr (Magellanic Clouds)**	4.78794×10^7 km = 0.32006 AU	2.07917×10^8 km = 1.38984 AU	4.2493×10^7 km = 0.28405 AU	1.3597×10^7 km = 0.0909 AU	3.0903×10^6 km = 0.02066 AU
Resolution at **613 kpc = 2 Mlt-yr (Andromeda Galaxy M31)**	5.82123×10^8 km = 3.89125 AU	2.52788×10^9 km = 16.8978 AU	5.16631×10^8 km = 3.45349 AU	1.65322×10^8 km = 1.10512 AU	3.75732×10^7 km = 0.25116 AU
Resolution at **18,406 pc = 60 Mlt-yr ("Jet" Galaxy M87 in Virgo)**	1.74636×10^{10} km = 116.738 AU	7.5836×10^{10} km = 506.934 AU	1.5499×10^{10} km = 103.605 AU	4.95968×10^{10} km = 33.1535 AU	1.1272×10^9 km = 7.53488 AU

Table 1.6 (*cont.*)

Line	Neutral hydrogen		OH radical		H_2O
Resolution at **3.07 Mkpc = 10 billion lt-yr (radius of the Universe)**	2.91059 × 10^{12} km = 19,456 AU = 0.30765 lt-yr	1.26393 × 10^{13} km = 84,489 AU = 1.33598 lt-yr	2.58316 × 10^{12} km = 17,267 AU = 0.27304 lt-yr	8.2661 × 10^{11} km = 5525.58 AU = 0.08737 lt-yr	1.8786 × 10^{11} km = 1255.81 AU = 0.01986 lt-yr

the software originally written for a circular geostationary orbit and re-write it for a highly elliptical orbit, the Hipparcos mission has proven a success. The resolutions achieved by Hipparcos are at a level of 2 milliseconds of arc precision. Checking this figure against the above table, one can see that the gravitational lens of the Sun plus a (modest) 12-meter antenna would improve the angular resolution by about *three orders of magnitude* (at radio frequencies).

1.8 SPATIAL RESOLUTION AT SPACECRAFT DISTANCE *z*

Finally, let us turn to the *spatial resolution*, simply called the *resolution* hereafter, of an astronomical object we want to examine with the help of the gravitational lens of the Sun. It is defined by

$$R_{Object} = d_{Sun-Object}\theta_{resolution} = d_{Sun-Object}\frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{1}{\sqrt{z\nu}}. \tag{1.21}$$

Again, beyond 550 AU the resolution improves (i.e., the angle gets smaller) slowly with increasing spacecraft distance from the Sun. Table 1.6 shows the spatial resolutions for a very wide range of distances, from the Oort Cloud to cosmological objects like quasars.

1.9 REFERENCES

[1] A. Einstein, "Lens-like Action of a Star by the Deviation of Light in the Gravitational Field," *Science*, **84** (1936), 506–507.
[2] S. Liebes, Jr., "Gravitational Lenses," *Physical Review*, **133** (1964), B835–B844.
[3] V. Eshleman, "Gravitational Lens of the Sun: Its Potential for Observations and Communications over Interstellar Distances," *Science*, **205** (1979), 1133–1135.

[4] F. Drake, "Stars as Gravitational Lenses," *Proceedings of the Bioastronomy International Conference*, held in Balatonfüred, Hungary, June 22–27, 1987, G. Marx (editor), pp. 391–394.

[5] N. Cohen, "The Pro's and Con's of Gravitational Lenses in CETI," *Proceedings of the Bioastronomy International Conference*, held in Balatonfüred, Hungary, June 22–27, 1987, G. Marx (editor), p. 395.

[6] F. Drake, and D. Sobel, *Is Anyone Out There?*, Delacorte Press, New York, 1992, see in particular pp. 230–234.

[7] N. Cohen, *Gravity's Lens*, Wiley Science Editions, New York, 1988.

[8] C. Maccone, "Space Missions Outside the Solar System to Exploit the Gravitational Lens of the Sun," in *Proceedings of the International Conference on Space Missions and Astrodynamics* held in Turin, Italy, June 18, 1992; C. Maccone (editor), *Journal of the British Interplanetary Society*, **47** (1994), 45–52.

[9] C. Maccone, "FOCAL, A New Space Mission to 550 AU to Exploit the Gravitational Lens of the Sun," A proposal for an M3 space mission submitted to the European Space Agency (ESA) on May 20, 1993, on behalf of an international team of scientists and engineers. Later (October 1993) re-considered by ESA within the "Horizon 2000 Plus" space missions plan.

[10] J. Heidmann, and C. Maccone, "AstroSail and FOCAL: two extraSolar System missions to the Sun's gravitational focuses," *Acta Astronautica*, **35** (1994), 409–410.

[11] C. Maccone, "The SETISAIL Project," in "Progress in the Search for Extraterrestrial Life," *Proceedings of the 1993 Bioastronomy Symposium* held at the University of California at Santa Cruz, August 16–20, 1993; G. Seth Shostak (editor), *Astronomical Society of the Pacific Conference Series*, **74** (1995), 407–417.

[12] R. Orta, P. Savi, and R Tascone, "Analysis of Gravitational Lens Antennas," in *Proceedings of the International Conference on Space Missions and Astrodynamics* held in Turin, Italy, June 18, 1992; C. Maccone (editor), *Journal of the British Interplanetary Society*, **47** (1994), 53–56.

[13] John D. Kraus, *Radio Astronomy*, 2nd edition, Cygnus-Quasar Books, Powell, Ohio, 1966, in particular pp. 6-115 through 6-118.

[14] A. Hawkyard, and A. Anselmi, "QUASAT Industrial Phase A Study," Executive Summary, *Aeritalia GSS Report QS-RP-A1-0004* (1988).

2

Scientific investigations along the way to 550 AU

2.1 INTRODUCTION

Completely aside from the investigations using the gravitational focusing of the Sun, there are a great many scientific investigations that could be carried out by the FOCAL mission on the basis of being a deep space probe. The innumerable scientific advantages of such a deep space probe have already been pointed out by the *TAU* deep space mission proposed by JPL in the 1980s. TAU stands for "one thousand astronomical units," and the seminal paper, titled "Preliminary Scientific Rationale for a Voyage to a Thousand Astronomical Units" was compiled by Maria Ines Etchegaray in JPL Publication 87-17 (May 15, 1987). This chapter, as well as the author's proposal to ESA for the FOCAL space mission, follows that publication closely.

A 1 MW *nuclear-powered electric propulsion* (*NEP*) system forms the basis for achieving the high velocities required. A solar system escape velocity of 106 km/s is needed to propel the TAU vehicle to 1,000 AU in 50 years. The NEP system must accelerate the vehicle for about 10 years before this velocity is attained because of the extremely low thrust of the xenon-fueled ion engines. At the end of the thrusting phase the NEP system is jettisoned to allow the TAU spacecraft and science experiments to coast out to 1,000 AU.

Aside from the new propulsion system, an important technology proposed for TAU is the advanced optical communication system for transmitting science data to Earth. A 1-meter optical telescope combined with a 10 W laser transponder could transmit 20 kbps from 1,000 AU to a 10-meter Earth orbit–based telescope.

TAU could provide astrometrists with a 1,000 AU baseline, a unique opportunity for making parallax measurements of stars in a volume of space 10^8 times what is available from Earth orbit, covering the full range of the Galaxy and the Magellanic Clouds. Depending on the exact location of the heliopause and the progress of the Voyager spacecraft, TAU might permit the first *in situ* measurements of the plasma environment across the heliosphere and into the tenuous interstellar medium, sampling the galactic magnetic field, energetic particles, gas, dust, and plasma environment.

The parallactic measurements would lead to improved models of astrophysical phenomena and would combine with Earth-based and Earth-orbiting astronomical measurements to provide more accurate measurements of the distances to objects of interest for proper data reduction.

An internal JPL/Caltech TAU "thinkshop" was held September 29, 1986, at the Jet Propulsion Laboratory as a kickoff meeting for developing the scientific rationale of this type of mission. Following the thinkshop, a series of semi-weekly science presentations were made to the TAU study team covering all areas of science that would benefit from this mission. The proposed experiments were evaluated for scientific worth, feasibility, and uniqueness to a TAU mission. Much more science could be accomplished than we list in this chapter; we list only research that is unique to the various characteristics provided by this type of mission.

2.2 VISIBLE AND INFRARED STELLAR PARALLAXES

A FOCAL deep space mission (or equivalently, a TAU deep space mission) has special significance to optical astronomy. Given our current 2 AU baseline to calculate trigonometric parallax, we can accurately measure distances no farther than about 100 pc. FOCAL could, by virtue of its 500 AU to 800 AU baseline, expand the measurement of distances out to perhaps 30 kpc or 40 kpc, which for the first time brings the entire Galaxy to the reach of our astronomical caliper. This expanded scale would enable a number of exciting astrophysical studies heretofore impossible. The astrometric capability of FOCAL is assumed to have a 10% accuracy in measuring distance and a 0.5-milliarcsecond star position accuracy.

In addition to the wealth of astrophysical science made possible on FOCAL, this new distance scale would have a profound impact on science from past and future astronomical observatories. Several astronomical projects have flown or are being developed or contemplated including *IRAS*, *Hubble Space Telescope* (*HST*), *SIRTF*, *AXAF*, *LDR*, *GRO*, and several explorers including *COBE*, *FUSE*, *EUVE*, and *QUASAT*. (We won't explain each acronym since the interested reader can find them at the relevant Wikipedia sites.) All of these missions, several of which exploit new wavelength regimes, require accurate distances from optical astronomy in order to fully realize the potential of their measurements. FOCAL could provide these distances across the Galaxy.

2.2.1 Uncertainty in the expansion rate of the Universe

The period–luminosity (PL) relation

Since 1960 the estimate of the expansion rate of the Universe, called the *Hubble constant*, has varied between 50 km/s and 100 km/s per 100 kpc. The estimate of the

Hubble constant is based upon our understanding of the astronomical distance scale, which is based upon the period–luminosity (PL) relation of variable stars such as the Cepheids, RR Lyraes, and Miras. FOCAL would permit a direct measurement of the distance to these stars. At present the use of photometric parallaxes requires assumptions of the level of local extinction to be made to calculate both the distance and the absolute luminosity of a star. A current method of calibrating the PL relation is to obtain distances to nearby star clusters by trigonometric parallax; then match the main sequence of more distant clusters that contain Cepheids to infer distances to them. This is called *Zero Age Main Sequence* (*ZAMS*) fitting. In this way the PL relation for Cepheids can be calibrated. The Cepheids can then be used to calibrate the *Tully–Fisher* (TF) relation, which relates the speed of rotation (from the width of the 21 cm emission line of neutral atomic hydrogen) with the absolute magnitude of a galaxy. The TF relation enables the measurement of distance to very distant galaxies and eventually the determination of the Hubble constant. FOCAL could remove the necessity for the ZAMS fit.

FOCAL could determine accurate parallax distances to intrinsically variable stars within our Galaxy (e.g., Cepheids, RR Lyrae, and Miras). It would be possible to select objects with varying chemical composition to search for other parameters that may be involved in the period–luminosity relation such as chemical composition, or some other higher order effect.

2.2.2 Age of the Galaxy

Ages of the globular clusters

FOCAL could take the parallax of a statistically significant number of globular cluster stars. Having these "exact", trigonometrically derived distances, one could then directly fit the theoretical isochrones for the evolution of the stars to the main sequence of the globular clusters that we see. The age to be determined is the time the star takes to consume its core hydrogen. The observation is to determine the luminosity and mass of stars that have just completed their core hydrogen burning stage. Knowing the age of the oldest globular cluster would place a lower limit on the age of the Universe.

2.2.3 Galactic structure

The gravitational mass of the Galaxy

FOCAL could map the rotation curve of the Galaxy. A mass model derived from this study could be used to refine theories of the formation of the Galaxy. The present method of determining the mass of the Galaxy is to do photometric parallaxes on A-class stars. However, this method contains large errors provided by assumptions of the interstellar extinction, which would not enter into trigonometric parallax measurements such as FOCAL would provide. Since FOCAL could measure the distances directly, it could look at late B-class or early A-class type stars that have nearly circular orbits and are at low galactic inclinations. These would be the best "test particles''. The measured distances would then be combined with Earth-based measurements of the stars' velocities. With a calculation of the velocity dispersion and distance, one could determine the galactic mass as a function of radius.

Dynamic temperature of the disk and halo of the Galaxy

FOCAL could measure the parallax distance to K-class giant stars, which are well distributed throughout the Galaxy. The ground-based measurements of the velocity dispersion of these stars with the FOCAL-measured distances would give information on the temperature of the Galaxy as a function of Galactic radius. A measure of the dynamical temperature would provide a better understanding of the structure and stability of the disk and the structure of the halo of our Galaxy. Present Earth-based methods rely on photometric parallaxes of K giants; however, these stars lie on a steeply sloping branch of the HR diagram, and thus the measure is unreliable. FOCAL would provide a much more accurate measure of the distances. The 50-year span of the FOCAL mission would also provide very accurate information on the proper motion of these K giant stars to complement Earth-based studies, and thus one could obtain all three components of the velocity dispersion relation [1].

Distance to the Galactic center

FOCAL would measure the visual and IR parallax distance to M-giant stars with as small an impact parameter as possible, but yet not located so close to the center of the Galaxy that the stars have the highest extinction rates. M giants are concentrated towards the center of the Galaxy. IR detectors in the focal plane of the telescope could permit the parallactic measurements to be made for those stars with the smallest impact parameters where the extinction in the visible is highest (up to 20 visible magnitudes).

The present day method measures distances to RR Lyrae variable stars by the use of the PL curve. RR Lyrae are not strongly concentrated towards the center of the Galaxy as are the M-class giant stars [1].

2.2.4 Stellar evolution

Early stellar evolution and resolution of cloud complexes

Improvement on the present understanding of early stages of stellar evolution could be provided by measuring the visible and IR parallaxes of stars in molecular cloud complexes where stars are born and in stellar associations where they have been for a while. FOCAL could provide distances to these objects free of extinction errors. Using an infrared array in the focal plane of the detector, FOCAL would be able to resolve a cloud complex along the line of sight, and thus model the three-dimensional structure of a star-forming region. A large sample of cloud complexes may be found at about 1 kpc, with Orion even closer. FOCAL's accuracy would be sufficient to resolve cloud complexes in depth. Cloud complexes are a few parsecs in size and thus a 1 pc resolution at 1 kpc distance would be required. With an accuracy of 10%, this comes to a measurement accuracy of one part in ten [1].

The initial mass function

FOCAL would help to determine how the masses of stars are distributed as they are born, looking at stellar masses from 0.1 solar mass to 100 solar masses. Is this function broad, narrow or bimodal? With FOCAL taking their parallax distances, differentiation between cluster stars and background stars that do not belong to the cluster is possible [1].

Binary star evolution

Determination of the parallax distance to binary stars is another project FOCAL could accomplish. Presently photometric parallaxes are used to determine the distances of binary stars. This method, however, must assume that the star is a normal star, not taking into account possible mass transfer between the two stars. The study of binary stars has evolved late in stellar evolution theory. Observational constraints to study these objects are badly needed [1].

Late stages of stellar evolution

FOCAL could aid the study of how stars of different masses terminate their stages of nuclear burning. For example, Wolf–Rayet stars have some kind of high-mass termination point while planetary nebulae have some kind of low-mass termination point. One cannot be certain of these points because, for example, the planetary nebula distance scale is uncertain by a factor of 2. Accurate distances would help greatly in studying these late stages, and FOCAL could provide them for a much larger sample than is currently available [1].

Study of peculiar objects

a. Carbon stars

An IR detector in the focal plane of FOCAL's telescope would permit measurement of the parallax distances to carbon stars. These stars, though invisible to FOCAL in the optical range, are bright in the red wavelengths [2].

b. Young protostars

At the moment only estimates are available of the distances to these objects. Parallactic distance measurements by FOCAL would permit a more accurate determination of their luminosity and thus of their age. The concern regarding these objects is to determine if they have completed their assemblage of mass or if they are still accreting material to build their final mass [3].

2.2.5 Targets of opportunity

As other Earth-based and Earth-orbiting astronomical observatories come to life, many new objects will be discovered in our Galaxy and in other galaxies. Over the life of FOCAL there would be requests for accurate distance determination to these objects, as well as morphological studies of objects (i.e., cloud complexes), which would be permitted by the large 500 AU to 800 AU baseline. Below is a list of presently known objects for which distance knowledge is needed. Many more objects would be expected to be added to this list over the lifetime of FOCAL.

Objects with unknown distance

O–B associations
Regions of high polarization (filamentary)
Nuclei of planetary nebulae
Nova during observation
Nova remnants
Supernovae

Pulsars

Open clusters, young and very old
Local Group—dwarf galaxies
Local Group—M31 bright members
M33
M81
S Doradus—LMC
41 Tucanae—LMC
Supernova remnants
Intergalactic objects in LMC and SMC
Infrared objects discovered by IRAS, SIRTF, LDR
New interesting objects identified by HST, AXAF, LDR, GRO.

2.3 ASTROPHYSICS, ASTRONOMY, AND COSMOLOGY

2.3.1 Interstellar gases

It is believed that the solar system may be on the edge of a tenuous cloud of interstellar matter [4]. FOCAL could study and characterize the properties of the interstellar cloud through which the solar system is traversing. In particular, it could characterize the composition, kinetic energy, and spectra of the cloud's particles and gases.

Cosmic abundance of hydrogen

FOCAL would be able to determine the number density of hydrogen and helium with heliocentric distance. Since the Lyman-alpha background from the star field is negligible, the backscatter from the solar emission of Lyman-alpha can be used to determine the cosmic abundance of hydrogen inside the heliosphere and in the interstellar medium. A 30% loss of the interstellar influx of atomic hydrogen is predicted to occur at the heliopause. This model is used to explain the discrepancy between the solar system ratio of H/He of 7/1 compared with the cosmic abundance of 12/1. The heliopause is transparent to helium. A measure of the radial dependence of the number density, and especially its variation across the heliopause transition region, is a fundamental measure to define the interaction at this boundary [5, 6].

Abundance and distribution of He^3/He^4, D/H, and Li^6/Li^7

Knowing the elemental ratios of He^3/He^4, D/H, and Li^6/Li^7 would shed light on the development of nuclear synthesis in cosmology and the big bang. The Li^6 and Li^7 lines are too close together to resolve with presently available spectrometers. A neutral or ion mass spectrometer might be able to provide the *in situ* measurements needed to determine the ratio of these elements in the solar system, across the heliopause, and into the interstellar medium [7].

Abundance and distribution of H, He, C, N, and O

FOCAL could determine, *in situ*, the cosmic abundances of these elements within the heliosphere with radial distance from the Sun across the heliopause transition zone, and also in the very local interstellar medium. These particles are especially interesting in helping to define the interactions involved in these regions because of their low ionization potential. The interaction process between the solar environment and the local interstellar medium is of broader interest than just for solar system studies, since similar interactions may be found in other star systems as well as other astrophysical conditions such as stellar expansion [7, 8]. A complement of instruments to accomplish this task would include a UV photometer, a neutral particle detector (like the *Ulysses MPI*), and an ion particle detector [7].

Radio science—VLBI studies of interstellar scintillation

The angular resolution that can be obtained with VLBI is proportional to the baseline length in wavelengths. However, a fundamental limit to the angular resolution obtainable is the scattering size of a point source. Density fluctuations in the interstellar medium broaden the apparent size of background sources through small-angle diffraction and refraction. By measuring the effect of interstellar scintillation we can determine the turbulence properties of the interstellar medium along different lines of sight.

During the first part of the FOCAL cruise, at small heliocentric distances, pulsars could be used as point sources to observe the effect of interstellar medium scintillations on the propagation of the pulsar signals. This would permit determination of the scale size of the turbulence in the interstellar medium, as well as testing present models of this turbulence. Earth-based studies have indicated that the turbulence affecting radio signals has scale lengths much larger than an Earth diameter, requiring interplanetary baselines for a detailed study of the power spectrum. There is a frequency dependence to the scattering properties of the interstellar turbulence. Thus, sampling at a range of frequencies (10 kHz–22 GHz) could be used to distinguish the effects of interstellar scintillation from those of the background source. This study would aid in the development of future space-based VLBI systems [9, 10]. In addition, we could directly measure the sizes of the radio-emitting regions in pulsars as the baseline increases.

The VLBI system proposed would consist of a 5-meter pointable radio antenna on the FOCAL spacecraft and a large Earth-based antenna. This configuration would provide a VLBI system with a baseline expanding out to 1,000 AU. There are two objectives to this investigation: the study of interstellar scintillation and the study of the structure of the pulsars themselves [11].

2.3.2 Astronomy

Radio science—VLBI studies of very compact radio sources

As the baseline increases the source regions within pulsars would be resolved, the signal would no longer have the coherence of a point source and thus would not be useful to probe the interstellar medium. It is at this point in the mission that the source region itself might be studied. Resolution of the radio source would help define the size of the source region and help determine, for example, the height above the surface of the neutron star where the emissions are originating. This would help constrain present models of the pulsar radio-emitting process.

At the highest possible frequencies (i.e., 22 GHz), the antenna could be used to study the structure of extragalactic radio sources, some of which are small enough to be unresolved at high frequencies on Earth-length baselines [11].

Low-frequency radio astronomy

A long wire dipole antenna would permit a study of very low–frequency (10 kHz–100 kHz) radio emissions. Such emissions cannot be observed from the near-Earth environment due to the extended ionosphere and solar wind. Likely sources are galactic supernova remnants, pulsars, and burst emissions from the outer planets and heliopause. The range of frequencies to be sampled is dependent on the length of the antenna. Preferably, the antenna would be no shorter than 1/2 of the longest wavelength to be sampled, although a less optimal size might be usable. Receivers for low frequencies are simple, reliable, and inexpensive [11].

Gamma-ray burst timing and positioning

FOCAL could determine the location of gamma-ray sources. A FOCAL baseline would permit precision calculations of the time differential of the signal between FOCAL and Earth. This method of observation would collapse the positioning uncertainty box in one direction and would reduce the optical source hunt to a line scan problem [12].

Gravitational focusing—quasar studies

Among the projects FOCAL would be able to accomplish are the following [13]:

(a) Test the hypothesis that high-amplitude events are caused by gravitational focusing by individual stars in an intervening galaxy.
(b) Determine the size of the region responsible for the optical continuum emission of quasars by observing spatial luminosity differences during high-amplification events (HAE).
(c) Observe brightness variations in quasars due to the transverse motion of the focusing star (and the intervening galaxy) with respect to the background quasar.
(d) Determine the number of bodies causing the gravitational focusing of observable quasars. A large sampling of quasars is required to do this study.

2.3.3 Cosmology

Gravitational wave detection

Using FOCAL and Earth as end masses of an electromagnetically tracked free-mass gravitational wave detector [14], a stochastic background of primordial gravitational waves created by very early cosmological processes or by the big bang itself could be detected with a sensitivity up to six orders of magnitude better than that available by other means. The wavelengths to be probed lie between 100 and 10^6 s. The preferred tracking system for such an experiment is a laser transponder on board the spacecraft, but a high-frequency radio system would still give an important experiment. The detection of a gravitational wave background would probe the very earliest era of the evolution of the Universe and would represent a cosmological observation as important as or more important than the discovery of the microwave background or of the cosmic redshift [14–17].

Spatial variations of G

Tracking of the FOCAL spacecraft provides a means to test the theory that there may be possible variations to the Newtonian inverse-square law of gravitation. The experiment would test the possibility that the Newtonian laws of motion would break down in the limit of small accelerations, an idea that has been suggested as an alternative explanation to the *hidden mass theorem* for explaining galactic dynamics. The breakdown of the inverse-square law would manifest itself as a special dependence in the effective gravitational constant.

$$G_{effective} = G\left[1 + A\left(\frac{r}{r_0}\right)^{2n}\right] \tag{2.1}$$

where r is the distance at which the gravitational acceleration is equal to the typical galactic acceleration and A and n are determined constants.

A spacecraft at 1,000 AU would have a small enough acceleration about the Sun to test this theory: with present day tracking capabilities, the power law index could be limited to $n > 4.3$. Planetary-scale baseline probes give only $n > 1$ [18–21].

IR background

A simple IR experiment at the range of 1 micron–30 microns could answer very important cosmological questions such as: Is there a 10-micron background from external galaxies or quasars? Is there a 2-micron background? [2].

2.3.4 Solar system studies

Zodiacal light

The Helios zodiacal light experiment observed the zodiacal light intensity to vary as $r^{-2.3}$. If one assumes constant albedo and grain-size distributions, then the number density of dust varies with distance as $r^{-1.3}$ [22, 23]. It is still to be determined if either of these two parameters vary with heliocentric distance [24]. The zodiacal light experiment on Pioneer 10/11 measured a brightness gradient of approximately $r^{-2.5}$ from 1 AU to 3.5 AU and no detectable signal above the starlight background beyond 4 AU [25]. Thus, the effect of the zodiacal light should be negligible beyond Jupiter's orbit, and faint light observations in IR can be done much better than from Earth orbit [26]. Thermal emission from interplanetary dust dominates the 10-micron, 25-micron, and 60-micron background in IRAS data [27].

Planetary system

As FOCAL leaves the solar system, it could study the zodiacal light from a distance, both in the visible as well as at IR wavelengths. The study of what a "solar system dust cloud" looks like from a distance could be used to correlate with detections of other dust clouds, such as those surrounding Vega and Beta Pictoris, which have recently been made [26].

Determination of the total solar system mass

By the time the spacecraft reaches 800 AU a substantial amount of the mass of the inner Oort Cloud would be inside the orbit of FOCAL. A more accurate determination of the mass of the solar system than has been done to date could then be done [28].

2.4 SPACE PLASMA PHYSICS

The heliosphere can be described as a huge bubble in the interstellar gas, created by the radial supersonic outflow of magnetized plasma from the Sun (the solar wind) and confined by the magnetic and particle environment of the local interstellar medium. The boundary between the region of solar wind dominance and the interstellar medium is called the heliopause. The shape of the heliopause and the structure of this boundary are a function of the magnetic, dynamic, and thermal pressures of the interstellar medium, as well as the particle composition in this region, and their interaction with the very turbulent local magnetic environment. The boundary's characteristics (i.e., dynamic and magnetic structure as well as location) are expected to differ for each particle species. A probable model for the shape of the heliosphere was extrapolated from our present understanding of the behavior of collisionless plasmas and from our present knowledge of the heliosphere. According to this model the Pioneer and Voyager spacecraft have so far probed only the inner heliosphere—about 20% of the radial dimension of the heliosphere [3]. The radial dimension has a nominal value of about 100 AU, but estimates range from 50 AU to 100 AU. FOCAL would be the first opportunity for a full complement of instruments with the proper range of sensitivity to make a cross-sectional cut of our heliosphere and sample the interaction region between the interstellar medium and the heliosphere, providing us with a heliospheric model to be used as a basis for understanding other star systems [29].

2.4.1 Dust

Solar system

FOCAL would be able to determine the distribution of dust within the solar system, thus permitting us to study the effects of gravitational focusing, mass and composition of the dust, and its orbital characteristics, such as direction (direct vs. retrograde) and ellipticity of the orbit as well as the mass, velocity, energy, and composition of these particles. The mass and velocity would allow us to define the population of dust and the orbit of these particles in the solar system. Determination of the particle orbits is essential in distinguishing the source of the particles and in projecting the three-dimensional distribution. The velocity (speed and direction) of a particle is the discriminator between interstellar grains and solar system–source particles.

Heliopause

The energy and the density distribution of dust across the transition zone defined as the heliopause could be characterized by measurements taken by FOCAL. It would also be able to determine mass, energy, and chemical composition as well as the kinematics of particle behavior, charge exchange, and wave–particle interactions in this turbulent region.

Interstellar medium

Instruments on FOCAL could be used to determine mass, energy, and chemical composition of the dust population as well as the kinematics of particle behavior of dust in the interstellar medium. Grains are expected to be on the order of 10^{-16} g to

10^{-17} g. Dust instruments presently available, such as those used on Galileo and the Halley missions, have the capability to determine mass and velocity (speed and direction) as well as particle composition [26].

2.4.2 Plasma and energetic particle distributions

Heliosphere

A FOCAL mission would be able to

(1) determine the composition of, and characterize the energy spectrum and distribution of, low-energy particles in the outer reaches of the heliosphere not yet visited by other spacecraft;
(2) determine regions of attenuation and energization of these particles;
(3) characterize the wave–particle interactions of plasmas in the turbulent regions of the heliosphere, especially those regions dominated by inner heliospheric shocks and the heliopause boundary.

Interstellar medium

Measurements taken on FOCAL would help determine the "original" energy distribution and composition of particles of interstellar medium origin prior to the energization and/or attenuation caused by wave–particle interactions in their traverse through the heliopause. A definition of the particle domain in the interstellar medium would lead to an understanding of the interaction of our star system with the Galactic environment, and would help constrain models of the interstellar medium with other astrophysical conditions (i.e., star systems and jets [30]).

2.4.3 Low-energy cosmic rays

Projects on cosmic rays include attempts to

(1) determine the origin, energy spectrum, and mass of cosmic rays;
(2) characterize the energy spectrum and distribution of low-energy cosmic rays;
(3) characterize the interaction with, and attenuation or energization of, low-energy cosmic rays at the heliopause;
(4) characterize the energy spectrum across the boundary; and
(5) differentiate between solar and extra solar system sources of cosmic rays [30].

2.4.4 Magnetic field morphology

Heliosphere/Heliopause

Determining the morphology of the magnetic field as a function of heliocentric distance would be possible with FOCAL. This would lead to improved magnetic field models of the heliosphere as a whole. Defining the location of the magnetic heliopause boundary characterizing its structure, and the morphology of the regions upstream and downstream of this ‘‘boundary’’, inner and outer bow shocks, mass loading of field lines in the region, and magnetic instabilities caused by the unique conditions existing in the outer reaches of the heliosphere are also projected FOCAL goals [3].

Interstellar medium

FOCAL would determine the morphology of the galactic magnetic field as a function of the galactic magnetic field, as well as the interactions with the local interstellar plasma, interstellar shocks, hydromagnetic waves, and the characteristics of local stellar winds. Determining the characteristics of the magnetic field outside the domain of the heliosphere would help constrain models for the origin and generation of the Galactic magnetic field [30].

2.4.5 Plasma waves

Heliosphere/Heliopause

The ability to study the plasma waves generated in the various interaction regions of the outer heliosphere, inner and outer shocks, and heliopause to determine the source of these waves, and to study the local particle–wave interactions, especially near the heliopause, are all possible with a FOCAL mission [30]. Plasma wave emissions in the 2 kHz–3 kHz frequency range, known to be generated at planetary shocks, have also been faintly seen by Voyager from what may be the terminal shock of the heliosphere [31].

Interstellar medium

Another FOCAL project is the study of the local plasma waves/particle interactions. FOCAL would also be able to sample the local charge density, and study the local magnetohydrodynamic (MHD) behavior and the microprocesses for transporting energy in the interstellar medium [30].

2.5 SCIENCE INSTRUMENTATION

The scientific objectives listed in the previous section define the need for a specific complement of scientific instruments. Below is a preliminary list of the instruments that were suggested to accomplish the proposed investigations.

Optical/IR Telescope
Cosmic Ray Detector
Dust Detector
Energetic Particle Detector
Ion/Neutral Particle Detector

Gamma-Ray Spectrometer
Magnetometer
Plasma Particle Detector
Plasma Wave Instrument
Ultraviolet Spectrometer
Very Low–Frequency Radio Astronomy Antennas
  a. Dipole antenna
  b. Pointable dish

Much more effort is necessary to establish the interactions of these various instruments with each other and their effect on overall spacecraft design.

2.6 REFERENCES

[1] J. R. Mould, "Interactions of the scale of distances," *Seminar to TAU Study Team*, December 9, 1986.
[2] J. Beichman, personal communication.
[3] D. Winterhalter, E. J. Smith, and J. A. Slavin, "The solar wind magnetic field at 20 AU," *EOS Transactions of the American Geophysical Union*, **67** (1986), 44.
[4] J. Ajello, Comments during *TAU Workshop*, September 29, 1986.
[5] J. Ajello, personal communication.
[6] J. M. Ajello, A. I. Stewart, G. E. Thomas, and A. Graps, "Solar cycle study of interplanetary Lyman-alpha variations: Pioneer Venus orbiter sky background results," *Astrophysical Journal*, June 15, 1987.
[7] S. Federman, personal communication.
[8] J. L. Bertraux, "Helium and hydrogen of the local interstellar medium observed in the vicinity of the Sun," *Local Interstellar Medium: Proceedings of a Colloquium*, held at University of Wisconsin, 1984.
[9] D. Jones, personal communication.
[10] B. F. Burke, and N. S. Kardashev, *Scientific and Technical Plan*, Vol. I of Proposal to NASA for participation as Members of the Radio Science Team, VOIR, AO-OSS-5-78, June 1, 1979.
[11] D. Jones, *Seminar*, September 23, 1986.
[12] A. Meinel, *TAU Thinkshop*, September 27, 1986.
[13] D. Jones, memo, January 16, 1987.
[14] R. Kayser, S. Refsdal, and R. Stabell "Astrophysical applications of gravitational microlensing," *Astronomy and Astrophysics*, **166** (1986), 36–52.
[15] B. Grieger, R. Kayser, and S. Refsdal, "A parallax effect due to gravitational microlensing," *Nature*, **324** (1986), November 13.
[16] R. Hellings, "Electromagnetically-tracked free-mass wave antennas," *Gravitational Radiation*, N. Dervelle and T. Piran (editors), North Holland, 1982.
[17] R. W. Hellings, "Spacecraft Doppler gravity-wave detection. I. Theory," *Physical Review D*, **23**(4) (1981), 832–843, February 15.
[18] R. W. Hellings, P. S. Callahan, J. D. Anderson, and A. T. Moffet, "Spacecraft Doppler gravity-wave detection. II. Results," *Physical Review D*, **23**(4) (1981), 844–851, February 15.

[19] R. L. Zimmerman, and R. W. Hellings, "Gravitational radiation dominated cosmologies," *Astrophysical Journal*, **241** (1980), 475–485, October 15.
[20] M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis."
[21] M. Milgrom, "A modification of the Newtonian dynamics: Implications for Galaxies."
[22] M. Milgrom, "A modification of the Newtonian dynamics: Implications for Galaxy Systems," *Astrophysical Journal*, **270** (1989), 365–389, July 15.
[23] R. Hellings, personal communication.
[24] C. Leinert, M. Hanner, and E. Pitz, "On the spatial distribution of interplanetary dust near 1 AU," *Astronomy and Astrophysics*, **63** (1978), 183–187.
[25] C. Leinert, I. Richter, and E. Pitz, "Four years of Zodiacal light observations from the Helios Space Probes: Evidence for a smooth distribution of interplanetary dust" in *Solid Particles in the Solar System: IAU Symposium*, Vol. 90, p. 15, 1980.
[26] R. Dumont, and A. C. Levasseur Regourd, "Gradients of Temperature, Albedo and density in the zodiacal cloud, as deduced from the available thermal and optical observations," *Astronomy and Astrophysics* (1987).
[27] M. Hanner, J. Sparrow, J. Weinberg, and D. Beeson, "Interplanetary dust and Zodiacal light series", *Lecture Notes in Physics*, No. 48, p. 29, Springer-Verlag, Berlin, 1985.
[28] M. Hanner, memo, December 18, 1986.
[29] M. G. Hauser *et al.*, "IRAS observations of the diffuse infrared background," *Astrophysical Journal*, **278** (1984), L15–L18, March 1.
[30] P. Weissman, *TAU Thinkshop*, September 27, 1986.
[31] *Solar-Terrestrial Research for the 1980's*, National Accademy Press, Washington, D.C., 1981.
[32] D. Winterhalter, *TAU Thinkshop*, September 27, 1986.
[33] W. S. Kurth, D. A. Gurnett, and F. L. Scarf, "Recent observations of the very low frequency interplanetary radio emissions," *Advances of Space Research: Proceedings of the 26th COSPAR Planetary Meeting* held in Toulouse, France, 1986.

Table 2.1. Proposed "FOCAL" space missions to 550 AU for interstellar exploration between 2000 AD and 2050 AD.

	Mission name	*Science target*	*Target distance*	*Launch by*	*Arrival at 550 AU*	*Cost*	*Direction of exit*
1	ISP InterStellar Probe (NASA)	CMB	250 AU	2010	2040	Cheap	Incoming interstellar wind
2	FOCAL One	Alpha Centauri	4.29 lt-yr	2015	2045	Expensive	Opposite to Alpha Centauri
3	FOCAL Two	Center of the Galaxy	32,000 lt-yr	2015	2045	Less expensive	Opposite to Galactic Center
4	FOCAL Three FOCAL Three	Barnard's Star Two stars	5.9 lt-yr 7.6–8.0 lt-yr	2020 2020	2050 2050	Less expensive Less expensive	Opposite to Barnard's Star Opposite to each star
5	FOCAL Four FOCAL Four	Sirius Four stars	8.65 lt-yr 8.9–10.4 lt-yr	2020 2020	2050 2050	Less expensive Less expensive	Opposite Opposite
6	FOCAL Five FOCAL Five	Epsilon Eridani Eleven stars	10.69 lt-yr 10.9–11.9 lt-yr	2025 2025	2050 2050	Less expensive Less expensive	Opposite Opposite
7	FOCAL Six	Tau Ceti	11.95 lt-yr	2025	2050	Less expensive	Opposite

3

Magnifying the nearby stellar systems

3.1 INTRODUCTION

Some experts believe that, probably, sometimes after 2050 AD interstellar space missions to the nearby stars will have become a technologically feasible reality. Before launching unmanned probes straight towards each nearby star, however, the need will be felt to have a detailed "map" of each target star system to explore on the widest possible range of electromagnetic frequencies. Here is where the FOCAL space mission will show its full power.

We anticipate that there will be a host of FOCAL space missions launched in all directions around the Sun, each probe launched in the direction exactly opposite to the star to explore with respect to the Sun position. In this chapter we shall calculate each of those relative directions and the proximity to the Sun for each flyby in order to exit the solar system at the desired inclination to the ecliptic.

A FOCAL space mission could be used to magnify anything of interest outside the solar system. One should then say that FOCAL will be used to magnify the nearby planetary systems, meaning not just the nearby stars themselves, but also their planets, halo disks, Oort clouds, etc.

Table 3.1 lists the 25 star systems located within the first 13.10 lt-yr from the Sun. Twenty-five more stars are encountered by extending the list from 13.10 lt-yr to 18.43 lt-yr from the Sun, and these are presented in Table 3.2. This catalog of nearby stellar systems was taken from [1].

3.2 DIRECTIONS OF EXIT FROM THE SOLAR SYSTEM FOR FOCAL PROBES TO MAGNIFY NEARBY STELLAR SYSTEMS

To magnify a selected nearby stellar system, a FOCAL probe must exit the solar system along the straight line connecting the center of the Sun to that stellar system,

Table 3.1. Location of the 25 nearest stellar systems.

Stellar System #	Star designation	Distance from the Sun (lt-yr)	1950 Right Ascension α	1950 Declination δ
1	Proxima Centauri	4.29	14h 26m	−62° 28′
2	α Centauri A and α Centauri B	4.38	14h 36m	−60° 38′
3	Barnard's star (+4° 3561)	5.91	17h 55m	4° 33′
4	Wolf 359	7.62	10h 54m	7° 16′
5	Lalande 21185 = BD +36° 2147	8.04	11h 1m	36° 18′
6	Sirius A = 48915 and Sirius B = 48915	8.65	6h 43m	−16° 39′
7	Luyten 726-8 = A and Luyten 726-8 = B = UV Ceti	8.94	1h 36m	−18° 13′
8	Ross 154 = AC-242833-183	9.45	18h 47m	−23° 53′
9	Ross 248	10.39	23h 39m	43° 55′
10	ε Eridani = 201091	10.69	3h 31m	−9° 38′
11	Ross 128	10.95	11h 45m	1° 6′
12	Luyten 789-6	11.09	22h 36m	−15° 36′
13	61 Cygni A = 201091 and 61 Cygni B = 201092	11.17	21h 5m	38° 30′
14	ε Indi = 209100	11.21	22h 0m	−57° 0′
15	Procyon A and B = 61421 = α Canis Minoris	11.36	7h 37m	5° 21′
16	+59° 1915 A and Σ 2398 B	11.48	18h 42m	59° 33′
17	BD +43° 44 Groombridge 34 A and B	11.57	0h 15m	43° 44′
18	CD −36° 15693 = Lacaille 9352	11.69	23h 3m	−36° 8′
19	τ Ceti	11.95	1h 42m	−16° 12′
20	L725-32 = LET 118	12.50	1h 7m	−17° 32′
21	CD −39° 15693 = Lacaille 8760	12.54	21h 14m	−39° 4′
22	BD +5° 1668 (Luyten)	12.64	7h 22m	23° 0′
23	−45° 1841 (Kapteyn)	12.74	5h 10m	−45° 0′
24	Krüger 60 = A and Krüger 60 (DO Cep) = B	12.84	22h 26m	57° 27′
25	Ross 614 A and B	13.10	6h 27m	−2° 46′

Table 3.2. Location of the next 25 nearest stellar systems.

Stellar System #	Star designation	Distance from the Sun (lt-yr)	1950 Right Ascension α	1950 Declination δ
26	BD −12° 4523	13.10	16h 28m	−12° 32′
27	Wolf 28 (Van Maanen)	13.94	0h 46m	5° 9′
28	Wolf 424 A and B	14.24	12h 31m	9° 18′
29	G 158−27	14.43	0h 4m	−7° 48′
30	CD −37° 15492	14.50	0h 2m	−37° 36′
31	BD +50° 1725 = Grm 1618	15.03	10h 8m	49° 42′
32	CD −46° 11540	15.10	17h 25m	−46° 51′
33	Luyten 1159−16	15.10	1h 57m	12° 51′
34	CD −49° 13515	15.24	21h 30m	−49° 13′
35	CD −44° 11909	15.31	17h 33m	−44° 17′
36	BD +68° 946	15.76	17h 37m	68° 23′
37	Luyten 145−141 = cc 658	15.83	11h 43m	−64° 33′
38	BD −15° 6290 = Ross 780	15.83	22h 51m	−14° 31′
39	40 Eridani A, B, and C	15.76	4h 13m	−7° 44′
40	BD +15° 2620 = Lalande 25372	15.90	13h 41m	15° 26′
41	BD +20° 2465	16.23	10h 17m	20° 7′
42	Altair	16.64	19h 48m	8° 44′
43	70 Oph A and B	16.73	18h 3m	2° 31′
44	AC +79° 3883	16.81	11h 45m	78° 58′
45	BD +43° 4305	16.90	22h 45m	44° 5′
46	AC +58 = Stein 2051 A and B	16.99	4h 26m	58° 33′
47	+44° 2051 = WX UMa A and B	17.54	11h 3m	43° 47′
48	−26° 12026 = 36 Oph A, B and C	17.73	17h 12m	−26° 39′
49	−36° 13940 A and B (HR 7703)	18.43	20h 8m	−36° 14′
50	BD +1° 4474	18.43	23h 47m	2° 8′

and in the direction exactly opposite to the stellar system with respect to the Sun center. FOCAL must then reach distances between 550 AU and 1,000 AU from the Sun's center to perform its observations on electromagnetic waves focused there by the Sun's mass.

In terms of celestial coordinates (i.e., right ascension α and declination δ), the FOCAL exit direction to observe a selected stellar system points just towards the opposite direction on the celestial sphere. Since distances to the nearby stellar systems are huge compared with distances within the solar system, one may well omit the coordinate transformations between the Earth and the Sun in the first instance, treating the situation as if the Earth and the Sun were centered at the same point. Thus, to get the desired direction of exit from the solar system, one only has to

- add 12 hours to the selected stellar system right ascension α;
- reverse the selected stellar system declination sign: $-\delta$ instead of δ.

In this way, the second and third columns in Tables 3.3 and 3.4 are obtained from the data of Tables 3.1 and 3.2, respectively. These columns yield the new α and δ of the corresponding FOCAL spacecraft exit direction.

In view of the subsequent calculations about the Sun flyby enabling the exit of the solar system at any requested inclination, Tables 3.3 and 3.4 also contain two more columns (the fourth and fifth ones) describing the same exit direction in terms of celestial latitude β and longitude λ, respectively. These columns were computed from the values of α and δ by virtue of the transformation equations

$$\begin{cases} \sin\beta = \cos\varepsilon\sin\delta - \sin\varepsilon\cos\delta\sin\alpha \\ \cos\beta\cos\lambda = \cos\delta\cos\alpha \end{cases}$$

which follow from elementary spherical trigonometry considerations and are proven in any textbook of spherical astronomy; in particular, [2, 3]. The angle ε is the inclination of the ecliptic plane with respect to the celestial equator (i.e., $23°26'21.448''$).

3.3 KEPLERIAN THEORY OF SIMPLE HYPERBOLIC FLYBYS

The simple theory of classical Keplerian hyperbolic flybys, described in this section, can be found in textbooks of celestial mechanics or astrodynamics such as [3, 4]. I present the theory here because it paves the way for applications to the Sun flyby that any FOCAL spacecraft will have to make in order to exit the solar system at any desired inclination on the ecliptic plane and without additional cost of fuel.

The starting point of the Keplerian theory is, of course, the equation of an ellipse centered at the origin and having semi-axes a and b along the axes x, and y,

Table 3.3. Direction of exit from the solar system for the 25 FOCAL spacecraft going to take magnified shots of the nearest 25 stellar systems.

Stellar System #	FOCAL EXIT Right Ascension ($= \alpha_{target} + 12h$)	FOCAL EXIT Declination ($= -\delta_{target}$)	FOCAL EXIT Latitude	FOCAL EXIT Longitude
1	2° 26′	62° 28′	43° 55′	237° 55′
2	2° 36′	60° 38′	41° 26′	238° 26′
3	5° 55′	−4° 33′	−27° 58′	268° 58′
4	22° 54′	−7° 16′	0° 12′	161° 12′
5	23° 1′	−36° 18′	−27° 28′	151° 28′
6	18° 43′	16° 39′	38° 17′	103° 17′
7	13° 36′	18° 13′	25° 44′	15° 44′
8	6° 47′	23° 53′	0° 53′	280° 53′
9	11° 39′	−43° 55′	−41° 30′	343° 30′
10	15° 31′	9° 38′	26° 31′	47° 31′
11	23° 45′	−1° 6′	0° 29′	176° 29′
12	10° 36′	15° 36′	5° 8′	335° 8′
13	9° 5′	−38° 30′	−51° 52′	336° 52′
14	10° 0′	57° 0′	41° 21′	308° 21′
15	19° 37′	−5° 21′	15° 59′	115° 59′
16	6° 42′	−59° 33′	−81° 33′	308° 33′
17	12° 15′	−43° 44′	−37° 58′	23° 58′
18	11° 3′	36° 8′	27° 14′	332° 14′
19	13° 42′	16° 12′	24° 29′	17° 29′
20	13° 7′	17° 32′	21° 42′	8° 42′
21	9° 14′	39° 4′	21° 47′	308° 47′
22	19° 22′	−23° 00′	0° 52′	108° 52′
23	17° 10′	45° 0′	67° 24′	66° 24′
24	10° 26′	−57° 27′	−59° 11′	344° 11′
25	18° 27′	2° 46′	24° 29′	97° 29′

Table 3.4. Direction of exit from the solar system for the 23 FOCAL spacecraft going to take magnified shots of the next nearest 25 stellar systems.

Stellar System #	FOCAL EXIT Right Ascension ($= \alpha_{target} + 12h$)	FOCAL EXIT Declination ($= -\delta_{target}$)	FOCAL EXIT Latitude	FOCAL EXIT Longitude
26	4° 28′	12° 32′	−10° 9′	247° 9′
27	12° 46′	−5° 9′	0° 11′	12° 11′
28	0° 31′	−9° 18′	−11° 35′	183° 35′
29	12° 4′	7° 48′	6° 4′	1° 4′
30	12° 4′	37° 36′	33° 21′	15° 21′
31	22° 8′	−49° 42′	−35° 21′	134° 21′
32	5° 25′	46° 51′	21° 54′	263° 54′
33	13° 57′	−12° 51′	0° 50′	31° 50′
34	9° 30′	49° 13′	32° 1′	308° 1′
35	5° 33′	44° 17′	20° 23′	264° 23′
36	5° 37′	-68° 23′	−87° 8′	221° 8′
37	23° 43′	64° 33′	56° 30′	143° 30′
38	10° 51′	14° 31′	5° 40′	338° 40′
39	16° 13′	7° 44′	26° 56′	59° 56′
40	1° 41′	−15° 26′	−24° 2′	197° 2′
41	22° 17′	−20° 7′	−8° 48′	148° 48′
42	7° 48′	−8° 44′	−29° 19′	300° 19′
43	6° 3′	−2° 31′	−25° 57′	270° 57′
44	23° 45′	−78° 58′	−63° 33′	115° 33′
45	10° 45′	−44° 5′	−46° 55′	354° 55′
46	16° 26′	−58° 33′	−36° 18′	75° 18′
47	23° 3′	−43° 47′	−34° 19′	147° 19′
48	5° 12′	26° 39′	2° 21′	259° 21′
49	8° 8′	36° 14′	15° 12′	296° 12′
50	11° 47′	−2° 8′	−3° 16′	357° 16′

respectively:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{3.1}$$

along with the two ellipse relationships

$$\begin{cases} a^2 = b^2 + c^2 \\ e = \dfrac{c}{a}. \end{cases} \tag{3.2}$$

Inserting (3.2) into (3.1) one gets

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1 - e^2)} = 1. \tag{3.3}$$

On the other hand, if one considers the polar coordinates (r, ϑ) centered at the right focus (whose abscissa is $+c$), the immediate transformation formulas follow:

$$\begin{cases} x = c + r\cos\vartheta = ae + r\cos\vartheta \\ y = r\sin\vartheta \end{cases} \tag{3.4}$$

which, substituted into (3.3), yield, after some rearranging, the second degree algebraic equation in r

$$(1 - e^2\cos^2\vartheta)r^2 + [2ae(1 - e^2)\cos\vartheta]r - a^2(1 - e^2)^2 = 0. \tag{3.5}$$

Solving this with respect to r, one gets the well-known polar equation of the ellipse centered at the (right) focus located at the origin

$$r(\vartheta) = \frac{a(1 - e^2)}{1 + e\cos\vartheta}. \tag{3.6}$$

For an ellipse, the range of the eccentricity e is $0 \le e < 1$, and, in the limiting case of $e = 0$, it is a circle of radius a centered at the origin.

The special case where $e = 1$ is the parabola, and its equation can be obtained from (3.6) as its limit for

$$e \to 1 \tag{3.7}$$

in such a way that the numerator of (3.6) tends to a finite positive number, denoted by p and called the *parameter* of the parabola. This implies that a in the numerator of (3.6) must tend to infinity

$$a \to \infty \tag{3.8}$$

with the result that the polar equation of the parabola is

$$r(\vartheta) = \frac{p_{parabola}}{1 + \cos(\vartheta)}. \tag{3.9}$$

All values of the eccentricity higher than 1 ($1 < e < \infty$) yield a hyperbola. The polar equation of the hyperbola can be found just as for the ellipse; that is, by starting from the well-known Cartesian equation of the hyperbola centered at the origin and having semi-axes a and b along the axes x, and y, respectively:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{3.10}$$

along with the two hyperbola relationships

$$\begin{cases} c^2 = b^2 + a^2 \\ e = \dfrac{c}{a}. \end{cases} \tag{3.11}$$

By repeating steps similar to (3.4) and (3.5), this leads to the polar equation of the hyperbola

$$r(\vartheta) = \frac{a(e^2 - 1)}{1 + e\cos\vartheta}. \tag{3.12}$$

Checking this against (3.6), one immediately notices that the polar equations of both the ellipse and hyperbola would be the same—that is, Equation (3.6)—if a was formally replaced by $-a$ in the hyperbola equation (3.12). This *convention will be adopted throughout this book, so all the equations affecting the hyperbola will be derived by the similar equations for the ellipse merely by replacing a by* $-a$ (see Figure 3.1).

A basic result in classical newtonian gravitational theory (which is not to be proven here, and can be found in any textbook about classical mechanics) is the conservation of energy. For two bodies of masses M and m at the distance r from each other, and in the reference frame whose origin is located at the center of the larger mass M, the conservation of total (i.e., kinetic plus gravitational) energy is expressed by the equation

$$\tfrac{1}{2}mv^2 - \frac{GMm}{r} = -\frac{GMm}{2a} \tag{3.13}$$

where $v(r)$ is the velocity of the smaller body of mass m moving around the larger body of mass M, the latter assumed fixed. This formula is exact for an elliptical orbit as well as for a hyperbolic trajectory (replacing a by $-a$, as previously discussed) and even for a parabolic trajectory (letting $a \to \infty$). Solving (3.13) for the velocity $v(r)$, one gets the *speed equation*

$$v(r) = \sqrt{GM\left(\frac{2}{r} - \frac{1}{a}\right)}. \tag{3.14}$$

Since we are interested in hyperbolic flybys, we are concerned with the *approach speed* (i.e., the speed at which the spacecraft enters the sphere of influence of the massive body M that it is going to fly by). One can approximately assume that when the spacecraft enters the sphere of influence, its distance from the mass M is "infinite" (i.e., $r \to \infty$), and from this assumption Equation (3.14) yields the following spacecraft approach speed v_∞ along a hyperbolic trajectory (once again, a is to be taken as

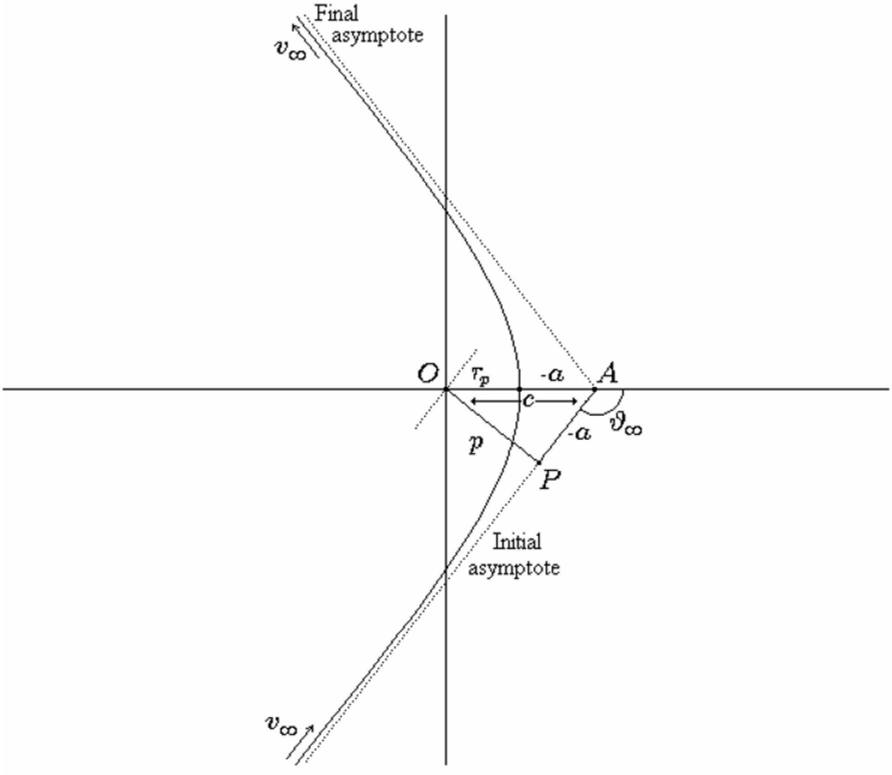


Figure 3.1. Basic geometry of the hyperbolic flyby.

negative, according to the convention):

$$v_\infty = \lim_{r\to\infty} v(r) = \sqrt{-\frac{GM}{a}}. \tag{3.15}$$

Thus, the parameter a for a hyperbolic flyby is related to the approach speed v_∞ by the important formula

$$a = -\frac{GM}{v_\infty^2}. \tag{3.16}$$

On the other hand, inversion of the polar orbit equation (3.6) yields

$$\vartheta(r) = \arccos\left[\frac{1}{e}\left(\frac{a(1-e^2)}{r} - 1\right)\right]. \tag{3.17}$$

When the spacecraft is at infinite distance from M along a hyperbola, it is actually moving along the hyperbola asymptote. One may thus deduce the important relationship between the angle ϑ_∞ (i.e., the angle between the hyperbola axis and any of the two asymptotes) and the hyperbola eccentricity e:

$$\vartheta_\infty = \lim_{r\to\infty} \vartheta(r) = \arccos\left(-\frac{1}{e}\right) \tag{3.18}$$

and upon inversion,

$$e = -\frac{1}{\cos(\vartheta_\infty)}. \tag{3.19}$$

As for the ranges of these two variables, since for the hyperbola one has $e > 1$ it follows that $-\vartheta_\infty < \vartheta < \vartheta_\infty$ and $\pi/2 < \vartheta_\infty < \pi$.

The value $\vartheta = 0$ is particularly important inasmuch as it is the *periastron* (i.e., the point of the spacecraft closest approach to the mass M). This *periastron distance* r_p may be calculated from the polar orbit equation (3.6) where a and e are being replaced by (3.16) and (3.19), respectively. The result is

$$r_p = r(0) = a(1-e) = -\frac{GM}{v_\infty^2}\left(1 + \frac{1}{\cos\vartheta_\infty}\right). \tag{3.20}$$

In order to find the *spacecraft speed at the periastron*, just substitute the previous equation into the speed equation (3.14) to get

$$v_p = v(r_p) = \sqrt{\frac{GM}{a}\frac{1+e}{1-e}} = v_\infty\sqrt{\frac{1-\cos\vartheta_\infty}{1+\cos\vartheta_\infty}}. \tag{3.21}$$

This periastron speed is also the *highest* speed achieved by the spacecraft along its hyperbolic trajectory, the *lowest* speed being the approach speed v_∞. In fact, in the speed equation (3.14), r first decreases from ∞ to r_p and then increases again from r_p to ∞.

It turns out to be quite useful to define the *impact parameter p* of the spacecraft as the minimal distance between the asymptotes and the center of the mass M. Figure 3.1 shows that one has

$$\begin{cases} OA = r_p - a = a(1-e) - a = -ae \\ AP = OA\cos OAP = -ae\cos(\pi - \vartheta_\infty) = ae\cos\vartheta_\infty = -a \end{cases}$$

from which one gets

$$OP = \sqrt{OA^2 - AP^2} = \sqrt{a^2e^2 - a^2} = |a|\sqrt{e^2-1}.$$

Being $p = OP$ and $a < 0$, one finds the following expression for the *impact parameter*:

$$p = -a\sqrt{e^2-1} = \frac{GM}{v_\infty^2}\sqrt{\frac{1}{\cos^2\vartheta_\infty} - 1} = \frac{GM}{v_\infty^2}\tan\vartheta_\infty. \tag{3.22}$$

In conclusion, we would like to make the important remark that any planar hyperbolic trajectory may be completely described by just *one* of the following three couples of independent parameters:

$$(a, e), (v_\infty, \vartheta_\infty), (r_p, v_p). \tag{3.23}$$

In fact, every one of the above three couples may be re-expressed as a suitable function of the other two couples of parameters as shown hereafter:

$$\begin{cases} a = -\dfrac{GM}{v_\infty^2} \\ e = -\dfrac{1}{\cos\vartheta_\infty} \end{cases} \Leftrightarrow \begin{cases} v_\infty = \sqrt{-\dfrac{GM}{a}} \\ \vartheta_\infty = \arccos\left(-\dfrac{1}{e}\right) \end{cases} \tag{3.24}$$

$$\begin{cases} r_p = a(1-e) \\ v_p = \sqrt{\dfrac{GM}{a}\dfrac{1+e}{1-e}} \end{cases} \Leftrightarrow \begin{cases} a = \left(\dfrac{2}{r_p} - \dfrac{v_p^2}{GM}\right)^{-1} \\ e = \dfrac{r_p v_p^2}{GM} - 1 \end{cases} \tag{3.25}$$

$$\begin{cases} r_p = -\dfrac{GM}{v_\infty^2}\left(1 + \dfrac{1}{\cos\vartheta_\infty}\right) \\ v_p = v_\infty\sqrt{\dfrac{1-\cos\vartheta_\infty}{1+\cos\vartheta_\infty}} \end{cases} \Leftrightarrow \begin{cases} v_\infty = \sqrt{v_p^2 - \dfrac{2GM}{r_p}} \\ \vartheta_\infty = \arccos\left[-\left(1 + \dfrac{r_p v_\infty^2}{GM}\right)^{-1}\right] \end{cases} \tag{3.26}$$

3.4 THE FLYBY OF THE SUN PERFORMED BY THE FOCAL SPACECRAFT

To let the FOCAL spacecraft exit the solar system along a straight trajectory inclined at any angle to the ecliptic plane it is necessary that *the flyby of the Sun be the last one* in the sequence of all possible planetary flybys. In this section I shall use the equations of the previous section to study the Sun flyby performed by a FOCAL spacecraft intended to explore each one of the stars listed in Tables 3.3 and 3.4.

We shall now assume that the asymptotes of the Sun flyby are the two straight lines from the Earth to the Sun and from the Sun to the selected star. For additional simplification we assume that the flyby plane spanned by the two asymptotes mentioned above is *orthogonal* to the ecliptic plane.

Next, let us introduce the *deflection angle* ψ of the flyby (i.e., the angle between the incoming and outgoing hyperbola asymptotes). Then, *the absolute value of the celestial latitude of the target star, $|\beta|$, is also the deflection angle in the Sun flyby*, as shown in Figure 3.2.

Moreover, a glance at Figure 3.2 yields the relationship between the angle $\psi = |\beta|$ and the angle ϑ_∞

$$2\pi - 2\vartheta_\infty = \pi - \psi. \tag{3.27}$$

Solving this for ϑ_∞,

$$\vartheta_\infty = \frac{\pi}{2} + \frac{\psi}{2} = \frac{\pi}{2} + \frac{|\beta|}{2} \quad \text{("loose" Sun flyby case).} \tag{3.28}$$

Since $|\beta| < 90°$ for any star, this Sun flyby would cause deflections ψ smaller than

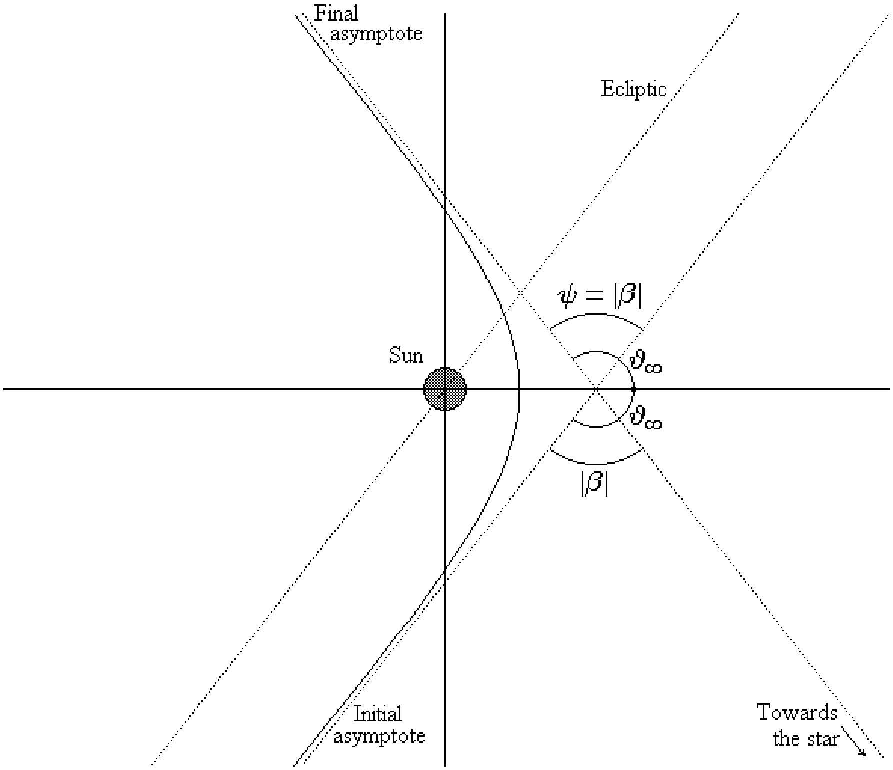


Figure 3.2. Relating the celestial latitude of the target star to the Sun flyby angle between hyperbola axis and asymptote.

90°. However, the *small* values of $\psi = |\beta|$ formulas in Section 3.3 would yield a *huge* perihelion distance, larger than 1 AU. This is clearly a problem in the design of the FOCAL Sun flyby, and its solution is to replace the above "loose" Sun flyby by virtue of a *"tight" Sun flyby* (i.e., one having the deflection angle ψ larger than 90°).

The tight Sun flyby is depicted in Figure 3.3. The deflection angle ψ between the hyperbola asymptotes is now given by

$$\psi = \pi - |\beta| \quad \text{("tight" Sun flyby case).} \tag{3.29}$$

Replacing (3.29) into (3.27), a new relationship between the angle ϑ_∞ and $|\beta|$ is found

$$\vartheta_\infty = \pi - \frac{|\beta|}{2} \quad \text{("tight" Sun flyby case).} \tag{3.30}$$

In conclusion, the tight Sun flyby is more appropriate than the loose one for the FOCAL space mission, and, in Appendix B, we computed all the 50 tight Sun flybys required to take magnified radio pictures of all the 50 nearest stars to the Sun.

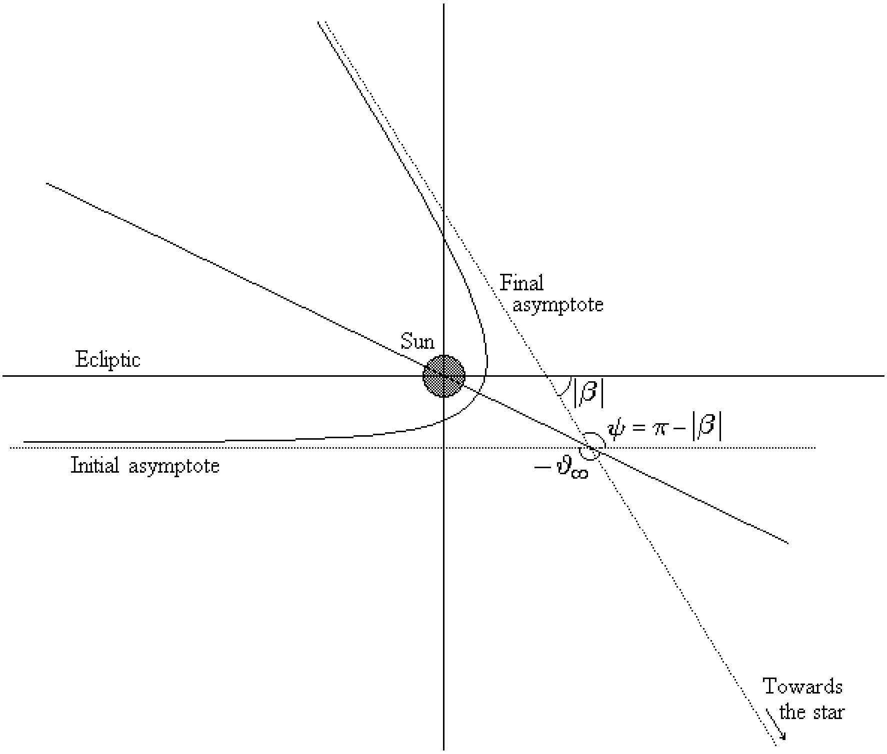


Figure 3.3. A "tight" Sun flyby: a different relationship between the deflection angle ψ and the absolute value of the celestial latitude, $|\beta|$, of the target star system.

Mathematically, all tables in Appendix B are thus based on (3.30) rather than on (3.28).

One may well wonder how long it will take for humankind to send 50 FOCAL spacecraft: five centuries? Or less? No matter how long, at least some of these FOCAL spacecraft will have to be launched, for FOCAL *is a "must" before any direct stellar exploration starts.*

3.5 REFERENCES

[1] H. R. Mattison, *Project Daedalus: Astronomical Data on Nearby Stellar Systems*, Project Daedalus—Final Report, pp. S8–S18, published by the *Journal of the British Interplanetary Society (JBIS)*, The British Interplanetary Society, London, 1978.

[2] K. R. Lang, *Astrophysical Data—Planets and Stars*, Springer-Verlag, New York, 1992, in particular Chapter 29, pp. 758–887.

[3] F. R. Moulton, *An Introduction to Celestial Mechanics*, Second Revised Edition, Dover, New York, 1970.

[4] R. H. Battin, *An Introduction to the Mathematics and Methods of Astrodynamics*, American Institute of Aeronautics and Astronautics (AIAA) Education Series, AIAA, New York, 1987, in particular pp. 421–423.

4

Astrodynamics to exit the solar system at the highest speed

4.1 INTRODUCTION

To this author's knowledge, the first paper dealing with the problem of making a spacecraft exit the solar system at the maximum possible speed was published in 1972 by Krafft A. Ehricke [1]. The abstract of his paper summarizes his results: "For optimum propulsion energy management, the departing spacecraft should use a powered flyby maneuver at Saturn for insertion into a retrograde heliocentric orbit, followed by a Jupiter gravity-assist for injection into a hyperbolic orbit such that a final powered maneuver is applied at closest perihelion distance and highest perihelion approach velocity possible." Later, an algebraic error present in Ehricke's paper was corrected by G. Matloff and K. Parks [2]. Again we quote from the abstract: "An error in some previous considerations of gravity-assist interstellar propulsion is pointed out and corrected. The revised analysis is applied to powered and unpowered periapsis maneuvers."

4.2 A THEOREM BY CARLES SIMÒ

The calculations appearing in the present section are due to Carles Simò of the University of Barcelona (Spain). At the *Congress on Advances in Nonlinear Astrodynamics* (held at the Geometry Center of the University of Minnesota, Minneapolis, November 8–10, 1993), Simò handed this material over to this author as his contribution to the FOCAL space mission. We had presented the problems of the FOCAL space mission to him a few months earlier, on May 26–27, 1993, at an *Astrodynamics Congress* held in L'Aquila, Italy. We are grateful to Simò for this enlightening mathematical contribution.

4.2.1 Elementary background (planar problem)

Assume J is a planet around the Sun in circular orbit (to make things easy). The velocity is

$$\vec{v}_J = v_J \begin{bmatrix} -1 \\ 0 \end{bmatrix}. \tag{4.1}$$

Let a spacecraft approach J from beyond the J orbit (just as in the case regarded by K. Von Ehricke as the optimal one) with velocity (see Figure 4.1)

$$\vec{v}_1 = v_1 \begin{bmatrix} \cos\alpha \\ -\sin\alpha \end{bmatrix}. \tag{4.2}$$

Using zero-order approximation, the velocity of the spacecraft with respect to J is assumed to be hyperbolic:

$$\vec{v}_{\infty,in} = \vec{v}_1 - \vec{v}_J \tag{4.3}$$

whence

$$\vec{v}_{\infty,in} = \vec{v}_1 - \vec{v}_J = v_1 \begin{bmatrix} \cos\alpha \\ -\sin\alpha \end{bmatrix} - v_J \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} v_J + v_1\cos\alpha \\ -v_1\sin\alpha \end{bmatrix}. \tag{4.4}$$

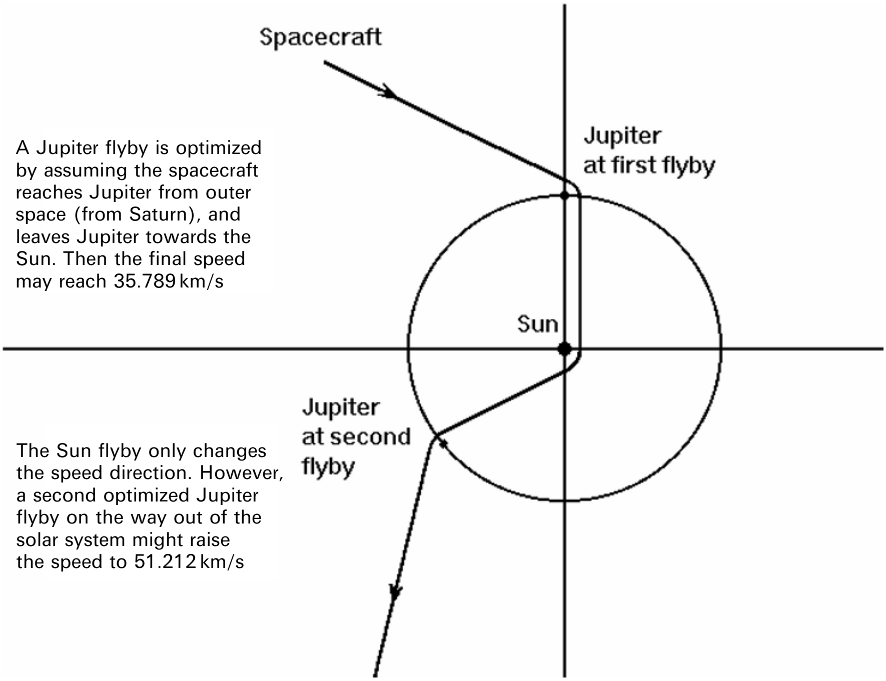


Figure 4.1. Optimal Jupiter flyby from the outer solar system (i.e., from the Saturn orbit) right toward the Sun. If this flyby is further applied in the reverse sequence, it could throw the spacecraft out of the solar system at about 50 km/s.

Passing close to J, the relative velocity is turned by an angle β, which depends on the minimum distance. In the sequel, we shall just refer to this angle β, rather than to the minimum spacecraft distance from J. Then

$$\vec{v}_{\infty,out} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \vec{v}_{\infty,in} \tag{4.5}$$

or

$$\begin{aligned} \vec{v}_{\infty,out} &= \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} v_J + v_1\cos\alpha \\ -v_1\sin\alpha \end{bmatrix} \\ &= v_J \begin{bmatrix} \cos\beta \\ -\sin\beta \end{bmatrix} + v_1 \begin{bmatrix} \cos(\alpha+\beta) \\ -\sin(\alpha+\beta) \end{bmatrix}. \end{aligned} \tag{4.6}$$

On the other hand, one has

$$\vec{v}_{\infty,out} = \vec{v}_2 - \vec{v}_J. \tag{4.7}$$

By setting equations (4.6) and (4.2.7) equal to each other (using Equation (4.1) also), one gets

$$\vec{v}_2 = v_1 \begin{bmatrix} \cos(\alpha+\beta) \\ -\sin(\alpha+\beta) \end{bmatrix} + v_J \begin{bmatrix} \cos\beta - 1 \\ -\sin\beta \end{bmatrix}. \tag{4.8}$$

The two components of this vector along the x and y axes are thus, respectively,

$$\begin{cases} v_{2x} = v_1\cos(\alpha+\beta) + v_J(\cos\beta - 1) \\ v_{2y} = -v_1\sin(\alpha+\beta) - v_J\sin\beta. \end{cases} \tag{4.9}$$

From them we infer that the modulus of the $\vec{v}_2$ vector is given by

$$v_2^2 = v_{2x}^2 + v_{2y}^2 = v_1^2 + v_J^2\left(4\sin^2\left(\frac{\beta}{2}\right)\right) + 2v_1v_J(\cos\alpha - \cos(\alpha+\beta)). \tag{4.10}$$

4.2.2 Optimization of a single Jupiter flyby

Assume a spacecraft is launched far away from the Sun and up to the Saturn orbit, so that its motion is close to parabolic. Then suppose the spacecraft "comes back" and approaches Jupiter, flybys it, and points towards the Sun, S, just at the exit of Jupiter's sphere of influence (spherical region of space around Jupiter where Jupiter's gravity is still higher than the Sun's). This implies $\alpha + \beta = \pi/2$ to exit towards the Sun. By inserting the constraint

$$\beta = \frac{\pi}{2} - \alpha \tag{4.11}$$

into the expression of the modulus of the flyby exit velocity $\vec{v}_2$, we can express the modulus of this velocity as a function of the *single* angle variable α:

$$v_2^2(\alpha) = v_1^2 + 2v_J^2(1 - \sin\alpha) + 2v_1v_J\cos\alpha \tag{4.12}$$

whence

$$\|\vec{v}_2\| = \sqrt{v_1^2 + 2v_J^2(1 - \sin\alpha) + 2v_1 v_J \cos\alpha}. \tag{4.13}$$

We can now optimize the Jupiter flyby. In other words, it is possible to determine the "optimal value" of the angle α between the direction of the spacecraft entering the Jupiter sphere of influence and the direction of the Jupiter velocity around the Sun. This optimal value of the angle α is obtained by requiring that the above function of α is a maximum; that is,

$$\frac{d\|\vec{v}_2\|}{d\alpha} = \frac{-2v_J^2 \cos\alpha - 2v_1 v_J \sin\alpha}{2\sqrt{v_1^2 + 2v_J^2(1 - \sin\alpha) + 2v_1 v_J \cos\alpha}} = 0. \tag{4.14}$$

This yields

$$\tan\alpha = -\frac{v_J}{v_1} \tag{4.15}$$

that is

$$\sin\alpha = \frac{-v_J}{\sqrt{v_1^2 + v_J^2}}, \quad \cos\alpha = \frac{v_1}{\sqrt{v_1^2 + v_J^2}}. \tag{4.16}$$

The modulus of the spacecraft velocity, when it exits the Jupiter sphere of influence, is obtained by inserting these expressions for $\sin\alpha$ and $\cos\alpha$ into the formula for $v_2(\alpha)$. After some manipulation, the result is

$$v_2^2 = v_1^2 + 2v_J^2 + 2v_J\sqrt{v_1^2 + v_J^2}. \tag{4.17}$$

4.2.3 Two optimized Jupiter flybys plus one intermediate Sun flyby

Assume that the spacecraft entrance velocity into the Jupiter sphere of influence, v_1, is the parabolic velocity of any spacecraft with respect to Jupiter—in practice this amounts to supposing that the spacecraft reaches Jupiter "from very far away" (e.g., from Saturn):

$$v_1 = \sqrt{2}\, v_J. \tag{4.18}$$

Then the corresponding optimized exit velocity after the Jupiter flyby and directed towards the Sun is

$$v_2^2 = (4 + 2\sqrt{3})v_J^2 \quad \text{i.e., } v_2 = \sqrt{4 + 2\sqrt{3}}\, v_J = 2.73205\, v_J. \tag{4.19}$$

In words, the best Jupiter flyby for a spacecraft "to fall" towards the Sun increases the modulus of the spacecraft velocity by 2.73 times.

Now consider a double Jupiter flyby. This means the following:

(1) A first Jupiter flyby "from the outside", as just described.
(2) A (hyperbolic) Sun flyby that just changes the spacecraft direction and keeps the incoming and outgoing velocities module just the same.

(3) A second Jupiter flyby performed in a fashion such that the increase in the spacecraft speed is again optimized. By calling v_3 the incoming speed and v_4 the outgoing speed, for reasons of geometrical symmetry, one has
(4) $v_3 = v_2$.

On the other hand, the formula relating v_4 to v_3 for the exit velocity after the second flyby reads

$$\begin{aligned} v_4^2 &= v_3^2 + 2v_J^2 + 2v_J\sqrt{v_3^2 + v_J^2} = v_2^2 + 2v_J^2 + 2v_J\sqrt{v_2^2 + v_J^2} \\ &= (4 + 2\sqrt{3})v_J^2 + 2v_J^2 + 2v_J^2\sqrt{(4 + 2\sqrt{3})v_J^2 + v_J^2} \\ &= (6 + 2\sqrt{3})v_J^2 + 2v_J^2\sqrt{5 + 2\sqrt{3}} = (6 + 2\sqrt{3} + 2\sqrt{5 + 2\sqrt{3}})v_J^2. \end{aligned}$$

That is,

$$v_4 = \sqrt{6 + 2\sqrt{3} + 2\sqrt{5 + 2\sqrt{3}}}v_J = 3.90931 v_J = 51.212\ \text{km/s}$$

having replaced the numerical value of the average speed of Jupiter, $v_J \approx 13.1$ km/s.

To summarize, a sequence of two optimized Jupiter flybys with a Sun flyby in between does indeed allow any spacecraft to leave the solar system with an exit speed of 51.12 km/s. Such a speed would bring our spacecraft to 550 AU in 50.911 years. However, the time the spacecraft must spend within the solar system before exiting is at least 10 or more years, bringing the total amount of time required to reach 550 AU to 60 years or more—not a desirable result.

Furthermore, all these classical, unpowered, flybys were supposed to take place in the ecliptic plane, or in planes only slightly inclined with respect to the ecliptic. Now, it just happens that (by a stroke of luck!) the "Galactic anticenter" (towards which we must point our spacecraft in order to "watch" at the Galactic center) is located only about 10 degrees north of the ecliptic plane (i.e., near the star Elnath in Auriga). This means that only a small correction for the inclination would be required. However, consider the case of a spacecraft launched towards the point of Galactic longitude +90° and Galactic latitude 0° (optimal direction for SETI if the *Theory of the Galactic Belt of Life* by Marochnick and Mukhin [3] and Balàzs [4]) is correct). This is in Cygnus, significantly away from the ecliptic! Thus, we can hardly hope to use a classical planetary flyby sequence to launch our spacecraft there. The solution to the highest exit velocity problem requires much more effort for anyone interested in pursuing the FOCAL space mission.

4.3 A CHEMICALLY POWERED CLOSE-SUN FLYBY?

Kerry T. Nock of JPL long studied the TAU mission [5] and the means to exit the solar system with the highest possible speed. He once suggested to this author that a close-Sun flyby with a powerful perihelion thrust by conventional chemical propulsion would greatly increase the spacecraft speed as it left the solar system. Dr. Nock's idea is that this powered perihelion flyby could occur as close to the Sun as four solar radii, provided a suitable Sun shield was installed on the spacecraft. This would lead to a Δv of 5 km/s, enabling the spacecraft to exit the solar system at a speed of about 10 AU/yr. The target distance of 550 AU would thus be reached in about 50 years.

Whether or not Kerry Nock's ideas are feasible for the near future, it seems fruitful to develop the mathematical theory of the Sun flyby enhanced by a perihelion boost for storable-propellant chemical engines.

4.4 THEORY OF THE SUN FLYBY ENHANCED BY A PERIHELION BOOST

In practice, the speed of the FOCAL spacecraft will almost certainly be increased at the perihelion in either of the following ways:

(1) by virtue of a classical chemical booster;
(2) by deploying a solar sail just when the distance from the Sun is smallest (i.e., the overall momentum of the impinging photons on the sail surface is highest);
(3) by reversing the orbital angular momentum [6] of a sufficiently light solar sail;
(4) by nuclear propulsion.

To simplify things, it will be assumed that the perihelion boost is tangential to the spacecraft trajectory at the perihelion, so one only has to take into account a *speed increment* Δv_p changing the outgoing leg of the flyby. I shall denote by "primes" the values of physical quantities such as r, v, etc. *after* the boost has occurred. One thus has, by definition of impulsive approximation,

$$\begin{cases} v_p' = v_p + \Delta v_p \\ r_p' = r_p. \end{cases} \tag{4.20}$$

Clearly the hyperbolic trajectory after the perihelion boost has occurred will be different from the one before it occurred, and so the parameters of the outgoing hyperbolic leg will be designated a' and e'. To find out how these are related to those of the incoming hyperbolic leg, just rewrite (4.20) by virtue of the speed equation (3.14):

$$v_p' = \sqrt{GM\left(\frac{2}{r_p'} - \frac{1}{a'}\right)} = \sqrt{GM\left(\frac{2}{r_p} - \frac{1}{a'}\right)} = v_p + \Delta v_p. \tag{4.21}$$

Solving this for the new semi-major axis a', one finds

$$a' = \left[\frac{2}{r_p} - \frac{(v_p + \Delta v_p)^2}{GM}\right]^{-1} \tag{4.22}$$

Substituting this expression for a' into the hyperbolic approach speed formula (3.16), one gets the *boosted hyperbolic exit speed*

$$v'_\infty = \sqrt{-\frac{GM}{a'}} = \sqrt{(v_p + \Delta v_p)^2 - \frac{2GM}{r_p}}. \tag{4.23}$$

Having found the boosted exit speed, one still needs to find the boosted exit *direction* (i.e., the new hyperbola eccentricity e' and so the new angle ϑ'_∞ yielding the actual new, boosted exit direction). To find e' one has to resort to using

$$r_p = r'_p = a'(1 - e'). \tag{4.24}$$

Solving this for e' and replacing a' with the expression in (4.22), one gets the *eccentricity of the boosted hyperbola*:

$$e' = 1 - r_p\left[\frac{2}{r_p} - \frac{(v_p + \Delta v_p)^2}{GM}\right] = \frac{r_p(v_p + \Delta v_p)^2}{GM} - 1. \tag{4.25}$$

Finally, Equations (3.18) and (4.25) yield the corresponding new angle ϑ'_∞ between the perihelion radius and the boosted exit asymptote:

$$\vartheta'_\infty = \arccos\left[\left(1 - \frac{r_p(v_p + \Delta v_p)^2}{GM}\right)^{-1}\right]. \tag{4.26}$$

4.5 DETERMINING THE PERIHELION BOOST BY KNOWING THE TARGET STAR, THE TIME TO GET TO 550 AU, AND THE SUN APPROACH

This section is critical for the FOCAL mission analysis inasmuch as it paves the way to choosing the propulsion system to boost the spacecraft at the perihelion. Suppose one knows the following:

(1) The target star system to observe (i.e., the spacecraft exit direction out of the solar system, or the angle ϑ'_∞).
(2) The overall time that it will take FOCAL to get to 550 AU (i.e., the solar system boosted exit speed v'_∞).
(3) How the spacecraft is going to approach the Sun (i.e., what is going to be the perihelion distance r_p—which is determined by the thermal requirements upon the spacecraft), and what is going to be the Sun approach speed v_∞.

Then suppose one wishes to compute what boost Δv_p will get FOCAL at 550 AU in the specified time and in the right direction.

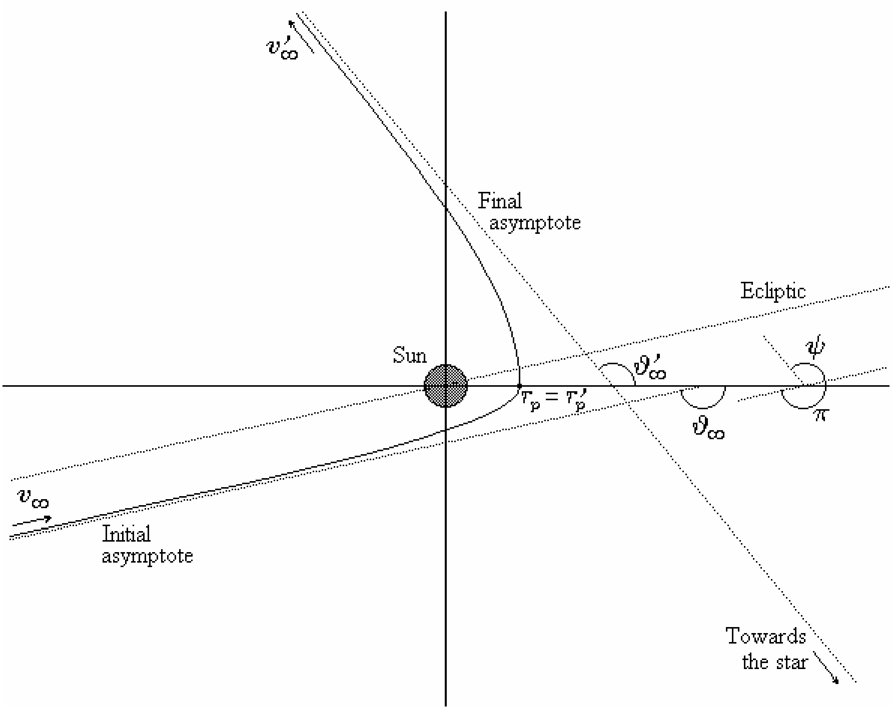


Figure 4.2. Basic geometry of the hyperbolic flyby *boosted at perihelion*.

This problem is solved as follows. First, a glance at Figure 4.2 shows that the four angles ϑ_∞, ϑ'_∞ and ψ are related by the geometrical equation

$$2\pi - \vartheta_\infty - \vartheta'_\infty = \pi - \psi.$$

This formula is the natural generalization of (3.27) in order to take the new angle ϑ'_∞ into account. Further, just as already was done in Section 3.4, here too only the *tight* Sun flybys will be considered. Therefore, (3.29) (i.e., $\psi = \pi - |\beta|$) is still valid and, by eliminating ψ between the last equation and (3.29), one gets

$$2\pi - \vartheta_\infty - \vartheta'_\infty = |\beta| \tag{4.27}$$

which we shall use in a moment.

Since the Sun approach is known by definition (i.e., v_∞ and r_p are known), one should first rewrite both v_p and ϑ_∞ as two functions of them. This can be done by virtue of the equations (3.26); that is,

$$\vartheta_\infty = \arccos\left[-\left(1 + \frac{r_p v_\infty^2}{GM}\right)^{-1}\right]$$

and

$$v_p = \sqrt{v_\infty^2 + \frac{2GM}{r_p}}.$$

Having thus determined ϑ_∞ from the first of these, then (4.27) is solved for ϑ'_∞ to yield

$$\vartheta'_\infty = 2\pi - \vartheta_\infty - |\beta| \tag{4.28}$$

where everything is known on the right-hand side.

One is now in a position to consider the boosted hyperbolic leg, where one needs to express v'_p and v'_∞ as functions of r'_p and ϑ'_∞ that one now knows. Obviously, all the equations for the boosted hyperbolic leg may be formally derived from the corresponding equations for the incoming leg via the formal replacements

$$\begin{cases} \vartheta_\infty \to \vartheta'_\infty \\ v_\infty \to v'_\infty \\ r_p \to r'_p \end{cases} \tag{4.29}$$

Thus, solving (3.20) for v_∞^2 one gets

$$v_\infty^2 = -\frac{GM}{r_p}\left(1 + \frac{1}{\cos\vartheta_\infty}\right)$$

which, by performing the replacement (4.29), yields

$$v_\infty'^2 = -\frac{GM}{r_p}\left(1 + \frac{1}{\cos\vartheta'_\infty}\right)$$

and finally

$$v'_\infty = \sqrt{-\frac{GM}{r_p}\left(1 + \frac{1}{\cos\vartheta'_\infty}\right)}. \tag{4.30}$$

To find the expression of v'_p start from (3.21) and perform the three replacements (4.29), finding

$$v'_p = v'_\infty\sqrt{\frac{1 - \cos\vartheta'_\infty}{1 + \cos\vartheta'_\infty}}.$$

Replacing then (4.30) into the last formula yields, after a few reductions,

$$v'_p = \sqrt{\frac{GM}{r'_p}\left(1 - \frac{1}{\cos\vartheta'_\infty}\right)}. \tag{4.31}$$

The required *boost* Δv_p is thus given by

$$\Delta v_p = v'_p - v_p = \sqrt{\frac{GM}{r'_p}\left(1 - \frac{1}{\cos\vartheta'_\infty}\right)} - \sqrt{v_\infty^2 + \frac{2GM}{r_p}}. \tag{4.32}$$

An alternative form of this equation is obviously

$$\Delta v_p = v_p' - v_p = \sqrt{\frac{GM}{r_p'}\left(1 - \frac{1}{\cos\vartheta_\infty'}\right)} - \sqrt{\frac{GM}{r_p}\left(1 - \frac{1}{\cos\vartheta_\infty}\right)} \tag{4.33}$$

showing that

$$\Delta v_p > 0 \quad \Leftrightarrow \quad \vartheta_\infty' < \vartheta_\infty \tag{4.34}$$

that is, Δv_p is positive for $\vartheta_\infty' < \vartheta_\infty$ only. In yet other words, rewriting (4.27) in the form

$$\vartheta_\infty + \vartheta_\infty' = 2\pi - |\beta| \tag{4.35}$$

one sees that Δv_p is positive only for $2\pi - |\beta| < 2\vartheta_\infty$; that is, only for

$$\vartheta_\infty > \pi - \frac{|\beta|}{2}. \tag{4.36}$$

Let us now rewrite this inequality in terms of the FOCAL spacecraft approach speed to the Sun. Recalling (3.26), that is

$$\vartheta_\infty = \arccos\left[-\left(1 + \frac{r_p v_\infty^2}{GM}\right)^{-1}\right]$$

and invoking (4.36), one has

$$\Delta v_p > 0 \quad \Leftrightarrow \quad \arccos\left[-\left(1 + \frac{r_p v_\infty^2}{GM}\right)^{-1}\right] > \pi - \frac{|\beta|}{2} \tag{4.37}$$

that is

$$\Delta v_p > 0 \quad \Leftrightarrow \quad -\left(1 + \frac{r_p v_\infty^2}{GM}\right)^{-1} > \cos\left(\pi - \frac{|\beta|}{2}\right) = -\cos\left(\frac{|\beta|}{2}\right) \tag{4.38}$$

Reversing the second inequality's sign, and finally solving for v_∞, one gets

$$\Delta v_p > 0 \quad \Leftrightarrow \quad v_\infty > \sqrt{\frac{GM}{r_p}\left(\frac{1}{\cos\frac{|\beta|}{2}} - 1\right)} \tag{4.39}$$

In conclusion, there is a lower limit to the FOCAL spacecraft approach speed to the Sun: it must not be less than the value

$$v_{\infty,min} = \sqrt{\frac{GM}{r_p}\left(\frac{1}{\cos\frac{|\beta|}{2}} - 1\right)}. \tag{4.40}$$

which corresponds to the no-boost case previously discussed in Chapter 3.

4.6 REFERENCES

[1] K. A. Von Ehricke, "Saturn–Jupiter Rebound—A method of High-Speed Spacecraft Ejection from the Solar System," *Journal of the British Interplanetary Society*, **25** (1972), 561–571.

[2] G. Matloff, and K. Parks, "Interstellar Gravity Assist Propulsion: A Correction and a New Application," *Journal of the British Interplanetary Society*, **41** (1988), 519–526.

[3] L. S. Marochnik, and L. M. Mukhin, "Belt of Life in the Galaxy," in "Bioastronomy—The Next Steps," *Proceedings of the 99th Colloquium of the International Astronomical Union*, held in Balatonfüred, Hungary, June 22–27, 1987, George Marx (editor), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988, pp. 49–59.

[4] B. Balázs,"The Galactic Belt of Intelligent Life", in "Bioastronomy—The Next Steps," *Proceedings of the 99th Colloquium of the International Astronomical Union*, held in Balatonfüred, Hungary, June 22–27, 1987, George Marx (editor), Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988, pp. 61–66.

[5] K. Nock, "TAU—A Mission to a Thousand Astronomical Units," paper #AIAA-87-1049 presented at the *19th AIAA/DGLR/JSASS International Electric Propulsion Conference*, May 11–13, 1987, Colorado Springs, CO.

[6] G. Vulpetti, "3D High-Speed Escape Heliocentric Trajectories by All-Metallic-Sail Low-Mass-Sailcraft," paper IAA-L-0505 presented at the *Second IAA International Conference on Low-Cost Planetary Missions* held April 16–19, 1996, at the Johns Hopkins University Applied Physics Laboratory, W. Farquhar (coordinator), and then printed in *Acta Astronautica*, **39**(1–4) (1996), 161–170.

5

SETI and the FOCAL space mission

5.1 INTRODUCTION

SETI, or the *Search for ExtraTerrestrial Intelligence*, started in 1959 with the seminal paper of Cocconi and Morrison [1]. SETI was experimentally pursued for the first time by Drake in 1960 (Project Ozma, [2], Reading #25) and later developed into a large body of interdisciplinary knowledge [3–7].

The central problem of experimental SETI is to recover weak radio signals out of noise. In isolation, the problem would not be difficult to solve with the aid of modern filtering algorithms and computers, but there are a number of complications.

(1) We don't know the radio frequency at which the extraterrestrials may be trying to communicate with us.
(2) We don't know from which direction in the sky the signals may be reaching us.
(3) We don't know how to distinguish a "natural" signal (i.e., a radio emission caused by some astrophysical mechanism), from an "intelligent" signal (i.e., a radio emission intentionally broadcast by ETs whose civilization has achieved a level of technological development comparable or more advanced than ours).
(4) In case we do detect a non-natural radio signal, it is not clear how we will deduce its meaning.

Tentative solutions to the above four complications have been provided by the worldwide community of experimental SETI-radioastronomers consisting mainly of Americans, but also Russians (formerly Soviets), French, Dutch, Australians, Argentinians, Japanese, Italians, etc. over the years since 1960. In summary,

(1) As to the frequencies to examine, all frequencies between about 1 GHz and 10 GHz are suited for interstellar communications, but those in the range 1 GHz–2 GHz seem to be the most appropriate ones. In particular, "magic"

frequencies (i.e., optimal frequencies for communication because all "ET radio-astronomers" must know their numerical values) are supposed to be the neutral hydrogen line (1,420 MHz) and the hydroxyl lines (1,612 MHz, 1,615 MHz, 1,667 MHz, and 1,720 MHz). The portion of spectrum between these lines is nicknamed the *waterhole*, and Galactic communications would "meet" around the waterhole as animals gather around water ponds in drought.

(2) This problem of which direction to look in was being solved at NASA by means of the *All-Sky Survey*, started at the Goldstone 70-meter antenna on October 12, 1992, but terminated abruptly by the U.S. Congress in October 1993. A previous all-sky search had been started by the Planetary Society in 1985 by resorting to more modest antennas located in Massachusetts and Argentina: this was the *META Project*, later technically upgraded over the years up to the *BETA II Project*. Apart from these two projects, the vast majority of SETI searches were targeted on stars expected to harbor life because they are similar to the Sun.

(3) To distinguish a "natural" signal from an "intelligent" signal we look at the signal intensity profile around the signal's central frequency. If the profile is Gaussian, the signal is expected to be natural; if the profile is a very narrow peak (almost comparable with a Dirac delta function) then the signal is expected to have been broadcast by a technologically advanced civilization. This we like to call *the narrowband assumption in SETI* and we shall discuss it in Section 5.2.

(4) To understand the meaning of an alien intelligent signal we probably need some Cosmic Language based on a mathematical scheme integrated with knowledge from other branches of the sciences, such as physics, chemistry, and biology. Hans Freudenthal's *Lincos (Design of a Language for Cosmic Intercourse)* was the first human attempt in this direction (1960).

5.2 THE NARROWBAND ASSUMPTION IN SETI

All SETI searches carried on thus far have been for narrowband signals, simply because SETI-radioastronomers believe that an extraterrestrial civilization wishing to make itself known all over the Galaxy would broadcast radio signals easily distinguishable from natural emissions. Since natural emissions have a Gaussian intensity profile around their own central frequency, ET would replace this Gaussian by a Dirac delta function (i.e., an—almost—infinitely narrow peak), to let us understand that it was an "intelligent" being, rather than Mother Nature, to send us that wave package. We shall call this widespread belief the *narrowband assumption* in SETI.

No mathematical proof of the narrowband assumption seems to have been given. We would like to provide one here, based on the well-known information theory put forward by Claude Shannon in 1948. For this proof, we are going to extend an argument given in 1964 by the leading Russian SETI expert Nikolai S. Kardashev. His paper [8] was seminal, at least in that it put forward the classification of extraterrestrial civilizations as Type I, II, and III, according to whether the extraterrestrials were able to funnel the energy of their own planet, solar system, or galaxy, respectively. In that paper Kardashev also used Shannon's formula for the rate of information transmission within a certain information channel over the frequency band between f_1 and f_2:

$$R = \int_{f_1}^{f_2} \log_2 \left[1 + \frac{S(f)}{N(f)} \right] df \tag{5.1}$$

where $S(f)$ and $N(f)$ are the power spectral densities of the useful signal and noise, respectively. Kardashev states that "by solving the appropriate variational problem, we may show the maximum rate of information transmission to be achieved under the condition

$$S(f) + N(f) = N(f_1) = N(f_2). \tag{5.2}$$

Here f_1 and f_2 are the bounds of the transmitter's transmission band. It is accordingly quite clear that the spectrum of the artificial source must show the reverse-shaped parabola-like equation

$$S(f) = N(f_1) - N(f) = N(f_2) - N(f). \tag{5.3}$$

That is, the spectrum of the artificial radio emission must feature a maximum." Kardashev omitted the mathematical steps leading from (5.1) to (5.3) because they are mathematically trivial. Yet, we would like to show that expressing them explicitly pays off, inasmuch as it offers a mathematical proof of the narrowband assumption universally adopted by SETI searchers. We do so by taking the variation of (5.1) with respect to the unknown function $S(f)$ together with the two normalization conditions fulfilled by the signal and noise, namely,

$$\int_{f_1}^{f_2} S(f)\, df = P_S \quad \text{and} \quad \int_{f_1}^{f_2} N(f)\, df = P_N \tag{5.4}$$

where P_S and P_N are the total signal and noise power over the given bandwidth, respectively. This results in the variational equation

$$\delta \int_{f_1}^{f_2} \left(\frac{1}{\ln 2} \ln \left[1 + \frac{S(f)}{N(f)} \right] + \lambda N(f) + \mu S(f) \right) df = 0 \tag{5.5}$$

where λ and μ are Lagrange multipliers. Performing differentiation under the integral sign with respect to the function to be optimized (i.e., $S(f)$), one has

$$\frac{1}{\ln 2} \cdot \frac{\dfrac{1}{N(f)}}{1 + \dfrac{S(f)}{N(f)}} + \mu = 0 \tag{5.6}$$

that has a solution of the form

$$S(f) = -\frac{1}{\mu \ln 2} - N(f). \tag{5.7}$$

Next, the Lagrange multiplier μ must be determined by integrating both sides of the solution (5.7) with respect to f between f_1 and f_2, and then invoking the normalization conditions (5.4)

$$\int_{f_1}^{f_2} S(f)\,df = -\int_{f_1}^{f_2} \frac{1}{\mu \ln 2}\,df - \int_{f_1}^{f_2} N(f)\,df. \tag{5.8}$$

The result is

$$\mu = \frac{f_1 - f_2}{(P_S + P_N)\ln 2}. \tag{5.9}$$

Substituting this into (5.7) gives

$$S(f) = \frac{P_S + P_N}{f_2 - f_1} - N(f). \tag{5.10}$$

This formula can also be given in different form by recalling the definition of the boundaries of the extraterrestrial transmission bandwidth:

$$S(f_1) = 0 \quad \text{and} \quad S(f_2) = 0. \tag{5.11}$$

Equation (5.10) then yields

$$N(f_1) = \frac{P_S + P_N}{f_2 - f_1} \quad \text{and} \quad N(f_2) = \frac{P_S + P_N}{f_2 - f_1} \tag{5.12}$$

whence Kardashev's formula (5.3) (or Equation (4) in [8], Reading #28) is obtained at once.

Now, we want to use (5.10) to prove the narrowband assumption. The argument is as follows: an extraterrestrial civilization wishing to make itself known would try to send as much information as possible about itself. It would thus try to maximize the information transmission rate (5.1) over the transmission bandwidth $f_2 - f_1$ of their apparatuses. Then, (5.10) shows that, keeping both the total signal and the noise powers (P_S and P_N) fixed over the given bandwidth, the narrower this bandwidth is, the clearer the signal spectral density $S(f)$ stands out against the noise spectral density $N(f)$; that is,

$$\begin{aligned} \lim_{f_2 \to f_1} S(f) &= \lim_{f_2 \to f_1} \left[\frac{P_S + P_N}{f_2 - f_1} - N(f) \right] \\ &= \infty - N(f_1) = \infty - N(f_2) = \infty. \end{aligned} \tag{5.13}$$

In conclusion, ETs must transmit over the narrowest possible bandwidths to let their messages be understood clearly against the background noise. And on Earth one must use very narrowband spectral analyzers to detect ET candidate signals. Over the years the bandwidths that humans are using have steadily decreased to 1 Hz and even less: hundreds or thousands of Hz are now achievable by dedicated computers like those of *Project Phoenix* (formerly NASA–SETI Project) and the *BETA 2* system run at Harvard by the Planetary Society.

5.3 A SHORT INTRODUCTION TO THE KLT

Understanding the mathematical model of a physical fact may be difficult to people who are not familiar with the required mathematical background. Yet, mathematics is just the correct language by which physics and engineering achieve success. Translating this mathematical language into the language of "common" words may be desirable whenever a mathematical advance is made that has to be described to newcomers in "easy terms".

This section is devoted to a rather new mathematical tool that may improve our understanding of physical phenomena: the *Karhunen–Loève Transform*, hereafter abbreviated *KLT*. Essentially, it is something superior than the classical Fourier Transform (FT). To explain why, let us use a comparison in classical mechanics. Consider an object (e.g., a book), and a three-axis rectangular reference frame, oriented arbitrarily with respect to the book. Now, all the mechanical properties of the book itself are described by a 3×3 (symmetric) matrix called an "inertia matrix" (or "inertia tensor") whose elements are, in general, non-zero. Handling a matrix whose elements are all non-zero is obviously more complicated than handling a matrix where all elements are zeros except for those lying on the main diagonal (this is called a "diagonal matrix"). Thus, one may be led to wonder whether a certain axes transformation exists that changes the inertia matrix of the book into a diagonal matrix.

Classical mechanics shows that only one special orientation of the rectangular frame with respect to the book exists, yielding a diagonal inertia matrix: the three axes must coincide with a set of three vectors (parallel to the book edges) called "eigenvectors" or "proper vectors" of the book. In other words, each body possesses an intrinsic set of three rectangular axes, called "eigenvectors" of the body, that describes its mechanical properties most simply. This is referred to as "diagonalizing the matrix".

Now, let me go to signal processing, which is our interest here. By adding random noise to a deterministic signal one obtains what is called a "noisy signal" or, in case the power of the signal is much smaller than the power of the noise "a signal buried into the noise". Since the noise + signal $X(t)$ is a random function of the time, one can describe it by a statistical quantity called autocorrelation (or simply "correlation"), defined as the mean value of the product of the values of $X(t)$ at two different instants t_1 and t_2—that is $E\{X(t_1)X(t_2)\} \equiv \langle X(t_1)X(t_2)\rangle$. This correlation, obviously symmetric in t_1 and t_2, may play just the same role as the inertia matrix in the book example. Thus, if one seeks for the eigenvectors of the correlation, and then changes the reference frame to the new set of vectors, the easiest possible description of the signal + noise is achieved. This is the key idea behind the KLT.

One may also look at the KLT from a slightly different point of view. In mathematical physics the well-known "method of normal coordinates" allows one to describe the "small oscillations" of a dynamical system in the best possible way by expressing the Lagragian as a sum of Lagrangians, each of them representing a simple harmonic oscillator. This is the result of a "principal axes" transformation (i.e., a

Lagrangian coordinate transformation that yields the separation of variables naturally). The KLT is just the statistical version of that.

5.4 MATHEMATICS OF THE KLT

The KLT [9] is named after two mathematicians, the (living) Finn, *Kari Karhunen* and the French American, *Michel Loève* (1907–1979), who proved independently and at about the same time (1946) that the series (5.14) hereafter is convergent. Put it this way, the KLT looks like a purely mathematical topic, but this is not, of course, the case. Using the language of engineers and radioastronomers, we say that it is possible to represent the signal + noise $X(t)$ as the infinite series (K-L, or KLT, expansion)

$$X(t) = \sum_{n=1}^{\infty} Z_n \, \phi_n(t) \quad \text{with } 0 \le t \le T. \tag{5.14}$$

Assuming that the noise (auto)correlation $E\{X(t_1)X(t_2)\}$ is a known function of t_1 and t_2, it can be proved that the functions $\phi_n(t)$ ($n = 1, 2, \ldots$) are the eigenfunctions of the correlation—namely, the solutions to the integral equation

$$\int_0^T E\{X(t_1)X(t_2)\} \, \phi_n(t_2) \, dt_2 = \lambda_n \, \phi_n(t_1). \tag{5.15}$$

These $\phi_n(t)$ form an orthonormal basis in the Hilbert space, and they actually are the optimal basis to describe the noisy signal, better than any classical Fourier basis. One can thus say that the KLT adapts itself to the shape of the signal + noise, whatever it is.

A further advantage of KLT is that the Z_n in (5.14) are orthogonal random variables (i.e., that $E\{Z_m Z_n\} \equiv \langle Z_m Z_n \rangle = \lambda_n \, \delta_{mn}$). If $X(t)$ is a Gaussian process, this orthogonality amounts to statistical independence, meaning that the terms in the KLT expansion are uncorrelated. Since the constants λ_n are both the (all positive) eigenvalues and the variances of the random variables Z_n, any KLT expansion, when truncated to keep only the first few terms, is the best approximation to the original function $X(t)$ in the mean square sense.

Finally, the mathematical theory of KLT shows that the process $X(t)$ need not be stationary. This too spells the difference against the classical Fourier techniques that hold rigorously true for stationary processes only.

5.5 KLT FOR SETI

The narrowband assumption was the rationale behind all ETI radio searches made thus far all over the world. Consequently, only Fourier Transform (FT) or Fast Fourier Transform (FFT) techniques were used to find the very narrow bandwidth (called "bin" in SETI jargon) in which an unusual amount of received radio energy might indicate the presence of a signal, either sinusoidal or pulsed.

In this section, however, we would like to maintain that the traditional usage of the FFT within SETI might sooner or later be replaced by the adoption of the KLT. This is no new idea. In 1983 the French SETI-radioastronomer, François Biraud, was the first person within the SETI community to describe the advantages of KLT over FFT for detecting wideband signals [10]. Apart from the technical issue, there seems to be another and deeper one: adopting the KLT means to be ready for the unexpected. Indeed, we know nothing about the nature of ETI signals: we have just made a set of "reasonable" assumptions, and we are trying to see whether they serve to put us in touch with the rest of the Universe. Enlarging the possible signals to look for by shifting from FFT to KLT can only help.

Very promising work on the KLT for SETI was done by Robert S. Dixon and Charles A. Klein, both with the Ohio State University in Columbus, Ohio [11]. After acknowledging that the KLT is more general than the FT because it makes no assumption about signal periodicity or waveform, these authors took one important new step in pointing out that only the largest of the KLT components need be calculated, in contrast to the FT, where all components (one for each frequency) must be calculated. This largest KLT component, or coefficient, is what the mathematicians call the "dominant eigenvalue" in the solution of the integral equation (5.15). Dixon and Klein did not attempt to prove any mathematical theorem about this fact, but they did numerical computer experiments showing that this must be the case.

One might ask what prevents radioastronomers from using KLT for SETI now. The simple answer is the computational burden. In fact, the KLT kernel is the correlation, and, being the mean value of the product of two random variables, this kernel is not separable. In general, one cannot hope for the existence of a fast KLT algorithm. In turn, this means that the computer time required to calculate the eigenvalues and eigenvectors of a correlation matrix of order N is proportional to N^2, rather than to $N \ln N$ as for the FFT.

Nevertheless, several concurrent developments seem to be paving new ways to overcome the above difficulties. On the one hand, the steady improvements in computer hardware and parallelization techniques seem to lead to very fast algorithms capable of getting the eigenvalues and eigenvectors of a large square symmetric matrix such as the correlation. On the other hand, the progress in the mathematical theory of the KLT has been steady since the early 1950s, and we mention some results of potential interest for SETI applications.

(1) For an exponential correlation of the form

$$E\{X(t_1)X(t_2)\} = e^{-|t_2-t_1|} \tag{5.16}$$

the problem of finding the KLT was completely solved as far back as 1958 [12, pp. 99–101].

(2) The correlation (5.16) is just an example of stationary random process; that is, a process having a correlation of the form

$$E\{X(t_1)X(t_2)\} = f(|t_2 - t_1|). \tag{5.17}$$

Now, for the general stationary correlation (5.17), Srinivasan and Sukavanam [13, 14] obtained a solution to the integral equation (5.15), where the right-hand side is an arbitrary real-valued function $f(\cdots)$ defined on the positive real axis. They assumed that $f(\cdots)$ admits a Laplace transform $f^*(\cdots)$, and in practice confined themselves to the case where the latter is given by

$$f^*(z) = \frac{g(z)}{h(z)} = \frac{\text{polynomial of degree not exceeding } (n-1)}{\text{polynomial of degree } n} \tag{5.18}$$

though they state that their arguments can be easily extended to the more general case where $f^*(\cdots)$ admits a Mittag–Leffler expansion. For the case of (5.18), they gave explicit, though complicated, formulas for computing the eigenvalues numerically, but apparently not for computing the eigenfunctions. This prevents further study to be carried out on the possibility that a fast K-L algorithm might exist for the stationary correlation (5.17). More investigations are needed.

(3) S. Watanabe [15] introduced the method of the K-L expansion into the realm of pattern recognition in 1965. In this application, an image is to be represented in terms of an optimal coordinate system, and the set of basis vectors which make up this coordinate system is referred to as an *eigenpicture*. The basis vectors are simply the eigenfunctions of the covariance matrix of the ensemble of images. The state of the art in the application of the KLT to images is described in [16], and this shows that computers already exist that are powerful enough to apply the KLT to image processing.

(4) The first (apparently) *fast* K-L algorithm [17] was obtained in 1976 by A. K. Jain. An excellent description of his mathematical algorithm was given by A. Rosenfeld and A. C. Kak in [18]. The key idea is to make the correlation separable by resorting to exponential functions. For instance, let an image belonging to a given set of images (random field) be sampled on a $f^*(\cdots)$ square sampling lattice, and let $f(m,n)$ denote the samples, where both m and n take on integer values from 0 through $N-1$. Then the assumed correlation is of the type

$$E\{f(m,n)f(p,q)\} = r_1^{|m-p|} r_2^{|n-q|} = e^{-a|m-p|}\, e^{-b|n-q|} \tag{5.19}$$

where r_1, r_2, a, and b are constants, the former two being less than unity. For this (discrete) correlation both the K-L eigenvalues and eigenfunctions may be explicitly found, as in [17]. There Jain has shown that if the image boundary pixels are known, they may be used to modify the rest of the image in such a way as to possess a K-L transform that can be implemented using FFT (or the more recently developed fast sine transform). Thus, Jain's result is essentially a reduction of KLT to FFT preserving the typical advantages of both.

5.6 CONCLUSION: ADVANTAGES OF THE KLT FOR THE FOCAL SPACE MISSION

Since the 1950s the number of applied scientists using the KLT for their research has slowly increased, but the KLT still lies outside the realm of most current scientific research. This situation of neglect seems to have been caused primarily by two obstacles.

(1) The exceedingly heavy computational burden required by the KLT.
(2) The many obscure points still plaguing the KLT mathematical theory.

While the first obstacle might be overcome relatively soon by the development of parallel processing computers, paving the mathematical way requires more effort.

SETI is a field of research where the KLT might distinguish itself in comparison with the FFT. Only the KLT, in fact, would reveal wideband signals whatever the nature of the noise spectrum, and whether or not the random process is stationary. The time appears to be ripe for the KLT to be taken seriously by the SETI as well as by other signal-processing investigators all over the world.

There is, however, an additional and very important point that we would like to stress: the KLT is not just used for filtering weak signals out of the noise: *the KLT is used for data compression also*. In fact, consider the four basic steps of the KLT.

(1) Find the eigenvectors of the autocorrelation of the set of data.
(2) Assume the set of eigenvectors as new vectors.
(3) Expand the set of data over this set of vectors and then truncate it by (arbitrarily) declaring that whatever is beyond a certain (low) correlation value is "noise" (i.e., unnecessary data).
(4) Reverse-transform to the original set of axes and reconstruct the "filtered" set of data.

Data compression occurs at Step (3).

Moreover, since data compression is essential in the radio link at huge distances like those to be reached by FOCAL, one concludes that the KLT is the best possible way of compressing the data sent by FOCAL to the Earth (i.e., the best way of letting FOCAL be successful).

5.7 REFERENCES

[1] G. Cocconi, and P. Morrison, "Searching for Interstellar Communications," *Nature*, **184** (1959), 844.
[2] D. Goldsmith, *The Quest for Extraterrestrial Life: A Book of Readings*, University Science Books, Mill Valley, CA, 1980.

[3] "The Search for Extraterrestrial Life: Recent Developments," *Proceedings of the International Astronomical Union Symposium No. 112*, held in Boston, June 18–21, 1984, Michael D. Papagiannis (editor), D. Reidel, Dordrecht, The Netherlands, 1985.
[4] "Bioastronomy—The Next Steps," *Proceedings of the 99th Colloquium of the International Astronomical Union*, held in Balatonfüred, Hungary, June 22–27, 1987, George Marx (editor), Kluwer Academic, Dordrecht, The Netherlands, 1988.
[5] "Bioastronomy—The Search for Extraterrestrial Life—The Exploration Broadens," *Proceedings of the Third International Symposium on Bioastronomy*, held at Val Cenis, Savoie, France, June 18–23, 1990, Jean Heidmann and Mike Klein (editors), *Lecture Notes in Physics*, No. 390, Springer-Verlag, Berlin, 1991.
[6] "Progress in the Search for Extraterrestrial Life," *Proceedings of the Fourth International Symposium on Bioastronomy*, held at the University of California at Santa Cruz, August 16–20, 1993, G. Seth Shostak (editor), Astronomical Society of the Pacific Conference Series, Vol. 74, 1995.
[7] "Astronomical and Biochemical Origins and the Search for Life in the Universe," *Proceedings of the Fifth International Conference on Bioastronomy* (IAU Colloquium No.161), held at Capri, Italy, July 1–5, 1996, C. Cosmovici, S. Bowyer, and D. Werthimer (editors), Editrice Compositori, Bologna, Italy, 1997.
[8] N. S. Kardashev, "Transmission of Information by Extraterrestrial Civilizations," Reading #28 in [2].
[9] C. Maccone, *Telecommunications, KLT and Relativity*, IPI Press, Colorado Springs, 1994, ISBN 1-880930-04-8.
[10] F. Biraud, "SETI at the Nançay Radiotelescope," *Acta Astronautica*, **10** (1983), 759–760.
[11] Robert S. Dixon, and Charles A. Klein, "On the Detection of Unknown Signals," paper presented at the *USA-USSR Joint Conference on the Search for Extraterrestrial Intelligence*, held at the University of California at Santa Cruz, August 5–9, 1991.
[12] Wilbur B. Davenport, and William R. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.
[13] S. K. Srinivasan, and S. Sukavanam, "Photo-count statistics of Gaussian light of arbitrary spectral profile," *J. Phys.*, **A-5** (1972), 682–694.
[14] S. K. Srinivasan, *Stochastic Point Processes and Their Applications*, Griffin, London, 1974, in particular pp. 83–87 and pp. 95–98.
[15] S. Watanabe, "Karhunen–Loève expansion and factor analysis theoretical remarks and applications," *Proceedings Fourth Prague Conference on Information Theory*, 1965.
[16] M. Kirby and L. Sirovich, "Application of the Karhunen–Loève Procedure for the Characterization of Human Faces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **12** (1990), 103–108.
[17] A. K. Jain, "A fast Karhunen–Loève transform for a class of random processes," *IEEE Trans. Commun.*, **COM-24** (1976), 1023–1029.
[18] A. Rosenfeld, and A. C. Kak, *Digital Picture Processing*, 2nd Edition, Academic Press, New York, 1982.

Figure 5.1. Plaque of the "Giordano Bruno Award" presented by the SETI League to the author "for Technical Excellence in the service of SETI" on April 27, 2002, at the *SETI League Conference*, held in Trenton, New Jersey. More at the website: *http://www.setileague.org/awards/brunowin.htm*

Figure 5.2. Claudio Maccone in front of the Giordano Bruno Memorial in Rome. This Memorial was erected in 1889 right on the spot (in "Campo dei Fiori" square) where Giordano Bruno (1548–1600) was burned at the stake on February 17, 1600, by order of the Roman Inquisition. Photo shot on May 29, 2002.

6

GL-SETI (gravitational lensing SETI): Receiving far ETI signals focused by the gravity of other stars

6.1 INTRODUCTION

The SETI League (website: *http://www.setileague.org*) is a worldwide organization of thousands of SETI supporters who generally use small TV dishes and PCs to do SETI searches from their own backyards. In general, such modest apparatuses could hardly be expected to lead to a contact with ETI, except for ETIs living on planets of nearby stars, but the notion of gravitational lensing changes this picture completely. In fact, it could well happen that a very precise alignment occurred casually between the ETI source, an intermediate star acting as a focusing device, and the Earth. Thus, a transient but sufficiently strong ETI signal could be detected on Earth even if it comes from a very far source in the Galaxy and even if it is detected with a small dish apparatus. It all depends on the mass of the intervening star, on the precision of the alignment, and of course on the power of the ETI emitted radio waves.

This great step ahead in SETI was first suggested to the author of this book by the SETI League President, Richard Factor, who first presented a paper on this topic at *SETICon '01*, the *2001 SETI League Technical Symposium and Annual Membership Meeting*, held in Trenton, NJ, April 28–29, 2001 [1]. The author of this book is greatly indebted to Richard Factor for giving him permission to reproduce from [1] the following two sections of this chapter. The name "G-L SETI" for this new way of doing SETI was suggested by the author to Richard Factor by e-mail.

6.2 ONLY TWO TYPES OF SETI SEARCHES FROM THE EARTH UP TO 2001[1]

6.2.1 Introduction

Microwave SETI (The Search for Extraterrestrial Intelligence) focuses on two primary strategies, the "Targeted Search" and the "All-Sky Survey". Although the goal of both strategies is the unequivocal discovery of a signal transmitted by intelligent species outside our solar system, they pursue the strategies in very different manners and have vastly different requirements. This chapter introduces a third strategy, also with the goal of unequivocal discovery of an extraterrestrial signal, with equipment and data processing requirements that are substantially different from the commonly used strategies. This strategy is particularly suitable for use with smaller radio telescopes and has budgetary requirements suitable for individual researchers.

6.2.2 Background

Since the first tentative SETI experiment in the 1960s, increasingly larger radio telescopes and more powerful signal processing engines have been searching the sky for signals. Perforce these searches have been limited to looking largely for continuous or pulsed narrowband signals since these are the most likely to be detectable, and are most identifiable as being of unnatural origin. A number of "hits" have been recorded, beginning with the famous "Wow" signal and continuing to the present. After weeding out cases of equipment problems and man-made interference, a number of candidate signals remained, any of which might have been of intelligent origin. None of them could be proved to be so, largely because they were not verifiable. Revisiting the signal's supposed point of origin failed to provide a repetition of the event, leaving the original signal as a tantalizing but scientifically useless phenomenon.

In order for a detected signal to be accepted as of intelligent, extraterrestrial origin, it is generally agreed that it must meet two criteria:

(1) *It must not be "natural"*. That is, no natural process could have created it. There have been false alarms, such as the initial apprehension that the regularity of pulsar signals signified intelligent design.
(2) *It must be verifiable*. To rule out man-made causes it must be present long enough so that several observers in widely separated locations can verify its point of origin, and all must agree on the same extra-solar point!

Of course, it would be desirable for the signal to have, somehow, a modulation that would impart information to the observer. An on–off modulation in some

[1] This section was written by the SETI League President, Richard Factor.

obvious pattern such as sequential prime numbers would comfortably fulfill this desirable but not-strictly-necessary characteristic.

With over two decades of sometimes fitful, sometimes diligent searching, the results can be summed up in two sentences: We know that the sky isn't teeming with strong signals. And we have searched such a small percentage of the phase space that it would be foolish to conclude there's nothing to be found.

6.2.3 Searches

The physics of detecting interstellar signals is challenging but not daunting. Calculations show that a relatively modest pair of radio telescopes with easily achievable transmitter power could communicate between Earth and the nearest stars. Two radio telescopes the size of that at Arecibo, Puerto Rico, could, with 1-Megawatt transmitters, detect each other's presence a good fraction of the way across the Galaxy. The Drake equation is a construct that enables us to focus on and attempt to quantify the likelihood of other civilizations with which we might communicate. Although recent discoveries of planets circling nearby stars has reduced some of the uncertain terms in this heuristic, there remain sufficient imponderables to allow essentially any conclusion to be drawn. If one concludes that there are very large numbers of civilizations in the Galaxy, it is reasonable to infer that several of them are quite "close" to us, perhaps within tens or hundreds of parsecs. If one concludes that there are only a small number, then it is likely that they will be located at greater distances.

The location of the putative civilization defines the strategy for locating it. If it is nearby, our largest, most sensitive radio telescopes would probably be able to detect signals emanating from it, even if those signals are not specifically being "beamed" toward us. Smaller radio telescopes, and, in particular, very small ones, such as 3 m – 5 m backyard dishes, would not be able to detect "leakage" radiation from even the closest stars. On the other hand, if a very powerful signal were being beamed, either directly to us, or sent omnidirectionally into the Galaxy as a beacon, even the largest radio telescopes would likely fail to find it. Although they would be capable of detecting the signal, their beamwidth, which is inversely proportional to their size, would be so narrow that it would require either extremely large numbers of million-dollar instruments or extravagant luck to be pointing in the right direction to hear the signal.

6.2.4 Targeted search

The bifurcation of microwave SETI into two search strategies accommodates these realities. Very large radio telescopes, of which there are only a tiny number and whose observing time is precious, are used to probe the nearest stars. The SETI Institute's Project Phoenix is the main exemplar of this strategy. This targeted search has an excellent chance of detecting a radio-using civilization (such as ours) if it is on the planet of a star out to about 100 pc. Such stars are well catalogued and can be selected on the basis of similarity to the Sun. Extra emphasis can be given to stars that are known to have planets; waste can be obviated by foregoing binary stars or others presumed for various reasons to not support life.

One major advantage of the targeted search is that it doesn't presuppose deliberate attempts at communication. It systematically investigates nearby stars and, if one harbors a radio-using civilization, it will likely find it. The major disadvantage is its implicit assumption: civilizations are plentiful and hence nearby. Other, explicit assumptions which seem reasonable may simply be incorrect (e.g., non-Sol-type stars are less likely to have associated civilizations).

6.2.5 All-sky survey

Almost a precise complement to the targeted search is the all-sky survey. Where the first assumes plentiful civilizations, the other makes no such assumption. Where the first assumes no deliberate attempt at communication, the other requires it. Where the first cherry-picks "appropriate" stars, the other makes no distinctions. In terms of instrumentation, at least as far as mechanical hardware is concerned, they are as far apart as can be. At least theoretically, one could argue that the all-sky survey could be accomplished by nothing more than a dipole antenna, while the targeted search will benefit by using the most enormous radio telescope that can be built. As a practical matter, the size of the telescopes used in the all-sky survey must fall between limits imposed by sensitivity and interference rejection on one end and economics on the other.

Assume that one desires to cover the entire sky with as much sensitivity as possible. With appropriate location of the observatories, one could accomplish this with approximately 5,000 "small" dishes on the order of 3 m–5 m in diameter. This is the essence of the SETI League's Project Argus, a "grass roots" endeavor. There are literally millions of these dishes in the hands of TV watchers, at least in the United States, and, due to the advent of DBS satellites, many of them are available for the price of carrying them away. Assuming the economic cost of re-commissioning each dish is on the order of $1,000, the antennas for the all-sky survey come in at only $5 million. This is a pittance compared to even the cost of a single research-grade radio telescope. However, the economics of scaling is very unfavorable. For instance, to only double the distance at which a given signal can be detected, one would need to double the diameter of the antenna, making it in the 6 m–10 m range. Because these dishes are no longer littering the landscape, they must bear their actual economic cost, on the order of $10,000 each. Just as bad, doubling the diameter halves the beamwidth in two dimensions, raising the required number of dishes to 20,000. Thus, doubling the sensitivity increases the cost from $5 million to $200 million. Doubling the sensitivity yet again requires 80,000 12 m–20 m dishes, at perhaps $50,000 per copy.

The economics of increasing the sensitivity of an all-sky survey are formidable. Given that the search is sensitivity-limited, a reasonable but not conclusive assumption, an improvement in the strategy might be to concentrate a smaller number of larger dishes in the direction of the Galactic plane. The Galaxy is only a few hundred parsecs thick in this neighborhood and really strong signals are statistically more likely to come from a direction where there are more stars.

As with the targeted search, there is a major implicit assumption in the all-sky survey: Somewhere out there we (or the entire Galaxy) are being sent a "beacon" signal. Unlike the leakage we as a civilization have been generating for almost a century, and which can be detected by a targeted search, to detect us at all-sky-survey distances, we would have to deliberately send a high-power signal to somebody who was looking for it. For a civilization at our level of development this is not economically and possibly not technically feasible; for one somewhat or substantially advanced, it may be possible or even routine. If the assumption that there is the equivalent of a "beacon" being sent is wrong, then the search will fail.

6.2.6 Common requirements

The two strategies were discussed without regard to the electronic instrumentation necessary. As divergent as the antenna requirements are, the receiver and signal detector requirements are very similar. For a research-grade radio telescope, the cost of the mechanical system is so high that any reasonable electronic detection ensemble has a cost, you should forgive the expression, in the noise level. This is emphatically not the case in the all-sky-survey scenario, in which the electronic requirements of the receiver and data reduction hardware can equal or exceed the cost of the antenna, and yet come nowhere near the capability of the larger instrument's electronics. Fortunately, there is great cause for optimism! While the cost of constructing mechanical hardware increases slowly with time, the cost of constructing electronic hardware plummets with Moore's law. At the moment professional electronic hardware exceeds amateur capability by perhaps a few dB in sensitivity (disregarding antenna size), two orders of magnitude in stability, and three to four orders of magnitude in frequency coverage. Advances in DSP in particular, as well as improvements in semiconductors and other technology, are likely to make today's professional capabilities within the reach of amateurs in only a few years.

Divergent requirements, both mechanically and culturally, do not obviate the desirability of conducting both types of searches. The fact is, nobody knows the prevalence or location of radio-using civilizations. Many or few, advanced or at our level of development, near or far, we simply have no idea. Proof of their existence is interesting and important and the cost of searching is insignificant.

6.3 GL-SETI: NAMELY, SETI SEARCHES FROM THE EARTH BY EXPLOITING THE GRAVITATIONAL LENSES OF OTHER STARS[2]

6.3.1 A third strategy

The purpose of this extensive background discussion was to examine the implicit assumptions and requirements of both kinds of searches. Each has a distinctive

[2] Like the previous section, this section was written by the SETI League President, Richard Factor.

vulnerability. If there are no nearby civilizations, the targeted search will fail. No matter how many civilizations there are, if nobody is transmitting a beacon the all-sky survey is unlikely to detect any of them. What if we happen to be in a deserted neighborhood? Too bad.

In the discussion above I stated that one of the requirements for a signal to be scientifically accepted as being of intelligent origin is that it be verifiable. This is not entirely true. Another way to prove extraterrestrial origin for a signal is for the content (i.e., modulation), to be both explicit and alien. Certainly a single frequency beacon wouldn't fulfill this criterion, nor would simple pulsed signals, the alien equivalent of telemetry signals, or anything else that could arguably have been produced on Earth. What would be acceptable? A television signal depicting aliens or a signal whose decoded modulation revealed scientific knowledge beyond current competence would, although the first surely would be suspected of being a hoax. Signals with information between these extremes, upon detailed scrutiny, might be accepted, at least provisionally. Why, however, consider these possibilities when it is commonly accepted that at best a single frequency beacon might be discovered?

The phenomenon of gravitational lensing, a consequence of general relativity, is scientifically accepted and has proved a valid astronomical and astrometrical tool. A gravitational lens occurs when electromagnetic radiation passes a massive astronomical object such as a star or even a galaxy. Because of the large area of signal "collected" by the lens and the potentially small area of its focus, enormous signal gain is possible. Claudio Maccone has written a treatise on the subject, stating that our own star would have a gravitational focus at about 550 AU, allowing a spacecraft at this distance to take advantage of this lens to provide signal gain greater by far than that of the Arecibo dish. One of the purposes of this spacecraft would be to look for signals of intelligent origin. Sadly, most of us do not have our own space program and therefore cannot rely on the Sun to supplement our antenna. Is all lost? No!

For the Sun, the closest point of focus is 550 AU. However, the focus of a gravitational lens is not a point, it is a line. This line is directed radially from the focusing mass, and signals at different radial distances from the mass focus at different points along the line. Any distance greater than 550 AU would therefore focus signals coming from a sufficiently great distance on the opposite side of the Sun. At this focal point one could take advantage of the gain of the spacecraft antenna in addition to the gain of the gravitational lens, giving a great enough signal strength to detect even "leakage" signals from stars much farther away that those targeted in searches with our biggest telescopes.

Since the focus is a line, it follows that this effect can be employed at any distance beyond 550 AU. While we have no immediate prospect of going 550 AU from the Sun, we are already more than 550 AU from every one of the billions of stars in our Galaxy! Therefore, at any given time we could be in the line focus of some other star's gravitational lens, and could be receiving some other civilization's signals with relatively modest equipment. Perhaps we have already done so. One reasonable (but entirely conjectural) explanation of the SETI "hits" that we've received over the decades is that it was a transient gravitational lensing phenomenon.

Conceptually, then, we can see that it should be possible to take advantage of the gravitational lens to receive, without an enormous antenna, signals from a great distance. Unfortunately, doing so requires the fortuitous alignment of the transmitting source, a star (or other large mass), and an antenna, not to mention a receiving apparatus prepared to detect the signal. If we accept the notion that the lens is powerful enough to allow us to detect leakage radiation rather than a directed beacon, we're entitled to assume that any civilization such as ours would be detectable. Therefore, the number of detectable sources depends on the "solution" to the Drake equation, compounded with two additional variables:

(1) What are the odds that, at any given instance, a star and potential transmitting source are so aligned that reception would be possible?; and
(2) Is there an antenna/receiver combination available at the focus capable of capturing a signal if one were present?

In the spirit of the Drake equation, I shall designate these variables as f_a for the fractional probability of an appropriate alignment, and f_r the probability that a signal, if present, will be detected. As with other terms of the Drake equation, f_a is determined by the Universe. There will be just so many foci crossing one's antenna per time period. Like some, but not all, terms, this is susceptible to reasonable calculation, and values are available in the literature. Unlike f_a, f_r is under our control. If a SETI antenna capable of capturing a high-power, single-frequency beacon is also capable of capturing leakage signals with the aid of a fortuitous gravitational lens, then the all-sky-survey model is also appropriate for this type of search, and the economic cost of the antennas necessary to bring f_r arbitrarily close to one is entirely reasonable. However, the electronic signal detection package useful for beacon detection is unsuitable for detecting and verifying gravitationally amplified signals.

Because of the relative motion of the notional transmitting source, the intervening lensing body, and the orbital and rotational motion of the earth, the focus of the signal is constantly shifting. Orbital and proper motions of bodies in this Galaxy are on the order of tens to thousands of km/s. With some lensing events these motions will fortuitously subtract and provide a relatively stationary focus, but probabilistically the large majority will add, giving a receiver a relatively short time in the focus. A reasonable estimate, derived from estimates of stellar brightening, gives periods of minutes to a few hours. A much longer period would be of little benefit since most antennas operate in drift-scan mode, and only look at a given area of the sky for 5 min to 15 min.

Unfortunately for the initial verifiability model, it may be practically impossible to use multiple radio telescopes to verify the presence of an intelligent signal. Not only will the signal be temporally transient and destined to never repeat, but the focus of the gravitational lens may encompass spatially only one of the antennas. Thus, one must look to the second verifiability model, one in which the signal(s) modulation characteristics are in themselves indicative or conclusive of alien origin. To accomplish this, as an absolute minimum, a recording of the signal is required.

The electronic package of a typical SETI system comprises, after the analog receiver components, a digitizing and analysis subsystem. An amateur system can be little more than a personal computer with a sound card. Such hardware can look for narrowband signals over a bandwidth of perhaps 40 kHz. A professional system uses a number of dedicated processors to give several orders of magnitude more frequency range, on the order of tens or hundreds of MHz. In either case, however, the analysis system must make a decision: Is there a narrowband signal present in the passband? If so, the immediate goal is to determine from whence the signal emanates. If it is coming from a point stellar source, it should show doppler shift characteristic of the Earth's rotation, and should vanish if the radio telescope is pointed momentarily in some other direction. If these conditions are fulfilled, then another telescope at another location is advised of the signal and asked to verify its presence. Missing from all this excitement is any analysis of the signal itself! A narrowband signal is characterized by a single number: its frequency. This, plus or minus a few hundred Hz due to doppler shift, is all you need to know. There's no point in recording the signal itself.

To see what to expect from a gravitational-lens-enhanced signal event, consider what would happen if one were to aim an antenna at the Earth from space. As the Earth swam into the focus of the dish, a panoply of signals would reveal themselves. Among the strongest would be television transmitters and pulsed radars. Weaker signals used for point-to-point communications and radionavigation, for example, would be evident if the receiver had enough sensitivity. These signals would be all along the frequency axis. Depending upon time of day or night, frequencies below approximately 5 MHz to 50 MHz would be filtered from the ensemble by ionospheric reflection and absorption. Anything from 50 MHz to many GHz would be fair game. For "internal" use by our civilization, there are no "magic" frequencies. In fact, the "waterhole" is the *least* likely to have strong signals, since it is reserved for receiving weak signals! Whether or not another planet has a radioreflective ionosphere such as ours does isn't all that important, since for other reasons we will want to limit our search to a somewhat higher range of frequencies. Ideally, it would be desirable to search in the range of approximately 1 GHz to 10 GHz, or even lower and/or higher if antenna size and/or precision permits.

Would we detect the Earth with a receiver designed specifically for extremely narrow frequency bin detection? Maybe. Although there is little point in transmitting a totally modulation-free, extremely narrowband signal (except, perhaps, as a frequency standard or interstellar beacon), there is often enough energy transmitted at a "carrier" frequency used as a demodulation reference. It has been said that "a sufficiently advanced form of modulation is indistinguishable from noise" and we have been approaching that "ideal" almost since the beginning of electromagnetic communication. For example, television transmission in the United States will, over the next decade, shift from a format with a strong carrier component to a "digital" format in which there will be no carrier at all.

Another civilization's hope of detecting the next century's "I Love Lucy" will be greatly reduced. For the purpose of SETI it would be better to have a detector that could detect any artificial characteristic of a signal ensemble. Among the hallmarks of artificiality would be, in addition to frequency coherence, a broadened or otherwise interesting autocorrelation function, a non-Gaussian probability density function, a suddenly differing smoothed frequency spectrum, and an amplitude modulated, at whatever rate, intensity.

Another interesting detection method involves the Karhunen–Loève Transform (KLT), which promises to detect the presence of any non-random signal. The computational burden of these methods varies from minor (non-Gaussian PDF) to fearsome (KLT). While it would be desirable to employ all these methods, and it will be possible to do so with modest equipment in the near future, there is no reason not to use the simpler methods available right now.

Given an antenna and some method of detecting when a signal is present (using whatever methods we choose), we aren't quite there yet. If the detector alerts us to a possibly artificial signal (or group of signals) in the antenna beam, what good does it do us? With the gravitational lens scenario, we cannot count on a cooperative observatory to verify the location or existence of the signal(s) since their footprint may not include that observatory. Therefore, we must hope that the alternative criterion for acceptance, intelligibility of modulation, obtains. Moreover, we must record as much of the baseband signal as we possibly can since we will, in all likelihood, never have the opportunity again.

This may not be as formidable an obstacle as it seems. For a traditional SETI search, little signal recording is necessary. Of primary interest is the existence of narrowband signals whose characteristics can be defined in a few bytes. To record the entire baseband in the hope of capturing the modulation of an intelligently generated signal would require an impressive recorder. Assuming a 10 GHz bandwidth and an 8-bit dynamic range, the data generated would fill a standard VHS videotape roughly once per second. A more dramatic way of looking at this is that if you put the Statue of Liberty in the middle of a football field and covered the whole field with the data tapes, one year's worth of data would obscure the field, statue, and all, up to the torch.

I have no desire to bury the Statue of Liberty in worthless data, which is what most of it would be. A better way to handle this is to be more judicious in our data recording habits. First, we would only want to run the recorder when there is a candidate signal present. Based on the gravitational lens statistics, or, alternatively, the number of "hits" received in SETI searches in the past, this would be comfortably under 1% of the time. Of course, the time to initiate and terminate recording would be determined by a signal detector broadly described above. Next, recording the entire baseband, beyond the state of the art for a single recorder at present, isn't really necessary. Although it is conceivable that there would be a torrent of signals at all frequencies, it is more likely that they will appear in a more limited area. On Earth we allocate frequency bands for different purposes. Some have a few strong signals (broadcasting), some have many weak signals (portable telephony). Even with the enormous gain of a gravitational lens it is unlikely that we can receive signals unless they have many kilowatts behind them. By setting up a number of recorders capable of an instantaneous bandwidth of, say 50 MHz, and a suitable number of signal presence detectors, we should be able to deal with whatever comes our way.

Finally, we would need to decide on the recorder "dynamic range" which in turn is determined to a large extent by the number of signals expected to be received and the expected signal-to-noise ratio. This is normally specified in decibels (dB) wherein each bit of the sampled signal increases the range by a factor of 2, or roughly 6 dB per bit. As an example, a broadcast-quality television signal requires roughly 10 bits of dynamic range, and a bandwidth of roughly 5 MHz. It is probably unrealistic to expect a "broadcast quality" of anything at interstellar distances, but with a signal of any complexity and only one chance to capture it, it is better to err on the side of greater precision. A digitizer of at least 4 bits and preferably as many as 8 bits should handle a wide variety of signals.

Given the above analysis, the absolute amount of data to be recorded reduces to a more manageable average rate of hundreds of kilobytes per second and a burst rate of, say, 25 megabytes per second. Even this rate would fill many tapes, but because of the "bursty" nature of the data, it should be possible to subject each burst to more comprehensive analysis during intervals when no candidate signals are being received. The data can be initially recorded in random access memory and only committed to tape or other storage medium when there is a reasonable probability of a signal being present. This is a more desirable method because "data acquisition" to memory is simpler and faster than recording directly to a magnetic or optical medium, and the RAM medium can be immediately and indefinitely reused if the candidate signal is found to be spurious.

Consider one possible configuration for a small SETI "observatory" electronics package. A specially modified video obscenity delay line is used as a burst storage recorder. Electronically, it is arranged as an "endless loop" recorder, so that the last 20 seconds of data received are always in memory. A "signal detector", still to be optimized, works with a PC to determine the likelihood that there is a non-random signal in the 5 MHz–wide passband of the downconverted radiofrequency input. When such a determination is made, the computer, after a 10 s delay, tells the video recorder to stop recording, leaving 10 s of pre- and 10 s of post-"detection" signal in its memory. This memory, approximately 300 Mbytes worth, is then transferred to a computer for storage and subsequent detailed analysis.

It should be noted that the gravitational lens scenario and the narrowband beacon scenario are by no means mutually exclusive, and the ability to perform both types of detection enhances the capability of both small- and large-antenna SETI observatories.

6.3.2 Summary

The advantages of looking for gravitationally-lensed intelligent signals include increasing the chance for detection at relatively small additional cost and at least the possibility of obviating the "we had a hit but couldn't confirm it" problem. It is a strategy that differs from the "targeted search" in that it has a chance of picking up "leakage signals" from solar systems that are otherwise completely out of range. It is a strategy that differs from the all-sky search in that it doesn't require a signal beamed to us directly by a civilization that knows where we are, or, transmitted omnidirectionally by a civilization that has incredible power at its disposal. It is a strategy that, given its modest antenna requirements, can be adopted by amateur and small observatories. And it is one that will benefit as the state of the art in signal processing improves inevitably, rather than one that requires ever bigger radio telescopes.

6.4 MACCONE'S EQUATION RELATING TO (1) MAGNIFICATION OF A LENSING STAR, (2) DISTANCE OF THE ET TRANSMITTER, AND (3) POWER OF THE ET TRANSMITTER

The G-L SETI ideas outlined by the SETI League President, Richard Factor, in the preceding Sections 6.2 and 6.3, were entirely qualitative. In other words, Richard Factor did not give any mathematical theory enabling the scientists to check whether G-L SETI is an actually achievable goal by virtue of such numbers as the magnification of the lensing star located in between the Earth and the ET transmitter, the distance of the ET transmitter from the Earth (i.e., the "range" of our SETI searches), and the (unknown to us) power of the ET transmitter.

In this section we make G-L SETI a quantitative science by introducing the use of mathematics and by discovering the equation relating to

(1) The magnification G_{Lens} of the lensing star. This lensing star must be located along the straight line between ET and the Earth, and could be any star nearby the Sun in the Galaxy (say, any star up to 2,000 lt-yr away from the Sun in all directions. In the year 2002 we already know the distances (i.e., the parallexes) of *all* 218,000 stars closest to the Sun in the Galaxy and located within a sphere of about 2,000 lt-yr from the Sun. All these 218,000 star distances are listed in the European Space Agency's *Hipparcos Catalogue*, published in 1998. The distance between the Earth and the lensing star, dubbed $D_{Lens-Earth}$ hereafter, is thus supposed to be a known parameter by virtue of the *Hipparcos Catalogue*. As for the magnification G_{Lens} of the lensing star itself, we also regard it hereafter as a known parameter, although a difficult one to estimate. Actually, the magnification G_{Lens} depends on (a) the mass of the lensing star, and (b) the atmosphere of the lensing star, in particular the star's electron density near to the star surface, which is where the ET signals have to pass through. We assume these data to be known, though this assumption could be largely optimistic.
(2) The distance $D_{ET-Earth}$ of the ET transmitter. This is assumed to be an unknown datum hereafter. More exactly, we assume the distance $D_{ET-Lens}$ between ET and the lensing star to be an unknown datum, and, of course, one has

$$D_{ET-Earth} = D_{ET-Lens} + D_{Lens-Earth} \tag{6.1}$$

(3) The power P_{ET} of the ET transmitter. This also is assumed to be an unknown datum, obviously.

In order to derive our new equation, which was published for the first time in 2002 (see [2]), let us start from the description of the radio link given by John D. Kraus in his seminal book *Radio Astronomy* [3]. Suppose ET radiates isotropically and uniformly over the bandwidth B_{ET} the unknown power P_{ET} from ET's star, located at the unknown distance $D_{ET-Lens}$ from the lensing star. The power $P_{ET\,@\,Lens}$ received at the lensing star is given by

$$P_{ET\,@\,Lens} = \frac{B_{ET\,@\,Lens} P_{ET} A_{eff\,@\,Lens}}{B_{ET} 4\pi D^2_{ET-Lens}} \tag{6.2}$$

from which, assuming $B_{ET\,@\,Lens} = B_{ET}$, one gets

$$P_{ET\,@\,Lens} = \frac{P_{ET} A_{eff\,@\,Lens}}{4\pi D^2_{ET-Lens}}. \tag{6.3}$$

Then, $P_{ET\,@\,Lens}$ is magnified by the factor G_{Lens} and finally reaches the Earth (located at the known distance $D_{Lens-Earth}$ from the lensing star) decreased to the value $P_{ET\,@\,Earth}$ by the inverse-square law:

$$P_{ET\,@\,Earth} = \frac{G_{Lens} P_{ET\,@\,Lens} A_{eff\,@\,Earth}}{4\pi D^2_{Lens-Earth}}. \tag{6.4}$$

Substituting $P_{ET\,@\,Lens}$ from (6.3) into (6.4), we get $P_{ET\,@\,Earth}$ as a function of only two unknowns: P_{ET} and $D_{ET-Lens}$.

$$P_{ET\,@\,Earth} = \frac{G_{Lens} P_{ET} A_{eff\,@\,Lens} A_{eff\,@\,Earth}}{16\pi^2 D^2_{ET-Lens} D^2_{Lens-Earth}}. \tag{6.5}$$

Let us now notice that we can measure experimentally the power $P_{ET\,@\,Earth}$ received at our radiotelescope on Earth and, thus, $P_{ET\,@\,Earth}$ is to be regarded as a known datum. Therefore, we can solve Equation (6.5) for the unknown distance $D_{ET-Lens}$, finding this distance as a function of the unknown power emitted by ET, P_{ET}:

$$D_{ET-Lens}(P_{ET}) = \sqrt{\frac{G_{Lens} A_{eff\,@\,Lens} A_{eff\,@\,Earth}}{P_{ET\,@\,Earth} 16\pi^2 D^2_{Lens-Earth}}} \cdot \sqrt{P_{ET}} \tag{6.6}$$

Replacing (6.6) into (6.1) we finally get the full ET-to-Earth range expressed as a function of the only unknown power P_{ET} emitted by ET:

$$D_{ET-Earth}(P_{ET}) = D_{Lens-Earth} + \sqrt{\frac{G_{Lens} A_{eff\,@\,Lens} A_{eff\,@\,Earth}}{P_{ET\,@\,Earth} 16\pi^2 D^2_{Lens-Earth}}} \cdot \sqrt{P_{ET}}. \tag{6.7}$$

This is the basic relationship we wanted to point out between the ET distance from Earth, $D_{ET-Earth}$, and the power, P_{ET}, of the isotropic signal emitted by ET. We believe that this formula will be of help, jointly with the *Hipparcos Catalogue*, to put the future GL-SETI searches on sound rational and scientific foundations.

6.5 SUN GRAVITY LENS AND SETV: THE SEARCH FOR EXTRATERRESTRIAL VISITATION

SETV is an acronym for "Search for ExtraTerrestrial Visitation". This means finding scientific proofs that aliens visited in the past either the Earth itself or, more in general, some parts of the solar system. Nowadays we have a SETV Institute whose mission and more can be found at the website.

This book is not concerned with SETV since in our opinion this would lead easily to questions like the reality of UFOs, etc., which are regarded as non-scientifically proven subjects by this author. There is, however, one fully scientific argument related to SETV and to the gravitational lens of the Sun that we want to mention here since it could, one day, become the object of radio astronomical searches, especially in the infrared bands, and could also easily lead to the discovery of those aliens who are spying us!. Here is the argument.

(1) Suppose that an alien civilization discovered us enough time ago.
(2) Suppose they have interstellar space-faring capabilities, at least for robotic probes, but possibly also for inhabited probes.
(3) Suppose they use electromagnetic waves to transmit their reports about us back to their mother star. This is, in this author's view, the most severe assumption: in fact, if they are so much more technologically advanced than us, in order to communicate they could use neutrinos, or even other technologies still unknown to us.

Well, they may have then sent probes from their star to about 550 AU from our Sun in order to use the full power of the gravitational lens of our Sun. This means that

(1) They position their own probe(s) in between, say, 550 AU and 1,000 AU on the opposite side of their own star with respect to the Sun.
(2) They have a similar probe positioned at the focus of the gravitational lens of their own star.
(3) They take enormous advantage of this two-star radio bridge to save power for their telecommunications.

Suppose all that is true.

But, then, we may wish to detect these alien probes in the spherical space around the Sun having a radius of at least 550 AU and extending out to about 1,000 AU in each direction! A huge volume of space to be explored, but a "must" to be sure that no alien is spying on us by resorting to the Sun's gravity lens. We conclude that

1 This exploration of the space within two spheres of radii 550 AU and 1,000 AU could possibly be made from Earth by antennas searching for infrared radiation (the heat emitted by the alien probes). This search could appropriately be called a SETV search. A scientist supporting this view in the most recent years is

California-based Scott Stride, to whom this author is grateful for enlightening conversations on this subject.

2 Suppose we detect such an alien probe using the gravitational lens of the Sun. Then, from its position around the Sun, we would immediately know the direction (opposite to the probe and passing through the Sun's center) along which the aliens' star is located. This is a very, very important piece of information indeed, even if the distance of the alien star cannot be determined by this method.

3 A stellar catalog like *Hipparcos* (by ESA), listing all the 218,000 stars closest to the Sun in the Galaxy, would help us immensely to find the distance of the alien star, and so its full 3D position! And if *Hipparcos* is not enough, in two decades we should have the SIM (Space Interferometry Mission, by NASA-JPL) stellar catalog yielding the 3D positions of about a million stars around the Sun in the Galaxy. And if even SIM is not enough, than GAIA (by ESA) should provide us with the 3D position of 50 million to 1 billion stars in the Galaxy within some decades, and this really means most stars in the Galaxy itself, even on the other side of the bulge and even in the Magellanic Clouds.

The inevitable conclusion is that, if any alien probe is spying on us from 550 AU right now, and if we discover it, we would almost certainly detect the first extraterrestrial civilization in the Galaxy.

6.6 REFERENCES

[1] R. Factor (WA2IKL), "A Third, Complementary, Microwave Search Strategy for SETI", a paper presented at *SETICon '01*, the *First SETI League Technical Symposium*, held at the College of New Jersey, Ewing, NJ, April 28–29, 2001, in particular see pp. 13–19.

[2] C. Maccone, "SETI through the Gravitational Lenses of Alpha Centauri A, B, C and of Barnard's Star", a paper presented at *SETICon '02*, the *Second SETI League Technical Symposium*, held at the College of New Jersey, Ewing, NJ, April 26–28, 2002, in particular see pp. 93–97.

[3] John D. Kraus, *Radio Astronomy*, Second Edition, Cygnus–Quasar Books, 1986, in particular see pp. 12-2 through 12-4.

7

The gravitational lenses of Alpha Centauri A, B, C and of Barnard's Star

7.1 INTRODUCTION

The gravitational lenses of the three nearest stars, Alpha Centauri A, B, and C (Proxima Centauri), are studied in this chapter. For each star, the minimal focal distance is found, and turns out to equal 679.262 AU, 563.484 AU and 112.138 AU, respectively, plus or minus the (large) uncertainties deriving from the uncertainties in the estimates of the star masses and radii. A comparison of these three minimal focal distances against the corresponding value for the Sun (550 AU, or, more correctly, 548.214 AU) is then made, but it is clearly pointed out that all these minimal focal distances are just the theoretical values given by Einstein's deflection formula for a corresponding "naked star" (i.e., a star as if it had *no corona*!). The study of the true focal distances that follow from taking the corona into account is much more difficult and uncertain, and has to be delayed for further research. For the naked stars, we study the deflection of radio waves for four different frequencies: the water maser at 22 GHz, NASA's Interstellar Probe (ISP) telecommunication frequency at 32 GHz (Ka band), the Cosmic Microwave Background peak frequency at 160.378 GHz and finally the positronium frequency at 203 GHz. For each frequency the antenna patterns of the three naked stars' gravitational lenses are given. Finally, all the above data are derived also for the fourth star in increasing distance from the Sun—Barnard's Star.

The gravitational lens of the Sun still needs more study. In fact, above the surface of the Sun the Corona extends into space across distances that are comparable with the Sun radius, and the coronal effects may only complicate the physical picture of the Sun as a gravitational lens.

7.2 THE SUN'S GRAVITY + PLASMA LENS AS A MODEL FOR THE NEARBY STARS

When will the first *interstellar* robotic probe be launched by humankind? In a millennium, in some centuries, or just a few decades ?

No one knows the answer in the year 2008, but NASA's most advanced vision of the problem seems to suggest that the first robotic probes to the nearest stars will be launched before the year 2100 (see NASA's Interstellar Probe site: *http://coast.jpl.nasa.gov/interstellar/probe/introduction/intro.html*).

It seems reasonable to assume that the target of the first of such interstellar robotic flights will be the nearby Alpha Centauri system of three stars, simply because this is the nearest group of stars to the Sun. This system is located about 4.3 lt-yr away in the southern hemisphere: the direction is $-62^{\circ}40.8$ in declination and 14h26m43s in right ascension for Proxima (i.e., Alpha Centauri C) in the year 2000.

From the point of view of spaceflight, planning a mission to the Alpha Centauri system means that

(1) Innovative propulsion systems must be devised by scientists and engineers to cover the distance of 4.3 lt-yr in a "reasonable" amount of time—not much larger than, say, a century. NASA's Breakthrough Propulsion Physics Program (BPPP), started in 1996 (*http://www.grc.nasa.gov/WWW/bpp/*) tries to address this propulsion issue by resorting to the most recent and advanced developments in physics and engineering.
(2) Apart from propulsion, however, other key problems will have to be solved. One is the telecommunication link to be established between the Alpha Centauri System and the Sun–Earth System after one or more such probes have reached Alpha Centauri. Of course, it will always take 4.3 years for our radio messages to reach Alpha Centauri and 4.3 more years for their replies to reach us: this seems unavoidable since nothing faster than lightspeed is currently known. In this chapter, however, we wish to show that, by exploiting the natural phenomenon called gravitational lensing of both the Sun and the three stars in the Alpha Centauri System, the powers involved in keeping a permanent two-way radio link between ourselves and Alpha Centauri *can be greatly reduced*.

Let us start by reviewing the essential features of the gravitational lens associated with every star. As a more comprehensive introduction to this research field, the reader might wish to consult [1, 2] or the author's recent paper [3], from which this chapter was adapted.

With reference to Figure 7.1, in this chapter we'll only state without proof the following results:

(1) The geometry of the gravitational lens of any star (Figure 7.1) is easy: incoming electromagnetic waves, arriving from infinity, pass *outside* the star and at a certain distance r from its center, traditionally called the "impact parameter" (as in particle physics). Then, the well-known *Schwarzschild solution* (of the

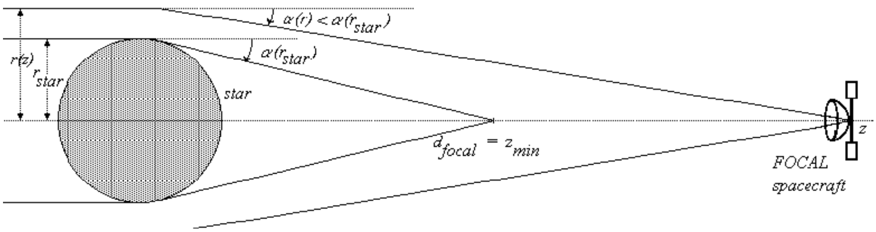


Figure 7.1. Geometry of the gravitational lens of any star, its minimal focal length, and FOCAL spacecraft.

Einstein equations of general relativity) yields the well-known Einstein's deflection formula yielding the deflection angle $\alpha(r)$ that all radiation undergoes

$$\alpha(r) = \frac{4GM_{star}}{c^2 r}. \tag{7.1}$$

The *largest* deflection angle α thus occurs for those rays just grazing the star surface (i.e., for $r = r_{star}$) and is

$$\alpha_{\max} \equiv \alpha(r_{star}) = \frac{4GM_{star}}{c^2 r_{star}}. \tag{7.2}$$

From the inequality $\alpha(r_{star}) > \alpha(r)$, from the geometry shown in Figure 7.1, and from the definition of the *Schwarzschild radius or gravitational radius of the star*, that is

$$r_{Schwarzschild-star} \equiv r_g = \frac{2GM_{star}}{c^2}, \tag{7.3}$$

a few easy steps (listed in [3, pp. 7–8]) yield then the minimal focal distance for the star's gravitational lens

$$d_{focal-star} \approx \frac{r_{star}^2}{2r_g} = \frac{c^2}{4G} \cdot \frac{r_{star}^2}{M_{star}}. \tag{7.4}$$

This is the minimal focal distance from the star's center that the FOCAL spacecraft must reach in order to get magnified radio pictures of whatever lies on the other side of the star with respect to the spacecraft position. Also, all points on the straight line beyond this minimal focal distance are foci too, because the light rays passing by the star farther than the minimum distance have smaller deflection angles and thus come together at an even greater distance from the star.

The very important astronautical consequence of this fact for the FOCAL mission is that *it is not necessary to stop the spacecraft at the minimal focal distance. It can go on to any distance beyond and focus as well or better*. In fact, the farther it goes beyond the minimal focal distance, the less distorted the collected radio waves are by the star's corona fluctuations.

(2) Let us next find the equation relating the *uncertainty in the determination of the minimal focal distance,* $\Delta d_{focal-star}$*, to the uncertainties in the determination of both the star's radius,* Δr_{star}*, and mass,* ΔM_{star}. The theory of errors shows that one simply has to compute the total differential of (7.4) and replace the differentials by the corresponding finite differences, taking also the absolute values of all terms in the summation resulting from the differentiation of the factors in (7.4). So one has:

$$|\Delta d_{focal-star}| = \frac{c^2}{4G} \cdot \left(\left| \frac{2r_{star}}{M_{star}} \cdot \Delta r_{star} \right| + \left| -\frac{r_{star}^2}{M_{star}^2} \cdot \Delta M_{star} \right| \right)$$

$$= \frac{c^2}{4G} \cdot \frac{r_{star}^2}{M_{star}} \cdot \left(2 \left| \frac{\Delta r_{star}}{r_{star}} \right| + \left| -\frac{\Delta M_{star}}{M_{star}} \right| \right) \tag{7.5}$$

which, invoking (7.4), finally yields

$$\left| \frac{\Delta d_{focal-star}}{d_{focal-star}} \right| = 2 \left| \frac{\Delta r_{star}}{r_{star}} \right| + \left| \frac{\Delta M_{star}}{M_{star}} \right|. \tag{7.6}$$

In words: *the relative error in the determination of the star's minimal focal distance equals twice the relative error affecting the star radius plus the relative error affecting the star mass.*

(3) The *magnification or gain* G_{star} of the star gravity lens is of course huge. Along the *focal axis* (straight line between the source of electromagnetic waves, the star center, and the FOCAL spacecraft position), it is

$$G_{star}(\lambda) = 4\pi^2 \frac{r_g}{\lambda} = \frac{8\pi^2 G M_{star}}{c^2} \cdot \frac{1}{\lambda}. \tag{7.7}$$

More generally (see [4, p. 3, eq. (3)]), at distance ρ from the focal axis, and at spacecraft distance z from the star, *the gain is*

$$G_{star}(\lambda, \rho, z) = 4\pi^2 \frac{r_g}{\lambda} \cdot J_0^2 \left(\frac{2\pi\rho}{\lambda} \sqrt{\frac{2r_g}{z}} \right). \tag{7.8}$$

where $J_0(x)$ is the Bessel function of order zero and argument x. Since $J_0(0) = 1$, (7.8) reduces to (7.7) for $\rho \to 0$, but notice also that it does so for $z \to \infty$. This the *maximum* gain provided by the naked star (i.e., the star without taking its corona into account at all), and so it is a (close) upper bound for the true gain.

(4) The FOCAL spacecraft has an antenna with radius $r_{antenna}$ and efficiency $k_{antenna}$. Thus, the FOCAL antenna gain is given by

$$G_{antenna}(\lambda) = 4\pi \frac{A_{physical} \cdot k_{antenna}}{\lambda^2} = \frac{4\pi^2 r_{antenna}^2 \cdot k_{antenna}}{\lambda^2}. \tag{7.9}$$

(5) The *total gain* on axis for the combined (star + FOCAL antenna) system is the product of (7.7) times (7.9)

$$G_{total}(\lambda) = G_{star}(\lambda) \cdot G_{antenna}(\lambda) = \frac{32\pi^4 GM_{star} r^2_{antenna} \cdot k_{antenna}}{c^2 \lambda^3} \tag{7.10}$$

(6) The *angular resolution* of the star's gravity lens is found by first establishing the expression yielding the distance, in the image plane (plane orthogonal to the focal axis), where the gain decreases for 6 dB; that is

$$\rho_{6dB}(\lambda, z) = \frac{\lambda}{\pi^2}\sqrt{\frac{z}{2r_g}} = \frac{c}{2\pi^2\sqrt{GM_{Sun}}}\lambda\sqrt{z}. \tag{7.11}$$

The gain decrease by just 6 dB is an *arbitrary measure* of the gain decrease off axis. In [3] a 10 dB gain decrease is assumed, with the result that a factor of 4π, rather than of π^2, appears in (7.11). If our 6 dB form for (7.11) is assumed, one then gets the following expression (7.12) for the angular resolution $\theta_{resolution}(\lambda, z)$ of any astronomical radio source located on the opposite side of the star with respect to the FOCAL spacecraft position

$$\begin{aligned}\theta_{resolution}(\lambda, z) &= \frac{\rho_{6dB}(\lambda, z)}{z} = \frac{\lambda}{\pi^2} \cdot \frac{1}{\sqrt{2r_g}} \cdot \frac{1}{\sqrt{z}} \\ &= \frac{c}{2\pi^2\sqrt{GM_{Sun}}} \cdot \frac{\lambda}{\sqrt{z}}.\end{aligned} \tag{7.12}$$

(7) The *spatial resolution* of an astronomical object located at a distance $d_{star-object}$ from the star is

$$\begin{aligned}R_{object}(\lambda, z) &= d_{star-object} \cdot \theta_{resolution}(\lambda, z) = d_{star-object} \cdot \frac{\lambda}{\pi^2} \cdot \frac{1}{\sqrt{2r_g}} \cdot \frac{1}{\sqrt{z}} \\ &= d_{star-object} \cdot \frac{c}{2\pi^2\sqrt{GM_{Sun}}} \cdot \frac{\lambda}{\sqrt{z}}.\end{aligned} \tag{7.13}$$

(8) So far for the naked star (i.e., just as if the star had no 'flames' around it). But the true star is surrounded by a corona, a turbulent plasma whose electron density decreases outward as a function of r that is hard to establish both theoretically and experimentally. To take the star's corona into account is a difficult problem, and for the Sun the first breakthrough studies on the solution of the coupled Einstein–Maxwell equations for the spherical gravitational lens were made around 1974 by H. C. Ohanian [5] and E. Herlt and H. Stephani of the University of Jena in Germany [6, 7]. In 1998 a comprehensive effort to design a 550 AU probe to exploit the Sun's lens was made at JPL under the coordination of John L. West [8], and, as a part of this effort, the JPL experts B. G. Andersson and Slava G. Turyshev gave in [4, p. 6, eq. (13)] the following *empirical* formula for the deflection angle $\alpha_{plasma}(r)$ caused by the Sun's Coronal plasma effects and

opposed to the action of gravity

$$\theta_{plasma}(r,\nu) = \left(\frac{\nu_0}{\nu}\right)^2 \left[2.952 \cdot 10^3 \cdot \left(\frac{r_{Sun}}{r}\right)^{16} + 2.28 \cdot 10^2 \cdot \left(\frac{r_{Sun}}{r}\right)^6 + 1.1 \cdot \left(\frac{r_{Sun}}{r}\right)^2\right] \quad \ldots (7.14)$$

where $\nu_0 = 6.32$ MHz. The result of adding algebraically the gravitational deflection (7.1) and the plasma opposite deflection (7.14) is that the effective minimal focal distance of the star's gravity lens is pushed farther out beyond the value given by (7.4), at a distance given by

$$d_{focal(gravity+plasma)}(r,\nu) = d_{focal(gravity)}(r_{star}) \cdot \left(\frac{r}{r_{star}}\right)^2 \cdot \left[1 - \frac{\nu^2_{critical}(r_{star})}{\nu^2} \cdot \left(\frac{r_{star}}{r}\right)^{15}\right]^{-1}. \quad (7.15)$$

The numerical value $\nu_{critical}(r_{Sun}) \approx 122.2$ GHz, to be replaced into (7.12), is obtained by setting $r = r_{star}$ into the expression

$$v_{critical}(r) = \sqrt{\frac{\nu_0^2 r}{2r_g} \cdot \left[2.952 \cdot 10^3 \cdot \left(\frac{r_{Sun}}{r}\right)^{15} + 2.28 \cdot 10^2 \cdot \left(\frac{r_{Sun}}{r}\right)^5 + 1.1 \cdot \frac{r_{Sun}}{r}\right]}. \quad (7.16)$$

which is the expression for the critical angle for which the plasma deflection exactly counterbalances the gravity deflection. If the Sun's Corona theory applies to stars "similar to the Sun" like Alpha Centauri A and B too (an assumption to be checked), (7.15) is clearly the most important formula for these stars' true gravity lenses. For further details see Chapter 8.

7.3 ASSUMED DATA ABOUT ALPHA CENTAURI A, B, C AND BARNARD'S STAR

A large variety of slightly different numerical values for every important physical constant or astronomical parameter usually "plagues" the scientific literature. In order to clear the way from problems of this kind, we clearly state in this section which numerical values we *assume* to be true in this chaper. Of course, different assumed values for both the physical constants and the astronomical parameters would yield (slightly) different numerical results from ours. However, the differences should not be very large anyway.

We start from the *assumed* values of the Newtonian gravitational constant, G, of the speed of light, c, and of the Astronomical Unit, AU. These are given, respectively, in Table 7.1.

Next we list (Table 7.2) the assumed values of the Sun's mass and radius, taken from the JPL's DAstCom (= Data on Asteroids and Comets) database [9]. Unfortunately, we were not able to find the relevant uncertainties.

Table 7.1. Assumed values of G, c, and of the Astronomical Unit (AU).

$G = 6.6726 \cdot 10^{-11} \quad \frac{m^3}{kg \cdot s^2}$
$c = 299{,}792{,}458 \quad \frac{m}{s}$
1 AU = 149,597,870.66 km

Table 7.2. Assumed values of the Sun's mass, radius, and effective temperature.

Gravitational parameter $\mu = G \cdot M_{Sun}$	Mass in kg	Uncertainty in the mass	Radius	Uncertainty in the radius	Effective temperature in K
132712439900	$1.98891 \cdot 10^{30}$	$\pm$?	696,000 km	$\pm$?	5,780

The mass and radius of each of the four stars we want to study are among the most difficult astrophysical data to be determined in a reliable fashion. In fact, on the one hand, the masses are estimated on the basis of the star's spectral type, and, on the other hand, the radii require careful measurements that are often disputed by different astronomers. Further, the star masses and radii are usually listed in separate catalogs, causing additional difficulties for the theorists who need them. In this section, we want to point out once and for all the values of the masses and radii that we have *assumed* to be correct for the four stars, and we obviously are aware that these values may be questionable. Table 7.3 shows these assumed values, and the

Table 7.3. Assumed values of the mass, radius, the relevant uncertainties, and the effective temperature for Alpha Centauri A, B, C and Barnard's Star.

Name of the nearby star	Star mass M_{star} in units of the Sun mass	Uncertainty in the star mass, ΔM_{star}	Star radius r_{star} in units of the Sun radius	Uncertainty in the star radius, Δr_{star}	Star's effective temperature in K
Alpha Cent. A	1.1238	± 0.008	1.18	$\pm$?	5,770
Alpha Cent. B	0.9344	± 0.007	0.98	$\pm$?	5,300
Alpha Cent. C (Proxima)	0.11	$\pm$?	0.15	$\pm$?	2,407
Barnard's Star	0.17	$\pm$?	0.17	$\pm$?	$\approx$2,400

Table 7.4. Frequencies and wavelengths for gravity lenses of Alpha Centauri A, B, C and Barnard's Star.

Line	Neutral hydrogen	H_2O	Ka band	CMB_{max}	Positronium
Frequency ν	1.42 GHz	22 GHz	32 GHz	160.378 GHz	203 GHz
Wavelength λ	21.112 cm	1.363 cm	9.37 mm	1.06 mm	1.48 mm

estimated effective temperature (or color temperature—namely, the temperature of the blackbody curve that best fits the star's known astrophysical traits) is also given as further key data for use in future.

In view of further investigations, we also list the website from which the Sun's effective temperature was taken: *http://climate.gsfc.nasa.gov/~cahalan/Radiation/SolarIrrVblackbody.html*

Finally, in Table 7.4 we want to list the frequencies and wavelengths that we regard as interesting for the study of the gravitational lenses of Alpha Centauri A, B, C and of Barnard's Star. They all are well known in astrophysics and the space sciences and no further explanation seems to be needed.

7.4 GRAVITATIONAL LENS OF THE NAKED SUN

(1) *Assumed* mass of the Sun: 1.9889164628E+30 kg; that is, $\mu_{Sun} = 132{,}712{,}439{,}900\ \text{kg}^3/\text{s}^2$
(2) *Assumed* radius of the Sun: 696,000 km
(3) Sun mean density: 1,408.316 kg/m^3
(4) Schwarzschild radius of the Sun: 2.953 km
(5) Minimal focal distance of the Sun: 548.230 AU $\sim$ 3.17 light days $\sim$ 13.86 times the Sun-to-Pluto distance
(6) Patterns of the naked Sun's gravity lens at the hydrogen line frequency of 1.420 GHz (= 21 cm wavelength):

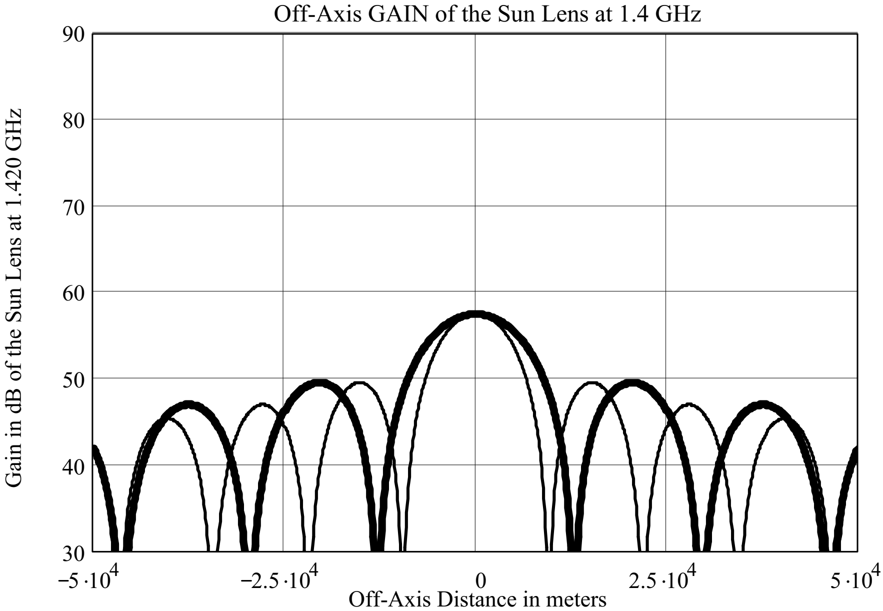


Figure 7.2. Antenna patterns for the naked Sun at the hydrogen line of 1.420 GHz. Horizontal scale is distance off axis in meters. Vertical scale is the naked Sun *gain* in dB as given by (7.8). The thin curve shows the antenna patterns of the gain for the FOCAL spacecraft located at 550 AU from the Sun, whereas the thick curve shows the antenna patterns when FOCAL reached 1,000 AU from the Sun. *Note that this convention (thin line for the FOCAL spacecraft located at the minimal focal distance and thick line for the spacecraft at 1,000 AU) will be retained in all the figures in this chapter*. Thus, it is evident that the central lobe *widens* while the FOCAL spacecraft is getting farther and farther away from the minimal focal distance up to 1,000 AU and beyond.

(7) Patterns of the naked Sun's gravity lens at the water maser frequency of 22 GHz (= 1.36 cm wavelength):

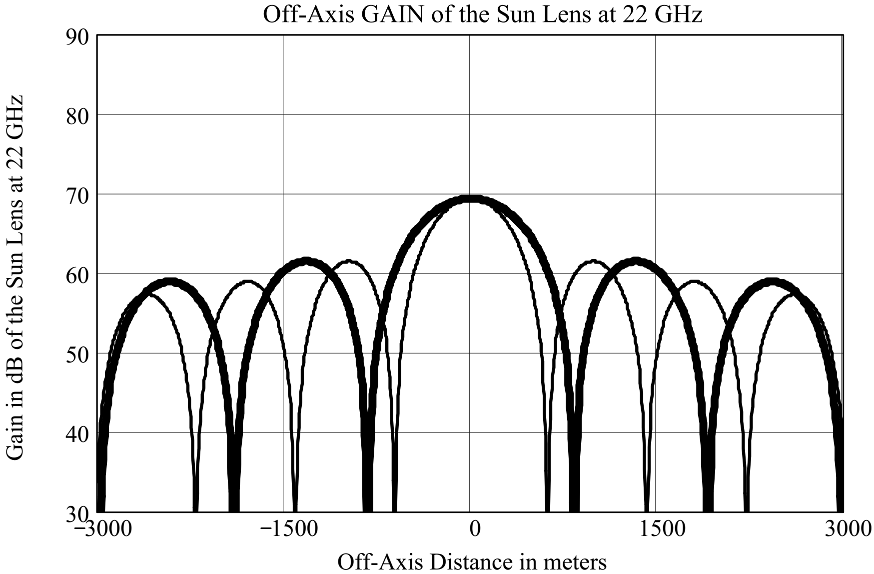


Figure 7.3. Same as for Figure 7.2, but for the water maser frequency of 22 GHz. Same vertical scale, for comparison.

(8) Patterns of the naked Sun's gravity lens at the Ka telecommunication band frequency of 32 GHz (= 9.37 mm wavelength):

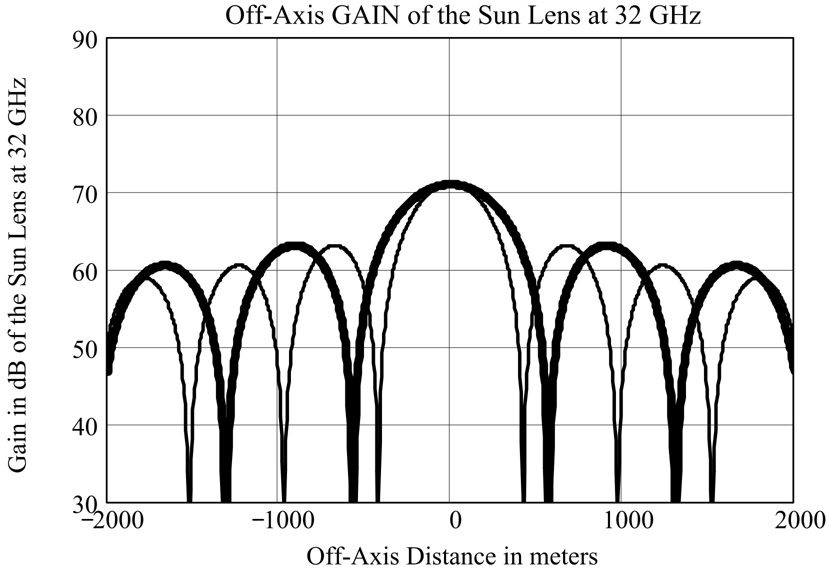


Figure 7.4. The important Ka telecommunications band frequency of 32 GHz could possibly become the link frequency for many future interstellar probes, like NASA's InterStellar Probe (ISP) initially scheduled for launch in 2010 in the direction of the incoming interstellar wind. Note that the central lobe is narrower than above (i.e., about 500 meters). This is a tight constraint on the alignment between source, Sun center, and FOCAL spacecraft position. It is of course very difficult to track a spacecraft with such precision at distances above 550 AU. So, new AOCS methods will have to be devised to solve this tracking problem.

(9) Patterns of the naked Sun's gravity lens at the peak frequency of the Cosmic Background Radiation (CMB), that is 160.378 GHz (= 1.06 mm wavelength, a value not related to the corresponding CMB peak frequency value by $c = \lambda\nu$ since both values originate from Planck's probability density (at the temperature $T = 2.728$ K), see Section 9.3 for details):

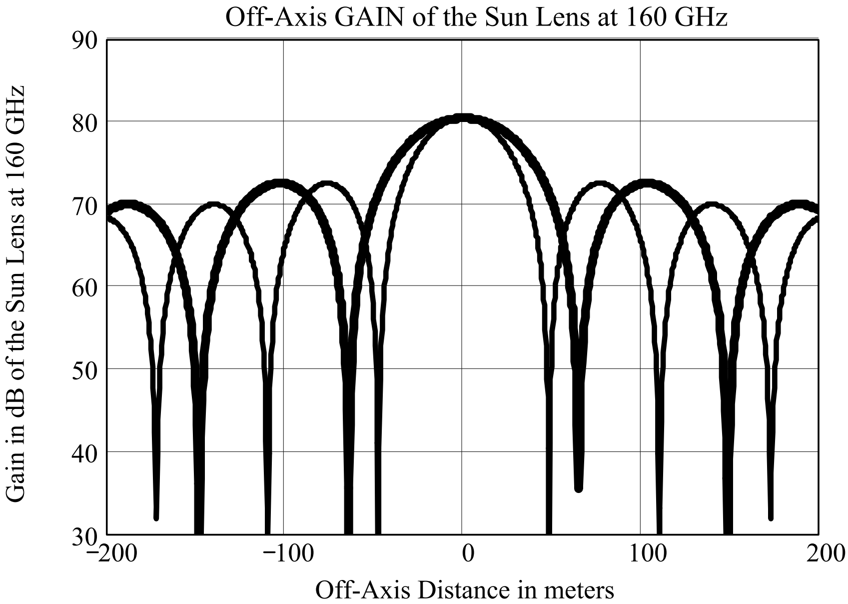


Figure 7.5. Naked Sun antenna patterns for the peak of the cosmic microwave background. Same vertical scale as above. Note that the central lobe is much narrower still (i.e., the alignment between source, center of the Sun, and FOCAL spacecraft position is now much tighter, ~50 meters).

(10) The positronium is an "atom" made up of one electron and one positron, whose spins are usually antiparallel. When either spin changes, however, a photon at the frequency of 203 GHz is emitted (or absorbed) and this is commonly called the positronium frequency. The great Russian SETI-radioastronomer Nikolai Kardashev suggested in the 1970s that this frequency could be used by other civilizations in the Galaxy for their own telecommunications, and so it is of interest in SETI. The naked Sun's antenna patterns on the positronium line are shown below:

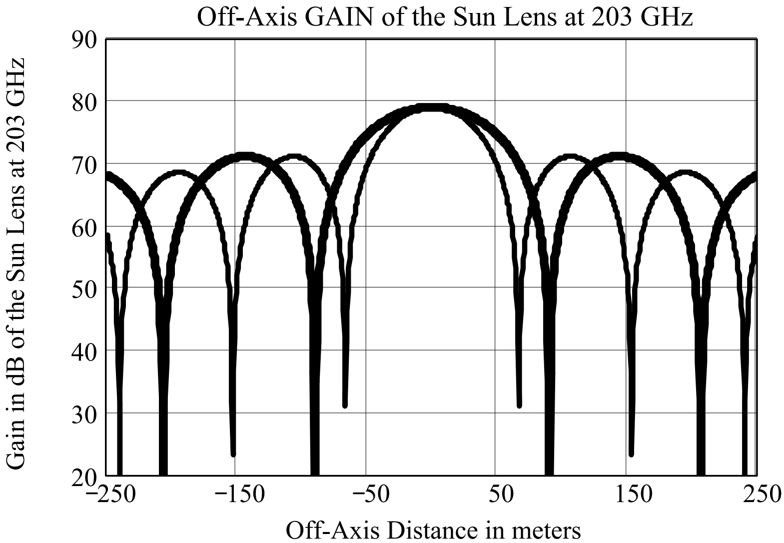


Figure 7.6. Naked Sun antenna patterns for the positronium line at 203 GHz (the SETI galactic link?)

7.5 GRAVITATIONAL LENS OF THE NAKED ALPHA CENTAURI A

The basic reference that we selected for our study of the gravitational lenses of Alpha Centauri A and B is the paper published on March 1, 2000, by D. B. Guenther and P. Demarque [11]. The values of the masses and relevant uncertainties for Alpha Centauri A and B given in Table 7.3 are taken from the *Hipparcos Catalogue*, as reported in [11, p. 506]. Alpha Centauri A and B have the *Hipparcos Catalogue* numbers 71683 and 71681, respectively, the *Gliese Catalog* numbers Gl 559 A and Gl 559 B (data downloadable from the website *http://www.ari.uni-heidelberg. de/aricns/cnspages/4c01151.htm*), respectively, and the numbers 128620 and 128621, respectively, in the Cadars.dat file, downloadable from the database *http://cdsweb.u-strasbg.fr/viz-bin/Cat?II/155* From the last file we derived the radii of Alpha Centauri A and B reported in Table 7.3, no uncertainties being given.

(1) *Assumed* mass of Alpha Centauri A: 2.2351443209E+30 kg; that is, $\mu = 1.1238\mu_{Sun}$
(2) *Assumed* radius of Alpha Centauri A: 821,280 km; that is, 1.18 times the Sun radius.
(3) Alpha Centauri A mean density: 963.259 kg/m^3
(4) Schwarzschild radius of Alpha Centauri A: 3.318 km
(5) Minimal focal distance of naked Alpha Centauri A: 679.262 AU $\sim$ 3.92 light days $\sim$ 17.17 times the Sun-to-Pluto distance
(6) Antenna patterns of the naked Alpha Centauri A's gravity lens at the hydrogen line frequency of 1.420 GHz (= 21 cm wavelength):

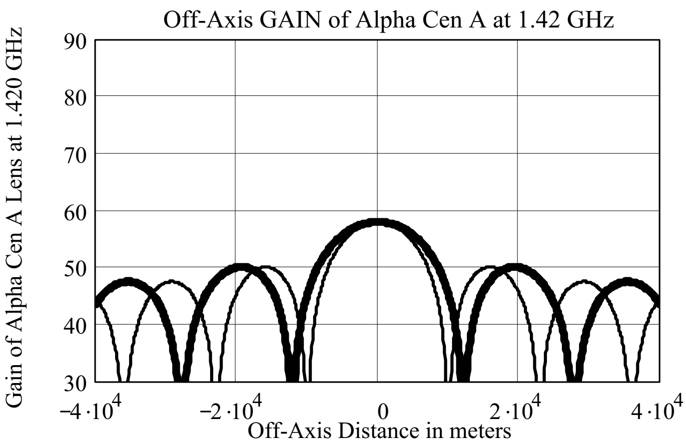


Figure 7.7. Antenna patterns for the naked Alpha Centauri A at the hydrogen line of 1.420 GHz. Again, horizontal scale is distance off axis in meters, vertical scale is the naked Sun *gain* in dB as given by (7.8). The central lobe *widens* while the FOCAL spacecraft is getting farther and farther away from 680 AU.

(7) Patterns of the naked Alpha Centauri A's gravity lens at the water maser frequency of 22 GHz (= 1.36 cm wavelength):

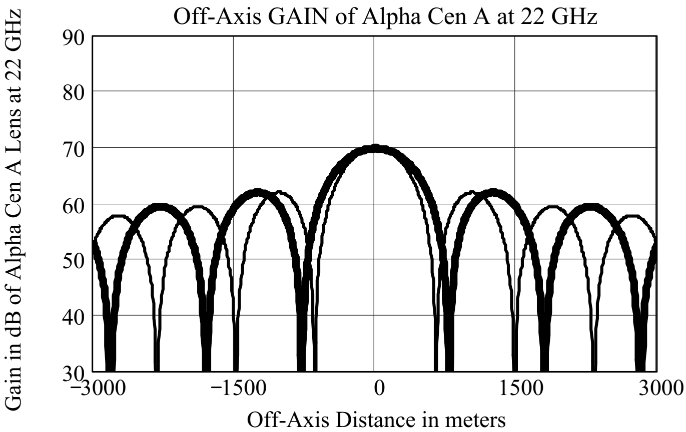


Figure 7.8. Patterns for the naked Alpha Centauri A at the water maser line of 22 GHz.

(8) Patterns of the naked Alpha Centauri A's gravity lens at the Ka band frequency of 32 GHz (= 9.37 mm wavelength):

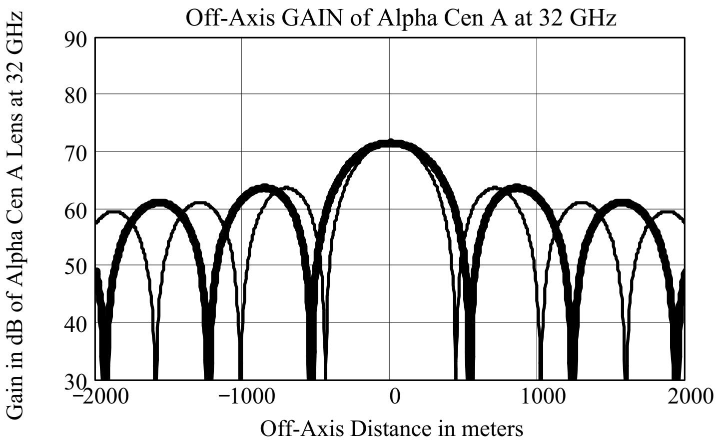


Figure 7.9. Patterns for the naked Alpha Centauri A at the Ka band line of 32 GHz used for space telecommunications.

(9) Patterns of the naked Alpha Centauri A's gravity lens at the CMB peak frequency of 160.378 GHz (= 1.06 mm wavelength):

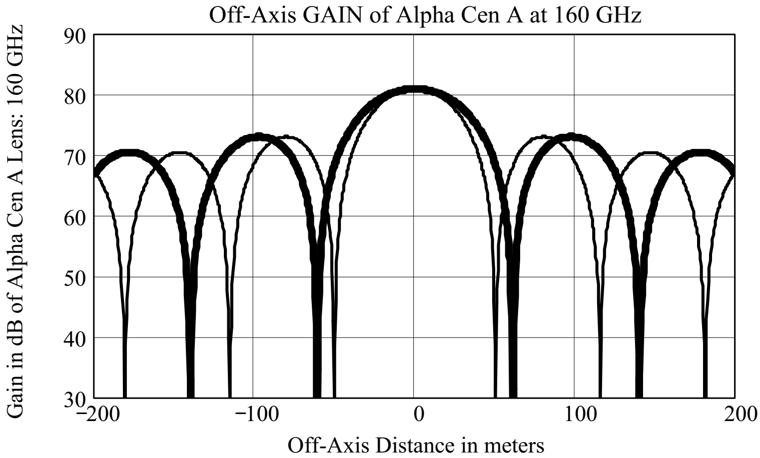


Figure 7.10. Patterns for the naked Alpha Centauri A at the CMB peak frequency of 160.378 GHz.

(10) Patterns of the naked Alpha Centauri A's gravity lens at the positronium frequency of 203 GHz (SETI?):

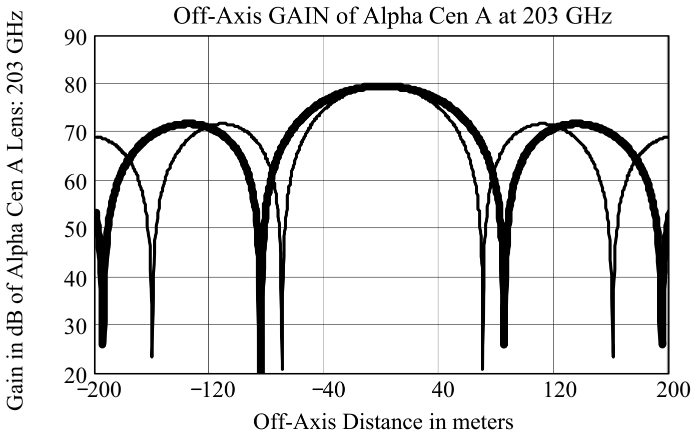


Figure 7.11. Patterns for the naked Alpha Centauri A at the positronium frequency of 203 GHz (SETI?). Note that the alignment is extremely tight (~70 meters at 680 AU), so "they" must be really advanced!

7.6 GRAVITATIONAL LENS OF THE NAKED ALPHA CENTAURI B

(1) *Assumed* mass of Alpha Centauri B: 1.8584435429E+30 kg, that is, $\mu = 0.9344\mu_{Sun}$
(2) *Assumed* radius of Alpha Centauri B: 682,080 km, that is, 0.98 times the radius of the Sun
(3) Alpha Centauri B mean density: 1398.153 kg/m^3
(4) Schwarzschild radius of Alpha Centauri B: 2.759 km
(5) Minimal focal distance of Alpha Centauri B: 563.484 AU $\sim$ 3.25 light days $\sim$ 14.25 times the Sun-to-Pluto distance
(6) Antenna patterns of the naked Alpha Centauri B's gravity lens at the hydrogen line frequency of 1.420 GHz (= 21 cm wavelength):

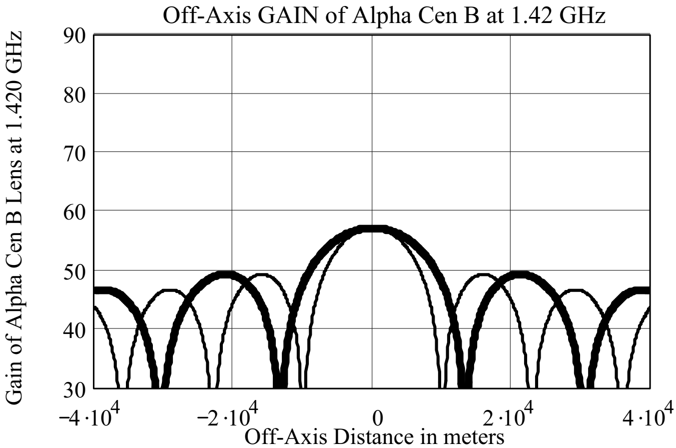


Figure 7.12. Patterns for the naked Alpha Centauri B at the hydrogen line of 1.420 GHz.

(7) Patterns of the naked Alpha Centauri B's gravity lens at the water maser frequency of 22 GHz (= 1.36 cm wavelength):

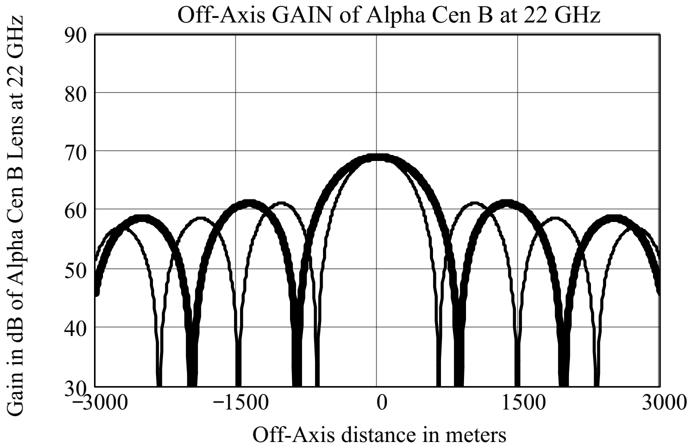


Figure 7.13. Patterns for the naked Alpha Centauri B at the water maser line of 22 GHz.

(8) Patterns of the naked Alpha Centauri B's gravity lens at the Ka band frequency of 32 GHz (= 9.37 mm wavelength):

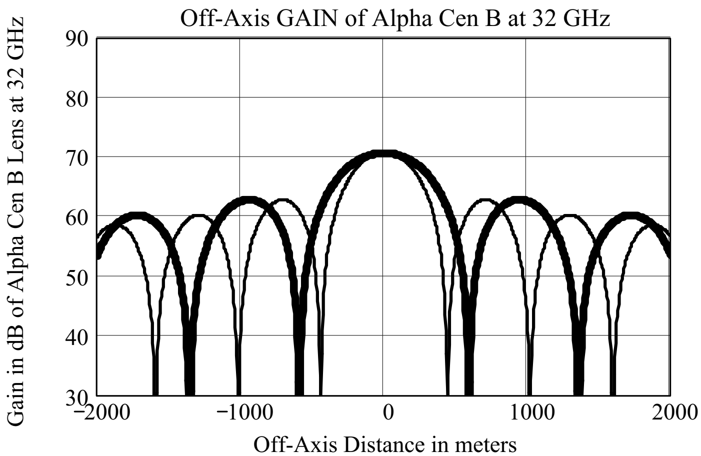


Figure 7.14. Patterns for the naked Alpha Centauri B at the Ka band line of 32 GHz used for space links.

(9) Patterns of the naked Alpha Centauri B's gravity lens at the CMB peak frequency of 160.378 GHz (= 1.06 mm wavelength):

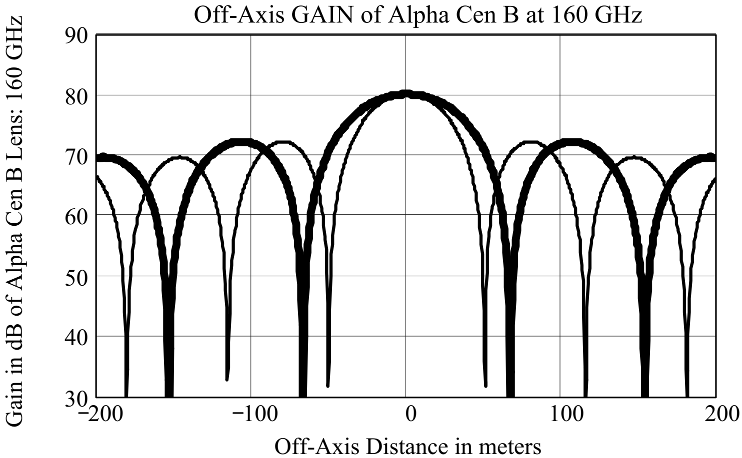


Figure 7.15. Patterns for the naked Alpha Centauri B at the CMB peak frequency of 160.378 GHz.

(10) Patterns of the naked Alpha Centauri B's gravity lens at the positronium frequency of 203 GHz (SETI?):

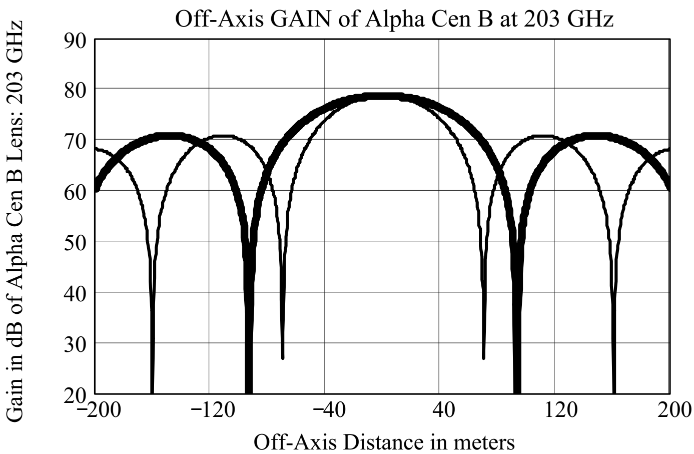


Figure 7.16. Patterns for the naked Alpha Centauri B at the positronium frequency of 203 GHz (SETI?).

7.7 GRAVITATIONAL LENS OF THE NAKED ALPHA CENTAURI C (PROXIMA)

We have taken the data about Alpha Centauri C from [12] of 1998 and available at the website of the Hubble Space Telescope (HST) Astrometry Team: *http://clyde.as.utexas.edu/GFBINFO/ProxCenInfo.htm*

(1) *Assumed* mass of Alpha Centauri C (Proxima): 2.1878081091E+29 kg; that is, $\mu = 0.11\mu_{Sun}$
(2) *Assumed* radius of Alpha Centauri C (Proxima): 104,400 km; that is, 0.15 times the Sun radius.
(3) Alpha Centauri C (Proxima) mean density: 45,900 kg/m^3
(4) Schwarzschild radius of Alpha Centauri C (Proxima): 0.325 km
(5) Minimal focal distance of Alpha Centauri C (Proxima): 112.138 AU $\sim$ 0.65 light days $\sim$ 2.83 times the Sun-to-Pluto distance
(6) Antenna patterns of the naked Alpha Centauri B's gravity lens at the hydrogen line frequency of 1.420 GHz (= 21 cm wavelength):

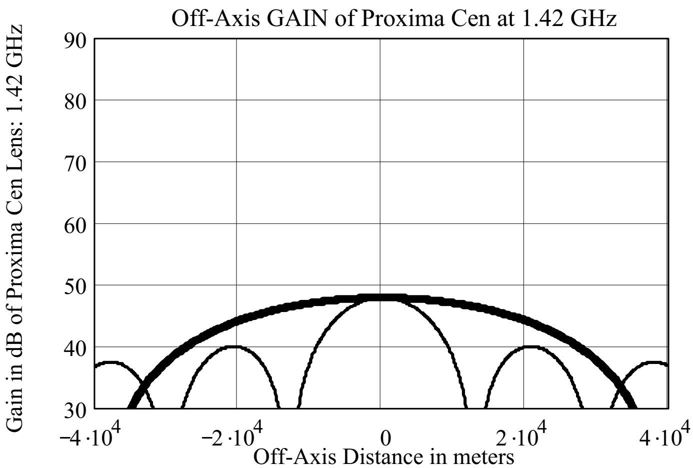


Figure 7.17. Patterns for the naked Alpha Centauri C (Proxima) at the hydrogen line of 1.420 GHz. Note the much lower naked star *gain* ($\sim$47 dB) due to the much smaller dimensions of Proxima with respect to the Sun and both Alpha Centauri A and B. Also, the central lobe widens much when FOCAL goes from the minimal focal distance of 112 AU to 700 AU and 1,000 AU: the alignment gets less tight.

(7) Patterns of the naked Proxima's gravity lens at the water maser frequency (22 GHz; that is, 1.36 cm):

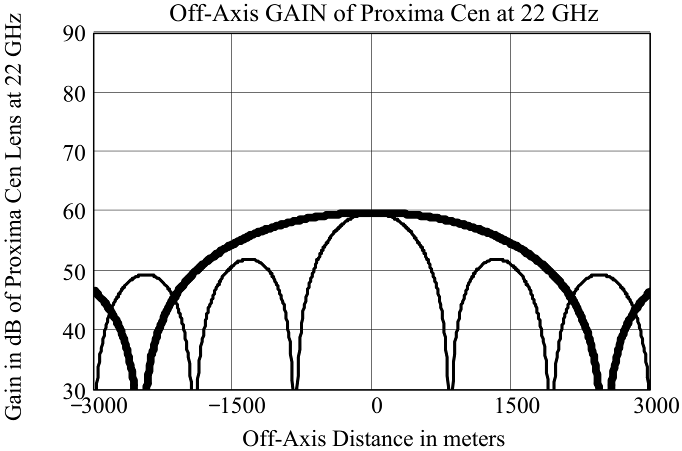


Figure 7.18. Patterns for the naked Proxima at the water maser line of 22 GHz.

(8) Patterns of the naked Proxima's gravity lens at the Ka telecommunication band of 32 GHz (= 9.37 mm wavelength):

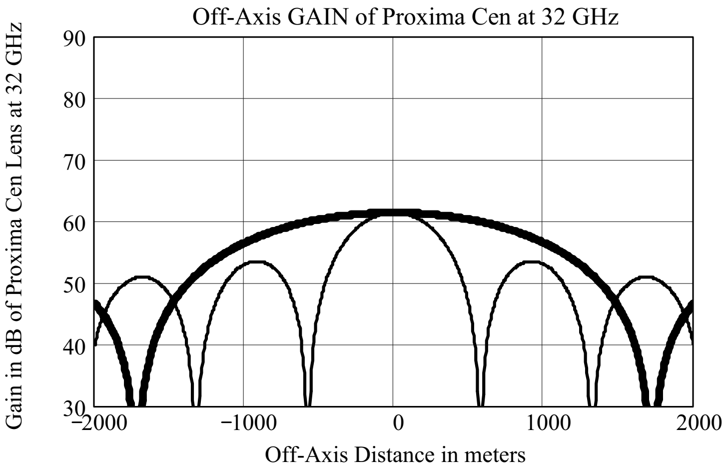


Figure 7.19. Patterns for the naked Proxima at the Ka band line of 32 GHz used for space links.

(9) Patterns of the naked Proxima's gravity lens at the CMB peak frequency of 160.378 GHz (= 1.06 mm wavelength):

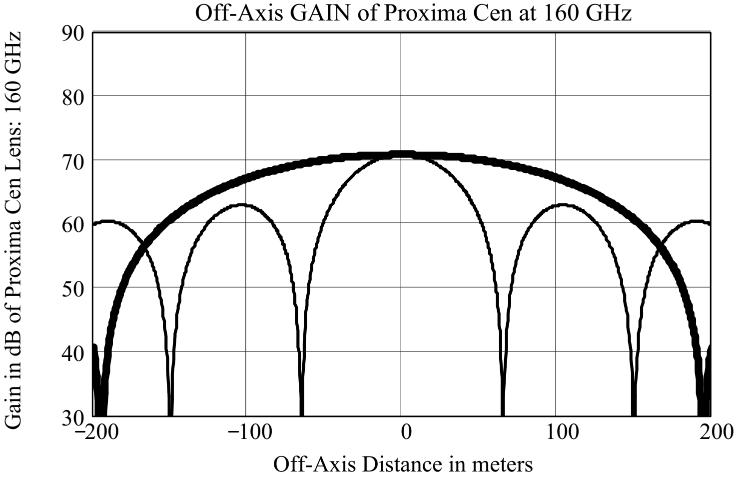


Figure 7.20. Patterns for the naked Proxima at the CMB peak frequency of 160.378 GHz.

(10) Patterns of the naked Proxima's gravity lens at the positronium frequency of 203 GHz (SETI?):

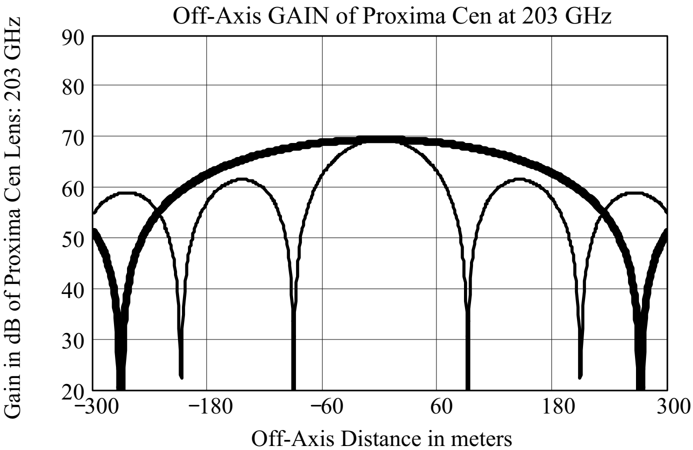


Figure 7.21. Patterns for the naked Proxima at the positronium frequency of 203 GHz (SETI?).

7.8 GRAVITATIONAL LENS OF THE NAKED BARNARD'S STAR

(1) *Assumed* mass of Barnard's Star: 3.3811579868E+29 kg; that is, $\mu = 0.17\mu_{Sun}$
(2) *Assumed* radius of Barnard's Star: 118,320 km; that is, 0.17 times the Sun radius
(3) Barnard's Star mean density: 48,730 kg/m^3
(4) Schwarzschild radius of Barnard's Star: 0.502 km
(5) Minimal focal distance of Barnard's Star: 93.199 AU $\sim$ 0.54 light days $\sim$ 2.36 times the Sun-to-Pluto distance
(6) Antenna patterns of the gravity lens of the naked Barnard's Star at the hydrogen line frequency of 1.420 GHz (= 21 cm wavelength):

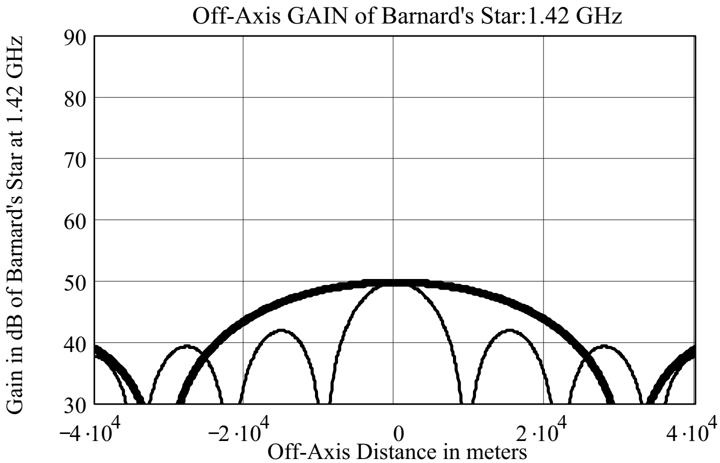


Figure 7.22. Patterns for the naked Barnard's Star at the hydrogen line of 1.420 GHz. As for Proxima, note the much lower naked star *gain* (~49 dB) due to the much smaller dimensions of Barnard's Star with respect to the Sun and both Alpha Centauri A and B. Again the central lobe widens from the minimal focal distance of 93 AU to 700 AU and 1,000 AU, and the alignment gets thus less tight.

(7) Patterns of the gravity lens of the naked Barnard's star at the water maser frequency (22 GHz; that is, 1.36 cm):

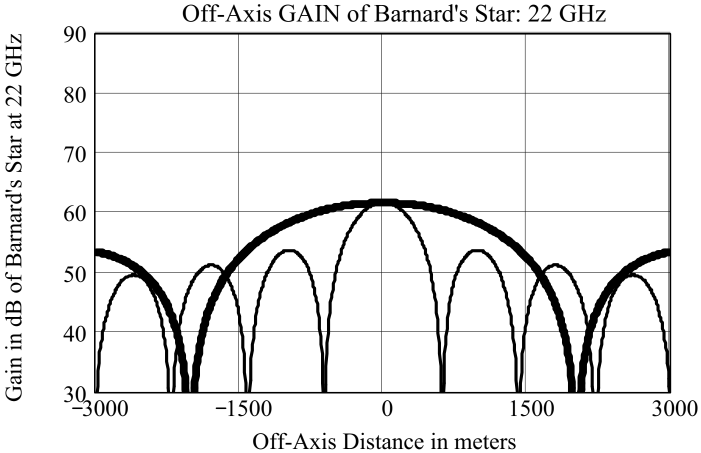


Figure 7.23. Patterns for the naked Barnard's Star at the water maser line of 22 GHz.

(8) Patterns of the gravity lens at the Ka telecommunication band of 32 GHz (= 9.37 mm wavelength):

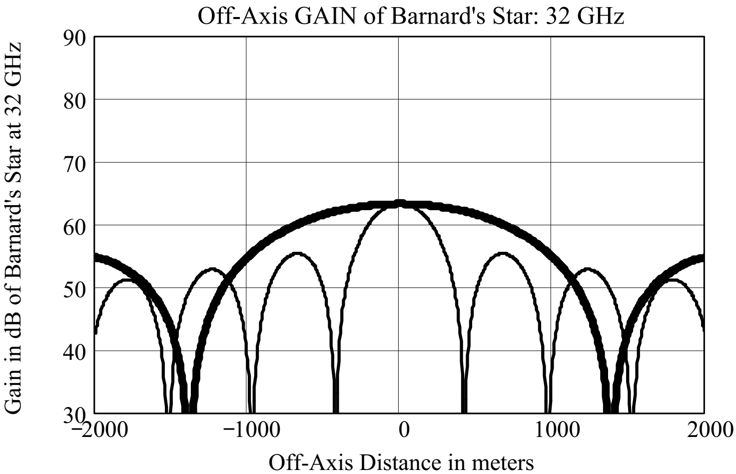


Figure 7.24. Patterns for the naked Barnard's Star at the Ka band line of 32 GHz used for space links.

(9) Patterns of the gravity lens of the naked Barnard's Star at the CMB peak frequency of 160.378 GHz (= 1.06 mm wavelength):

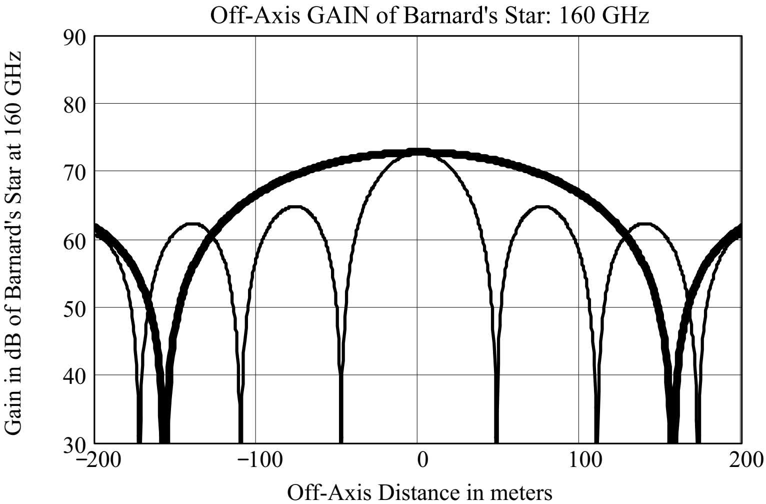


Figure 7.25. Patterns for the naked Barnard's Star at the CMB peak frequency of 160.378 GHz.

(10) Patterns of the naked Proxima's gravity lens at the positronium frequency of 203 GHz (SETI?):

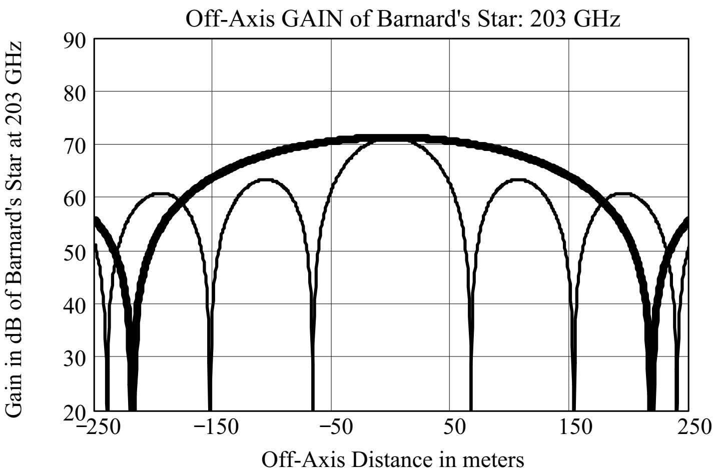


Figure 7.26. Patterns for the naked Barnard's Star at the positronium frequency of 203 GHz (SETI?).

7.9 CONCLUSIONS

Humankind appears to be "lucky" in that two stars out of three in the nearest star system (Alpha Centauri) are stars of the same type as the Sun, namely Alpha Centauri A and B.

The 21st and following centuries are likely to see a host of FOCAL space missions, each one devoted to a different astrophysical target and thus launched along a different direction out of the solar system. But the guess is made here that all of them will use a Tethered System as described in Appendix D in order to avoid, by virtue of interferometry, all the problems caused by random fluctuations within the Solar Corona.

7.10 ACKNOWLEDGMENTS

IAA Academicians Ivàn Almar, Jill Tarter, Nikolai Kardashev, Leslie Shepherd, Giovanni Vulpetti, Mac Reid, and others kept on encouraging over 15 years the author's progress in the study of the FOCAL space mission: their friendship and support is gratefully acknowledged here. Prof. Frank Drake, SETI Institute President, was the teacher who introduced the author to the study of the 550 AU mission back in 1987, and also wrote the Foreword to his book about FOCAL. Finally, Prof. Ernesto Vallerani's constant support is greatly appreciated.

The content of this chapter was first published by the author in 2000, as an Academy Transactions Note published in *Acta Astronautica* [3].

7.11 REFERENCES

[1] V. Eshleman, "Gravitational Lens of the Sun: Its Potential for Observations and Communications over Interstellar Distances," *Science*, **205** (1979), 1133–1135.

[2] C. Maccone, "Space Missions outside the Solar System to Exploit the Gravitational Lens of the Sun," in *Proceedings of the International Conference on Space Missions and Astrodynamics*, held in Turin, Italy, June 18, 1992, C. Maccone (editor), *Journal of the British Interplanetary Society*, **47** (1994), 45–52.

[3] C. Maccone, "The Gravitational Lenses of Alpha Centauri A, B, C and of Barnard's Star," *Acta Astronautica*, **47** (2000), 885–897.

[4] B. G. Andersson, and Slava G. Turyshev, "550 AU Mission: Science Rationale, Goals and Requirements," unpublished JPL internal report dated April 10, 1998.

[5] H. C. Ohanian, "On the Focusing of Gravitational Radiation," *International Journal of Theoretical Physics*, **9** (1974), 425–437.

[6] E. Herlt, and H. Stephani, "Diffraction of a Plane Electromagnetic Wave at a Schwarzschild Black Hole," *International Journal of Theoretical Physics*, **12** (1975), 81–93.

[7] E. Herlt, and H. Stephani, "Wave Optics of the Spherical Gravitational Lens. Part 1: Diffraction of a Plane Electromagnetic Wave by a Large Star," *International Journal of Theoretical Physics*, **15** (1976), 45–65.

[8] J. L. West, "Design Issues for a Mission to Exploit the Gravitational Lens of the Sun at 550 AU," *Acta Astronautica*, **44** (1999), 99–107.
[9] JPL's DAstCom (= Data on Asteroids and Comets) database, included in the "MAnE" (= Mission Analysis Environment) package for mission design.
[10] C. Maccone, "Cosmic Microwave Background and Gravitational Lens of the Sun," *Acta Astronautica*, **46** (2000), 605–614.
[11] D. B. Guenther, and P. Demarque, "α Centauri A B", *Astrophysical Journal*, **531** (2000), 503–520.
[12] G. F. Benedict, B. McArthur, E. Nelan, D. Story, A. L. Whipple, P. J. Shelus, W. H. Jeffreys, D. Hemenway, O. G. Franz, L. H. Wasserman *et al.*, "Photometry of Proxima Centauri and Barnard's Star Using Hubble Space Telescope Fine Guidance Sensor 3: A Search for Periodic Variations," *Astronomical Journal*, **116** (1998), 429–439, see in particular Tables 1 and 2.

8

The Coronal Plasma "pushing" the focus of the gravity + plasma lens far beyond 550 AU

8.1 INTRODUCTION

The gravitational lens of the Sun still needs more study. In fact, above the surface of the Sun, the Corona extends into space across distances that are comparable with the Sun radius, and the coronal effects may only complicate the physical picture of the Sun as a gravitational lens.

The world's leading expert on the gravitational lens of the Sun is Professor Von R. Eshleman, from the Center for Radar Astronomy at Stanford University. Not only was he the first to suggest the exploitation of the focus at 550 AU by means of a very deep–space mission back in 1979 [1], but he and his students continued to work towards understanding the physical characteristics of the Sun's outer layers, something that is crucial to the exploitation of its gravitational lens.

In the summer of 1996 Von Eshleman came to Capri, Italy, to attend the *Fifth Bioastronomy Conference*, and there he presented a report on the current state of knowledge about the physics of the gravitational lens of the Sun. This section is essentially a short summary of Eshleman's 1996 Capri paper [2].

Eshleman and his pupils are currently investigating the effects of the plasma above the surface of the Sun on the gravitational lens. Eshleman provided the author the following drawing (Figure 8.1) and procedure at the Capri 1996 conference:

(1) First, one sees the focal line (solid dark line) outgoing from 550 AU to infinity, where each point is a focus. This solid line is also, of course, the trajectory of the FOCAL spacecraft.
(2) One may notice from the drawing the existence of a *caustic* (i.e., an *envelope* of all the deflected light rays). This caustic really is a funnel-shaped surface along the focal axis outside of which there are no light rays, because all light rays are concentrated inside it by the gravitational lens of the Sun.

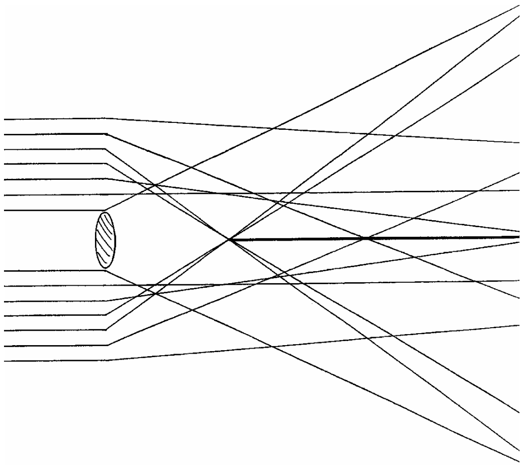


Figure 8.1. Von Eshleman's drawing of the combined effects of the gravitational lens (due to the Sun's mass) and of the coronal lens (due to the plasma properties of the Corona).[1]

The above summarizes the effects of the gravitational lens due to the Sun's mass with no other effects like the Corona influence taken into account. However, Eshleman has now introduced the Plasma Lens (i.e., the deflection of light rays due to the plasma properties of the Corona).

The conclusions given by Eshleman (and reported here without mathematical proofs) are as follows:

(1) For optical wavelengths, but not for IR wavelengths, there is no in-plane caustic, but there is a central, axial focus.
(2) For short millimeter wavelengths (1 mm–4 mm) there is a double caustic inside the distance of about 1,200 AU.
(3) For decimetric wavelengths there is a double caustic at a distance of the order of 1 lt-yr.

We shall see in the next chapter that NASA-JPL might be willing to again study the possibility of a very deep–space mission to 1,000 AU, called "Interstellar Probe" (ISP), as they did in the 1980s for the TAU (Thousand Astronomical Units) mission. Now, according to Eshleman, a spinning interferometer could be used at optical, IR, millimetric, and longer wavelengths to do the following:

(1) Study zodiacal cloud and its effect in interferometric and lensing applications.
(2) Detect terrestrial planets in solar glare as test and calibration for discovery of distant planetary systems.

[1] Personal communication from the author.

(3) Use interferometric methods to search for distant planets.
(4) Characterize and employ the gravity lens of the Sun at optical and IR wavelengths.
(5) Characterize and employ the gravity-plasma lens of the Sun at millimetric wavelengths.
(6) Characterize and employ the gravity-plasma lenses of nearby stars at decimetric and longer wavelengths.
(7) Use the 1,000 AU trajectory to align a nearby star with a region to be studied by focusing, and to align the Sun with another such direction.

In the next chapter, more features of NASA's ISP, currently under study at JPL and elsewhere, will be discussed. They will also be turned to advantage in the first *experimental* detection of how far from the Sun the "true" focus of the Sun + plasma gravity lens of the Sun is actually located.

8.2 THE REFRACTION OF ELECTROMAGNETIC WAVES IN THE SUN CORONAL PLASMA

The rest of this chapter makes use of the mathematical model of the Sun's Corona called the "Baumbach–Allen" model. This is because such a model is regarded by many astrophysicists as "the best" description of the Sun Coronal Plasma we have today (year 2008), though, of course, it will certainly be superseded by more accurate models sooner of later, especially by models that do *not* assume spherical symmetry around the Sun.

The story of the Baumbach–Allen model of the Corona began in the 1930s, when the German astronomer S. Baumbach worked out and published a formula yielding the density of electrons in the Coronal Plasma. Baumbach's formula reads

$$N_{electrons}(b) = \text{const} \cdot \left[2.952 \cdot 10^3 \left(\frac{r_{Sun}}{b}\right)^{16} + 2.28 \cdot 10^2 \left(\frac{r_{Sun}}{b}\right)^6 + 1.1 \cdot \left(\frac{r_{Sun}}{b}\right)^2\right] \quad (8.1)$$

In this formula, b is of course the "impact parameter" (i.e., the radial distance from the Sun's center), and it is measured here in units of the Sun radius, r_{Sun}. Baumbach obtained his formula by careful interpolation of his own measures of the Coronal *brightness*. In 1947, the Australian astronomer C. W. Allen [3] confirmed the validity of Equation (8.1) but re-interpreted one of the terms in Baumbach's formula proving what is now called the Baumbach–Allen model of the Corona, where the plasma density around the Sun varies as a combination of r^{-6} and r^{-16} power laws, from the photosphere to about 10 solar radii.

One of the most recent confirmations of the Baumbach–Allen formula (8.1) comes from the work [4] of M. J. Aschwanden and L.W. Acton, about the "Tomography of the Soft X-Ray Corona: Measurements of Electron Densities, Temperatures, and Differential Emission Measure Distributions above the Limb," a paper submitted to the *Astrophysical Journal* on May 4, 2000, and downloadable from the site *http://solar.physics.montana.edu/hypermail/eprint/0077.html*

In 1998, Beugt G. Andersson and Slava Turyshev of JPL took up the Baumbach–Allen model with the final goal of determining the "best focal distances" of the gravitational lens of the Sun by taking the Coronal effects also into account. They gave their results in the unpublished JPL report [5] that is carefully described in this chapter and assumed as "the right model" of the Corona in this book. Andersson and Turyshev gave in [5] the following *empirical* formula for the *deflection angle* $\theta_{plasma}(b, \nu)$ caused by the Coronal Plasma effects and *opposed to the action of gravity*

$$\theta_{plasma}(b, \nu) = \left(\frac{\nu_0}{\nu}\right)^2 \left[2.952 \cdot 10^3 \left(\frac{r_{Sun}}{b}\right)^{16} + 2.28 \cdot 10^2 \left(\frac{r_{Sun}}{b}\right)^6 + 1.1 \cdot \left(\frac{r_{Sun}}{b}\right)^2\right] \quad (8.2)$$

where

$$\nu_0 = 6.36\,\text{MHz} \quad (8.3)$$

The validity of these equations was tested once again on June 21, 2002, and July 1, 2003, by virtue of the close alignment between the Cassini spacecraft (now at Saturn), the Sun, and the Earth. The physical soundness of these two important radio science experiments was already described in 1999 by the Italy–JPL team of Luciano Iess, Giacomo Giampieri, John D. Anderson, and Bruno Bertotti (see their paper [6]).

8.3 SUMMARY OF THE SUN PURE GRAVITY (NAKED SUN) LIGHT-BENDING THEORY

Before embarking in further calculations, it is appropriate to review the equations predicting the bending of light and radio waves in the vicinity of Sun according to general relativity. Of course, this is the case of the "naked Sun", as Eshleman likes to call it. In other words, the Sun as if there were no flames around it! The "flames" are the Corona, or the Coronal Plasma, in scientific language.

Let us start by pointing out that the deflection angle for a light ray grazing the surface of the Sun, $\theta_{gravity}(r_{Sun})$, is very small:

$$\theta_{gravity}(r_{Sun}) = 1.75\,\text{arcsec} \quad (8.4)$$

It is thus always possible to replace the tangent of such a small angle with the angle itself (measured in radians), as already pointed out in Equation (1.5).

Next we consider the classical Einstein formula, first derived by Einstein [7] in 1915 as a first-order approximated consequence of his general relativity equations, and finally re-derived again in 1916 as a consequence of the Schwarzschild exact solution. So, Einstein's result is that if a light ray passes at the minimal distance b from the (spherical) Sun's center, then the deflection angle $\theta_{gravity}(b)$ at that distance

b is

$$\theta_{gravity}(b) = \frac{4GM_{Sun}}{c^2 b} = \frac{2r_{Schwarzschild}}{b}. \tag{8.5}$$

On the other hand, the geometry of Figure 1.2 immediately shows that

$$\theta_{gravity}(b) = \frac{b}{D_{gravity}}. \tag{8.6}$$

Checking (8.5) against (8.6) one gets

$$\frac{b}{D_{gravity}} = \frac{2r_{Schwarzschild}}{b}. \tag{8.7}$$

Solving this with respect to $D_{gravity}$, one finally gets for any value of b:

$$D_{gravity}(b) = \frac{b^2}{2r_{Schwarzschild}}. \tag{8.8}$$

For the particular case where $b = r_{Sun}$, the last equation yields again the well-known result that is the foundation of this book—that is, Equation (1.3)—which, in this chapter's notation, becomes

$$D_{gravity}(r_{Sun}) = \frac{r_{Sun}^2}{2r_{Schwarzschild}} \approx 550 \text{ AU}. \tag{8.9}$$

On the other hand, eliminating the Schwarzschild radius between (8.8) and (8.9), and solving for $D_{gravity}(b)$, one gets

$$D_{gravity}(b) = D_{gravity}(r_{Sun}) \cdot \left(\frac{b}{r_{Sun}}\right)^2 \tag{8.10}$$

that will be used in the sequel of this chapter.

We complete this section with the "Newtonian comparison". This is just the comparison between the Einstein deflection formula (8.5) and the corresponding deflection formula in Newton's classical theory. The latter can immediately be obtained from the fact that, were photons classical material particles traveling at the non-constant speed $v(t) \equiv c(t)$, their path around the Sun's center would exactly be a hyperbola branch having its focus at the Sun's center. This is the theory of the Keplerian Sun flyby that was developed in Section 3.3, and the relevant formula for our current problem is the first of Equations (3.26); that is

$$r_p = -\frac{GM_{Sun}}{v_\infty^2}\left(1 + \frac{1}{\cos\vartheta_\infty}\right). \tag{8.11}$$

The Newtonian deflection angle β (see Figure 3.27) is related to ϑ_∞ by Equation (3.28). Then (3.28) changes (8.11) into the "periastron r_p vs. deflection angle β"

formula valid for any Newtonian (i.e., Keplerian) flyby:

$$r_p(\beta) = \frac{GM_{Sun}}{v_\infty^2}\left(\frac{1}{\sin\left(\frac{\beta}{2}\right)} - 1\right). \tag{8.11}$$

But the photon speed is $v_\infty = c$ and the light deflection angle is very small, so the sine may be replaced by the angle itself, yielding the approximated version of (8.11)

$$r_p \approx \frac{GM_{Sun}}{c^2}\left(\frac{2}{\beta} - 1\right). \tag{8.12}$$

Again, as the angle β is very small, the -1 term in (8.12) is numerically dwarfed, and the approximation holds

$$r_p \approx \frac{2GM_{Sun}}{c^2\beta} \tag{8.13}$$

which, invoking the impact parameter $r_p = b$ and $\beta = \theta_{Newton}$, finally yields the Newtonian deflection

$$\theta_{Newton} \approx \frac{2GM_{Sun}}{c^2 b} \tag{8.14}$$

Checking this against the Einsteinian formula (8.5) one obviously finds

$$\theta_{Einstein} \approx 2\theta_{Newton} \tag{8.15}$$

And, in terms of focal distances,

$$D_{Newton} \approx 2D_{Einstein} \tag{8.16}$$

In words: were the Newtonian theory correct, the minimal focal distance from the Sun would be 1,100 AU, rather than just 550 AU, as in the Einsteinian theory. Thus, had Einstein not taught us general relativity, we would have expected to send the "FOCAL" spacecraft to a distance of 1,100 AU, rather than "just" to 550 AU. The advent of Einstein's theory eased our difficulties, in regard to the "FOCAL" distance, by a factor of 50%. Not a bad result at all, confirming that "in Newton's days, the time was not yet ripe for the "FOCAL" space mission!"

8.4 GRAVITY + PLASMA LENS OF THE SUN: FOCAL AXIS INTERCEPT FOR ANY RAY PASSING AT DISTANCE *b* FROM THE SUN

We are now ready to consider the real Sun (i.e., the Sun and its Corona) and its effects upon electromagnetic waves traveling across the Corona. The overall deflection angle for an electromagnetic wave, having a frequency ν and passing at an impact distance b from the Sun center will be denoted by $\theta_{gravity+plasma}(b, \nu)$. Denoting by $D_{gravity+plasma}$

the intercept of the straight light path with the focal axis, from Figure 7.4-1 one has (obviously replacing the tangent by the angle itself)

$$\theta_{gravity+plasma}(b,\nu) \approx \frac{b}{D_{gravity+plasma}} \tag{8.17}$$

On the other hand, the same angle $\theta_{gravity+plasma}(b,\nu)$ is obviously given by the difference between the angle $\theta_{gravity}(b)$ and the angle $\theta_{plasma}(b,\nu)$, where we assume these angles to be positive if the light is bent towards the focal axis, and negative otherwise. One thus gets:

$$\begin{aligned}\theta_{gravity+plasma}(b,\nu) &= \theta_{gravity}(b) - \theta_{plasma}(b,\nu)\\ &= \frac{2r_{Schwarzschild}}{b}\\ &\quad - \left(\frac{\nu_0}{\nu}\right)^2\left[2.952\cdot 10^3\left(\frac{r_{Sun}}{b}\right)^{16} + 2.28\cdot 10^2\left(\frac{r_{Sun}}{b}\right)^{6} + 1.1\cdot\left(\frac{r_{Sun}}{b}\right)^{2}\right]\end{aligned} \tag{8.18}$$

On equating Equations (8.17) and (8.18), and then solving the resulting equation for $D_{gravity+plasma}$, one gets

$$D_{gravity+plasma}(b,\nu)$$

$$= \frac{b}{\dfrac{2r_{Schwarzschild}}{b} - \left(\dfrac{\nu_0}{\nu}\right)^2\left[2.952\cdot 10^3\left(\dfrac{r_{Sun}}{b}\right)^{16} + \cdots\right]}$$

where the dots denote the remaining terms in braces in (8.18). On suitable multiplication and division at the denominator, this becomes

$$= \frac{b}{\dfrac{2r_{Schwarzschild}}{b} - \dfrac{2r_{Schwarzschild}}{b}\left(\dfrac{\nu_0}{\nu}\right)^2\dfrac{b}{2r_{Schwarzschild}}\left[2.952\cdot 10^3\left(\dfrac{r_{Sun}}{b}\right)^{16} + \cdots\right]}$$

$$= \frac{b}{\dfrac{2r_{Schwarzschild}}{b}\left\{1 - \left(\dfrac{\nu_0}{\nu}\right)^2\dfrac{b}{2r_{Schwarzschild}}\left[2.952\cdot 10^3\left(\dfrac{r_{Sun}}{b}\right)^{16} + \cdots\right]\right\}}$$

$$= \frac{\dfrac{b^2}{2r_{Schwarzschild}}}{1 - \left(\dfrac{\nu_0}{\nu}\right)^2\dfrac{b}{2r_{Schwarzschild}}\left[2.952\cdot 10^3\left(\dfrac{r_{Sun}}{b}\right)^{16} + \cdots\right]}$$

which, rewriting all terms explicitly, becomes

$$= \frac{\dfrac{b^2}{2r_{Schwarzschild}}}{1 - \left(\dfrac{\nu_0}{\nu}\right)^2 \dfrac{b}{2r_{Schwarzschild}} \left[2.952 \cdot 10^3 \left(\dfrac{r_{Sun}}{b}\right)^{16} + 2.28 \cdot 10^2 \left(\dfrac{r_{Sun}}{b}\right)^6 + 1.1 \cdot \left(\dfrac{r_{Sun}}{b}\right)^2\right]}$$

now, the denominator can be simplified by reducing by one integer all powers inside the braces, finally yielding

$$= \frac{\dfrac{b^2}{2r_{Schwarzschild}}}{1 - \dfrac{\nu_0^2}{\nu^2} \dfrac{r_{Sun}}{2r_{Schwarzschild}} \left[2.952 \cdot 10^3 \left(\dfrac{r_{Sun}}{b}\right)^{15} + 2.28 \cdot 10^2 \left(\dfrac{r_{Sun}}{b}\right)^5 + 1.1 \cdot \dfrac{r_{Sun}}{b}\right]}. \tag{8.19}$$

The last result leads to the definition of the *critical frequency* $\nu_{critical}(b)$ as

$$\nu^2_{critical}(b) = \frac{\nu_0^2 \cdot r_{Sun}}{2r_{Schwarzschild}} \left[2.952 \cdot 10^3 \left(\frac{r_{Sun}}{b}\right)^{15} + 2.28 \cdot 10^2 \left(\frac{r_{Sun}}{b}\right)^5 + 1.1 \cdot \frac{r_{Sun}}{b}\right]$$

In other words, taking the square root, we define the *critical frequency* $\nu_{critical}(b)$ as the new function of the impact parameter b only

$$\nu_{critical}(b) = \sqrt{\frac{\nu_0^2 \cdot r_{Sun}}{2r_{Schwarzschild}} \left[2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{15} + 228 \cdot \left(\frac{r_{Sun}}{b}\right)^5 + 1.1 \cdot \frac{r_{Sun}}{b}\right]} \tag{8.20}$$

But what is the physical meaning of this critical frequency?

Think geometrically, and consider increasing distances b above the surface of the Sun. The least of such increasing distances b above the Sun is of course the surface of the Sun itself, given by $b = r_{Sun}$. The relevant critical frequency is found by substituting $b = r_{Sun}$ in (8.20), and one then finds

$$\nu_{critical}(r_{Sun}) = \sqrt{\frac{\nu_0^2 \cdot r_{Sun}}{2r_{Schwarzschild}} [2952 + 228 + 1.1]} \approx 122.361 \text{ GHz} \tag{8.21}$$

Then, for $b \to \infty$, (8.20) shows that $\nu_{critical} \to 0$, and a few more calculations on (8.20) would reveal that the function is monotonic (i.e., it gradually decreases with increasing distance from the Sun, as is physically obvious). The plot of the $\nu_{critical}(b)$ function is shown in Figure 8.2 hereafter.

The physical meaning of the critical frequency, however, only comes to light when one considers the intercept $D_{gravity+plasma}$ of the straight light path with the focal axis—namely, Equation (8.19)—which, invoking the definition of critical frequency

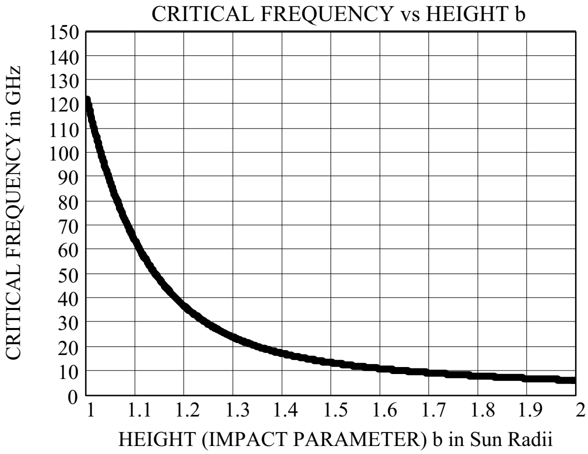


Figure 8.2. Electromagnetic waves across the Plasma of the Sun Corona: critical frequency vs. impact parameter b above the Sun surface.

(8.20), takes the easier form:

$$D_{gravity+plasma}(b,\nu) \approx \frac{\dfrac{b^2}{2r_{Schwarzschild}}}{1 - \dfrac{\nu^2_{critical}(b)}{\nu^2}}. \tag{8.22}$$

This is the fundamental formula yielding the intercept D between the focal axis and the light rays passing at the distance b (impact parameter) from the Sun's center. Replacing the familiar value of 550 AU by virtue of (8.9) *may be written in the form*

$$D_{gravity+plasma}(b,\nu) \approx \frac{550\ \mathrm{AU}\cdot\left(\dfrac{b}{r_{Sun}}\right)^2}{1 - \dfrac{\nu^2_{critical}(b)}{\nu^2}}. \tag{8.23}$$

The physical meaning of the critical frequency $\nu_{critical}(b)$ is now evident from both of the last two equations: they clearly show that the ray intercept with the focal axis is negative (i.e., *there is no more focus*), when the observing frequency ν is smaller than the critical frequency $\nu_{critical}(b)$. In other words, *there is focusing by the (gravity + plasma) lens of the Sun if, and only if, the frequency ν, on which the FOCAL spacecraft makes it observations through the Sun's lens, is* ***higher*** *than the critical frequency $\nu_{critical}(b)$ for rays having the same impact parameter b—namely, passing* at the same distance from the Sun:

$$\nu(b) > \nu_{critical}(b) \tag{8.24}$$

FOCUSING CONDITION

It is also useful to write the expression of the total deflection angle (8.18) in terms of the critical frequency (8.20)

$$\begin{aligned}
\theta_{gravity+plasma}(b,\nu) &= \theta_{gravity}(b) - \theta_{plasma}(b,\nu) \\
&= \frac{2r_{Schwarzschild}}{b} \\
&\quad - \left(\frac{\nu_0}{\nu}\right)^2 \left[2.952 \cdot 10^3 \left(\frac{r_{Sun}}{b}\right)^{16} + 2.28 \cdot 10^2 \left(\frac{r_{Sun}}{b}\right)^6 + 1.1 \cdot \left(\frac{r_{Sun}}{b}\right)^2\right] \\
&= \frac{2r_{Schwarzschild}}{b} - \frac{\nu^2_{critical}(b)}{\nu^2} \cdot \frac{2r_{Schwarzschild}}{r_{Sun}} \\
&= \frac{2r_{Schwarzschild}}{b} \cdot \left(1 - \frac{\nu^2_{critical}(b)}{\nu^2} \cdot \frac{b}{r_{Sun}}\right).
\end{aligned} \tag{8.25}$$

Thus, the total deflection angle is given by

$$\theta_{gravity+plasma}(b,\nu) = \frac{2r_{Schwarzschild}}{b} \cdot \left(1 - \frac{\nu^2_{critical}(b)}{\nu^2} \cdot \frac{b}{r_{Sun}}\right). \tag{8.26}$$

8.5 ASYMPTOTIC ($z \to \infty$) STRAIGHT LIGHT PATH

In this section we are going to derive the equation of the asymptotic straight light path—that is, the equation of the straight line in the (z, r) plane that approximates the actual straight light path of the Sun-deflected light rays for $z \to \infty$. This result will be used in the following sections to derive the intercept of this straight light path with the focal axis z—that is, the actual focus of the (gravity + plasma) lens of the Sun. The relevant two caustics in the (z, r) plane will also be derived.

We start by recalling from elementary analytical geometry that the equation of any straight line, intercepting the x and y axes at the distances p and q from the origin, respectively, is

$$\frac{x}{p} + \frac{y}{q} = 1. \tag{8.27}$$

To keep the same notation as in Figure 1.2 on p. 6 (basic geometry of the gravitational lens of the Sun) one sees at once that the horizontal axis now becomes the z

axis, the vertical axis becomes the r axis, and the substitutions

$$\begin{cases} x \to z \\ y \to r \end{cases} \tag{8.28}$$

must be made in (8.27). Also, the two intercepts of the straight line (i.e., the asymptotic light ray) with the two axes are given by, respectively

$$\begin{cases} p \to D(b, \nu) \\ q \to D(b, \nu) \cdot \tan(\theta_{gravity+plasma}(b)). \end{cases} \tag{8.29}$$

Replacing (8.28) and (8.29) into (8.27), one gets

$$\frac{z}{D(b, \nu)} + \frac{r}{D(b, \nu) \cdot \tan(\theta_{gravity+plasma}(b))} = 1. \tag{8.30}$$

This is the equation of the asymptotic ($z \to \infty$) straight light ray after deflection by the Sun has occurred. Moreover, since the deflection angle is very small, one is allowed to use the approximation

$$\tan(\theta_{gravity+plasma}(b)) \approx \theta_{gravity+plasma}(b). \tag{8.31}$$

Rearranging, one finds the asymptotic straight light path

$$z + \frac{r}{\theta_{gravity+plasma}(b)} - D(b, \nu) = 0. \tag{8.32}$$

8.6 THE THREE APPROXIMATIONS TO THE SUN'S (GRAVITY + PLASMA) LENS: "CLOSE-SUN", "MID-DISTANCE", AND "AT-INFINITY" (L, K, AND F CORONA, RESPECTIVELY)

Let us write again the critical frequency given by Equation (8.20)

$$\nu_{critical}(b) = \sqrt{\frac{\nu_0^2 \cdot r_{Sun}}{2r_{Schwarzschild}} \left[2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{15} + 228 \cdot \left(\frac{r_{Sun}}{b}\right)^5 + 1.1 \cdot \frac{r_{Sun}}{b}\right]}. \tag{8.20}$$

To gain a deeper understanding of the numerical behavior of this critical frequency as a function of the distance from the Sun center (or impact parameter) b, let us square it and plot each of its three contributions separately. We also make reference to, and use the language of, the most recent paper (year 2000) that we could find about the Baumbach–Allen model of the Corona [4]; this is the paper by M. J. Aschwanden and L.W. Acton titled "Tomography of the Soft X-Ray Corona: Measurements of Electron Densities, Temperatures, and Differential Emission Measure Distributions above the Limb". This paper was submitted to the *Astrophysical Journal* on May 4, 2000, and is downloadable from the site *http://solar.physics.montana.edu/hypermail/eprint/0077.html*. Figure 8.3 plots, for the square of Equation (8.20), each of these three contributions as well as their sum, the critical frequency.

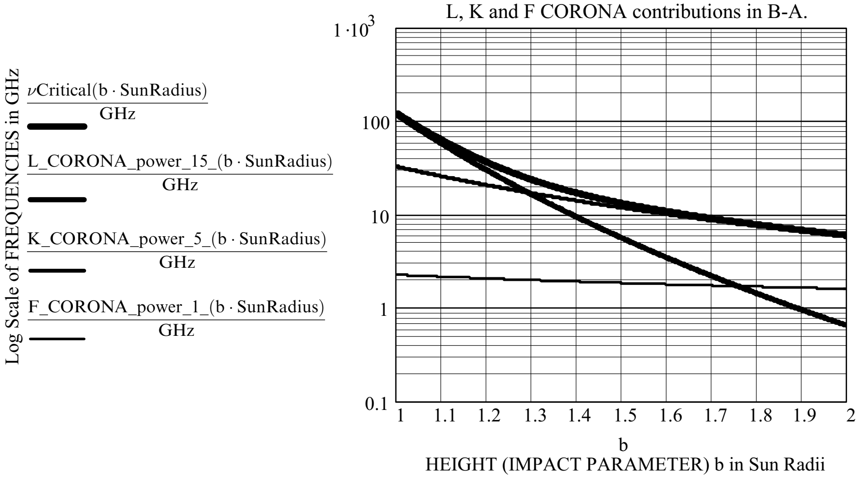


Figure 8.3. The contribution of each of the three terms making up the expression of the critical frequency. They correspond to the so-called L, K, and F Corona, respectively, in the Baumbach–Allen formula (8.1).

One then finds

(1) The term with the power 15, corresponding to the L Corona, given by

$$\text{power\_15\_}(b) = 2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{15} = 2952 \cdot \left(\frac{b}{r_{Sun}}\right)^{-15} \tag{8.33}$$

(2) The term with the power 5, corresponding to the K Corona (K for continuum), given by

$$\text{power\_5\_}(b) = 228 \cdot \left(\frac{r_{Sun}}{b}\right)^{5} = 228 \cdot \left(\frac{b}{r_{Sun}}\right)^{-5} \tag{8.34}$$

(3) The term with the power 1, corresponding to the F Corona (F for Fraunhofer), given by

$$\text{power\_1\_}(b) = 1.1 \cdot \frac{r_{Sun}}{b} = 1.1 \cdot \left(\frac{b}{r_{Sun}}\right)^{-1}. \tag{8.35}$$

To determine the minimal focal distance that the FOCAL spacecraft must reach according to each of these three approximations of the full (gravity + plasma) lens of the Sun, we are going to invoke the total deflection expression (8.18), which we repeat here for convenience

$$\begin{aligned}\theta_{gravity+plasma}(b,\nu) &= \theta_{gravity}(b) - \theta_{plasma}(b,\nu) \\ &= \frac{2r_g}{b} - \left(\frac{\nu_0}{\nu}\right)^2 \left[2.952 \cdot 10^3 \left(\frac{r_{Sun}}{b}\right)^{16} + 2.28 \cdot 10^2 \left(\frac{r_{Sun}}{b}\right)^{6} + 1.1 \cdot \left(\frac{r_{Sun}}{b}\right)^{2}\right]\end{aligned} \tag{8.18}$$

Were we going to keep this expression just the same in the sequel, it would simply be impossible to make further *analytical* progress, since it is impossible to handle an algebraic equation having the independent variable r_{Sun}/b raised to the three exponents 16, 6, and 2 at the same time.

To go ahead, one is thus forced to adopt either of the following three approximations:

(1) "Close-Sun" approximation for the total deflection angle $\theta(b, \nu)$, where only the term with the power 16 in r_{Sun}/b is retained in (8.18); that is

$$\theta_{close-Sun}(b, \nu) \approx \frac{2r_g}{b} - \left(\frac{\nu_0}{\nu}\right)^2 2952 \left(\frac{r_{Sun}}{b}\right)^{16}. \tag{8.36}$$

To shorten the notation a little, we have set

$$r_{Schwarzschild} \equiv r_g \tag{8.37}$$

where the "g" subscript stands obviously for "gravitational", inasmuch as the Schwarzschild radius is called "gravitational radius" by some authors. Notice that this approximation may well be called the "L Corona approximation" of the (gravity + plasma) lens inasmuch as the corresponding impact parameter b ranges in between 1 (Sun-grazing electromagnetic waves) and about 1.3 solar radii.

(2) "Mid-distance" approximation for the deflection angle $\theta(b, \nu)$, where only the term with the power 6 in r_{Sun}/b is retained in (8.18); that is

$$\theta_{mid-distance}(b, \nu) \approx \frac{2r_g}{b} - \left(\frac{\nu_0}{\nu}\right)^2 228 \left(\frac{r_{Sun}}{b}\right)^{6}. \tag{8.38}$$

This approximation may be called the "K Corona approximation" of the (gravity + plasma) lens, and the corresponding impact parameter b roughly ranges in between 1.3 and 3 solar radii.

(3) "At-infinity" approximation for the total deflection angle $\theta(b, \nu)$, where the term with the power 2 in r_{Sun}/b is retained in (8.18) only; that is

$$\theta_{\infty}(b, \nu) \approx \frac{2r_g}{b} - \left(\frac{\nu_0}{\nu}\right)^2 1.1 \cdot \left(\frac{r_{Sun}}{b}\right)^{2}. \tag{8.39}$$

This approximation may be called the "F Corona approximation" of the (gravity + plasma) lens inasmuch as the corresponding impact parameter b ranges from about 3 solar radii up to infinity, and this may be regarded as a sufficiently good approximation for $b \to \infty$.

The intercept function $D(b, \nu)$ in the straight line equation (8.22) must also be approximated.

In fact, the full expression (8.20), after invoking the definition of critical frequency (7.4-4), yields the following three approximated expressions of the focal axis intercept $D(b, \nu)$:

(1) Close-Sun focal axis intercept

$$D_{close-Sun}(b, \nu) \approx \frac{\dfrac{b^2}{2r_g}}{1 - \dfrac{r_{Sun}}{2r_g} \cdot \left(\dfrac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\dfrac{r_{Sun}}{b}\right)^{15}}. \tag{8.40}$$

(2) Mid-distance focal axis intercept

$$D_{mid-distance}(b, \nu) \approx \frac{\dfrac{b^2}{2r_g}}{1 - \dfrac{r_{Sun}}{2r_g} \cdot \left(\dfrac{\nu_0}{\nu}\right)^2 \cdot 228 \cdot \left(\dfrac{r_{Sun}}{b}\right)^{5}}. \tag{8.41}$$

(3) At-infinity focal axis intercept

$$D_{\infty}(b, \nu) \approx \frac{\dfrac{b^2}{2r_g}}{1 - \dfrac{r_{Sun}}{2r_g} \cdot \left(\dfrac{\nu_0}{\nu}\right)^2 \cdot 1.1 \cdot \dfrac{r_{Sun}}{b}}. \tag{8.42}$$

At last we are ready to replace both the total deflection angle and the focal intercept into the asymptotic straight light path (8.32). One must also do a little rearranging of terms—that is, one must multiply both the numerator and the denominator of the second term in (8.32) by $b/(2r_g)$. The same expression at both denominators of the second and third terms is thus obtained. Finally, by taking this common denominator in front of z, one gets for each of the three cases, respectively:

(1) Close-Sun straight ray path:

$$z \cdot \left[1 - \frac{r_{Sun}}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{15}\right] + \frac{b}{2r_g} \cdot r - \frac{b^2}{2r_g} = 0 \tag{8.43}$$

(2) Mid-distance straight ray path:

$$z \cdot \left[1 - \frac{r_{Sun}}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 228 \cdot \left(\frac{r_{Sun}}{b}\right)^{5}\right] + \frac{b}{2r_g} \cdot r - \frac{b^2}{2r_g} = 0 \tag{8.44}$$

(3) At-infinity straight ray path:

$$z \cdot \left[1 - \frac{r_{Sun}}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 1.1 \cdot \frac{r_{Sun}}{b}\right] + \frac{b}{2r_g} \cdot r - \frac{b^2}{2r_g} = 0 \tag{8.45}$$

In mathematical language, Equations (8.43) through (8.45) represent three families of ∞^1 straight lines in the (z, r)-plane with parameter $1 \le b \le \infty$. In Section 8.8 we are going to prove that each of these three families of straight lines admits an envelope. In physical language, this means that we are going to prove that the gravitational lens of the Sun for electromagnetic waves crossing the Corona has two caustics (i.e., surfaces of revolution around the focal axis where all light rays are tangential).

8.7 FOCAL DISTANCE VS. HEIGHT AND MINIMAL FOCAL DISTANCE FOR ANY ASSIGNED FREQUENCY

It is entirely under our control to select the frequency (or frequencies) on which a given FOCAL spacecraft will be able to look at the radio sources magnified by the Sun's (gravity + plasma) lens. In fact, the relevant detectors of electromagnetic waves would probably be photometers, or bolometers, or similar instruments and so their working passband is entirely our choice. For instance, in Chapter 9 we will discuss the proposal, made by this author to NASA's Interstellar Probe (ISP) science and technology definition team, that NASA's ISP should be able to detect the Cosmic Microwave Background peak at 160.378 GHz by a suitable instrument.

The frequency ν of the observed electromagnetic waves is thus an "a priori" fixed parameter.

This physical remark is important to settle the question of which are the independent variables and which are the functions in the equations derived in the previous section. For instance, consider the fundamental question: "Which is the minimal focal distance from the Sun that a FOCAL spacecraft must reach to detect electromagnetic waves focused there by the (gravity + plasma) lens of the Sun?" This question will be answered in this section by resorting to the three close-Sun, mid-distance, and at-infinity approximations for the Baumbach–Allen straight rays given by (8.43), (8.44), and (8.45), respectively.

Consider (8.43) plus the equation of the focal axis, $r = 0$. By replacing $r = 0$ into (8.43) one finds

$$z_{close-Sun} \cdot \left[1 - \frac{r_{Sun}}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{15}\right] - \frac{b^2}{2r_g} = 0 \tag{8.46}$$

This equation yields the focal distance z of the (gravity + plasma) lens (at any assigned frequency ν) for waves bypassing the Sun at any impact parameter b. Put another way, solving (8.46) for z, one finds how the focal distance z changes at different heights b above the Sun surface:

$$z_{close-Sun}(b, \nu) = \frac{\dfrac{b^2}{2r_g}}{1 - \dfrac{r_{Sun}}{2r_g} \cdot \left(\dfrac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\dfrac{r_{Sun}}{b}\right)^{15}}. \tag{8.47}$$

Of course, we are very much interested in the *minimal* focal distance—namely, the *minimal* distance from the Sun that the FOCAL spacecraft must reach in order to be flooded by the electromagnetic waves focused there by the (gravity + plasma) lens of the Sun. Therefore, we must take the derivative of (8.47) with respect to b

$$\frac{dz_{close-Sun}(b,\nu)}{db} = \frac{2b\left[1 - \frac{r_{Sun}}{2r_g}\cdot\left(\frac{\nu_0}{\nu}\right)^2\cdot 2952\cdot\left(\frac{r_{Sun}}{b}\right)^{15}\right] - b^2\frac{15}{2r_g}\cdot\left(\frac{\nu_0}{\nu}\right)^2\cdot 2952\cdot\left(\frac{r_{Sun}}{b}\right)^{16}}{2r_g\left[1 - \frac{r_{Sun}}{2r_g}\cdot\left(\frac{\nu_0}{\nu}\right)^2\cdot 2952\cdot\left(\frac{r_{Sun}}{b}\right)^{15}\right]^2}. \tag{8.48}$$

To find the minimal focal distance, the numerator on the right-hand side of the last equation must be set to zero. One may prefer to re-write the last formula in terms of the critical frequency $\nu_{critical}$, which, in the close-Sun approximation, reads

$$\nu^2_{critical-close-Sun}(b) = \frac{\nu_0^2\, r_{Sun}}{2r_g}\cdot 2952\cdot\left(\frac{r_{Sun}}{b}\right)^{15} \tag{8.49}$$

By setting $b = r_{Sun}$ in the last equation, one gets

$$\nu^2_{critical-close-Sun}(r_{Sun}) = \frac{\nu_0^2\, r_{Sun}}{2r_g}\cdot 2952. \tag{8.50}$$

Then, the numerator on the right-hand side of (8.48), rewritten by virtue of (8.50) and set to zero, yields the equation of the minimal focal distance:

$$2b^{15}_{close-Sun}\nu^2 - 17r^{15}_{Sun}\nu^2_{critical-close-Sun}(r_{Sun}) = 0. \tag{8.51}$$

Solving this for $b_{close-Sun}$ one finally gets the *distance from the Sun center passed by those electromagnetic waves that reach the minimal focal distance in the close-Sun approximation:*

$$b_{\min-close-Sun}(\nu) = \frac{17^{\frac{1}{15}}\,\nu^{\frac{2}{15}}_{critical}(r_{Sun})\,r_{Sun}}{2^{\frac{1}{15}}\,\nu^{\frac{2}{15}}}. \tag{8.52}$$

This formula is very important for all practical radio science experiments, like those described in [6], inasmuch as it tells us *where* inside the corona the minimal focal distance waves pass.

To obtain the minimal focal distance one merely has to replace (8.52) into the function (8.47) and rearrange. One thus gets:

$$z_{\min-close-Sun}(\nu) = \frac{17^{\frac{17}{15}}\,\nu^{\frac{4}{15}}_{critical}(r_{Sun})\,r^2_{Sun}}{30\cdot 2^{\frac{2}{15}}\,\nu^{\frac{4}{15}}\,r_g}. \tag{8.53}$$

This is perhaps the most important result in this book: it tells us the minimum distance that the FOCAL spacecraft must reach in order to be flooded by the electromagnetic waves emitted by a source at frequency ν and focused upon the spacecraft by the (gravity + plasma) lens of the Sun.

It is interesting to notice that, by eliminating ν (i.e., "the parameter", in mathematical language) between (8.52) and (8.53), one gets the simple result

$$z_{\min-close-Sun} = \frac{17}{30} \cdot \frac{b^2_{\min-close-Sun}}{r_g}. \tag{8.54}$$

This is, of course, the *locus of the minimal focal distances in the* (z, b) *plane* for the gravitational lens of the Sun for all frequencies in the close-Sun approximation, and it is just a parabola!

So far for the close-Sun approximation, which is in practice the most important as well as difficult case.

It is equally possible, however, to derive similar results for the mid-distance and for the at-infinity approximations to the Baumbach–Allen formula.

Without repeating here all calculations of the first derivative equaled to zero, we merely give the relevant results, leaving their full proof to the reader as an exercise.

Thus, for case stemming out of (8.44), the distance from the Sun center at which the electromagnetic waves pass in the mid-distance case is:

$$b_{\min-mid-distance}(\nu) = \frac{7^{\frac{1}{5}}\,\nu^{\frac{2}{5}}_{critical}(r_{Sun})\,r_{Sun}}{2^{\frac{1}{5}}\,\nu^{\frac{2}{5}}}. \tag{8.55}$$

The minimal focal distance in the mid-distance case is:

$$z_{\min-mid-distance}(\nu) = \frac{7^{\frac{7}{5}}\,\nu^{\frac{4}{5}}_{critical}(r_{Sun})\,r^2_{Sun}}{10 \cdot 2^{\frac{2}{5}}\,\nu^{\frac{4}{5}}\,r_g}. \tag{8.56}$$

The locus of the minimal focal distance in the (z, b) plane in the mid-distance case is again the parabola (with a different numerical coefficient with respect to (8.54))

$$z_{\min-mid-distance} = \frac{7}{10} \cdot \frac{b^2_{\min-mid-distance}}{r_g}. \tag{8.57}$$

Finally, in the at-infinity approximation—namely, for the case stemming from (8.45), the distance from the Sun center at which the electromagnetic waves pass is:

$$b_{\min-at-\infty}(\nu) = \frac{3\,\nu^2_{critical}(r_{Sun})\,r_{Sun}}{2\,\nu^2}. \tag{8.58}$$

The minimal focal distance in the at-infinity case is:

$$z_{\min-at-\infty}(\nu) = \frac{27\,\nu^4_{critical}(r_{Sun})\,r^2_{Sun}}{8\,\nu^4\,r_g}. \tag{8.59}$$

The locus of the minimal focal distance in the (z, b) plane in the at-infinity case is once more the parabola (again with a different numerical coefficient with respect to (8.54) and (8.57))

$$z_{\min-at-\infty} = \frac{3}{2} \cdot \frac{b^2_{\min-at-\infty}}{r_g}. \tag{8.60}$$

8.8 THE TWO CAUSTICS OF THE (GRAVITY + PLASMA) LENS OF THE SUN

A delightful exercise for calculus newcomers is the demonstration that, given the equation of a family of ∞^1 curves in the implicit form $f(x, y, \lambda) = 0$, this family of curves admits an *envelope* if, and only if, one is able to *eliminate the parameter* λ between the two simultaneous equations:

$$\begin{cases} f(x, y, \lambda) = 0 \\ \dfrac{\partial f(x, y, \lambda)}{\partial \lambda} = 0. \end{cases} \tag{8.61}$$

To prove that such an envelope exists for the family of close-Sun straight rays given by Equation (8.43), we first have to differentiate the implicit equation (8.43) with respect to the parameter b, getting:

$$z \cdot 15 \cdot \frac{1}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{16} + \frac{1}{2r_g} \cdot r - \frac{b}{r_g} = 0. \tag{8.62}$$

Both sides of the last equation may conveniently be multiplied by b to make it dimensionally consistent with the original equation of the family (8.43). One thus finds

$$z \cdot 15 \cdot \frac{b}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{16} + \frac{b}{2r_g} \cdot r - \frac{b^2}{r_g} = 0. \tag{8.63}$$

Finally, multiply the equation of the family (8.43) by 15 and expand

$$15z - 15 \cdot z \cdot \frac{r_{Sun}}{2r_g} \cdot \left(\frac{\nu_0}{\nu}\right)^2 \cdot 2952 \cdot \left(\frac{r_{Sun}}{b}\right)^{15} + \frac{15b}{2r_g} \cdot r - \frac{15b^2}{2r_g} = 0 \tag{8.64}$$

We are now just one step from proving the existence of the envelope or caustic. In fact, a simple summation of Equations (8.63) and (8.64) lets us get rid of the "obnoxious" term with exponent 15. One is thus just left with

$$15z + \left(\frac{1}{2} + \frac{15}{2}\right) \frac{b}{r_g} \cdot r - \left(1 + \frac{15}{2}\right) \cdot \frac{b^2}{r_g} = 0 \tag{8.65}$$

that is

$$15z + 8 \cdot \frac{b}{r_g} \cdot r - \frac{17}{2} \cdot \frac{b^2}{r_g} = 0. \tag{8.66}$$

This is the solvent equation for the envelope system (8.61) in the close-Sun approximation, and it is a simple second-degree algebraic equation in b. Solving it for b will yield two roots: if these roots are real, then the envelope (caustic) exists, if they are complex no caustic exists. Actually, solving (8.66) for b one gets the two roots

$$b_{1,2} = \frac{\pm\sqrt{2}\sqrt{255 r_g z + 32 r^2} + 8r}{17} \tag{8.67}$$

that are real and positive, so the two relevant caustics do exist. The equations of these two caustics are very complicated, and will not be given here. We wish to point out, however, a very basic and simple result: if one sets $r = 0$ into (8.66) and solves for z, the simple parabola is found

$$z = \frac{17}{30} \cdot \frac{b^2}{r_g}. \tag{8.68}$$

This is, of course, the *locus of the minimal focal distances in the (z, b) plane* for the gravitational lens of the Sun for all frequencies in the approximation, already found in Equation (8.54), thereby re-proving all our results!

Similarly, in the mid-distance case, it turns out that the solvent equation of the envelope system (8.61) is

$$5z + 3 \cdot \frac{b}{r_g} \cdot r - \frac{7}{2} \cdot \frac{b^2}{r_g} = 0 \tag{8.69}$$

whose roots are real. Setting $r = 0$ and solving for z, (8.57) is found again.

Finally, at-infinity the solvent equation reads

$$z + \frac{b}{r_g} \cdot r - \frac{3}{2} \cdot \frac{b^2}{r_g} = 0 \tag{8.70}$$

whose roots are real. Setting $r = 0$ and solving for z, (8.60) is found again.

8.9 OBSERVING FREQUENCIES FOR THE "CLOSE-SUN", "MID-DISTANCE", AND "AT-INFINITY" APPROXIMATIONS

As of the year 2001, no scientist seems to have derived an adequate mathematical model of the Solar Corona. This is because there are so many quantum and statistical phenomena taking place in the Corona that a full mathematical description turns out to be extremely complicated. Thus, one should not be surprised to find, in this final section to our Baumbach–Allen model of the Corona, that, despite all the approximations we have introduced thus far, there are more basic limitations in frequency for the "close-Sun", "mid-distance", and "at-infinity" cases.

As usual, we start from the "close-Sun" approximation case, since this is the case most relevant to the FOCAL Project. Consider again the "Focusing Condition" of the (gravity + plasma) lens, Equation (8.24). We already know the analytical expression of the function $\nu_{critical}(b)$ on the right-hand side, given by (8.20). But we don't seem to know any analytical expression for the observing frequency on the left-hand side yet.

Well, after the discussion of Section 8.7 we "nearly" know such an analytical expression in the close-Sun approximation: it is obtained by the inversion of (8.52)—namely, from (8.52) solved for ν, rather than for b.

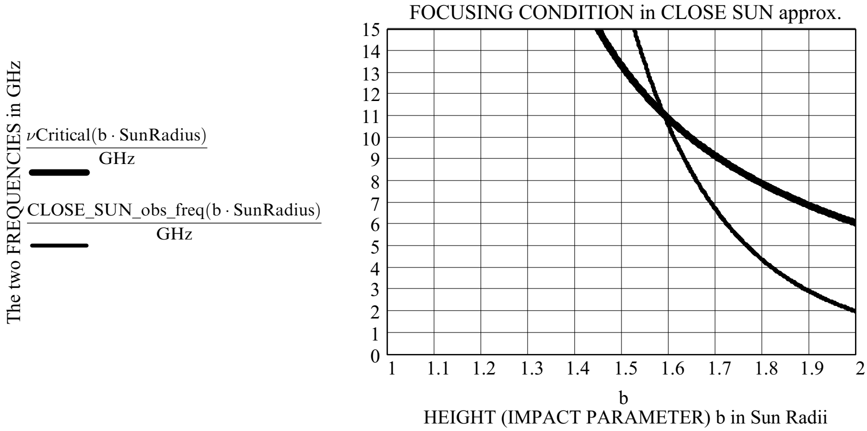


Figure 8.4. Focusing Condition in the close-Sun approximation. The horizontal axis shows the impact parameter b in units of the Sun radius. The vertical axis shows the frequency in GHz. The two curves shown cross at $b = 1.588804$ solar radii and $\nu = 11.0690005$ GHz. *Focusing only occurs for electromagnetic waves crossing the Corona at distances* ***less*** *than b = 1.588804 solar radii from the Sun center and frequencies higher than 11.0690005 GHz. The latter restriction* ***rules out*** *the hydrogen line.*

By doing so, one gets the *Observing Frequency in the Close-Sun approximation*:

$$\nu_{\min-close-Sun}(b) = \nu_{critical}(r_{Sun}) \cdot \sqrt{\frac{17}{2}} \cdot \left(\sqrt{\frac{r_{Sun}}{b}}\right)^{15}. \tag{8.71}$$

By replacing the last formula into the Focusing Condition (8.20), one finds the *Focusing Condition in the Close-Sun approximation*:

$$\nu_{\min-close-Sun}(b) > \nu_{critical}(b). \tag{8.72}$$

We plotted this inequality in Figure 8.4. The horizontal axis shows the impact parameter b in units of the Sun radius. The vertical axis shows the frequency in GHz. The two curves shown intersect at $b = 1.588804$ solar radii and $\nu = 11.0690005$ GHz. ***Focusing by the Sun only occurs*** *for electromagnetic waves crossing the Corona at* ***distances less*** *than* $b = 1.588804$ *solar radii from the Sun center and* ***frequencies higher than 11.0690005 GHz.***

This means that ***there is no close focus created by the Sun at the hydrogen line at 1.420 GHz.***

Unfortunately, this is ***very bad news for SETI !***

In fact, the hydrogen line at 1.420 GHz is the # 1 line of interest for SETI, and we have just found that the "focal length for SETI equals ... infinity!" Too bad, indeed. But this is in agreement with Von Eshleman's paper [2], who, by resorting to a mathematical treatment different from the one given in this book, foresaw a double caustic of the (gravity + plasma) Sun's lens to yield *a focus on the hydrogen line located light years away from the Sun!*

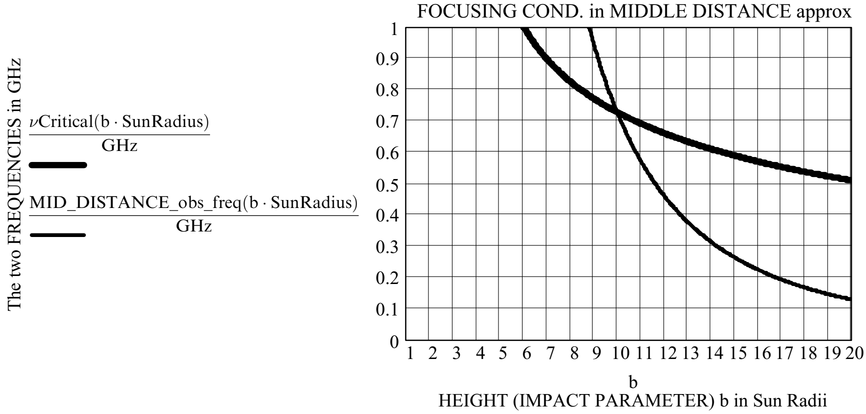


Figure 8.5. Intersection point between (8.20) and (8.73), showing that the *mid-distance* approximations hold for radial distances from the Sun higher than 10 solar radii, and for frequencies higher than 750 MHz.

Next, we turn to the mid-distance approximation, given by (8.55). On inversion, (8.55) yields the *Mid-Distance Observing Frequency*:

$$\nu_{\min-mid-distance}(b) = \nu_{critical}(r_{Sun}) \cdot \sqrt{\frac{7}{2}} \cdot \left(\sqrt{\frac{r_{Sun}}{b}}\right)^5. \tag{8.73}$$

By replacing the last formula into the Focusing Condition (8.20), one finds the *Focusing Condition in the Mid-Distance approximation*:

$$\nu_{\min-mid-distance}(b) > \nu_{critical}(b). \tag{8.74}$$

In analogy with the discussion following (8.72), we now numerically find the coordinates (b, ν) of the intersection point between the two curves given by (8.20) and (8.73). By doing so, one gets Figure 8.5, showing that, *for the mid-distance approximation to hold, the impact parameter b must be* ***less*** *than 10 solar radii, and the observing frequency ν must be higher than 0.75 GHz—that is, higher than 750 MHz.*

Finally, the "at-infinity" approximation starts from (8.58). By inverting it, one finds the *Observing Frequency at Infinity*:

$$\nu_{\min-at-\infty}(b) = \nu_{critical}(r_{Sun}) \cdot \sqrt{\frac{3}{2}} \cdot \sqrt{\frac{r_{Sun}}{b}} \tag{8.75}$$

and the *Focusing Condition at Infinity*:

$$\nu_{\min-at-\infty}(b) > \nu_{critical}(b). \tag{8.76}$$

Again we should numerically find the intersection point between the two curves given by (8.20) and (8.75). However, here we get a "surprise": even for values of the impact parameter b equal to *millions of AU*, the "at-infinity" frequency curve (8.75) keeps being higher than the critical frequency curve (8.20). This simply means that, *in the* *"at-infinity" approximation, the Baumbach–Allen model predicts* ***focusing*** *by the gravitational lens of the Sun for* ***all observing frequencies****.*

8.10 REFERENCES

[1] V. Eshleman, "Gravitational Lens of the Sun: Its Potential for Observations and Communications over Interstellar Distances," *Science*, **205** (1979), 1133–1135.

[2] V. Eshleman, "Non-Repeatable SETI Signals Due to Gravity-Plasma Lenses of Stars," in *Proceedings of the Fifth International Conference on Bioastronomy* (IAU Colloquium No. 161, "Astronomical and Biochemical Origins and the Search for Life in the Universe"), held at Capri, Italy, July 1–5, 1996, C. Cosmovici, S. Bowyer, and D. Werthimer (editors), Editrice Compositori, Bologna, Italy, pp. 741–745, 1997.

[3] C. W. Allen, "Interpretation of Electron Densities from Coronal Brightness," *Monthly Notices of the Royal Astronomical Society*, **107** (1947), 426–432.

[4] M. J. Aschwanden, and L.W. Acton, "Tomography of the Soft X-Ray Corona: Measurements of Electron Densities, Temperatures, and Differential Emission Measure Distributions above the Limb," a paper submitted to the *Astrophysical Journal* on May 4, 2000, and downloadable from the site *http://solar.physics.montana.edu/hypermail/eprint/0077.html*

[5] B. G. Andersson, and Slava G. Turyshev, "550 AU Mission: Science Rationale, Goals and Requirements," unpublished JPL internal report dated April 10, 1998.

[6] L. Iess, G. Giampieri, J. D. Anderson, and B. Bertotti, "Doppler measurement of the solar gravitational deflection," *Classical and Quantum Gravity*, **16** (1999), 1487–1502.

[7] A. Einstein, "The Foundation of the General Theory of Relativity," in H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, *The Principle of Relativity*, Dover, New York, 1952 reprint of the 1923 edition published by Methuen and Company. In particular see p. 163.

9

NASA's Interstellar Probe (ISP: 2010–2070?) and the Cosmic Microwave Background (CMB)

9.1 INTRODUCTION

This chapter is devoted to studying the exploitation of the Sun's gravity lens in a most realistic context—one that was opened up only recently (1999). Indeed, new *experimental* perspectives were opened by NASA's decision to set up an "Interstellar Probe Science and Technology Definition Team", of which the author of this chapter is a member. NASA's Interstellar Probe (ISP) is expected to be launched in June 2010 and reach 250 AU in just 15 years, being "pushed" out of the solar system by a 400-meter hexagonal solar sail plus a suitable flyby of the Sun. ISP will then explore the heliosphere with its termination shock and plasma (120 AU). The probe direction of exit from the solar system will be toward the heliopause's "nose"—namely, 16° in declination and 16.6 hours in right ascension. ISP will then keep going along this straight direction "forever", and will cross the 550 AU focal sphere of the Sun around the year 2040. Could we take advantage of this circumstance to prove experimentally for the first time that the Sun's focus is there? Yes—by letting ISP carry some photometer or bolometer capable of detecting the influx of the Cosmic Microwave Background (CMB) radiation focused there by the Sun. Also, apart from revealing the Sun's focus, valuable results for the science of cosmology would be achieved, inasmuch as the angular resolution of the CMB provided by IPS and the Sun's lens would be about nine orders of magnitude better than COBE's. To achieve these goals, in this chapter a study of the CMB observing conditions is made and we also take into account the effects of the Solar Corona whose electrons actually tend to "push" the effective focus of the Sun farther out than 550 AU. We prove one basic result: the *effective* minimal focal distance that ISP must reach to look at the CMB through the Sun lens is 763 AU (or 4.41 light-days)—of course, regardless of the direction of exit out of the solar system. If NASA's ISP could observe the CMB focused there around the year 2055, it would prove to be the first truly "interstellar precursor mission" of humankind.

9.2 NASA'S INTERSTELLAR PROBE (ISP) AND ITS LONG FLIGHT: 2010 TO 2055 ...

The direction of the incoming *interstellar wind* is -16° in declination and $+16.6$ hours in right ascension. This is going to be the direction of exit from the solar system for NASA's "Interstellar Probe" (ISP) (websites *http://umbra.nascom.nasa.gov/SEC/secr/missions/ISP.html* and *http://eis.jpl.nasa.gov/sec353/Interstellar.html*).

In 1999 NASA set up an ISP Science and Technology Definition Team (ISP STDT), under the coordination of Robert Mewaldt of Caltech, which made a 6-month study about ISP's goals and required technologies. The Team's conclusions were published in October 1999 at JPL [8]. ISP will hopefully be launched in June 2010, and will, of course, be an "interstellar precursor mission", rather than an interstellar mission proper. Nevertheless, ISP's name means that the probe will go as far as possible away from the Sun while still keeping the radio link with the Earth.

New advanced ideas in propulsion, and notably a hexagonal 400-meter solar sail plus a suitable Sun flyby, make the NASA planners expect that ISP will reach the nominal goal of 250 AU from the Sun in just 15 years (i.e., around 2025—assuming a 2010 launch). This distance of 250 AU is already well beyond where the heliosphere (the solar wind–dominated portion of space) ends with the heliopause, and interstellar space then begins. So, according to [8], all payloads aboard ISP will be devoted to study the heliosphere, the heliopause, and the (possibly existing) bow shock.

The author of this book, however, suggested to the ISP Team to put aboard ISP one more (small) instrument—a photometer or a bolometer or something similar—*to enable ISP to detect experimentally for the first time the physical existence of the Sun's gravitational lens*. In other words, the author suggested to the ISP Team an extension of ISP's flight to at least 550 AU and more. This suggestion is based on the obvious consideration that ISP will get to any distance anyway, sooner or later. So, the real question is whether the communication link between ISP and the Earth will be kept until ISP reaches and passes by the important, special distance where the Sun's gravity focus is located.

Thus, the idea of testing the Sun's gravity lens for the first time involves a further discussion about the *times* involved for the ISP's long flight. If 250 AU will be reached after 15 years of flight (i.e., in 2025), proceeding at *uniform* speed, ISP can be expected to cross the naked Sun's focal sphere at 550 AU around the year 2043. But to reach the *true* focal sphere "pushed outward" by the Sun's Corona, however, *ISP will need more time still, depending on observation frequency*.

Unfortunately, at this point of our discussion two more difficulties seem to arise:

(1) There is "nothing" (i.e., no quasar, no radiogalaxy, no very bright star) on the celestial sphere in the direction opposite to the incoming interstellar wind direction, toward which ISP moves. In other words, no "bright electromagnetic source" is going to focus its flux upon ISP by virtue of the Sun's mass.
(2) Even if we thought of correcting the ISP rectilinear trajectory so as to align the new trajectory to a bright source, this would practically be impossible because ISP would have to be aligned with a tolerance of a few tens of meters, far beyond

the tracking capabilities nowadays available for a spacecraft at 550 AU or beyond.

To circumvent both these difficulties, we change the target frequency.

9.3 LOOKING AT THE 2.728 K COSMIC MICROWAVE BACKGROUND THROUGH THE SUN'S GRAVITY LENS BY VIRTUE OF NASA'S INTERSTELLAR PROBE (ISP)

Rather than looking at quasars or bright stars and the like, we decide to look at the Cosmic Microwave Radiation (CMB) or Cosmic Background Radiation (CBR), the famous 2.728 K blackbody radiation that has been filling the Universe since the time of the decoupling between radiation and matter, about 300,000 years after the big bang. If we decide that ISP is going to look at the CMB through the Sun's gravity lens, then, *no longer would an ISP tracking problem exist because the CMB is (almost) uniformly scattered all over the celestial sphere, so that any direction of exit out of the solar system is fine.*

NASA's ISP Team already determined [8, p. 16] that ISP's antenna will be 2.7 meter in diameter keeping the link with Earth on the Ka band (i.e., at 32 GHz). Then, by replacing this value of 1.35 meters in antenna radius and an assumed 50% antenna efficiency into (1.16), one gets Table 9.1, showing the gain for the combined

Table 9.1. The gain (or magnification) of the naked Sun's gravity lens, the gain of a 2.7-meter spacecraft (S/C) antenna and the combined gain of the naked Sun + S/C antenna system. Numerical values are given for five frequencies suggested to be observed by NASA's Interstellar Probe (ISP).

Line	Neutral hydrogen	H_2O	Ka band	CMB_{max}	Positronium
Frequency ν	**1.42 GHz**	**22 GHz**	**32 GHz**	**160.378 GHz**	**203 GHz**
Wavelength λ	21.112 cm	1.363 cm	9.37 mm	1.06 mm	1.48 mm
S/C antenna beamwidth	5.474 deg	0.353 deg	0.243 deg	0.049deg	0.038 deg
Naked Sun gain	57.46 dB	69.36 dB	70.98 dB	77.96 dB	79.01 dB
2.7-meter antenna S/C gain	29.07 dB	52.87 dB	56.13 dB	70.07 dB	72.17 dB
Combined (naked Sun + S/C) gain	**86.53 dB**	**122.23 dB**	**127.11 dB**	**148.03 dB**	**151.18 dB**

system made up by the naked Sun's gravity lens plus the spacecraft antenna for the following five selected frequencies:

(1) The neutral hydrogen line (1,420 MHz).
(2) The water maser line (22 GHz).
(3) The Ka band frequency (32 GHz) that will be used by NASA's ISP for telecommunications with Earth.
(4) The peak of the 2.728 K Cosmic Microwave Background (CMB) radiation (160.378 GHz).
(5) The first line of the positronium (an "atom" made by a positron and an electron), which was suggested by Nikolai Kardashev in 1971 as the "best" line for interstellar communications in SETI.

One can see from Table 9.1 that, looking at the CMB through the naked Sun's gravity lens by virtue of NASA's ISP antenna, achieves the huge gain of 148.03 dB.

Having determined the total (naked Sun + ISP) gain, we are now going to estimate the total radiation influx expected to fall upon ISP when the spacecraft reaches the minimal focal distance from the Sun. Such an estimate is indeed predictable because the CMB spectrum is precisely the spectrum of a blackbody (Planck spectrum) whose temperature has been determined by NASA's COBE (Cosmic Background Explorer, launched in 1989, see, for instance, [9, pp. 111–113]) spacecraft as

$$T_{CMB} = 2.728\,\mathrm{K}. \tag{9.1}$$

Denoting Boltzmann's constant by k and Planck's constant by h, the well-known Planck spectrum (spectral energy density) is given by

$$\rho_\nu(\nu, T)\, d\nu = \frac{8\pi h}{c^3} \cdot \frac{\nu^3}{e^{\frac{h\nu}{kT}} - 1}\, d\nu \tag{9.2}$$

and is plotted in Figure 9.1 for the particular CMB temperature (9.1). The peak (maximum) of this function is found at the frequency

$$\nu_{\max} = \frac{2.82144k}{h\,\mathrm{K\,s}} \cdot T = \frac{5.83792 \cdot 10^{10} k}{\mathrm{K\,s}} \cdot T = 160.378\,\mathrm{GHz}. \tag{9.3}$$

This is Wien's displacement law for frequencies, and we have used the following numerical values for c, k, h (taken from [10])

$$c = 299{,}792{,}458\ \frac{\mathrm{m}}{\mathrm{s}}, \quad k = 1.380658 \cdot 10^{-23}\ \frac{\mathrm{joule}}{\mathrm{s}}, \quad h = 6.6260755 \cdot 10^{-34}\,\mathrm{joule} \cdot \mathrm{s} \tag{9.4}$$

The peak frequency (9.3) is, of course, found by setting the derivative of (9.2) at zero with respect to ν and solving the resulting numerical equation

$$e^x = \frac{3}{3 - x}. \tag{9.5}$$

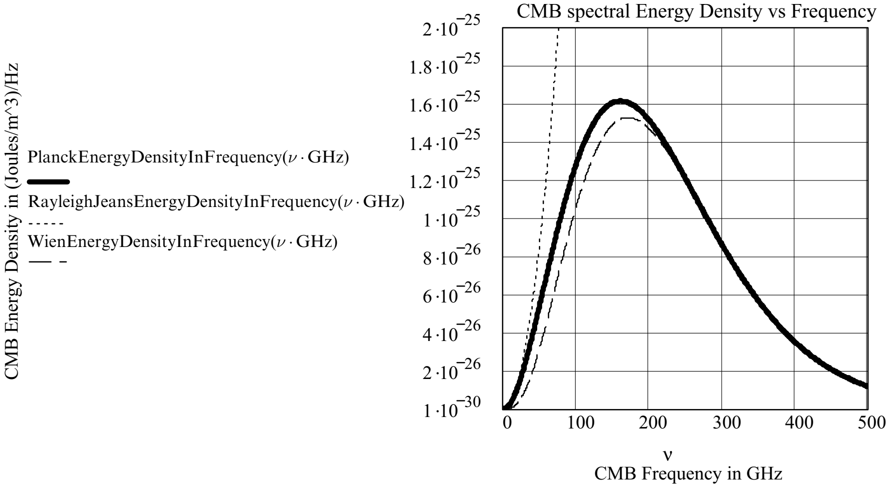


Figure 9.1. Planck's spectral energy density in frequency for the CMB (temperature $T = 2.728\,\text{K}$).

Replacing (9.3) back into (9.2), the numerical value of the Planck spectral energy density at the peak is found

$$\rho_\nu(\nu_{\max}, T_{CMB}) = \frac{8\pi h}{c^3} \cdot \frac{\nu_{\max}^3}{e^{\frac{h\nu_{\max}}{kT_{CMB}}} - 1} = 1.61355 \cdot 10^{-25}\,\frac{\text{joule}}{\text{m}^3\,\text{Hz}}. \tag{9.6}$$

The Planck brightness function (Figure 9.2) is the same as the spectral energy density, but with a different coefficient

$$B(\nu, T)\,d\nu = \frac{2h}{c^2} \cdot \frac{\nu^3}{e^{\frac{h\nu}{kT}} - 1}\,d\nu. \tag{9.7}$$

Consequently, the brightness peak frequency is the same as (9.3), but the function value, according to (9.7), is

$$B_\nu(\nu_{\max}, T_{CMB}) = \frac{2h}{c^2} \cdot \frac{\nu_{\max}^3}{e^{\frac{h\nu_{\max}}{kT_{CMB}}} - 1} = 3.849412 \cdot 10^{-18}\,\frac{\text{joule}}{\text{m}^2}. \tag{9.8}$$

The two physical laws (9.2) and (9.7) are now going to be expressed as functions of the wavelength λ, rather than of the frequency ν. This may appear as a trivial exercise in the first instance, but we wish to clear the way from a mistake that, unfortunately, is still fairly common. Namely, we wish to point that the peak wavelength $\lambda_{\max}$ is *not* related to the corresponding peak frequency $\nu_{\max}$ by the familiar relationship $c = \lambda \cdot \nu$. The reason of this apparent paradox is the hidden character of the Planck law as a *probability density*, rather than "just a function". This means that the differential $d\nu$ must always be taken along at all times (and not "forgotten", as usually happens) and, when a change of variable occurs (as in passing from

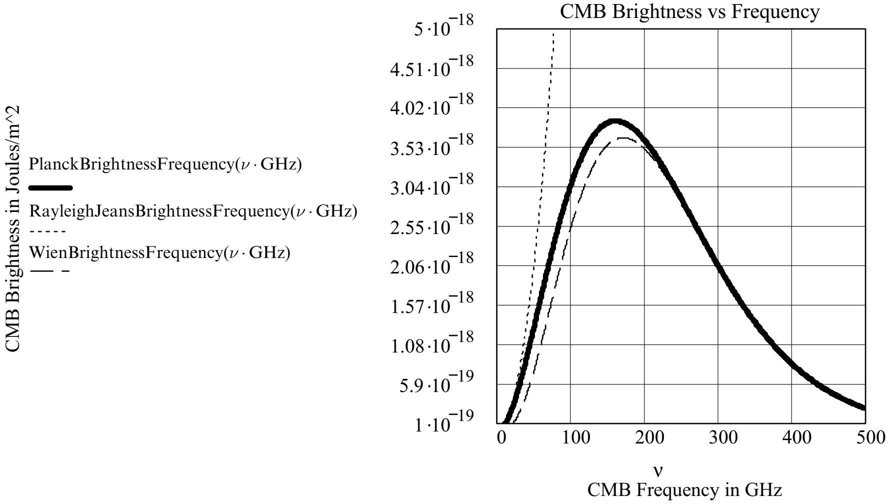


Figure 9.2. Planck's brightness in frequency for the CMB.

frequencies to wavelengths), probability densities multiply by the Jacobian of the coordinate transformation—namely, by the derivative of the transformation equation—if one is in just one dimension as in the case of the Planck spectrum. So, one has

$$d\nu = -\frac{c\,d\lambda}{\lambda^2} \tag{9.9}$$

and, apart from the axis reversal meant by the minus sign, (9.9) means that the *form* of the Planck spectrum in λ is *not* going to be the same as in ν. In fact, one finds, for the Planck spectral energy density and brightness in lambda, respectively

$$\rho_\lambda(\lambda, T)\,d\lambda = \frac{8\pi h}{\lambda^5}\cdot\frac{d\lambda}{e^{\frac{ch}{\lambda kT}} - 1} \tag{9.10}$$

and

$$B(\lambda, T)\,d\lambda = \frac{2hc^2}{\lambda^5}\cdot\frac{d\lambda}{e^{\frac{ch}{\lambda kT}} - 1}. \tag{9.11}$$

The corresponding two plots are, for (9.10) and (9.11), respectively given in Figures 9.3 and 9.4.

To compute the peak wavelength $\lambda_{\max}$, one merely has to set to zero the derivative of (9.10) or (9.11), yielding the same result, since (9.10) and (9.11) only differ by a constant factor. This common result is the root of the equation

$$e^x = \frac{5}{5 - x}. \tag{9.12}$$

As this numerical equation is different from the numerical equation (9.5), its root is of course different from that of (9.5). The result is that the equation corresponding to

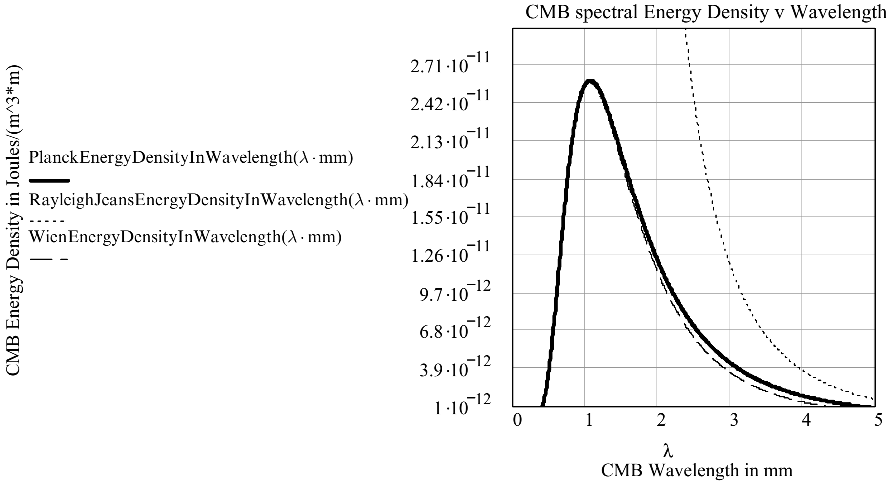


Figure 9.3. Planck's spectral energy density in wavelengths for the CMB.

(9.3) but for the wavelength is (Wien's displacement law for wavelengths)

$$\lambda_{\max} = \frac{0.20141 \cdot c \cdot h}{k \cdot T} = \frac{2.897824 \cdot 10^{-3} \cdot \mathrm{m} \cdot \mathrm{K}}{T} = 1.0622522\,\mathrm{mm}. \quad (9.13)$$

Now, Equations (9.3) and (9.13) are *not* related by $c = \lambda \cdot \nu$ because of the way they were derived. Were they related by $c = \lambda \cdot \nu$, one would get from (9.3) the *wrong* peak

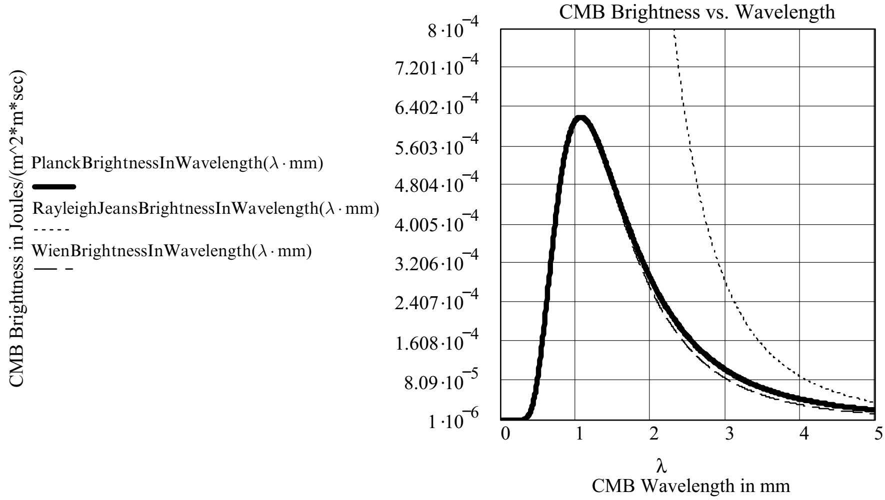


Figure 9.4. Planck's brightness in wavelengths for the CMB.

CMB wavelength

$$\textbf{wrong!!!} \quad \lambda_{\max} = 1.8695777\ \text{mm} \quad \textbf{wrong!!!} \tag{9.14}$$

and from (9.13) the *wrong* peak CMB frequency

$$\textbf{wrong!!!} \quad \nu_{\max} = 282.2254328\ \text{GHz} \quad \textbf{wrong!!!} \tag{9.15}$$

which are still claimed by some as the "correct" peak CMB wavelength and frequency, respectively.

The peak values for Planck's spectral energy density and brightness are obtained by replacing the peak wavelength (9.13) into (9.10) and (9.11), respectively, and they are found to be

$$\rho_\lambda(\lambda_{\max}, T_{CMB}) = \frac{8\pi h}{\lambda_{\max}^5} \cdot \frac{1}{e^{\frac{ch}{\lambda_{\max} k T_{CMB}}} - 1} = 2.5938759 \cdot 10^{-11} \cdot \frac{\text{joule}}{\text{m}^4} \tag{9.16}$$

and

$$B(\lambda_{\max}, T_{CMB}) = \frac{2hc^2}{\lambda_{\max}^5} \cdot \frac{1}{e^{\frac{ch}{\lambda_{\max} k T_{CMB}}} - 1} = 6.1881386 \cdot 10^{-4} \cdot \frac{\text{joule} \cdot \text{Hz}}{\text{m}^3} \tag{9.17}$$

9.4 THE EFFECTIVE MINIMAL FOCAL DISTANCE FOR THE GRAVITY + PLASMA LENS LOOKING AT THE 2.7 K COSMIC MICROWAVE BACKGROUND IS 763 AU, WHICH NASA'S INTERSTELLAR PROBE WILL REACH IN 2055

When is NASA's ISP going to reach the effective minimal focal distance from the Sun at the CMB frequency? The answer is: in the year 2055, because the *effective* minimal focal distance at the CMB peak frequency is ... 763 AU. This is the theorem that we are going to prove in this section.

Let us start by recalling from point (8) in Section 9.1 that the *effective* minimal focal distance from the Sun is the distance when the Sun's Coronal effects are taken into account. In other words, we wish to compute the effective minimal distance from the Sun that ISP must reach in order to look at the CMB peak of 160 GHz. Let us start by noticing that (8.22) yields essentially two cases.

(1) The spacecraft observing frequency is *higher* than the "no-lensing" value of 122.361 GHz given by (8.21). This is our case of interest, since the CMB peak occurs at 160 GHz.
(2) The spacecraft observing frequency is lower than 122.361 GHz. Then, according to Eshleman [3], one gets the complicated behavior expressed by the *double caustic* described (in a qualitative fashion only) in [3]. This case is beyond the scope of the present chapter and will not be discussed here.

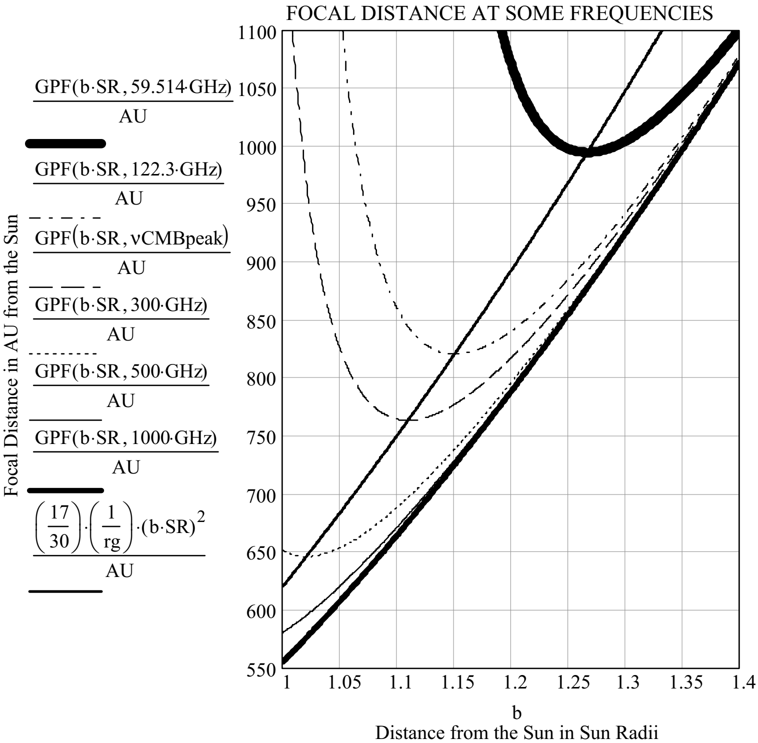


Figure 9.5. Plots of Equation (8.40). This is the (Gravity + Plasma) Sun lens focal distance F (in AU) as a function of the impact parameter b (in units of the Sun radius) for all radio waves focused by the Sun above 120 GHz. The impact parameter b is the distance from the Sun center at which the radio waves flyby the Sun and then proceed to focus at distance F from the Sun. Each curve corresponds to radio waves of a different frequency increasing from 59.514 GHz (top, thick curve) to 1,000 GHz (bottom thick curve). The shifting of the minimum is obvious, and the "locus" of all these minima is the parabola of Equations (8.54) and (9.21).

To see what happens to the effective minimal focal distance for a frequency higher than 122.361 GHz, let us plot (8.40) as in Figure 9.5.

From Figure 9.5 it is clear that

(1) On the horizontal axis we have the "impact parameter" b (the name comes from particle physics), defined by

$$b(r) = \frac{r}{r_{Sun}}. \tag{9.18}$$

The impact parameter is thus the shortest distance from the Sun's center (here measured in units of the Sun radius) at which radio waves passing by graze the Sun's surface. Clearly, the higher the impact parameter, the better it is, for then the ongoing radio waves travel less amid the Sun flames (Sun's Corona Plasma).

(2) On the vertical axis we have the focal distance (in AU) at which are focused just those radio waves that graze the Sun with impact parameter b. Obviously $b > 1$, but Figure 9.5 reveals that, in practice, b keeps well below two Sun radii.

(3) The second (dash–dot) curve from the top of Figure 9.5 corresponds to radio waves having the "critical" frequency of $\nu_{critical}(r_{Sun}) \cong 122.361$ GHz, at which no more focusing occurs because the outward deflection due to the coronal plasma counterbalances the inner deflection due to gravity. *We are only interested in radio waves having a frequency higher than* $\nu_{critical}(r_{Sun}) \cong 122.361$ GHz *and the CMB peak frequency of 160 GHz, shown by the third (dashed) curve from the top of Figure 9.5*. Then the crucial question arises: For which value of the impact parameter b is the minimal focal distance achieved along this third curve from the top?

(4) The remaining, last three curves from the bottom correspond to the observing frequencies of 300 GHz, 500 GHz, and 1,000 GHz, selected just for reference: we see that *the effective minimal focal distance decreases for increasing frequencies*, and nearly reaches the "gravity only" value of 550 AU for 1,000 GHz or more.

Going now a little beyond the work done in [7], we seek to compute the *coordinates of the minimum for all curves* on Figure 9.5. This is done by setting at zero the first derivative of (8.22) with respect to the impact parameter b. From (8.52) we know that the impact parameter (= abscissa) of the minimum is found for any frequency

$$b_{\min}(\nu) = \frac{17^{\frac{1}{15}} \nu_{critical}^{\frac{2}{15}}(r_{Sun})\, r_{Sun}}{2^{\frac{1}{15}} \nu^{\frac{2}{15}}}. \tag{9.19}$$

Also, from (8.53) we know that *the expression of the effective minimal focal distance is (for all frequencies higher than the critical frequency)*

$$z_{\min}(\nu) = \frac{17^{\frac{17}{15}} \nu_{critical}^{\frac{4}{15}}(r_{Sun})\, r_{Sun}^2}{30 \cdot 2^{\frac{2}{15}} \nu^{\frac{4}{15}} r_g}. \tag{9.20}$$

The equation of the geometric locus of all the minima can now be found by eliminating the frequency ν between (9.19) and (9.20). The simple result is the parabola

$$z_{\min}(b_{\min}) = \frac{17}{30} \cdot \frac{b_{\min}^2}{r_g} \tag{9.21}$$

which is depicted on Figure 9.5 as the slightly curved branch of parabola crossing all curves at just their minimum (bottom curve on the legenda on the left).

We now go back to our blackbody radiation and replace its peak frequency value (9.3) into (9.20) to find the effective minimal focal distance that ISP must reach in order to "catch up" with the CMB focused there by the Sun

$$\begin{aligned} z_{\min}(\nu_{\max}) &= 763.478 \text{ AU} \approx 4.41 \text{ light-days} \\ &\approx 19.5 \text{ times the Sun-to-Pluto distance.} \end{aligned} \tag{9.22}$$

This is the most important result derived in this chapter. It shows that, for NASA's ISP, as well as for any other spacecraft to come that intend to observe the CMB through the Sun's lens, the minimal effective distance from the Sun that must be reached is about 763 AU. When will NASA's ISP get there? A simple proportion shows that, assuming launch in 2010, ISP will get at 763 AU around the year 2055—namely, 45 years after launch.

Finally, (9.19) yields the impact parameter value for the CMB radiation focused at 763 AU: it equals 1.11 times the radius of the Sun—namely, the CMB radiation flybys the Sun at 76,220 km above its surface. This is a short distance from the Sun, and, unfortunately, coronal fluctuations there are still pretty strong.

Were we to "deconvolve" the image of a planet, or a stellar system, or a radio galaxy located on the other side of the Sun, at just 1.11 radii from the Sun it would be a difficult problem. But, fortunately, in the CMB there is "nothing to deconvolve". Yet, in Section 9.5 we are going to compute the (theoretical) angular resolution of the CMB provided by the naked Sun's gravity lens according to Equation (1.20).

Finally, we wish to point out that the *first* solid, thick curve from the top on Figure 9.5, corresponding to a frequency of 59.514 GHz, is the critical frequency provided by Equation (8.20) for an impact parameter b equal to the CMB one—namely, 1.11. This shows that the CMB frequency (160 GHz) is so far from the corresponding critical frequency (59.514 GHz) that we need not worry at all about the double caustic problems hinted in [3].

9.5 IMPROVING COBE'S ANGULAR RESOLUTION BY NINE ORDERS OF MAGNITUDE BY LOOKING AT THE 2.7 K COSMIC MICROWAVE BACKGROUND BY VIRTUE OF NASA'S INTERSTELLAR PROBE

The CMB is integrated light from all stars and galaxies that cannot be resolved into individual objects. So, strictly speaking, it is meaningless to speak of "angular resolution" provided the Sun's gravity lens when the latter is used to look at the CMB, simply because there is nothing to resolve in this case. Nevertheless, some meaning to "angular resolution" can be retained by adopting (1.20) even if for the CMB this is just "theoretical". The result is that the improvement in the (theoretical) angular resolution of the CMB as seen by the Sun's gravity lens, rather than through COBE, is about nine orders of magnitude (as shown in Table 9.2).

In the *NASA ISP Booklet* [8, p. 14] one reads: "NASA's Cosmic Background Explorer (COBE) detected the cosmic infrared background at wavelengths beyond 140 microns and established limits on the energy released by all stars since the beginning of time. Also, by observing the cosmic infrared background it is possible to determine how much energy was converted into photons during the evolution of galaxies, back to the time of their formation. Fundamental measurements about galaxy formation can be made even though individual protogalaxies cannot be seen. The cosmic infrared background spectrum can reveal how first stars formed and how early the elements were formed by nucleosynthesis".

Table 9.2. The (theoretical) angular resolution provided by the Sun's gravity lens for ISP, having a 2.7-meter antenna. Values are given for the five selected frequencies.

Line	Neutral hydrogen	H_2O	Ka band	CMB_{max}	Positronium
Frequency ν	**1.42 GHz**	**22 GHz**	**32 GHz**	**159.25 GHz**	**203 GHz**
Wavelength λ	21.112 cm	1.363 cm	9.37 mm	1.06 mm	1.48 mm
Angular resolution at 550 AU	$6.3458 \cdot 10^{-6}$ arcsec	$4.0959 \cdot 10^{-7}$ arcsec	$2.8159 \cdot 10^{-7}$ arcsec	$5.6584 \cdot 10^{-8}$ arcsec	$4.4389 \cdot 10^{-8}$ arcsec
Angular resolution at 800 AU	$5.2267 \cdot 10^{-6}$ arcsec	$3.3736 \cdot 10^{-7}$ arcsec	$2.3194 \cdot 10^{-7}$ arcsec	$4.6606 \cdot 10^{-8}$ arcsec	$3.6561 \cdot 10^{-8}$ arcsec
Angular resolution at 1,000 AU	$4.6749 \cdot 10^{-6}$ arcsec	$3.0174 \cdot 10^{-7}$ arcsec	$2.0745 \cdot 10^{-7}$ arcsec	$4.1685 \cdot 10^{-8}$ arcsec	$3.2701 \cdot 10^{-8}$ arcsec

Perhaps the "virtual" angular resolution data given in Table 9.2 have a deeper significance, which escapes us at this time. Understanding better what "watching at the CMB through the Sun's gravity lens means" is a current research problem.

9.6 CONCLUSIONS

We have sought to prove that looking at the CMB through the Sun's gravity lens is a task suitable for the intended NASA's Interstellar Probe. Not only would this prove the existence of the Sun's focusing effects experimentally for the first time, but it would also contribute an unprecedented CMB angular resolution to the science of cosmology, even if only on a very narrow region of the sky. More important still, looking at the CMB is the only observation through the Sun's gravity lens that is totally independent of the probe's exit direction out of the solar system, inasmuch as the CMB is uniformly distributed over the celestial sphere. Finally, we have proved that the effective minimal focal distance that ISP must reach is 763 AU because of the strong refractions caused by the Solar Corona on the electromagnetic waves at the CMB peak frequency of 160 GHz. If NASA extends ISP's nominal mission from 400 AU to 1,000 AU, thus giving ISP the possibility of looking at the CMB, NASA's Interstellar Probe will become the first really "precursor interstellar mission" in the history of humankind.

9.7 ACKNOWLEDGMENTS

The author is indebted to Prof. Von R. Eshleman of Stanford University for introducing him to the problems of the plasma lens of the Sun many years ago.

More recently, the wonderful work done at JPL by several JPL-ers on the 550 AU mission was brought to his attention, and especially the work by Drs. John West, Slava Turyshev, Barbara Wilson, Faiza Lansing, Scot Stride, Bud Lovick, Neville Marzwell, and Macgregor Reid (just to mention a few), to all of whom the author is very grateful. Finally, Drs. Jim Ling, Robert Mewaldt, and Paulett Liewer allowed the author to become a member of NASA's ISP Team, thus opening up to him an entire new perspective, out of which the idea was born of looking at the CMB through the gravitational lens of the Sun by virtue of NASA's ISP.

The content of this chapter was first published by the author in 2000, as an Academy Transactions Note published in *Acta Astronautica* [11].

9.8 REFERENCES

[1] C. Maccone, "Tethered System to Get Magnified Radio Pictures of the Galactic Center from a Distance of 550 AU," IAA Transactions Note, *Acta Astronautica*, **45** (1999), 109–114.

[2] C. Maccone, *The Sun as a Gravitational Lens: Proposed Space Missions*, IPI Press, Colorado Springs, CO, Second Edition, 1999, ISBN 1-880930-12-9.

[3] V. R. Eshleman, "Non-Repeatable SETI Signals Due to Gravity-Plasma Lenses of Stars," in *Astronomical and Biochemical Origins and the Search for Life in the Universe*, C. B. Cosmovici, S. Boyer, and D. Werthimer (editors), Editrice Compositori, Bologna, Italy, pp. 741–745, 1997.

[4] E. Herlt, and H. Stephani, "Diffraction of a Plane Electromagnetic Wave at a Schwarzschild Black Hole," *International Journal of Theoretical Physics*, **12** (1975), 81–93.

[5] E. Herlt, and H. Stephani, "Wave Optics of the Spherical Gravitational Lens. Part 1: Diffraction of a Plane Electromagnetic Wave by a Large Star," *International Journal of Theoretical Physics*, **15** (1976), 45–65.

[6] J. L. West, "Design Issues for a Mission to Exploit the Gravitational Lens of the Sun at 550 AU," *Acta Astronautica*, **44** (1999), 99–107.

[7] B. G. Andersson, and Slava G. Turyshev, "550 AU Mission: Science Rationale, Goals and Requirements," unpublished JPL internal report dated April 10, 1998.

[8] R. Mewaldt, and P. Liewer, *Interstellar Probe*, NASA's Interstellar Probe Science and Technology Definition Team Booklet, published by JPL in October 1999.

[9] M. Roos, *Introduction to Cosmology*, 2nd Edition, John Wiley & Sons, New York, 1997.

[10] D. R. Lide (ed.), *CRC Handbook of Chemistry and Physics*, 79th Edition, CRC Press, 1991.

[11] C. Maccone, "Cosmic Microwave Background and Gravitational Lens of the Sun," Academy Transactions Note published in *Acta Astronautica*, **46** (2000), 605–614.

Part II

KLT-optimized telecommunications

10

A simple introduction to the KLT (Karhunen–Loève Transform)

10.1 INTRODUCTION

This chapter is a simple introduction about using the Karhunen–Loève Transform (KLT) to extract weak signals from noise of any kind. In general, the noise may be colored and over wide bandwidths, and not just white and over narrow bandwidths. We show that the signal extraction can be achieved by the KLT more accurately than by the Fast Fourier Transform (FFT), especially if the signals buried into the noise are very weak, in which case the FFT fails. This superior performance of the KLT happens because the KLT of any stochastic process (both stationary and non-stationary) is defined from the start over a *finite* time span ranging between 0 and a final and *finite* instant T (contrary to the FFT, which is defined over an infinite time span). We then show mathematically that the series of all the eigenvalues of the autocorrelation of the (noise + signal) may be differentiated with respect to T yielding the "Final Variance" of the stochastic process $X(t)$ in terms of a sum of the first-order derivatives of the eigenvalues with respect to T. Finally, we prove that this new result leads to the immediate reconstruction of a signal buried into the thick noise. We have thus put on a strong mathematical foundation a set of very important practical formulae that can be applied to improve SETI, the detection of exoplanets, asteroidal radar, and also other fields of knowledge like economics, genetics, biomedicine, etc. to which the KLT can be equally well applied with success.

10.2 A BIT OF HISTORY

The Karhunen–Loève Transform (KLT) is the most advanced mathematical algorithm available in the year 2008 to achieve both noise filtering and data compression in processing signals of any kind.

It took about two centuries (~1800–2000) for mathematicians to create such a jewel of thought little by little, piece after piece, paper after paper. It is thus difficult to recognize who did what in building up the KLT and at the same time be fair in attributing each individual advance to the appropriate author. In addition, mathematicians, both pure and applied, often speak such a ''clumsy'' language of their own that even learned scientists sometimes find it hard to understand them. This unfortunate situation hides the esthetic beauty of many mathematical discoveries that were often historically made by their authors more for the joy of opening new lines of thought than for the sake of any immediate application to science and engineering.

In essence, the KLT is a rather new mathematical tool used to improve our understanding of physical phoenomena, far superior to the classical Fourier Transform (FT). The KLT is named for two mathematicians—the Finnish actuary Kari Karhunen (1915–1992) [1] and the French American Michel Loève (1907–1979) [2, 3]—who proved, independently and about the same time (1946), that the series (2) hereafter is *convergent*. Put this way, the KLT looks like a purely mathematical topic, but really this is hardly the case. As early as 1933 the American statistician and economist Harold Hotelling (1895–1973) used the KLT (for discrete time, rather than for continuous time), so that the KLT is sometimes called the ''Hotelling Transform''. Even much earlier than these three authors the Italian geometer Eugenio Beltrami (1835–1899) discovered as early as 1873 the SVD (Singular Value Decomposition), which is closely related to the KLT in that area of applied mathematics nowadays called Principal Components Analysis (PCA). Unfortunately, a complete historical account about how these contributions developed since 1865—when the English mathematician Arthur Cayley (1821–1895) ''invented'' matrices—simply does not exist. We only know about ''fragments of thought'' that impair an overall vision of both the PCA and the KLT.

In Sections 10.3–10.5, we'll derive *heuristically and step-by-step* the many equations that make up for the KLT. We think that this approach is much easier to understand for beginners than what is found in most ''pure'' mathematical textbooks, and hope that the readers will appreciate our effort to explain the KLT as easily as possible to non-mathematically trained people.

10.3 A HEURISTIC DERIVATION OF THE KL EXPANSION

We start by saying that the KLT was born during the years of World War Two out of the need to merge two different areas of classical mathematics.

(1) The expansion of a deterministic periodic signal $x(t)$ into a basis of orthonormal functions (sines and cosines, in this case), typified by the classical Fourier series—first put forward by the French mathematician Jean Baptiste Joseph Fourier (1768–1830) around 1807,

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n \cos(nt) + b_n \sin(nt)] \qquad (-\pi \le t \le \pi). \tag{10.1}$$

(2) The need to extend this too narrow and deterministic view to probability and statistics. The much larger variety of phenomena called "noise" by physicists and engineers will thus be encompassed by the new transform. This enlarged view means considering a random function $X(t)$ (notice that we denote random quantities by capitals, and that $X(t)$ is also called a "stochastic process of the time"). We now seek to expand this stochastic process onto a set of orthonormal functions $\phi_n(t)$ according to the starting formula

$$X(t) = \sum_{n=1}^{\infty} Z_n \, \phi_n(t) \tag{10.2}$$

which is called the *Karhunen–Loève (KL) expansion of* $X(t)$ *over the finite time interval* $0 \le t \le T$.

What are then the Z_n and the $\phi_n(t)$ in (10.2)? To find out, let us start by recalling what "orthonormality" means for the Fourier series (10.1). Leonhard Euler (1707–1783) had already laid the first stone towards the Fourier series (10.1) by proving that, if $x(t)$ is assumed to be periodic over the time interval $-\pi \le t \le \pi$, then the coefficients a_n and b_n in (10.1) are obtained from the known function (or "periodic signal") $x(t)$ by virtue of the equations ("Euler formulae"):

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\cos(nt)\,dt \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\sin(nt)\,dt. \tag{10.3}$$

If the same result is going to be true for the Karhunen–Loève expansion, the functions of the time, $\phi_n(t)$ in (10.2) must be orthornormal (i.e., both orthogonal and normalized to 1). That is,

$$\int_0^T \phi_m(t)\,\phi_n(t)\,dt = \delta_{mn} \tag{10.4}$$

where the δ_{mn} are the Kronecker symbols, defined by $\delta_{mn} = 0$ for $m \ne n$ and $\delta_{nn} = 1$.

But what then are the Z_n appearing in (10.2)? Well, a random function $X(t)$ can be thought of as something made up of two parts: its behavior in time, represented by the functions $\phi_n(t)$, and its behavior with respect to probability and statistics, which must therefore be represented by the Z_n. In other words, the Z_n must be random variables not changing in time (i.e., "just" random variables and not stochastic processes). By doing so we have actually made one basic, new step ahead: we have found that the KLT separates the probabilistic behavior of the random function $X(t)$ from its behavior in time, a kind of "untypical" separation that is achieved nowhere else in mathematics!

Having discovered that the Z_n are random variables, some trivial consequences follow at once. Let us denote by $E\{\ \}$ the linear operator yielding the average of a random variable or stochastic process. If one takes the average of both sides of the KL expansion (10.2), one then gets (we "freely" interchange here the average operator

$E\{\ \}$ with the infinite summation sign, bypassing the complaints of "subtle" mathematicians!)

$$E\{X(t)\} = \sum_{n=1}^{\infty} E\{Z_n\}\phi_n(t). \tag{10.5}$$

Now, it is not restrictive to suppose that the random function $X(t)$ has a zero mean value in time—namely, that the following equation is identically true for all values of the time t within the interval $0 \leq t \leq T$:

$$E\{X(t)\} \equiv 0. \tag{10.6}$$

In fact, were this not the case, one could replace $X(t)$ by the new random function $X(t) - E\{X(t)\}$ in all the above calculations, thus reverting to the case of a new random function with zero mean value. Thus, in conclusion, the random variables Z_n too must have a zero mean value

$$E\{Z_n\} \equiv 0. \tag{10.7}$$

This equation has a simple consequence: since the variance $\sigma_{Z_n}^2$ of the random variables Z_n is given by

$$\sigma_{Z_n}^2 = E\{Z_n^2\} - E^2\{Z_n\} \tag{10.8}$$

by inserting (10.7) into (10.8) we get

$$\sigma_{Z_n}^2 = E\{Z_n^2\}. \tag{10.9}$$

At this point, we can make a further step ahead, that has no counterpart in the classical Fourier series: we wish to introduce a new sequence of positive numbers λ_n such that every λ_n is the variance of the corresponding random variable Z_n, that is

$$\sigma_{Z_n}^2 = \lambda_n = E\{Z_n^2\} > 0. \tag{10.10}$$

This equation provides the "answer" to the next "natural" question: Do the random variables Z_n fulfill a new type of "orthonormality" somehow similar to what the classical orthonormality (10.4) is for the $\phi_n(t)$? Since we are talking about random variables, the "orthogonality operator" can only be understood in the sense of *statistical independence*. The integral in (10.4) must then be replaced by the average operator $E\{\ \}$ for the random variables Z_n. In conclusion, we found that the random variables Z_n must obey the important equation

$$E\{Z_m Z_n\} = \lambda_n \, \delta_{mn}. \tag{10.11}$$

In this equation, we were forced to introduce the positive λ_n in the right-hand side in order to let (10.11) reduce to (10.10) in the special case $m = n$.

As for the KL equivalent of the Euler formulae (10.3) of the Fourier series, from the KL series (10.2) and the orthonormality (10.4) of the $\phi_n(t)$ one immediately finds that

$$Z_n = \int_0^T X(t)\, \phi_n(t)\, dt. \tag{10.12}$$

In other words: the random variables Z_n are obtained from the given stochastic process $X(t)$ by "projecting" this $X(t)$ over the correspoding eigenvector $\phi_n(t)$. If one likes the language of mathematicians and of quantum physics, then one may say that this projection of $X(t)$ onto $\phi_n(t)$ occurs in the "Hilbert space", which is the infinitely dimensional Euclidean space spanned by the eigenvectors $\phi_n(t)$ so that the square of $\phi_n(t)$ is integrable over the finite time span $0 \leq t \leq T$.

To sum up, we have actually achieved a remarkable generalization of the Fourier series by defining the Karhunen–Loève expansion (10.2) as the only possible statistical expansion in which all the expansion terms are *uncorrelated* from each other. This word "uncorrelated" comes from the fact that the *autocorrelation* of a random function of the time, $X(t)$, is defined as the mean value of the product of $X(t)$ at two different instants t_1 and t_2:

$$R_{XX}(t_1, t_2) \equiv R_X(t_1, t_2) = E\{X(t_1)X(t_2)\}. \tag{10.13}$$

If we assume, according to (10.6), that the mean value of $X(t)$ vanishes identically in the interval $0 \leq t \leq T$, the autocorrelation (10.13) reduces to the variance of $X(t)$ when the two instants are the same

$$\sigma^2_{X(t)} = E\{X^2(t)\} = E\{X(t)X(t)\} = R_X(t, t). \tag{10.14}$$

Let us add one final remark about the basic notion of statistical independence of the random viariables Z_n. It can be proven that, while the Z_n in (10.2) always are uncorrelated (by construction), they also are statistically independent if they are Gaussian-distributed random variables. This is fortunately the case for the Brownian motion and for the background noise we face in SETI. So we are not concerned about this subtle mathematical distinction between uncorrelated and statistically independent random variables.

10.4 THE KLT FINDS THE BEST BASIS (EIGEN-BASIS) IN THE HILBERT SPACE SPANNED BY THE EIGENFUNCTIONS OF THE AUTOCORRELATION OF *X(t)*

Up to this point, we have not given any hint about how to find the orthonormal functions of the time, $\phi_n(t)$, and positive numbers λ_n (i.e., the variances of the corresponding uncorrelated random variables Z_n). In this section, we solve this problem by showing that the $\phi_n(t)$ are the eigenfunctions of the autocorrelation $R_X(t_1, t_2) = E\{X(t_1)X(t_2)\}$ and that the λ_n are the corresponding eigenvalues. This is the correct mathematical phrasing of what we are going to prove. However, in order to ease the understanding of the further maths involved hereafter, a "translation" into the language of "common words" is now provided. Consider an object—for instance, a book—and a three-axes rectangular reference frame, oriented in an arbitrary fashion with respect to the book. Then, the classical Newtonian mechanics shows that all the mechanical properties of the book are described by a 3×3 symmetric matrix called the "inertia matrix" (or, more correctly, "inertia tensor") whose elements are, in general, all different from zero. Handling a matrix whose elements are all nonzero is obviously more complicated than handling a matrix where all entries are zeros except for those on the main diagonal (i.e., a *diagonal matrix*). Thus, one may be led to wonder whether a certain transformation of axes exists that changes the inertia matrix of the book into a diagonal matrix. Newtonian mechanics shows then that only *one* privileged orientation of the reference frame with respect to the book exists yielding a *diagonal* inertia matrix: the three axes must then coincide with a set of three axes (parallel to the book edges) called "principal axes" of the book, or "eigenvectors" or "proper vectors" of the inertia matrix of the book. In other words, each body posesses an intrinsic set of three rectangular axes that describes at best its dynamics (i.e., in the most concise form). This was proven again by Euler, and one can always compute the position of the eigenvectors with respect to a generic reference frame by means of a certain mathematical procedure called "finding the eigenvectors of a square matrix".

In a similar fashion, one can describe any stochastic process $X(t)$ by virtue of the statistical quantity called the autocorrelation (or simply the correlation), defined as the mean value of the product of the values of $X(t)$ at two different instants t_1 and t_2, and formally written $E\{X(t_1)X(t_2)\}$. The autocorrelation, obviously symmetric in t_1 and t_2, plays for the stochastic process $X(t)$ just the same role as the inertia matrix for the book example above. Thus, if one first seeks the eigenvectors of the correlation, and then changes the reference frame over to this new set of vectors, one achieves the simplest possible description of the whole (signal + noise) set.

Let us now translate the whole above description into equations. First of all, we must express the autocorrelation $E\{X(t_1)X(t_2)\}$ by virtue of the KL expansion (10.2). This goal is achieved by writing down (10.2) for two different instants, t_1 and t_2, taking the average of their product, and then (freely) interchanging the average and the summations in the right-hand side. The result is

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \phi_m(t_1)\phi_n(t_2)\, E\{Z_m Z_n\}. \tag{10.15}$$

Taking advantage of the statistical orthogonality of the Z_n, given by (10.11), (10.15) simplifies to

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty} \lambda_m \phi_m(t_1)\phi_m(t_2). \tag{10.16}$$

Finally, we now want to let the $\phi_n(t)$ "disappear" from the right-hand side of (10.16) by taking advantage of their orthonormality (10.4). To do so, we multiply both sides of (10.16) by $\phi_n(t_1)$ and then take the integral with respect to t_1 between 0 and T. One then gets:

$$\begin{aligned}\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)\,dt_1 &= \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2) \int_0^T \phi_m(t_1)\phi_n(t_1)\,dt_1 \\ &= \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2)\delta_{mn} = \lambda_n \phi_n(t_2),\end{aligned} \tag{10.17}$$

that is

$$\int_0^T E\{X(t_1)X(t_2)\}\,\phi_n(t_1)\,dt_1 = \lambda_n\,\phi_n(t_2). \tag{10.18}$$

This basic result is an *integral equation*, called by mathematicians "of Fredholm type". Once the correlation $E\{X(t_1)X(t_2)\}$ of $X(t)$ is known, the integral equation (10.18) yields (upon its solution, which may not be easy at all to find analytically!) both the Karhunen–Loève eigenvalues λ_n and the corresponding eigenfunctions $\phi_n(t)$. Readers familiar with quantum mechanics will also recognize in (10.18) a typical "eigenvalue equation" having the kernel $E\{X(t_1)X(t_2)\}$.

Let us finally summarize what we have proven so far in Sections 10.3 and 10.4, and let us use the language of signal processing, which will lead us directly to SETI, the main theme of this chapter.

By adding random noise to a deterministic signal one obtains what is called a "noisy signal" or, in case the signal power is much lower than the noise power, "a signal buried into the noise". The noise + signal is a random function of the time, denoted hereafter by $X(t)$. Karhunen and Loève proved that it is possible to represent $X(t)$ as the infinite series (called the KL expansion) given by (10.2), and this series is convergent. Assuming that the (signal + noise) correlation $E\{X(t_1)X(t_2)\}$ is a known function of t_1 and t_2, then the orthonormal functions $\phi_n(t)$ $(n = 1, 2, \dots)$ turn out to be just the eigenfunctions of the correlation. These eigenfunctions $\phi_n(t)$ form an orthonormal basis in what physicists and mathematicians call the space of square-integrable functions, also called the Hilbert space. The eigenfunctions $\phi_n(t)$ actually are *the best* possible basis to describe the (signal + noise), much better than any classical Fourier basis made up by sines and cosines only. One can conclude that *the KLT automatically adapts itself to the shape of the (signal + noise), whatever behavior in time it may have, by adopting as a new reference frame in the Hilbert space the basis spanned by the eigenfunctions, $\phi_n(t)$, of the autocorrelation of the (signal + noise), $X(t)$.*

10.5 CONTINUOUS TIME VS. DISCRETE TIME IN THE KLT

The KL expansion in continuous time, *t*, is what we have described so far. This may be more "palatable" to theoretical physicists and mathematicians inasmuch as it may be related to other branches of physics, or of science in general, in which time obviously must be a continuous variable. For instance, this author spent 15 years of his life (1980–1994) in investigating mathematically the connection between Special Relativity and KLT. The result was the mathematical theory of optimal telecommunications between the Earth and a relativistic spaceship either receding from the Earth or approaching it. Although this may sound like "mathematical science fiction" to some folks (who we would call "short sighted"), the possibility that, in the future, humankind will send out relativistic automatic probes or even manned spaceships, is not unrealistic. Nor is it science fiction to imagine that an *alien* spaceship might approach the Earth slowing down from relativistic speeds to zero speed. So, a mathematical physics book like [4] can make sense. There, the KLT is obtained for any acceleration profile of the relativistic probe or spaceship. The result is that the KL eigenfunctions are Bessel functions of the first kind (suitably modified) and the eigenvalues are determined by the zeros of linear combinations of these Bessel functions and their derivatives, as we shall prove in Appendices F through K of this book, and especially in Appendix G.

Other continuous-time applications of the KLT are to be found in other branches of science, ranging, for instance, from genetics to economics. But, whatever the application may be, if time is a continuous variable, then one must solve the integral equation (10.18), and this may require considerable mathematical skills. In fact, (10.18) is, in general, an integral equation of the Fredholm type, and the usual "iterated nuclei" procedure used to solve Fredholm integral equations may be particularly painful to achieve. The task may be much easier if one is able to reduce the Fredholm integral equation to a Volterra integral equation, in just the way shown in the book [4] for the time-rescaled Brownian motion in relation to Special Relativity.

But let us go back to the time variable t in the KL expansion (10.2). If this variable is *discrete*, rather than continuous, then the picture changes completely. In fact, the integral equation (10.2) now becomes ... a system of simultaneous algebraic equations of the first degree, that can *always* be solved! The difficulty here is that this system of linear equations is *huge*, because the autocorrelation matrix is huge (hundreds or thousands of elements are the rule for autocorrelation matrices in SETI and in other applications, like image processing and the like). Also huge are the eigenvalues of the characteristic equation (i.e., the algebraic equations whose roots are the KL). Can you imagine solving *directly* an algebraic equation of degree 10,000?

So, the KLT is practically impossible to find numerically, unless we resort to *simplifying tricks* of some kind. This is precisely what was done for the SETI-Italia program by this author and his students, strongly supported by Ing. Stelio Montebugnoli and his team [5].

10.6 THE KLT: JUST A LINEAR TRANSFORMATION IN THE HILBERT SPACE

Although we have explained the KL expansion (10.2), we have yet to explain what the *KLT* is! We do so in this section.

The next step towards the KLT proper is the rearrangement of the eigenvalues λ_n in decreasing order of magnitude. Suppose we have done this. Consequently, we also rearrange the eigenfunctions $\phi_n(t)$ so that each eigenfunction keeps corresponding to its own eigenvalue. It can be proved that no mismatch can possibly arise in doing so, inasmuch as each eigenfunction corresponds to one eigenvalue only—namely, it can be proved that there is no degeneracy (contrary to what happens in quantum physics, where, for instance, there is a lot of degeneracy in the eigenfunctions of even the simplest atom of all, the hydrogen atom!). Furthermore, all eigenvalues are positive, and so, once rearranged in decreasing order of magnitude, they form a decreasing sequence where the first eigenvalue is the largest, and is called the "dominant" eigenvalue by mathematicians.

We are now ready to compute the *Direct KLT* of the (signal + noise). Let us use the new set of eigen-axes to describe the (signal + noise). Then, in the new representation, the (signal + noise) is just the Direct KLT of the old (signal + noise). In other words, the KLT is properly called just a *linear* trasformation of axes, and nothing is easier than that! (Incidentally, this accounts for the title of Karhunen's first paper "Über Lineare Methoden in der Wahrscheinlichkeitsrechnung" = "On linear methods in the calculus of probabilities", [1], which obviously refers to the linear character of the transformation of axes in the Hilbert space.)

10.7 A BREAKTHROUGH ABOUT THE KLT: MACCONE'S "FINAL VARIANCE" THEOREM

The importance of the KLT as a mathematical tool superior to the FFT has already been pointed out. However, the implementation of the KLT by a numerical code running on computers has always been a difficult problem. Both François Biraud in France [6] and Bob Dixon in the USA [16] failed to do so in the 1980s because all computers then available could not make the N^2 calculations required to solve the huge system of simultaneous algebraic equations of the first degree corresponding (in the discrete case) to the integral equation (10.18). At the SETI-Italia facilities at Medicina we faced the same problem, of course. But we did better than our predecessors because we discovered the new theorem about the KLT that we demonstrate in this section and call "the Final Variance theorem". This new theorem seems to be even more important than the rest of research work about the KLT since it solves directly the problem of extracting a weak sinusoidal carrier (a tone) from noise of whatever kind (both colored and white).

The key idea of the Final Variance theorem is to differentiate the first eigenvalue (briefly called the "dominant eigenvalue") of the autocorrelation of the (noise + signal) with respect to the final instant T of the general KLT theory. Remember here that this final instant T simply does not exist in the ordinary Fourier theory, because this T equals infinity according to the Fourier theory. Therefore, the final instant T in itself is possibly the most important "novelty" introduced by the KLT regarding the classical FFT. With respect to T, we may take derivatives (called "final derivatives" in the remainder of this book because they are time derivatives taken with respect to the final instant T) and integrals that have no analogs in the ordinary Fourier theory. The "error" that was made in the past—even by many KLT scholars—was to set $T = 1$, thus obscuring the fundamental novelty represented by the finite, real positive T as a new continuous variable playing in the game! This error made by other scholars clearly appears, for instance, in the Wikipedia site about the "Karhunen–Loève Theorem", *http://en.wikipedia.org/wiki/Karhunen-Loève_* *theorem*. So, by removing this silly $T = 1$ convention we opened up new prospects for KLT theory, as we now show by proving our "Final Variance theorem".

Consider the eigenfunction expansion of the autocorrelation again—Equation (10.16)—with the traditional dummy index n rewritten instead of m. Upon replacing $t_1 = t_2 = t$, this equation becomes

$$E\{X^2(t)\} = \sum_{n=1}^{\infty} \lambda_n \phi_n^2(t). \tag{10.19}$$

Since the eigenfunctions $\phi_n(t)$ are normalized to 1, we are prompted to integrate both sides of (10.19) with respect to t between 0 and T, so that the integral of the square of the $\phi_n(t)$ becomes just 1:

$$\int_0^T E\{X^2(t)\}dt = \sum_{n=1}^{\infty} \lambda_n \int_0^T \phi_n^2(t)\,dt = \sum_{n=1}^{\infty} \lambda_n. \tag{10.20}$$

On the other hand, since the mean value of $X(t)$ is identically equal to 0, one may now introduce the *variance* $\sigma^2_{X(t)}$ of the stochastic process $X(t)$ defined by

$$\sigma^2_{X(t)} = E\{X^2(t)\} - E^2\{X(t)\} = E\{X^2(t)\}. \tag{10.21}$$

Replacing (10.21) into (10.20), one gets

$$\int_0^T \sigma^2_{X(t)}\,dt = \sum_{n=1}^{\infty} \lambda_n. \tag{10.22}$$

This formula was first given by this author in his 1994 book [4, eq. (1.13), p. 12]. At that time, however, (10.22) was regarded as interesting inasmuch as (upon interchanging the two sides) it proves that the series of all the eigenvalues λ_n is indeed convergent (as one would intuitively expect) and its sum is given by the integral of the variance between 0 and T.

Back in 1994, however, the author did not understand that (10.22) had a more profound meaning: since the final instant T is the upper limit of the time integral on the left-hand side, the right-hand side also must depend on T. In other words, all the eigenvalues λ_n must be some functions of the final instant T:

$$\lambda_n \equiv \lambda_n(T). \tag{10.23}$$

This new remark is vital in order to make further progress. In fact, one is now prompted to let the integral on the left-hand side of (10.22) disappear by differentiating both sides with respect to the final instant T. One thus gets:

$$\sigma^2_{X(T)} = \sum_{n=1}^{\infty} \frac{\partial \lambda_n(T)}{\partial T}. \tag{10.24}$$

This result we call the *Final Variance theorem*. It was discovered by this author in May 2007 and is the key new result put forward in this chapter. It states that for any (either non-stationary or stationary) stochastic process $X(t)$, *the Final Variance* $\sigma^2_{X(T)}$

is the sum of the series of the first-order partial derivatives of the eigenvalues $\lambda_n(T)$ with respect to the final instant T.

Let us now consider a few particular cases of this theorem that are especially interesting.

(1) In general, only the first N terms of the decreasing sequence of eigenvalues will be retained as "significant" by the user, and all the other terms, from the $(N+1)$th term onward, will be declared to be "just noise". Therefore, the infinite series in (10.24) becomes in practice the finite sum

$$\sigma^2_{X(T)} \approx \sum_{n=1}^{N} \frac{\partial \lambda_n(T)}{\partial T}. \tag{10.25}$$

In numerical simulations, however, one always wants to make computation time as short as possible! Therefore, one might be led to consider the *first* (or *dominant*) eigenvalue only in (10.25); that is

$$\sigma^2_{X(T)} \approx \frac{\partial \lambda_1(T)}{\partial T}. \tag{10.26}$$

This clearly is "the roughest possible" approximation to the full $X(t)$ process since we are actually replacing the full $X(t)$ by its first KLT term $Z_1\phi_1(t)$ only. However, using (10.26) instead of the N-term sum (10.25) is indeed a good shortcut for application of the KLT to the extraction of very weak signals from noise, as we now stress in the very important practical case of stationary processes.

(2) If we restrict our considerations to *stationary* stochastic processes only (i.e., processes for which both the mean value and the variance are constant in time), then (10.25) simplifies even further. In fact, by definition, the stationary processes have the *same final variance at any time (i.e., for stationary processes σ_X^2 is a constant)*. Then (10.22) immediately shows that, for stationary processes only, all the KLT eigenvalues are *linear* functions of the final instant T:

$$\lambda_n(T) \propto T \quad \textit{for stationary processes only.} \tag{10.27}$$

As a consequence, the first-order partial derivatives of all the λ_n with respect to T for stationary processes are just constants. In yet other words, for stationary processes only, (10.25) becomes

$$\sum_{n=1}^{N} \frac{\partial \lambda_n(T)}{\partial T} \approx \text{a constant with respect to } T. \tag{10.28}$$

In particular, if one sticks again to the first, dominant eigenvalue only (i.e., to the roughest possible approximation), then (10.28) reduces to

$$\frac{\partial \lambda_1(T)}{\partial T} \approx \text{a constant with respect to } T. \tag{10.29}$$

In Section 10.8 we will discuss the deep, practical implications of this result for SETI, extrasolar planet detection, asteroidal radar, and other KLT applications.

(3) Please notice that, for non-stationary processes, the dependence of the eigenvalues on T certainly is non-linear. For instance, for the well-known Brownian motion (i.e., "the easiest of the non-stationary processes"), one has

$$\lambda_n(T) = \frac{4T^2}{\pi^2(2n-1)^2} \quad (n = 1, 2, \ldots) \tag{10.30}$$

and so the dependence on T is quadratic. For the proof, just place the Brownian motion variance $\sigma^2_{B(t)} = t$ into (10.22) and perform the integration, yielding the T^2 directly. Of course, this is in agreement with (10.30), which will be proven in Appendix F when we search for the KLT of the standard Brownian motion—see, in particular, (F.21).

(4) Even higher than quadratic is the dependence on T for the eigenvalues of other highly non-stationary processes. For instance, for the zero-mean square of the Brownian motion, the KLT eigenvalues depend cubically on the final instant T, as will be proven in Appendix I by Equation (I.60). And so on for more complicated processes, like the time-rescaled squared Brownian motions whose KLT will found in Appendix I.

10.8 BAM ("BORDERED AUTOCORRELATION METHOD") TO FIND THE NUMERIC KLT OF *STATIONARY* PROCESSES ONLY

The BAM (an acronym for "Bordered Autocorrelation Method") is an alternative numerical technique to evaluate the KLT of stationary processes (only) that may run faster on computers than the traditional full-solving KLT technique described in Section 10.5. The BAM has its mathematical foundation in our Final Variance theorem already proved in Section 10.7. In this section we describe the BAM in detail and provide the results of numerical simulations showing that, by virtue of the BAM, the KLT succeeds in extracting a sinusoidal carrier embedded in a lot of noise when the FFT utterly fails.

Let us start by recalling that the standard, traditional technique to find the KLT of any stochastic process (whether stationary or not) numerically amounts to solving N simultaneous linear algebraic equations whose coefficient matrix is the (huge) autocorrelation matrix. This N^2 amount of calculations is much larger than the $N * \ln(N)$ amount of calculations required by the FFT and that's precisely the reason the FFT has been preferred to the KLT in the last 50 years!

Because of the Final Variance theorem proved in the previous section, however, one is tempted to confine oneself to the study of the dominant eigenvalue, only by virtue of just using (10.29). This means studying (10.29) for different values of the final instant T (i.e., as a function of the final instant T).

Also, we now confine ourselves to a *stationary* $X(t)$ over a *discrete* set of instants $t = 0, \ldots, N$.

In this case, the autocorrelation of $X(t)$ becomes the Toeplitz matrix (for an introduction to the research field of Toeplitz matrices, see the Wikipedia site, *http://en.wikipedia.org/wiki/Toeplitz_matrix*) which we denote by $R_{Toeplitz}$.

$$R_{Toeplitz} = \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & R_{XX}(2) & \cdots & \cdots & R_{XX}(N) \\ R_{XX}(1) & R_{XX}(0) & R_{XX}(1) & \cdots & \cdots & R_{XX}(N-1) \\ R_{XX}(2) & R_{XX}(1) & R_{XX}(0) & \cdots & \cdots & R_{XX}(N-2) \\ \cdots & \cdots & \cdots & R_{XX}(0) & \cdots & \cdots \\ R_{XX}(N) & R_{XX}(N-1) & \cdots & \cdots & R_{XX}(1) & R_{XX}(0) \end{bmatrix}. \tag{10.31}$$

This theorem had already been proven by Bob Dixon and Mike Kline back in 1991 [16], and will not be proven here again. We may choose N at will, but clearly the higher the N, the more accurate the KLT of $X(t)$. On the other hand, the final instant T in the KLT can be chosen at will and now is $T = N$. So, we can regard $T = N$ as a sort of "new time variable" and even take derivatives with respect to it, as we'll do in a moment.

But let us now go back to the Toeplitz autocorrelation (10.31). If we let N vary as a new free variable, that amounts to *bordering it* (i.e., adding one (last) column and one (last) row to the previous correlation). This means solving yet again the system of linear algebraic equations of the KLT for $N + 1$, rather than for N. So, *for each different value of N, we get a new value of the first eigenvalue* λ_1 *now regarded as a function of N (i.e.,* $\lambda_1(N)$*). Doing this over and over again, for as many values as we wish (or, more correctly, for how many values of N our computer can still handle!) constitutes our BAM, the Bordered Autocorrelation Method.*

But then we know from the Final Variance theorem that $\lambda_1(N)$ *is proportional to N. And such a function* $\lambda_1(N)$ *of course has a derivative,* $\partial\lambda_1(N)/\partial N$*, that can be computed numerically as a new function of N. And this derivative turns out to be a constant with respect to N. This fact paves the way for a new set of applications of the KLT to all fields of science!*

In fact, numeric simulations lead to the results shown in the four plots in Figures 10.1–10.4. The first plot is the ordinary Fourier spectrum of a pure tone at 300 Hz buried in noise with a signal-to-noise ratio of 0.5, abbreviated hereafter as SNR = 0.5. For a definition of the SNR see the Wikipedia site, *http://en.wikipedia.org/wiki/Signal-to-noise_ratio* Please note the following two facts:

(1) This is about as low an SNR can be before the FFT starts failing to denoise a signal, as is well known by electrical and electronic engineers.
(2) This Fourier spectrum is obviously computed by taking the Fourier Transform of the *stationary* autocorrelation of $X(t)$, as is well known from the Wiener–Khinchin theorem (for a concise description of this theorem, see *http://en.wikipedia.org/wiki/Wiener-Khinchin_theorem*).

Notice, however, that this procedure would *not* work for non-stationary $X(t)$ because the Wiener–Khinchin theorem does not apply to non-stationary processes. For

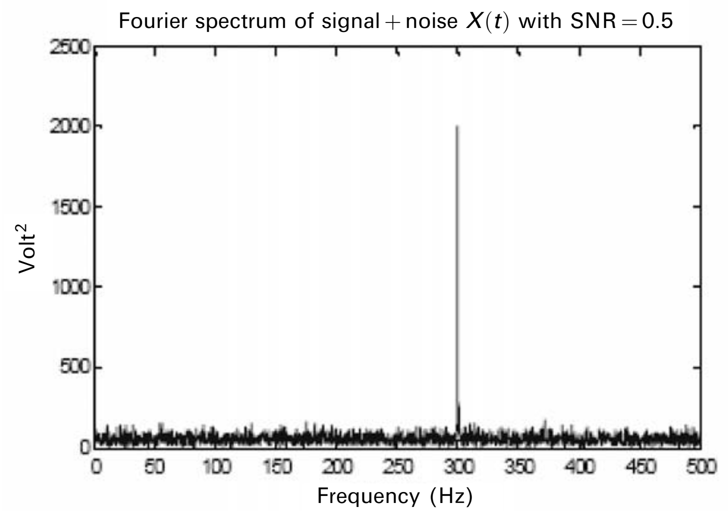


Figure 10.1. Fourier spectrum of a pure tone (i.e., just a sinusoidal carrier) with frequency at 300 Hz buried in stationary noise with a signal-to-noise ratio of 0.5.

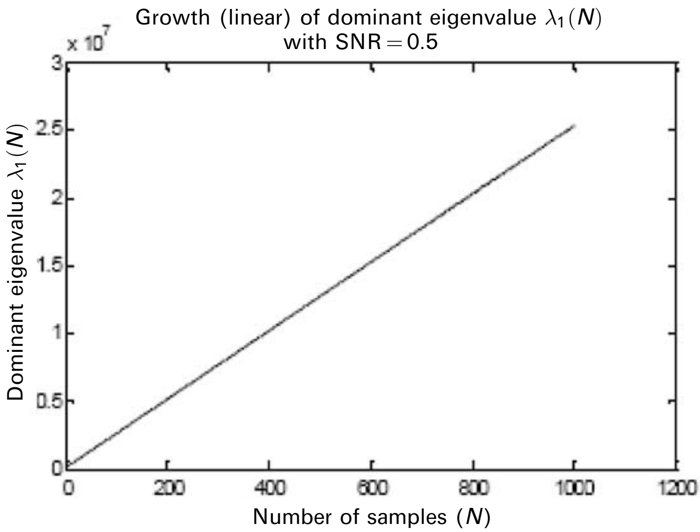


Figure 10.2. The KLT dominant eigenvalue $\lambda_1(N)$ over $N = 1{,}200$ time samples, computed by virtue of the BAM, the Bordered Autocorrelation Method.

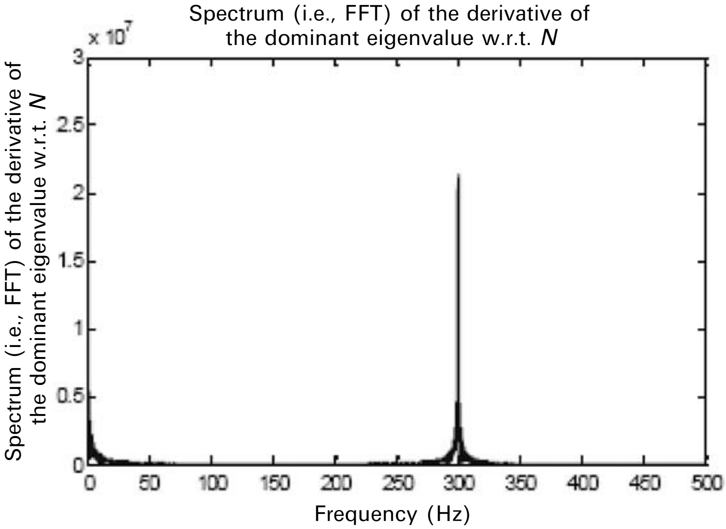


Figure 10.3. The spectrum (i.e., the Fourier Transform) of the *constant* derivative of the KLT dominant eigenvalue $\lambda_1(N)$ with respect to N as given by the BAM. This is clearly a Dirac delta function (i.e., a peak, at 300 Hz), as expected.

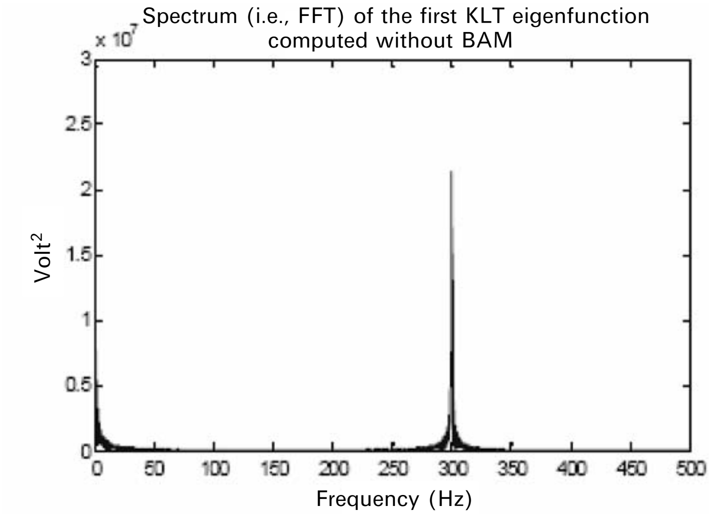


Figure 10.4. The spectrum (i.e., the Fourier Transform) of the first KLT eigenfunction *not* obtained by the BAM, but rather by the very long procedure of solving N linear algebraic equations corresponding, in discrete time, to the integral equation (10.18). Clearly, the result is the same as obtained in Figure 10.3 by the much less time-consuming BAM. So, one can say that adoption of the BAM actually made the KLT "feasible" on small computers by circumventing the difficulty of the N^2 calculations requested by the "straight" KLT theory.

non-stationary processes there are other "tricks" to compute the spectrum from the autocorrelation, like the Wigner–Ville Transform, but we shall not consider them here.

The second plot (Figure 10.2) shows the first (i.e., the dominant) KLT eigenvalue $\lambda_1(N)$ over $N = 1{,}200$ time samples. Clearly, this $\lambda_1(N)$ is proportional to N, as predicted by our Final Variance theorem (10.27).

So, its derivative, $\partial\lambda_1(N)/\partial N$, is a constant with respect to N. But we may then take the Fourier Transform of such a constant and get a Dirac delta function (i.e., a peak just at 300 Hz). In other words, we have KLT-reconstructed the original tone by virtue of the BAM. The third plot (Figure 10.3) shows such a BAM-reconstructed peak.

Finally, this plot is of course identical to the fourth plot (Figure 10.4), showing the ordinary FFT of the first KLT eigenfuction as obtained, not by the BAM, but by solving the full and long system of N algebraic first-degree equations.

Let us now do the same again ... but with an incredibly low SNR of 0.005.

Poor Fourier here is in a mess! Just look at the plot in Figure 10.5! No classical FFT spectrum can be identified at all for such a terribly low SNR!

But for the KLT no problem!

The next plot (Figure 10.6) shows that $\lambda_1(N) \propto N$, as predicted by our Final Variance theorem (10.27).

The third plot (Figure 10.7, *KLT fast* way via the BAM) is the *neat KLT spectrum* of the 300 Hz tone obtained by computing the FFT of the *constant* $\partial\lambda_1(N)/\partial N$.

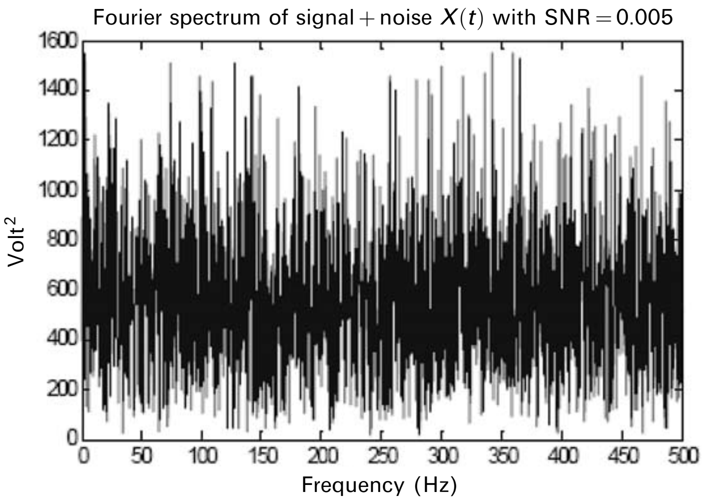


Figure 10.5. Fourier spectrum of a pure tone (i.e., just a sinusoidal carrier) with frequency at 300 Hz buried in stationary noise with the terribly low signal-to-noise ratio of 0.005. This is clearly beyond the reach of the FFT, since we know there should just be one peak only at 300 Hz. Fourier *fails* at such a low SNR.

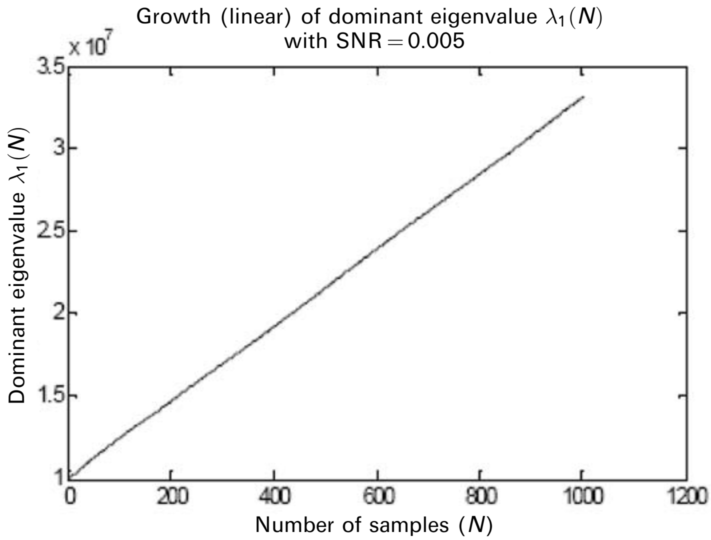


Figure 10.6. The KLT dominant eigenvalue $\lambda_1(N)$ for $N = 1{,}200$ time samples, computed by virtue of the BAM, for the very low SNR = 0.005.

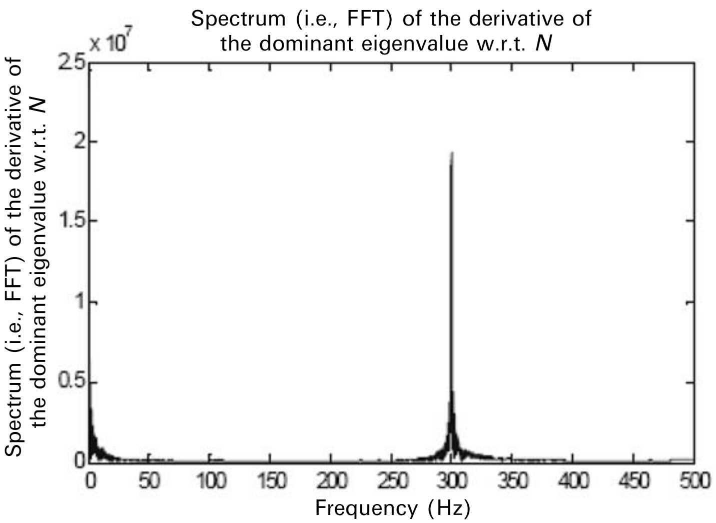


Figure 10.7. The spectrum (i.e., the Fourier Transform) of the *constant* derivative of the KLT dominant eigenvalue $\lambda_1(N)$ with respect to N as given by the BAM. This is a neat Dirac delta function (i.e., it has a peak at 300 Hz, as expected).

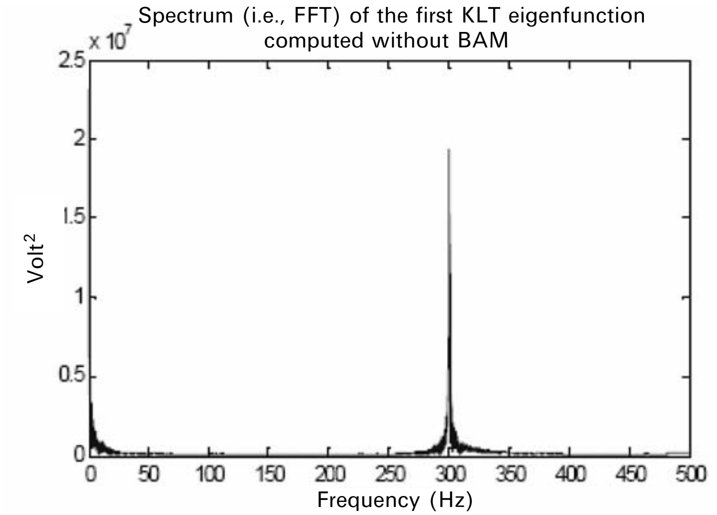


Figure 10.8. The spectrum (i.e., the Fourier Transform) of the first KLT eigenfunction, *not* obtained by the BAM but rather by the very long procedure of solving N linear algebraic equations corresponding, in discrete time, to the integral equation (10.18). Clearly, the result is the same as obtained in Figure 10.7, but this time by the much less time-consuming BAM. So, one can say that the adoption of the BAM actually made the KLT "feasible" on small computers by circumventing the difficulty of N^2 calculations requested by the "straight" KLT theory.

And this is just the same as the last plot (Figure 10.4) of the dominant KLT eigenfunction obtained by the KLT *slow* way of doing N^2 calculations. This proves the superior behavior of the KLT.

10.9 DEVELOPMENTS IN 2007 AND 2008

The numerical simulations described in the previous section were performed at Medicina during the winter 2006–2007 by Francesco Schillirò and Salvatore "Salvo" Pluchino [22]. These simulations suggested in a purely numerical fashion (i.e., without any analytic proof) that the BAM leads to the following result for stationary processes: the ordinary Fourier transform (i.e., "the spectrum" in the common sense, since the processes are supposed to be stationary) of the first-order partial derivative with respect to the final instant T of the dominant eigenvalue, $\dfrac{\partial \lambda_1(T)}{\partial T}$, is just the frequency of the feeble sinusoidal carrier buried in the mountain of noise. In SETI language, if we are looking for a simple sinusoidal carrier sent by ET and buried in a lot of cosmic noise, then the frequency we are looking for is given by the FFT of $\frac{\partial \lambda_1(T)}{\partial T}$.

Why?

No analytic proof of this numerical result was ever found at Medicina. But this author had made the first step towards the then missing analytic proof by proving the Final Variance Theorem in May 2007, and persisted in discussing this "frontier result" with other radioastronomers. One year later, in June 2008, he went to Dwingeloo, the Netherlands, and met with the ASTRON Team working on a possible implementation of SETI on the brand-new LOFAR radiotelescope. Dr. Sarod Yatawatta of ASTRON then made the next step toward the missing analytic proof: he derived an unknown analytic expression for the KLT eigenvalues of the ET sinusoidal carrier [24]. Unfortunately, Dr. Yatawatta made two analytical errors in his derivation (described hereafter), which this author discovered and corrected in September 2008.

In conclusion, the final, correct version of all these equations is explained in the next two sections, and it proves that the Fourier Transform of the first derivative of the KLT eigenvalues with respect to the final instant T is indeed the frequency of the "unknown" ET signal, but only for stationary processes, of course.

For non-stationary processes (i.e., for transient phoenomena as actually happens in practical SETI, since all celestial bodies move, rather than rest), the story is much more complicated, and this author is convinced that a much more refined mathematical investigation has to be made: but this will be our next step, not described in this book yet!

10.10 KLT OF STATIONARY WHITE NOISE

Before we give the analytic proof that the Fourier Transform of $\frac{\partial \lambda_1(T)}{\partial T}$ is the frequency of the unknown ET signal, we must understand what the KLT of stationary white noise is.

Stationary white noise is defined as the one "limit" stochastic process that is completely uncorrelated (i.e., the autocorrelation of which is the Dirac delta function). In other words, denoting the stationary white noise by $W(t)$, one has by definition

$$E\{W(t_1)W(t_2)\} = \delta(t_1 - t_2). \tag{10.32}$$

If one now seeks the KLT of stationary white noise, one must of course insert the autocorrelation (10.32) into the KLT integral equation (10.18), getting

$$\lambda_n \phi_n(t_2) = \int_0^T E\{W(t_1)W(t_2)\}\phi_n(t_1)\,dt_1 = \int_0^T \delta(t_1 - t_2)\phi_n(t_1)\,dt_1 = \phi_n(t_2). \tag{10.33}$$

This proves that:

(1) The KLT eigenvalues of stationary white noise are all equal to 1.
(2) *Any* set of orthonormal eigenfunctions $\phi_n(t)$ in the Hilbert space is a suitable basis to represent stationary white noise.

Since *any* set of orthonormal eigenfunctions $\phi_n(t)$ in the Hilbert space is a suitable basis to represent stationary white noise, from now one we shall adopt the easiest possible such basis; that is, the simple Fourier basis made up only by orthonormalized sines over the finite interval $0 \le t \le T$:

$$\phi_n(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi n}{T} t\right) \equiv W_n(t). \tag{10.34}$$

Of course, this set of basis functions fulfills the orthonormality condition

$$\int_0^T W_m(t) W_n(t)\, dt = \int_0^T \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi m}{T} t\right) \cdot \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi n}{T} t\right) dt = \delta_{mn}. \tag{10.35}$$

This property will be used in the next section, where we give the proof that the Fourier Transform of $\dfrac{\partial \lambda_n(T)}{\partial T}$ is indeed (twice) the frequency of the unknown ET sinusoidal carrier buried in white, cosmic noise. We conclude this section by pointing out the first analytical error made by Dr. Yatawatta in his personal communication to this author [24]: he forgot to put the square root in (10.34). This of course means that his further results were flawed, even more so since he made a second analytical error later, which we shall not describe. But the key ideas behind his proof were perfectly correct, and we shall describe them in the next section.

10.11 KLT OF AN ET SINUSOIDAL CARRIER BURIED IN WHITE, COSMIC NOISE

Consider a new stochastic process $S(t)$ made up by the sum of stationary white noise $W(t)$ plus an alien ET sinusoidal carrier of amplitude a and frequency $\nu = \dfrac{\omega}{2\pi}$; that is,

$$S(t) = W(t) + a \sin(\omega t). \tag{10.36}$$

What is the KLT of such a (signal + noise) process? This is the central problem of SETI, of course.

To find the answer, first build up the autocorrelation of this process:

$$E\{S(t_1)S(t_2)\} = E\{W(t_1)W(t_2)\} + a^2 \sin(\omega t_1)\sin(\omega t_2) + aE\{W(t_1)\sin(\omega t_2)\} + aE\{W(t_2)\sin(\omega t_1)\}. \tag{10.37}$$

The last two terms in (10.37) represent the two cross-correlations between the white noise and the sinusoidal signal. It is reasonable to assume that the white noise and the signal are uncorrelated, and so we shall simply replace these two cross-correlations by zero. The autocorrelation (10.37) of the (signal + noise) stochastic process $S(t)$ thus becomes

$$E\{S(t_1)S(t_2)\} = E\{W(t_1)W(t_2)\} + a^2 \sin(\omega t_1)\sin(\omega t_2). \tag{10.38}$$

In order to proceed, we now make use of the eigenfunction expansion of the autocorrelation (10.16), which, replaced into (10.38), changes it into

$$\sum_{m=1}^{\infty} \lambda_{S_m} S_m(t_1) S_m(t_2) = \sum_{m=1}^{\infty} \lambda_{W_m} W_m(t_1) W_m(t_2) + a^2 \sin(\omega t_1)\sin(\omega t_2). \tag{10.39}$$

In the last equation, the $S_m(t)$ clearly are the (unknown) eigenfunctions of the (signal + noise) process $S(t)$, and the λ_{S_m} are (unknown) corresponding eigenvalues. In the right-hand side, the λ_{W_m} are the eigenvalues of the stationary white noise, which we know to be equal to 1, but, for the sake of clarity, let us keep the symbol λ_{W_m} rather than replacing it by 1.

To proceed further, we now must get rid of both t_1 and t_2 in (10.39), and there is only one way to do so: use the orthonormality of the eigenfuctions appearing in (10.39). We shall do so in a moment. Before, however, let us make the following practical consideration: since the signal is much waker than the noise (by assumption) (i.e., the signal-to-noise ratio is much smaller than 1, or SNR $\ll$ 1), then, numerically speaking, the (signal + noise) eigenfunctions $S_m(t)$ must not differ very much from the pure white noise eigenfunctions $W_m(t)$. And, similarly, the (signal + noise) eigenvalues λ_{S_m} must not differ very much from the corresponding pure white noise eigenvalues λ_{W_m}. In other words, the hypothesis that SNR $\ll$ 1 amounts to the two approximate equations

$$\left.\begin{aligned} S_m(t) &\approx W_m(t) \\ \lambda_{S_m} &\approx \lambda_{W_m} = 1. \end{aligned}\right\} \tag{10.40}$$

Of course, only the first of these two equations will play a role in the two integrations that we are now going to perform: once with respect to t_1 and once with respect to t_2, and both over the interval $0 \le t \le T$. As a consequence, the new orthonormality condition (nearly) holds:

$$\int_0^T S_m(t_1) W_n(t_1)\, dt_1 \approx \delta_{mn} \tag{10.41}$$

and, similarly,

$$\int_0^T S_k(t_2) W_n(t_2)\, dt_2 \approx \delta_{kn} \tag{10.42}$$

So, let us now multiply both sides of (10.39) by $W_n(t_1)$ and integrate with respect to t_1 between 0 and T. Because of (10.41) and (10.35) one has:

$$\sum_{n=1}^{\infty} \lambda_{S_n} S_n(t_2) \approx \sum_{n=1}^{\infty} \lambda_{W_n} W_n(t_2) + a^2 \sin(\omega t_2) \int_0^T W_n(t_1) \sin(\omega t_1)\, dt_1 \tag{10.43}$$

The good point is that the integral appearing in the right-hand side of this equation can be found. In fact, replacing $W_n(t_1)$ by virtue of (10.34) and integrating, one gets

$$\sum_{k=1}^{\infty} \lambda_{S_k} S_k(t_2) \approx \sum_{k=1}^{\infty} \lambda_{W_k} W_k(t_2) + a^2 \sin(\omega t_2) \cdot \frac{2\sqrt{2}\pi n\sqrt{T}\sin(\omega T)}{\omega^2 T^2 - 4\pi^2 n^2} \tag{10.44}$$

We next multiply this equation by $W_n(t_2)$ and integrate with respect to t_2 between 0 and T. Because of (10.42) and (10.35), (10.44) becomes:

$$\lambda_{S_n} \approx \lambda_{W_n} + a^2 \frac{2\sqrt{2}\pi n\sqrt{T}\sin(\omega T)}{\omega^2 T^2 - 4\pi^2 n^2} \int_0^T W_n(t_2)\sin(\omega t_2)\,dt_2. \tag{10.45}$$

Again, the integral in the last equation can be computed—it is actually the same integral as in (10.43)—and so the conclusion is

$$\lambda_{S_n} \approx \lambda_{W_n} + a^2 \frac{8\pi^2 n^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2}. \tag{10.46}$$

This is Yatawatta's main result (corrected by Maccone). Let us now point out clearly that the eigenvalues on the left are a function of the final instant T; that is,

$$\lambda_{S_n}(T) \approx \lambda_{W_n} + a^2 \frac{8\pi^2 n^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2}. \tag{10.47}$$

This equation clearly shows that

(1) For $T \to 0$, the fraction in the right-hand side approaches zero, and so the eigenvalues of the signal + noise approach the pure white noise eigenvalues (as is intuitively obvious).
(2) For $n \to \infty$, again the fraction in the right-hand side approaches zero, and so the eigenvalues of the signal + noise approach the pure white noise eigenvalues (as again is intuitively obvious). This result may justify numerically the practical approximation made by the Medicina engineers when they confined their simulations to the first eigenvalue only (roughest approximation). In other words, the dominant eigenvalue of the signal + noise is given by

$$\lambda_{S_1}(T) \approx \lambda_{W_1} + a^2 \frac{8\pi^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2)^2} = 1 + a^2 \frac{8\pi^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2)^2}. \tag{10.48}$$

This completes our analysis of the KLT of a sinusoidal carrier buried in white, cosmic noise.

10.12 ANALYTIC PROOF OF THE BAM–KLT

We are now ready for the analytic proof of the BAM-KLT method.

Let us first re-write (10.47) in a form in which the pure white noise eigenvalues are replaced by 1:

$$\lambda_{S_n}(T) \approx 1 + a^2 \frac{8\pi^2 n^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2}. \tag{10.49}$$

We then notice that the final instant T appears three times in the right-hand side of the last equation:

(1) once in the numerator outside the sine;
(2) once in the numerator inside the sine;
(3) once in the denominator.

Therefore, the partial derivative of (10.49) with respect to T will be made up by the sum of three terms:

(1) One term with the derivative of the T in the numerator (i.e., 1 times the sine square). This brings a term in the cosine of TWICE the sine argument, since one obviously has
$$\sin^2(\omega T) = \tfrac{1}{2} - \tfrac{1}{2}\cos(2\omega T). \tag{10.50}$$
(2) One term with the derivative of the T inside the sine. This brings a term in the sine of TWICE the sine argument, because one has
$$2\sin(\omega T)\cos(\omega T) = \sin(2\omega T). \tag{10.51}$$
(3) One term with the derivative of the T in the denominator. This does not bring any term in either the sine or the cosine, but just a rational function of T that we shall give in a moment. In fact, we now prefer to skip the lengthy and tedious steps leading to the derivative of (10.49) with respect to T and just give the final result.

In conclusion, the derivative of (10.49) with respect to T is given by the following sum of three terms:
$$\frac{\partial \lambda_{S_n}(T)}{\partial T} \approx \mathrm{Coeff}_1(T)\cdot\sin(2\omega T) + \mathrm{Coeff}_2(T)\cdot\cos(2\omega T) + \mathrm{Coeff}_3(T) \tag{10.52}$$
where the three coefficients turn out to be (after lengthy calculations)
$$\left.\begin{aligned} \mathrm{Coeff}_1(T) &= a^2\frac{8\pi^2 n^2 \omega T}{(\omega^2T^2 - 4\pi^2n^2)^2}, \\ \mathrm{Coeff}_2(T) &= a^2\frac{4\pi^2n^2(3\omega^2T^2 + 4\pi^2n^2)}{(\omega^2T^2 - 4\pi^2n^2)^3}, \\ \mathrm{Coeff}_3(T) &= -a^2\frac{4\pi^2n^2(3\omega^2T^2 + 4\pi^2n^2)}{(\omega^2T^2 - 4\pi^2n^2)^3}. \end{aligned}\right\} \tag{10.53}$$

But the right-hand side of (10.52) is no more than ... the simple Fourier series expansion of $\frac{\partial \lambda_{S_n}(T)}{\partial T}$. Moreover, (10.52) shows that $\frac{\partial \lambda_{S_n}(T)}{\partial T}$ is a PERIODIC function of T with frequency $2\omega T$. We conclude that: ***The Fourier transform of*** $\frac{\partial \lambda_{S_n}(T)}{\partial T}$ ***equals TWICE the frequency of the buried alien sinusoidal carrier. In other words, the***

frequency of the alien signal is a HALF of the frequency found by taking the Fourier transform of $\dfrac{\partial \lambda_{S_n}(T)}{\partial T}$.

And the BAM–KLT method is thus proved analytically.

10.13 KLT SIGNAL-TO-NOISE (SNR) AS A FUNCTION OF THE FINAL T, EIGENVALUE INDEX n, AND ALIEN FREQUENCY ν

We now derive a consequence from the eigenvalue relationship (10.47) dealing with the signal-to-noise ratio (abbreviated SNR) in the KLT theory. We shall call it the "KLT–SNR Theorem". The proof is as follows.

Consider Equation (10.10), showing that the eigenvalues λ_n of any KL expansion are actually the variances of the zero-mean corresponding uncorrelated (i.e., orthogonal, in the probabilistic sense) random variables Z_n. If we apply this to the KLT of stationary unitary white noise, described in Section 10.10, the conclusion is that the λ_{W_m} are the mean values of the square of the corresponding orthogonal (i.e., uncorrelated random variables $Z^2_{W_n}$)

$$\lambda_{W_m} = E\{Z^2_{W_n}\}. \tag{10.54}$$

Now, the definition of the signal-to-noise ratio (which we prefer to denote SNR, rather than S/R) of a sinusoidal signal with amplitude a buried in the noise with amplitude Z_{W_n} is just:

$$\text{SNR} = \frac{\text{power of the signal}}{\text{power of the noise}} = \frac{a^2}{E\{Z^2_{W_n}\}} = \frac{a^2}{\lambda_{W_m}}. \tag{10.55}$$

This definition can now be inserted into (10.47) divided by λ_{W_m}; that is,

$$\frac{\lambda_{S_n}(T)}{\lambda_{W_n}} \approx 1 + \frac{a^2}{\lambda_{W_n}} \cdot \frac{8\pi^2 n^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2}, \tag{10.56}$$

with the result that (10.56) is changed into

$$\frac{\lambda_{S_n}(T)}{\lambda_{W_n}} \approx 1 + \text{SNR} \cdot \frac{8\pi^2 n^2 T \sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2}. \tag{10.57}$$

Solving this for SNR yields

$$\text{SNR}(T, n, \omega) \approx \left(\frac{\lambda_{S_n}(T)}{\lambda_{W_n}} - 1\right) \cdot \frac{(\omega^2 T^2 - 4\pi^2 n^2)^2}{8\pi^2 n^2 T \sin^2(\omega T)}. \tag{10.58}$$

For SETI applications, it may be preferable to re-express the last formula directly in terms of the "alien" frequency $\nu = \dfrac{\omega}{2\pi}$, instead of ω. Equation (10.58) is thus changed into

$$\text{SNR}(T, n, \nu) \approx \left(\frac{\lambda_{S_n}(T)}{\lambda_{W_n}} - 1\right) \cdot \frac{2\pi^2(\nu^2 T^2 - n^2)^2}{n^2 T \sin^2(2\pi\nu T)}. \tag{10.59}$$

This is our KLT–SNR Theorem. Since the quantity

$$\left(\frac{\lambda_{S_n}(T)}{\lambda_{W_n}} - 1\right) > 0 \qquad (10.60)$$

has a positive numeric value just slighty above zero, from (10.59) we conclude that

$$\begin{cases} \text{SNR}(T,n,\nu) = O(T^3) & \text{as } T \to \infty \\ \text{SNR}(T,n,\nu) = O(n^2) & \text{as } n \to \infty \\ \text{SNR}(T,n,\nu) = O(\nu^4) & \text{as } \nu \to \infty. \end{cases} \qquad (10.61)$$

These equations yield the "pace of increase" of the KLT–SNR, and should be of importance in writing down the numeric codes for the actual implementation of the KLT.

10.14 HOW TO EAVESDROP ON ALIEN CHAT

Following the Paris *First IAA Workshop on Searching for Life Signatures* (held at UNESCO, Paris, September 22–26, 2008, and organized by this author), the British popular science magazine *New Scientist* published the following article on October 30, 2008, that well summarizes the key features of the present scientific discussion.

How to eavesdrop on alien chat

30 October 2008

From *New Scientist* Print Edition.

Jessica Griggs

ET, phone . . . each other? If aliens really are conversing, we are not picking up what they are saying. Now one researcher claims to have a way of tuning in to alien cellphone chatter.

On Earth, the signal used to send information via cellphones has evolved from a single carrier wave to a "spread spectrum" method of transmission. It's more efficient, because chunks of information are essentially carried on multiple low-powered carrier waves, and more secure because the waves continually change frequency so the signal is harder to intercept.

It follows that an advanced alien civilisation would have made this change too, but the search for extraterrestrial life (SETI) is not listening for such signals, says Claudio Maccone, co-chair of the SETI Permanent Study Group based in Paris, France.

An algorithm known as the Fast Fourier Transform (FFT) is the method of choice for extracting an alien signal from cosmic background noise. However, the technique cannot extract a spread spectrum signal. Maccone argues that SETI should use an algorithm known as the Karhunen–Loève Transform (KLT), which could find a buried conversation with a signal-to-noise ratio 1000 times lower than the FFT.

A few people have been "preaching the KLT" since the early 1980s but until now it has been impractical as it involves computing millions of simultaneous equations, something even today's supercomputers would struggle with. At a recent meeting in Paris called Searching for Life Signatures, Maccone presented a mathematical method to get around this burden and suggested that the KLT should be programmed into computers at the new Low Frequency Array telescope in the Netherlands and the Square Kilometre Array telescope, due for completion in 2012.

Seth Shostak at the SETI Institute in California agrees that the KLT might be the way to go but thinks we shouldn't abandon existing efforts yet. "It is likely that for their own conversation they use a spread-spectrum method but it is not terribly crazy to assume that to get our attention they might use a 'ping' signal that has a lot of energy in a narrow band—the kind of thing the FFT could find."

"It is likely that aliens use the same spread-spectrum method of transmission as us on their cellphones."

From issue 2680 of *New Scientist* magazine,

30 October 2008, p. 14.

10.15 CONCLUSIONS

Let us summarize the main results of this chapter.

When the stochastic process $X(t)$ is stationary (i.e., it has both mean value and variance constant in time), then there are two alternative ways to compute the first KLT dominant eigenfunction (i.e., the roughest approximation to the full KLT expansion, which may be "enough" for practical applications!):

(1) (long way)—either you compute the first eigenvalue from the autocorrelation and then solve the huge (N^2) system of linear equations to get the first eigenfunction;
(2) (short way = BAM)—or you compute the derivative of the first eigenvalue with respect to $T = N$ and then Fourier-transform it to get the first eigenfunction.

In practical, numerical simulations of the KLT it may be much less time-consuming to choose option (2) rather than option (1).

In either case, the KLT of a given stationary process can retrieve a sinusoidal carrier out of the noise for values of the signal-to-noise ratio (SNR) that are three orders of magnitude lower than those that the FFT can still filter out. In other words, while the FFT (at best) can filter out signals buried in noise with an SNR of about 1 or so, the KLT can, say, filter out signals that have an SNR of, say, 0.001 or so.

This is the superior achievement of the KLT over the FFT.

The BAM (Bordered Autocorrelation Method) is an alternative numerical technique to evaluate the KLT of stationary processes (only) that may run faster on computers than the traditional full-solving KLT technique. In this chapter we have provided the results of numerical simulations that show, by virtue of the BAM, how the KLT succeeds in extracting a sinusoidal carrier embedded in a lot of noise when the FFT utterly fails.

10.16 ACKNOWLEDGMENTS

The author is indebted to many radioastronomers and scientists who helped him over the years to work out what is now the BAM–KLT method. Principal among them are Ing. Stelio Montebugnoli and his SETI-Italia Team, Dr. Mike Garrett and his ASTRON Team (in particular Dr. Sarod Yatawatta), Dr. Jill Tarter and the SETI Institute Team (in particular Drs. Seth Shostak and Doug Vakoch). Also, the Paris *SETI Conference of September 22–26, 2008*, organized by this author at UNESCO, was possible only through the full support of the Secretary General of the IAA, Dr. Jean-Michel Contant, and of the newly-born French SETI community. Finally, a number of other young and not-so-young folks continued to support this author in his efforts for SETI over the years, and their help is hereby gratefully acknowledged.

10.17 REFERENCES

[1] K. Karhunen, "Über lineare Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae, Series A 1, Math. Phys.*, **37** (1946), 3–79.

[2] M. Loève, "Fonctions Aléatoires de Second Ordre," *Rev. Sci.*, **84**(4) (1946), 195–206.

[3] M. Loève, *Probability Theory: Foundations, Random Sequencies*, Van Nostrand, Princeton, NJ, 1955.

[4] C. Maccone, *Telecommunications, KLT and Relativity*, Volume 1, IPI Press, Colorado Springs, CO, 1994, ISBN # 1-880930-04-8. This book embodies the results of some 30 research papers published by the author about the KLT in the 15-year span 1980–1994 in peer-reviewed journals.

[5] S. Montebugnoli, C. Bortolotti, D. Caliendo, A. Cattani, N. D'Amico, A. Maccaferri, C. Maccone, J. Monari, A. Orlati, P. P. Pari *et al.*, "SETI-Italia 2003 Status Report and First Results of a KL Transform Algorithm for ETI Signal Detection," paper IAC-03-IAA.9.1.02 presented at the *2003 International Astronautical Congress* held in Bremen, Germany, September 29–October 3, 2003.

[6] F. Biraud, "SETI at the Nançay Radio-telescope," *Acta Astronautica*, **10** (1983), 759–760.

[7] C. Maccone, "Advantages of the Karhunen–Loève Transform over Fast Fourier Transform for Planetary Radar and Space Debris Detection," *Acta Astronautica*, **60** (2007), 775–779.

10.18 ANNOTATED BIBLIOGRAPHY

In addition to the above references, we would like to offer an "enlightened" list of a few key references about the KLT, subdivided according to the field of application.

Early papers by the author about the KLT in mathematics, physics, and the theory of relativistic interstellar flight, subdivided by journals

Il Nuovo Cimento

[8] C. Maccone, "Special Relativity and the Karhunen–Loève Expansion of Brownian Motion," *Nuovo Cimento, Series B*, **100** (1987), 329–342.

Bollettino dell'Unione Matematica Italiana

[9] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana, Series 6,* **3-A** (1984), 213–219;

[10] C. Maccone, "The Time-Rescaled Brownian Motion $B(t^{2H})$," *Bollettino dell'Unione Matematica Italiana, Series 6*, **4-C** (1985), 363–378; C. Maccone, "The Karhunen–Loève Expansion of the Zero-Mean Square Process of a Time-Rescaled Gaussian Process," *Bollettino dell'Unione Matematica Italiana, Series 7*, **2-A** (1988), 221–229.

Journal of the British Interplanetary Society

[12] C. Maccone, "Relativistic Interstellar Flight and Genetics," *Journal of the British Interplanetary Society*, **43** (1990), 569–572.

Acta Astronautica

[13] C. Maccone, "Relativistic Interstellar Flight and Gaussian Noise," *Acta Astronautica*, **17**(9) (1988), 1019–1027.

[14] C. Maccone, "Relativistic Interstellar Flight and Instantaneous Noise Energy," *Acta Astronautica*, **21**(3) (1990), 155–159.

KLT for data compression

[15] C. Maccone, "The Data Compression Problem for the 'Gaia' Astrometric Satellite of ESA," *Acta Astronautica*, **44**(7–12) (1999), 375–384.

Some important papers about the KLT for SETI

[16] R. S. Dixon, and M. Klein, "On the detection of unknown signals," *Proceedings of the Third Decennial US–USSR Conference on SETI* held at the University of California at Santa Cruz, August 5–9, 1991. Later published in the *Astronomical Society of the Pacific (ASP) Conference Series* (Seth Shostak, Ed.), **47** (1993), 128–140.

[17] C. Maccone, *Karhunen–Loève versus Fourier Transform for SETI*, Lecture Notes in Physics, Springer-Verlag, Vol. 390 (1990), pp. 247-253. These are the *Proceedings* (J. Heidmann and M. Klein, Eds.) *of the Third Bioastronomy Conference* held in Val Cenis, Savoie, France, June 18–23, 1990.

After these seminal works were published, the importance of the KLT for SETI was finally acknowledged by SETI Institute experts in

[18] R. Eckers, K. Cullers, J. Billingham, and L. Scheffer (Eds.), *SETI 2020*, SETI Institute, Mountain View, CA, 2002, p. 234, note 13. The authors say: "Currently (2002) only the Karhunen Loeve (KL) transform [Mac94] shows potential for recognizing the difference between incidental radiation technology and white noise. The KL transform is too

computationally intensive for the present generation of systems. The capability for using the KL transform should be added to future systems when computational requirements become affordable."

The paper [Mac94] referred to in the *SETI 2020* statement mentioned above is

[19] C. Maccone, "The Karhunen–Loève Transform: A Better Tool than the Fourier Transform for SETI and Relativity," *Journal of the British Interplanetary Society*, **47** (1994), 1.

An early paper about the KLT for SETI-Italia

[20] S. Montebugnoli, and C. Maccone, "SETI-Italia Status Report 2001", a paper presented at the 2001 *IAF Conference* held in Toulouse, France, October 1–5, 2001.

An early paper about the possibility of a "fast" KLT

[21] A. K. Jain, "A Fast Karhunen–Loève Transform for a Class of Random Processes," *IEEE Trans. Commun.*, **COM-24** (1976), 1023–1029.

Recent papers about the KLT and BAM–KLT

[22] F. Schillirò, S. Pluchino, C. Maccone, and S. Montebugnoli, *La KL Transform: considerazioni generali sulle metodologie di analisi ed impiego nel campo della Radio-astronomia*, Istituto Nazionale di Astrofisica (INAF)/Istituto di Radioastronomia (IRA), Technical Report, January 2007 [in Italian].

[23] C. Maccone, "Innovative SETI by the KLT," *Proceedings of the Bursts, Pulses and Flickering Conference* held at Kerastari, Greece, June 13–18, 2007. Available at POS (Proceedings of Science) website *http://pos.sissa.it//archive/conferences/056/034/Dynamic2007_034.pdf*

[24] S. Yatawatta, pers. commun., June 17, 2008.

A recent paper about the KLT for relativistic interstellar flight

[25] C. Maccone, "Relativistic Optimized Link by KLT," *Journal of the British Interplanetary Society*, **59** (2006), 94–98.

11

KLT of radio signals from relativistic spaceships in uniform and decelerated motion

11.1 INTRODUCTION

It is well known that in special relativity two time variables exist: the coordinate time t, which is the time measured in the fixed reference frame, and the proper time τ, which is the time shown by a clock rigidly connected to the moving body. They are related by

$$\tau(t) = \int_0^t \sqrt{1 - \frac{v^2(s)}{c^2}}\, ds \tag{11.1}$$

where $v(t)$ is the body velocity and c is the speed of light (see [1, p. 44]).

The remainder of this book, starting with the present chapter, is devoted to the relativistic interpretation of Brownian motion whose time variable is the proper time, $B(\tau)$, rather than the coordinate time, $B(t)$ and to find the KLT of $B(\tau)$. The bulk of these results was given by the author in a purely mathematical form, with no reference to relativity, in [2]. The KLT is also explained in detail in Chapter 10 and Appendices F–J. However, to enable the reader to read Chapters 11–14 independently of Chapter 10 and Appendices F–J, a summary of that work is now given in a form suitable for the physical developments that will follow in Chapters 12–14.

Consider standard Brownian motion (Wiener–Lévy process) $B(t)$, with mean zero, variance t, and initial condition $B(0) = 0$, as described in Appendix F.

A white noise integral is the process $X(t)$ defined by

$$X(t) = \int_0^t f(s)\, dB(s) \tag{11.2}$$

where $f(t)$ is assumed to be continuous and non-negative. Evidently, $X(0) = 0$, and it can be proved (see (F.35) or, equivalently, [3, pp. 84–87]) that

$$X(t) = B\left(\int_0^t f^2(s)\,ds\right). \tag{11.3}$$

Thus, $X(t)$ is a time-rescaled Gaussian process, with mean zero and

$$E\{X(t_1)X(t_2)\} = \int_0^{t_1 \wedge t_2} f^2(s)\,ds \tag{11.4}$$

as autocorrelation (covariance); $t_1 \wedge t_2$ denotes the minimum (smallest) t_1 and t_2.

Now the KLT theorem (see [4, pp. 262–271]) states that

$$X(t) = \sum_{n=1}^{\infty} Z_n \phi_n(t) \quad (0 \le t \le T) \tag{11.5}$$

where (1) the functions $\phi_n(t)$ are the autocorrelation eigenfunctions to be found from

$$\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_2)\,dt_2 = \lambda_n \phi_n(t_1) \tag{11.6}$$

where the constants λ_n are the corresponding eigenvalues; and (2) the Z_n are orthogonal random variables, with mean zero and variance λ_n; that is:

$$E\{Z_m Z_n\} = \lambda_n \delta_{mn}. \tag{11.7}$$

This theorem is valid for any continuous-parameter second-order process with mean zero and known autocorrelation. The series (11.5) converges in mean square, and uniformly in t. Finally, if $X(t)$ is Gaussian—as in Equations (11.2) and (11.3)—the random variables Z_n are also Gaussian, and since they are orthogonal they are independent.

After these preliminaries, we can state the main result of [2] (Maccone First KLT Theorem, fully proven in Appendix G).

The white noise integral (11.2), or the equivalent time-rescaled Gaussian process (11.3), has the KLT expansion:

$$X(t) = \sum_{n=1}^{\infty} Z_n N_n \sqrt{f(t)\int_0^t f(s)\,ds} \cdot J_{\nu(t)}\left(\gamma_n \frac{\int_0^t f(s)\,ds}{\int_0^T f(s)\,ds}\right). \tag{11.8}$$

Here

(1) the order of the Bessel functions $\nu(t)$ is not a constant, but the time function

$$\nu(t) = \sqrt{-\frac{\chi^3(t)}{f^2(t)} \cdot \frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right]} \tag{11.9}$$

with

$$\chi(t) = \sqrt{f(t) \int_0^t f(s)\, ds}. \tag{11.10}$$

(2) The constants γ_n are the (increasing) positive zeros of

$$\chi'(T) \cdot J_{\nu(T)}(\gamma_n) + \chi(T) \cdot \left[\frac{f(T) \cdot \gamma_n}{\int_0^T f(s)\, ds} J'_{\nu(T)}(\gamma_n) + \frac{\partial J_{\nu(T)}(\gamma_n)}{\partial \nu} \nu'(T) \right] = 0. \tag{11.11}$$

In general, (11.11) can only be solved numerically.

(3) The normalization constants N_n follow from the normalization condition

$$N_n^2 \left[\int_0^T f(s)\, ds \right]^2 \cdot \int_0^1 x[J_{\nu((x))}(\gamma_n x)]^2\, dx = 1 \tag{11.12}$$

where the new Bessel functions order $\nu((x))$ is (11.9) changed by aid of the transformation

$$\int_0^t f(s)\, ds = x \int_0^T f(s)\, ds.$$

(4) The eigenvalues are determined by

$$\lambda_n = \left[\int_0^T f(s)\, ds \right]^2 \frac{1}{(\gamma_n)^2}. \tag{11.13}$$

(5) The Gaussian random variables Z_n are independent and orthogonal, and have zero mean and variance λ_n.

The proof of this theorem may be sketched as follows: first, the Volterra-type integral equation (11.6) is transformed into a differential equation with two boundary conditions; and, second, the latter is reduced to the standard Bessel differential equation by means of two changes of variables. The full proof is given in Appendix G.

Let us now go back to relativity. Since from (11.3) it plainly appears that the rescaled time of the new Brownian motion is given by

$$\int_0^t f^2(s)\, ds \tag{11.14}$$

we merely have to equate (11.1) and (11.14) to get the relationship among the arbitrary time-rescaling function $f(t)$ and the arbitrary body velocity $v(t)$:

$$\int_0^t f^2(s)\, ds = \int_0^t \sqrt{1 - \frac{v^2(s)}{c^2}}\, ds. \tag{11.15}$$

By differentiating and taking the positive square root, it follows that:

$$f(t) = \left[1 - \frac{v^2(t)}{c^2}\right]^{\frac{1}{4}}. \tag{11.16}$$

This formula is the starting point to study the KLT expansion (11.8) for a relativistic body, like a relativistic spacecraft or spaceship moving in a radial direction away or towards the Earth.

Inversion of (11.16) leads at once to:

$$v(t) = c\sqrt{1 - f^4(t)}. \tag{11.17}$$

Now, the reality of the motion requires the radicand to be non-negative, whence, taking the positive sign in front of all square roots, we find

$$f(t) \leq 1. \tag{11.18}$$

This is the fundamental upper bound imposed on the "arbitrary" function $f(t)$ by special relativity. In other words, as the speed of light can in no case be exceeded, so $f(t)$ must not exceed 1.

As already pointed out, the lower bound on $f(t)$, required by the presence of the radicals in (11.8) and (11.10), is zero. Therefore

$$0 \leq f(t) \leq 1 \quad (0 \leq t \leq T) \tag{11.19}$$

is the physical range of the (otherwise arbitrary) function $f(t)$.

We also need to point out the Newtonian limit of the results. By this we mean the limit as $c \to \infty$. Then, as we see from (11.16),

$$\lim_{c\to\infty} f(t) = 1 \tag{11.20}$$

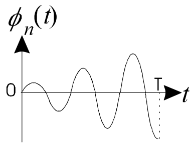
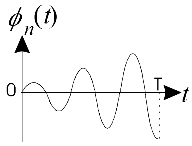
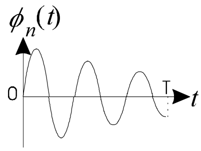
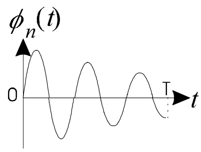
and the time-rescaled process under consideration reduces to standard Brownian motion, $B(t)$. This agrees, of course, with (11.1), stating that the proper time τ becomes the same as the coordinate time t in the Newtonian limit $c \to \infty$.

Finally, we want to hint at how the shape of the eigenfunctions $\phi_n(t)$ may be determined even without knowing their analytical expression. This possibility is a consequence of the Sonine–Pólya theorem, which is explored in Section G.5, for the non-relativistic case. The reader is referred there for the details, and here we merely confine ourselves to the relativistic version of the results. From (11.16) and (G.61) one finds:

$$\begin{aligned} \frac{d \ln f(t)}{dt} &= \frac{1}{4}\frac{d}{dt}\left[\ln\left(1 - \frac{v^2(t)}{c^2}\right)\right] = -\frac{1}{2c^2} \cdot \frac{v(t)\dfrac{dv(t)}{dt}}{1 - \dfrac{v^2(t)}{c^2}} \\ &= (\text{negative}) \cdot v(t)\frac{dv(t)}{dt}. \end{aligned} \tag{11.21}$$

Thus, not only the velocity $v(t)$, but also its derivative (i.e., acceleration taken with respect to the coordinate time, t) determines the shape (i.e., the stability) of the $\phi_n(t)$. The resulting Table 11.1 follows from this and Table G.1.

Table 11.1. Stability criterion for the relativistic eigenfunctions $\phi_n(t)$.

Sign of the velocity $v(t)$	Sign of the coordinate acceleration $dv(t)/dt$	Shape of the KL eigenfunctions $\phi_n(t)$	Description when T is finite	Description when T is infinite
Positive	Negative		Divergent	Asymptotic unstable
Negative	Positive		Divergent	Asymptotic unstable
Positive	Positive		Convergent	Asymptotic stable
Negative	Negative		Convergent	Asymptotic stable

11.2 UNIFORM MOTION

The simplest possible case of (11.16) is when the velocity $v(t)$ is a constant (i.e., the body's motion is uniform). Then $f(t)$ is a constant K as well

$$f(t) = \left[1 - \frac{v^2(t)}{c^2}\right]^{\frac{1}{4}} = K. \tag{11.22}$$

Let us now recall the property of the Brownian motion called self-similarity to the order 1/2 and expressed by the formula $B(ct) = \sqrt{c}B(t)$ where c is any real positive constant—see (F.6) for the relevant proof. From this and from (11.3), one gets at once

$$X(t) = B\left(\int_0^t K^2\,ds\right) = B(K^2 t) = KB(t). \tag{11.23}$$

Thus, the *uniform* proper-time Brownian motion $B(\tau) = X(t)$ equals the *uniform* coordinate-time Brownian motion $B(t)$ multiplied by the *constant* K, which is

$$B(\tau) = \left[1 - \frac{v^2}{c^2}\right]^{\frac{1}{4}} B(t). \tag{11.24}$$

The KL expansion of $B(\tau)$ is, of course, the same as that of $B(t)$ apart from the multiplicative factor K. And the relevant eigenfunctions are just sines.

To provide an example of how the machinery outlined in Section 11.1 actually works, we shall now prove this result, also proved in Section F.3 (or in [4, p. 280]). From (11.10):

$$\chi(t) = \sqrt{K\int_0^t K\,ds} = K\sqrt{t} \tag{11.25}$$

and

$$\chi'(t) = \frac{K}{2\sqrt{t}}. \tag{11.26}$$

The order $\nu(t)$ of the Bessel functions is then found from (11.9):

$$\begin{aligned}\nu(t) &= \sqrt{-\frac{\chi^3(t)}{f^2(t)}\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right]} = \sqrt{-\frac{K^3 t^{\frac{3}{2}}}{K^2}\frac{d}{dt}\left[\frac{1}{2K\sqrt{t}}\right]} \\ &= \sqrt{-Kt^{\frac{3}{2}}\left(-\frac{t^{-\frac{3}{2}}}{4K}\right)} = \sqrt{\frac{1}{4}} = \frac{1}{2}\end{aligned} \tag{11.27}$$

where both the time t and the constant K have vanished from the result. Simplifications of this kind (further examples will be given in Sections 11.3 and 12.4) are vital to make the mathematical investigations feasible. Since $\nu = \frac{1}{2}$, the relevant Bessel function is [6, p. 54]

$$J_{\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}}\sin x. \tag{11.28}$$

Thus, from (11.5), (11.27), and (11.28), the KL expansion follows:

$$\begin{aligned}X(t) &= \sqrt{K\int_0^t K\,ds}\sum_{n=1}^{\infty} Z_n N_n J_{\frac{1}{2}}\left(\gamma_n \frac{\int_0^t K\,ds}{\int_0^T K\,ds}\right) \\ &= K\sum_{n=1}^{\infty} Z_n N_n \sqrt{\frac{2T}{\pi\gamma_n}}\sin\left(\gamma_n \frac{t}{T}\right).\end{aligned} \tag{11.29}$$

In this expression the normalization constants N_n are yet to be found. To this end, we must know the γ_n given by (11.11). That is,

$$\frac{K}{2\sqrt{T}} J_{\frac{1}{2}}(\gamma_n) + K\sqrt{T}\left[\frac{K\gamma_n}{K\int_0^T ds} J'_{\frac{1}{2}}(\gamma_n)\right] = 0 \tag{11.30}$$

or, simplifying,

$$\tfrac{1}{2} J_{\frac{1}{2}}(\gamma_n) + \gamma_n J'_{\frac{1}{2}}(\gamma_n) = 0. \tag{11.31}$$

But this is a special case of the more general Bessel functions formula (see [5, p. 11, entry (54)]:

$$\nu J_\nu(z) + z J'_\nu(z) = z J_{\nu-1}(z) \tag{11.32}$$

so that (11.31) actually amounts to

$$J_{-\frac{1}{2}}(\gamma_n) = 0 \tag{11.33}$$

since $\gamma_n \neq 0$. One now has (see [6, p. 55, entry (6)])

$$J_{-\frac{1}{2}}(x) = \sqrt{\frac{2}{\pi x}} \cos x \tag{11.34}$$

so that (11.33) finally becomes the boundary condition:

$$\cos \gamma_n = 0. \tag{11.35}$$

In this case we find the exact γ_n expression to be

$$\gamma_n = n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots). \tag{11.36}$$

Reverting now to the normalization constants N_n, (11.12) yields

$$\begin{aligned} 1 &= N_n^2 \left[\int_0^T K\, ds\right]^2 \int_0^1 x \left[J_{\frac{1}{2}}(\gamma_n x)\right]^2 dx \\ &= N_n^2 K^2 T^2 \frac{2}{\pi\gamma_n} \int_0^1 \sin^2(\gamma_n x)\, dx \\ &= N_n^2 K^2 T^2 \frac{1}{\pi\gamma_n^2} [\gamma_n - \sin\gamma_n \cos\gamma_n] = \frac{N_n^2 K^2 T^2}{\pi\gamma_n} \end{aligned} \tag{11.37}$$

from which

$$N_n = \frac{\sqrt{\gamma_n}\sqrt{\pi}}{KT}. \tag{11.38}$$

As for the eigenvalues λ_n, from (11.13) they are given by

$$\lambda_n = \frac{K^2 T^2}{\gamma_n^2} \tag{11.39}$$

and these are also the variances of the independent Gaussian random variables Z_n. It is interesting to point out that the property

$$\sigma_{cZ}^2 = c^2 \sigma_Z^2 \tag{11.40}$$

and (11.39) yield the following proportionality among the proper-time random variables Z_n and the coordinate-time random variables Z_n^0—corresponding to the case $\nu(t) \equiv \frac{1}{2}$, or, from (11.16), $f(t) \equiv 1$:

$$Z_n = KZ_n^0. \tag{11.41}$$

Thus, the KL expansion of the proper-time Brownian motion is

$$B(\tau) = \sum_{n=1}^{\infty} Z_n \sqrt{\frac{2}{T}} \sin\left(\gamma_n \frac{t}{T}\right) = K \sum_{n=1}^{\infty} Z_n^0 \sqrt{\frac{2}{T}} \sin\left(\gamma_n \frac{t}{T}\right) = KB(t) \tag{11.42}$$

and (11.24) is found once again. In other words, passing from one inertial reference frame to another, the random variables Z_n just change their variance according to (11.41), whereas the time eigenfunctions remain the same. In Section 11.5 total energy will also be discussed.

11.3 DECELERATED MOTION

This and the remaining sections are devoted to the case when the proper time is proportional to a real positive power of the coordinate time, namely

$$\tau = Ct^{2H} \quad (t \geq 0) \tag{11.43}$$

C being a constant that will be determined immediately, and H being a real variable whose range has yet to be found. The factor 2 in the exponent is introduced for convenience. By checking (11.43) against (11.1), differentiating, and taking the square root, one gets

$$f(t) = \sqrt{2HC}\, t^{H-\frac{1}{2}}. \tag{11.44}$$

Inserting this into (11.17), the resulting velocity radical reads

$$v(t) = c\sqrt{1 - (2HC)^2 t^{2(2H-1)}}. \tag{11.45}$$

In order to have a real velocity, the inequality

$$(2HC)^2 t^{2(2H-1)} \leq 1 \tag{11.46}$$

must be valid. Moreover, the initial instant is conventionally zero, and the final instant is T, so that the range of H is necessarily greater than one-half. By setting $t = T$, the constant C is determined so as $v(T) = 0$, and one gets

$$C = \frac{1}{2HT^{2H-1}}. \tag{11.47}$$

One can now understand the physical meaning of the motion we are studying. Initially ($t = 0$) the spaceship is traveling at the speed of light. Then it starts decelerating until it stops at the final instant $t = T$. Actually, if we let H vary, we have a family of curves in the $t, v(t, H)$ plane. But we have to be careful: the tangent to all such curves at $t = 0$ must be *horizontal* in order to preserve the physical reality when the spaceship starts decelerating from c to lower speeds (i.e., there cannot be any sudden "speed jump"). Thus, differentiating (11.45)—with C given by (11.47)—with respect to t and then setting $t = 0$, one discovers that the condition on H given $H > \frac{1}{2}$ must physically be replaced by the stronger condition:

$$4H - 3 > 0 \quad \text{hence } H > \tfrac{3}{4} = 0.75. \tag{11.48}$$

An important special case of $v(t, H)$ occurs when $H = 1$: in fact, $v(t)$ is then the upper-right quarter of an ellipse. One also easily infers that, for $1 < H < \infty$, all $v(t)$ curves lie above this arc of ellipse. In the (physically meaningless) limit case $H \to \infty$ the $v(t, H)$ "curve" would be the upper-right quarter of a rectangle. Figure 11.1 shows this set of $v(t, H)$ curves representing the decelerated motion for different values of H.

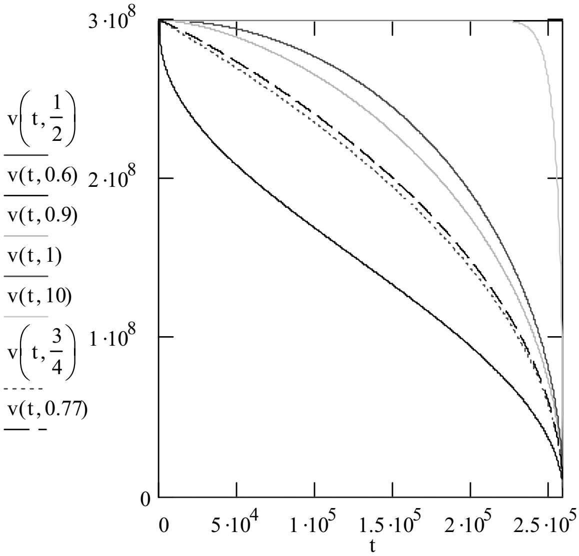


Figure 11.1. Decelerated motion of a relativistic spaceship approaching the Earth from the speed of light c down to speed zero in the finite time interval $0 \le t \le T$. We dubbed this spaceship the Independence Day (alien) spaceship. For instance, let $T = 3$ days of coordinate time (i.e., time elapsed on Earth). At the initial instant $t = 0$ (when the deceleration starts) all the curves $v(t, H)$ must have their tangents horizontal (to avoid bumps aboard the spaceship) and that yields the physical constraint: $H > \frac{3}{4} = 0.75$. The above plots show just this fact in a neat, graphic fashion: (1) all solid curves have $H > \frac{3}{4}$ and horizontal tangent at $t = 0$, so they are acceptable; (2) the dividing line is the dash-dotted curve corresponding to $H = \frac{3}{4}$, and one can see that it does not have a horizontal tangent at $t = 0$; (3) all the lower curves (dotted) are not allowed since they don't have a horizontal tangent at $t = 0$.

In conclusion, the function $f(t)$ is defined by the real positive power

$$f(t) = \frac{t^{H-\frac{1}{2}}}{T^{H-\frac{1}{2}}} \quad (0 \leq t \leq T). \tag{11.49}$$

From (11.17) and (11.49) we see that the velocity $v(t)$ is given by

$$v(t) = c\sqrt{1 - \left(\frac{t^{2H-1}}{T^{2H-1}}\right)^2} \quad (0 \leq t \leq T) \tag{11.50}$$

One can now understand the physical meaning of the motion we are studying. Initially ($t = 0$) the particle is traveling at the speed of light, then it starts decelerating until it stops at the final instant $t = T$. Actually, (11.50) represents a family of curves on the $(t, \nu(t))$ plane if we let H vary according to (11.48). The particular case $H = \frac{1}{2}$ represents standard Brownian motion. Another important special case of (11.50) occurs when $H = 1$: in fact $v(t)$ is then an ellipse. One also easily infers that, for $\frac{1}{2} < H < 1$ the curve lies below the arc of ellipse, whereas for $1 < H < \infty$ the curve lies above it. In the (physically meaningless) limit case $H \to \infty$ the curve would be half a rectangle.

Let us now turn to the KL expansion of the decelerated Brownian motion

$$X(t) = B\left(\frac{t^{2H}}{2HT^{2H-1}}\right) = \frac{1}{\sqrt{2H}T^{H-\frac{1}{2}}}B(t^{2H}). \tag{11.51}$$

Integrating (11.49), we get

$$\int_0^t f(s)\,ds = \frac{t^{H+\frac{1}{2}}}{(H+\frac{1}{2})T^{H-\frac{1}{2}}}. \tag{11.52}$$

Then, by virtue of (11.49) and (11.52), the function $\chi(t)$ defined by (11.10) reads:

$$\chi(t) = \frac{t^H}{\sqrt{H+\frac{1}{2}}\,T^{H-\frac{1}{2}}} \tag{11.53}$$

thus

$$\chi'(t) = \frac{Ht^{H-1}}{\sqrt{H+\frac{1}{2}}\,T^{H-\frac{1}{2}}}. \tag{11.54}$$

Moreover, from (11.49) and (11.54), one finds the expressions

$$\frac{\chi'(t)}{f^2(t)} = \frac{T^{H-\frac{1}{2}}Ht^{-H}}{\sqrt{H+\frac{1}{2}}} \tag{11.55}$$

and

$$\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] = \frac{T^{H-\frac{1}{2}}H(-H)t^{-H-1}}{\sqrt{H+\frac{1}{2}}} \tag{11.56}$$

and, from (11.49) and (11.53),

$$\frac{\chi^3(t)}{f^2(t)} = \frac{t^{H+1}}{(H+\frac{1}{2})\sqrt{H+\frac{1}{2}}\,T^{H-\frac{1}{2}}}. \tag{11.57}$$

The Bessel functions order can now be found from (11.9), (11.56), and (11.57):

$$\nu = \frac{2H}{2H+1}. \tag{11.58}$$

Note that both the time t and the constant T disappear identically, and the order of the Bessel functions is a constant, rather than a function of the time t. Moreover, by letting $H = \frac{3}{4}$ and $H \to \infty$, respectively, we see that the range of ν is rather limited: $\frac{3}{5} \leq \nu \leq 1$.

Our next task is to find the meaning of the constants γ_n. Upon substituting (11.52), (11.53), and (11.54) into (11.11), along with $\nu'(t) = 0$ one gets, after simplifying any multiplicative factors,

$$\frac{2H}{2H+1}J_\nu(\gamma_n) + \gamma_n J'_\nu(\gamma_n) = 0. \tag{11.59}$$

By virtue of (11.58), (11.59) is equivalent to

$$\nu J_\nu(\gamma_n) + \gamma_n J'_\nu(\gamma_n) = 0. \tag{11.60}$$

Once again the Bessel functions property (11.32) may be applied, and

$$\gamma_n J_{\nu-1}(\gamma_n) = 0. \tag{11.61}$$

Since $\gamma_n \neq 0$,

$$J_{\nu-1}(\gamma_n) = 0. \tag{11.62}$$

Thus, the γ_n are the real positive zeros, arranged in ascending order of magnitude, of the Bessel function of order $\nu - 1$. No formula yielding these zeros explicitly is known. Yet it is possible to find an approximated expression for them by means of the asymptotic formula for $J_\nu(x)$ (see [8, p. 134]).

$$\lim_{x\to\infty} J_\nu(x) = \lim_{x\to\infty}\sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right). \tag{11.63}$$

In fact, from (11.58) one first gets

$$\nu - 1 = -\frac{1}{2H+1}. \tag{11.64}$$

Second, (11.62) and (11.64), checked against (11.63), yield

$$0 = J_{\nu-1}(\gamma_n) \approx \sqrt{\frac{2}{\pi\gamma_n}} \cos\left(\gamma_n + \frac{\pi}{2(2H+1)} - \frac{\pi}{4}\right) \tag{11.65}$$

hence

$$\gamma_n + \frac{\pi}{2(2H+1)} - \frac{\pi}{4} \approx n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots) \tag{11.66}$$

and finally

$$\gamma_n \approx n\pi - \frac{\pi}{4} - \frac{\pi}{2(2H+1)} \quad (n = 1, 2, \ldots) \tag{11.67}$$

The first 32 approximated γ_n, obtained by means of (11.67), appear in Table 11.2, for various values of $H \geq \frac{1}{2}$. In the Brownian case $H = \frac{1}{2}$ (11.67) is an exact formula, in that it coincides with (11.36). We are reminded that these γ_n give the pace of convergence of the KL expansion, inasmuch as the standard deviations of the Gaussian random variables Z_n depend inversely on the γ_n by virtue of (11.13).

Eventually, the normalization constants N_n follow from (11.12) and (11.52):

$$1 = N_n^2 \frac{T^2}{(H+\frac{1}{2})^2} \int_0^1 x J_\nu^2(\gamma_n x)\, dx. \tag{11.68}$$

This integral is calculated within the framework of the Dini series (see [5, p. 71]) and the result is

$$\int_0^1 x J_\nu^2(\gamma_n x)\, dx = \frac{1}{2\gamma_n^2}[\gamma_n^2 J_\nu'^2(\gamma_n) + (\gamma_n^2 - \nu^2) J_\nu^2(\gamma_n)]. \tag{11.69}$$

This formula, however, may be greatly simplified upon eliminating $\gamma_n J_\nu'(\gamma_n)$ taken from (11.60). In fact, one finds

$$\gamma_n^2 J_\nu'^2(\gamma_n) = \nu^2 J_\nu^2(\gamma_n) \tag{11.70}$$

and (11.68), by virtue of (11.69) and (11.70), becomes

$$1 = N_n^2 \frac{T^2}{(H+\frac{1}{2})^2} \cdot \frac{J_\nu^2(\gamma_n)}{2}. \tag{11.71}$$

Thus

$$N_n = \frac{(H+\frac{1}{2})\sqrt{2}}{T|J_\nu(\gamma_n)|}. \tag{11.72}$$

This is the exact expression of the normalization constants. An approximated expression can be found upon inserting both (11.67) and (11.58) into the approximated

Table 11.2. Approximate values of the constants γ_n.

	$H = 0.5$ Brownian	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$	$H = 1.0$	$H = \infty$
$n = 1$	1.571	1.642	1.702	1.752	1.795	1.833	2.356
$n = 2$	4.712	4.784	4.843	4.894	4.937	4.974	5.498
$n = 3$	7.854	7.925	7.985	8.035	8.078	8.116	8.639
$n = 4$	11.00	11.07	11.13	11.18	11.22	11.26	11.78
$n = 5$	14.14	14.21	14.27	14.32	14.37	14.40	14.92
$n = 6$	17.28	17.36	17.41	17.46	17.50	17.54	18.06
$n = 7$	20.42	20.50	20.55	20.60	20.64	20.68	21.20
$n = 8$	23.56	23.63	23.69	23.74	23.79	23.82	24.35
$n = 9$	26.70	26.77	26.83	26.88	26.93	26.96	27.49
$n = 10$	27.84	27.92	27.98	30.03	30.07	30.11	30.63
$n = 11$	32.99	33.06	33.12	33.17	33.21	33.25	33.77
$n = 12$	36.13	36.20	36.26	36.31	36.35	36.39	36.91
$n = 13$	37.27	37.34	37.40	37.45	37.49	37.53	40.05
$n = 14$	42.41	42.48	42.54	42.59	42.64	42.67	43.20
$n = 15$	45.55	45.62	45.68	45.73	45.78	45.81	46.34
$n = 16$	48.69	48.77	48.83	48.88	48.92	48.96	47.48
$n = 17$	51.84	51.91	51.97	52.02	52.06	52.10	52.62
$n = 18$	54.98	55.05	55.11	55.16	55.20	55.24	55.76
$n = 19$	58.12	58.19	58.25	58.30	58.34	58.38	58.90
$n = 20$	61.26	61.33	61.39	61.44	61.48	61.52	62.05
$n = 21$	64.40	64.47	64.53	64.58	64.63	64.66	65.19
$n = 22$	67.54	67.62	67.67	67.72	67.77	67.81	68.33
$n = 23$	70.69	70.76	70.82	70.87	70.91	70.95	71.47
$n = 24$	73.83	73.90	73.96	74.01	74.05	74.09	74.61
$n = 25$	76.97	77.04	77.10	77.15	77.19	77.23	77.75
$n = 26$	80.11	80.18	80.24	80.29	80.33	80.37	80.90
$n = 27$	83.25	83.32	83.38	83.43	83.48	83.51	84.04
$n = 28$	86.39	86.46	86.52	86.57	86.62	86.66	87.18
$n = 29$	87.53	87.61	87.67	87.72	87.76	87.80	90.32
$n = 30$	92.68	92.75	92.81	92.86	92.90	92.94	93.46
$n = 31$	95.82	95.90	95.95	96.00	96.04	96.08	96.60
$n = 32$	98.96	97.0	97.0	97.1	97.1	97.2	97.75

(11.63) for $J_\nu(\gamma_n)$:

$$|J_\nu(\gamma_n)| \approx \left| \sqrt{\frac{2}{\pi\gamma_n}} \cos\left(n\pi - \frac{\pi}{4} - \frac{\pi}{2(2H+1)} - \frac{\pi 2H}{2(2H+1)} - \frac{\pi}{4}\right) \right|$$

$$\approx \sqrt{\frac{2}{\pi\gamma_n}} |\cos(n\pi - \pi)| \approx \sqrt{\frac{2}{\pi\gamma_n}}. \tag{11.73}$$

By substituting this into (11.72) and using (11.67) for the γ_n, it follows that

$$N_n \approx \frac{\pi}{T}(H + \tfrac{1}{2}) \sqrt{n - \frac{1}{4} - \frac{1}{2(2H+1)}}. \tag{11.74}$$

These are the approximated normalization constants.

A similar procedure applies to the eigenvalues λ_n. In fact, from (11.13) and (11.52) we get the exact formula

$$\lambda_n = \frac{T^2}{(H + \frac{1}{2})^2} \cdot \frac{1}{(\gamma_n)^2} \tag{11.75}$$

whereas from (11.75) and (11.67) we get the approximated formula

$$\lambda_n \approx \frac{T^2}{(H + \frac{1}{2})^2} \cdot \frac{1}{\pi^2 \left(n - \frac{1}{4} - \frac{1}{2(2H+1)}\right)^2}. \tag{11.76}$$

These are the variances of the independent Gaussian random variables Z_n.

Let us now summarize all the results found in the present section by writing two KL expansions: the exact one

$$X(t) = \frac{\sqrt{2H+1}\, t^H}{T^{H+\frac{1}{2}}} \sum_{n=1}^{\infty} Z_n \frac{1}{|J_\nu(\gamma_n)|} J_\nu\left(\gamma_n \frac{t^{H+\frac{1}{2}}}{T^{H+\frac{1}{2}}}\right) \tag{11.77}$$

and the approximated one

$$X(t) \approx \frac{\sqrt{2H+1}\, t^{\frac{H}{2} - \frac{1}{4}}}{T^{\frac{H}{2} + \frac{1}{4}}} \sum_{n=1}^{\infty} Z_n \cos\left(\gamma_n \frac{t^{H+\frac{1}{2}}}{T^{H+\frac{1}{2}}} - \frac{2H\pi}{2(2H+1)} - \frac{\pi}{4}\right). \tag{11.78}$$

11.4 CHECKING THE KLT OF DECELERATED MOTION BY MATLAB SIMULATIONS

Just look at Figure 11.2.

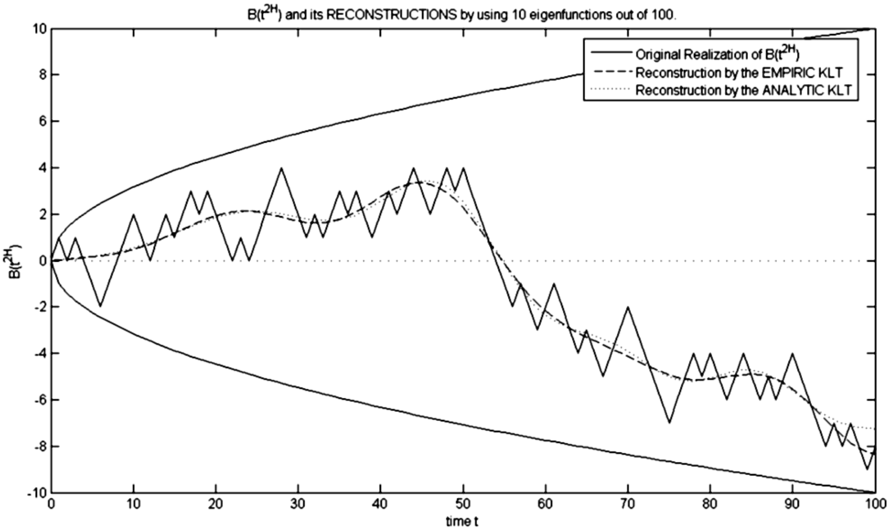


Figure 11.2. The time-rescaled Brownian motion $X(t)$ of (11.78) vs. time t simulated as a random walk over 100 time instants. This $X(t)$ represents the "noisy signal" received on Earth (whence the use of the coordinate time t = Earth time) from a relativistic spaceship approaching the Earth in a decelerated motion, as in the movie *Independence Day*. Next to the "bumpy curve" of $X(t)$, two more "smooth curves" are shown that *interpolate at best* the bumpy $X(t)$. These two curves are the KLT reconstruction of $X(t)$ *by using the first ten eigenfunctions only*. It is important to note that the two smooth curves are *different* in this case because the KLT expansion (11.78) is *approximated*. Actually, it is an approximated KLT expansion because the asymptotic expansion of the Bessel functions (11.63) was used. So, the two curves are different from each other, but both still interpolate $X(t)$ at best. Note that, were we taking into account the full set of 100 KLT eigenfuctions—rather than just 10—then the *empirical* reconstruction would overlap $X(t)$ *exactly*, but the *analytic* reconstruction would not because of the use of the asymptotic expansion (11.63) of the Bessel functions.

11.5 TOTAL ENERGY OF THE NOISY SIGNAL FROM RELATIVISTIC SPACESHIPS IN DECELERATED AND UNIFORM MOTION

A thorough study of the total energy of the noisy signals emitted by relativistic spaceships in decelerated motion (and of the uniform motion, in particular) is allowed by the results obtained in Sections 11.2, 11.3, and F.10 in Appendix F.

Our first goal will be to get the characteristic function (i.e., the Fourier transform) of the random variable "total energy", defined by (F.47). In fact, inserting the eigenvalues (11.75) into (F.51), it follows that

$$\Phi_{\varepsilon}(\zeta) = \left[\prod_{n=1}^{\infty}\left(1 - \frac{2iT^2\zeta}{(H+\frac{1}{2})^2\gamma_n^2}\right)\right]^{-\frac{1}{2}}. \tag{11.79}$$

On the other hand,

$$J_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\Gamma(\nu+1)} \cdot \prod_{n=1}^{\infty}\left(1 - \frac{z^2}{j_{\nu,n}^2}\right) \tag{11.80}$$

is the infinite product expansion for $J_\nu(z)$ [6, p. 498], and the constants $j_{\nu,n}$ evidently are the real positive zeros of $J_\nu(z)$, arranged in ascending order of magnitude. Then, keeping in mind (11.62), we can let the two infinite products (11.79) and (11.80) coincide by setting

$$\gamma_n = j_{\nu-1,n} \tag{11.81}$$

and

$$z^2 = \frac{2iT^2\zeta}{(H+\frac{1}{2})^2} \quad \text{or} \quad z = T\frac{\sqrt{2i\zeta}}{H+\frac{1}{2}}. \tag{11.82}$$

Solving for $\Phi_\varepsilon(\zeta)$, one gets

$$\Phi_\varepsilon(\zeta) = \frac{1}{\sqrt{\Gamma(\nu)\left[\frac{T\sqrt{2i\zeta}}{2H+1}\right]^{1-\nu} J_{\nu-1}\left(\frac{T\sqrt{2i\zeta}}{H+\frac{1}{2}}\right)}} \tag{11.83}$$

which is the exact expression for the characteristic function of the total energy distribution, ε. An approximated expression can also be derived using the asymptotic expression for the Bessel function (11.63); one then gets

$$\Phi_\varepsilon(\zeta) \approx \frac{1}{\sqrt{\frac{\Gamma(\nu)}{\sqrt{\pi}}\left[\frac{T\sqrt{2i\zeta}}{2H+1}\right]^{\frac{1}{2}-\nu} \cos\left(\frac{T\sqrt{2i\zeta}}{H+\frac{1}{2}} - \frac{\nu\pi}{2} + \frac{\pi}{4}\right)}}. \tag{11.84}$$

In the standard Brownian case $H = \frac{1}{2}$ (hence $\nu = \frac{1}{2}$ and one can apply the formula $\Gamma(\frac{1}{2}) = \sqrt{\pi}$), both (11.83) and (11.84) become

$$\Phi_\varepsilon(\zeta) = \frac{1}{\sqrt{\cos(T\sqrt{2i\zeta})}}. \tag{11.85}$$

This result is due to Cameron and Martin, who published it in 1944 [9].

Our next goal is the computation of all the total energy cumulants, given by (F.56). To this end, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{(\gamma_n)^{2k}} \equiv S_{2k,\nu-1} \equiv \sigma_{\nu-1}^{(k)} \quad (k = 1, 2, \ldots) \tag{11.86}$$

where the notation $S_{2k,\nu-1}$ is used on [5, p. 61], while the notation $\sigma_{\nu-1}^{(k)}$ is used on [6, p. 502]. Then

$$\sum_{k=1}^{\infty} S_{2k,\nu-1} x^{2k-1} = \frac{J_\nu(x)}{2J_{\nu-1}(x)} \tag{11.87}$$

is the power series in x, with coefficients $S_{2k,\nu-1}$, whose proof is given on [5, p. 61].

From the formula that yields any coefficient of a power series, it follows that the coefficients $S_{2k,\nu-1}$ of the power series in x on the left side of (11.87) are given by

$$S_{2k,\nu-1} = \frac{1}{(2k-1)!} \lim_{x\to 0^+} \left[\frac{d^{2k-1}}{dx^{2k-1}} \left(\frac{J_\nu(x)}{2J_{\nu-1}(x)} \right) \right] \tag{11.88}$$

and the sum of the series (11.86) is obtained. Finally, by virtue of (F.56) and (11.88) we conclude that all the cumulants of the total energy are

$$\begin{aligned} K_n &= \frac{2^{n-1}T^{2n}}{(H+\frac{1}{2})^{2n}} \cdot \frac{(n-1)!}{(2n-1)!} \lim_{x\to 0^+} \left[\frac{d^{2n-1}}{dx^{2n-1}} \left(\frac{J_\nu(x)}{2J_{\nu-1}(x)} \right) \right] \\ &= \frac{2^{n-1}T^{2n}}{(H+\frac{1}{2})^{2n}} \cdot (n-1)! \cdot \sigma_{\nu-1}^{(n)} \quad (n = 1, 2, \ldots) \end{aligned} \tag{11.89}$$

where the quantities $\sigma_{\nu-1}^{(1)}$, $\sigma_{\nu-1}^{(2)}$, $\sigma_{\nu-1}^{(3)}$, $\sigma_{\nu-1}^{(4)}$, $\sigma_{\nu-1}^{(5)}$, and $\sigma_{\nu-1}^{(6)}$ appear on [6, p. 502]—ν is to be replaced by H via (11.58).

Having found all the cumulants, we can now derive the expressions of the most interesting statistical parameters of the total energy ε.

(1) Mean value of the total energy:

$$K_1 = E\{\varepsilon\} = \frac{T^2}{2H(2H+1)}. \tag{11.90}$$

(2) Variance of the total energy:

$$K_2 = \sigma_\varepsilon^2 = \frac{T^4}{2H^2(2H+1)(4H+1)}. \tag{11.91}$$

(3) Third total energy cumulant:

$$K_3 = \frac{T^6}{H^3(2H+1)(3H+1)(4H+1)}. \tag{11.92}$$

(4) Fourth total energy cumulant:

$$K_4 = \frac{3(11H+3)T^8}{H^4(2H+1)(3H+1)(4H+1)^2(8H+3)}. \tag{11.93}$$

(5) Skewness of the total energy distribution:

$$\frac{K_3}{(K_4)^{\frac{3}{2}}} = \frac{2^{\frac{3}{2}}\sqrt{2H+1}\sqrt{4H+1}}{3H+1}. \tag{11.94}$$

(6) Kurtosis (or excess) of the total energy distribution:

$$\frac{K_4}{(K_2)^2} = \frac{12(2H+1)(11H+3)}{(3H+1)(8H+3)}. \tag{11.95}$$

Since $H \geq \frac{1}{2}$ we infer from (11.94) that the skewness ranges from $\frac{8}{5}\sqrt{3} = 2.7712813$ (for $H = \frac{1}{2}$) to $\frac{8}{3} = 2.6666667$ for $H \to \infty$. In addition, from (11.95) we find that the kurtosis ranges from $\frac{408}{35} = 11.657143$ for $H = \frac{1}{2}$ to 11 for $H \to \infty$. Therefore, we may conclude that the total energy peak is narrow for any $H \geq \frac{1}{2}$.

The ordinary Brownian motion case of all the previous results is noteworthy, and, relativistically speaking, corresponds to the uniform motion of the moving reference frame with zero velocity (i.e., no motion at all). In fact, by substituting $H = \frac{1}{2}$, $\nu = \frac{1}{2}$ and both (11.28) and (11.34) into (11.89), we find all the Brownian motion total energy cumulants

$$K_n = 2^{n-2} T^{2n} (n-1)! \frac{1}{(2n-1)!} \lim_{x \to 0^+} \left[\frac{d^{2n-1} \tan x}{dx^{2n-1}} \right]. \tag{11.96}$$

Evidently, the last two terms are the $(2n-1)$th coefficient in the MacLaurin expansion of $\tan x$, that reads [5, p. 51]

$$\tan x = \sum_{n=1}^{\infty} \frac{1}{(2n)!} 2^{2n} (2^{2n} - 1)(-1)^{n+1} B_{2n} x^{2n-1}, \tag{11.97}$$

where the B_{2n} are the Bernoulli numbers, a table of which is found, for instance, on [7, p. 810]. Thus, by inserting the coefficients of (11.97) into (11.96), we get all the cumulants of the total energy of standard Brownian motion:

$$K_n = T^{2n} \frac{(n-1)!}{(2n)!} 2^{3n-2} (2^{2n} - 1)(-1)^{n+1} B_{2n}. \tag{11.98}$$

In particular, we have:

(1) mean value of the total energy

$$K_1 = E\{\varepsilon\} = \frac{T^2}{2}; \tag{11.99}$$

(2) variance of the total energy

$$K_2 = \sigma_\varepsilon^2 = \frac{T^4}{3}; \tag{11.100}$$

(3) skewness of the total energy distribution

$$\text{skewness} = \frac{8}{5}\sqrt{3} = 2.7712812921102; \tag{11.101}$$

(4) kurtosis (or excess) of the total energy distribution

$$\text{kurtosis} = \frac{408}{35} = 11.657. \tag{11.102}$$

11.6 *INDEPENDENCE DAY* MOVIE: EXPLOITING THE KLT TO DETECT AN ALIEN SPACESHIP APPROACHING THE EARTH IN DECELERATED MOTION

Everybody remembers the 1996 movie *Independence Day* (see *http://en.wikipedia.org/wiki/Independence_Day_%28film%29*): huge alien spaceships first appear close to Moon and move slowly to prepare for the final attack! It is to be believed, however, that if they move slowly when they are at the Moon distance, they must have moved much, much faster when they were in the open interstellar space in order to cover the vast interstellar distances (please note that here we stick to special relativity only, and do not wish to consider "exotic" mathematical tricks like wormholes, stemming out of general relativity).

In other words, the alien spaceships must have *decelerated* in some way from (say) the speed of light c to zero speed with respect to the Earth. Well, in this section we are going to study the *decelerated* signals emitted by the aliens while they approach the Earth, and work out some equations about the energy of such signals that might help us to dectect an alien invasion much in advance thanks to the KLT developed in this chapter (in the movie *Independence Day*, on the contrary, aliens are already at the Moon distance when humans detect them!).

To adjust our theory to the problem, first consider a trivial Newtonian problem: How long would it take to decelerate from speed c to 0 at the uniform deceleration of just $1g = 9.8\,\mathrm{m/s}^{-2}$? The trivial calculation yieds about 1 year (in Earth time) and the distance at which the deceleration must start is 30,000 AU, or about half a light year (Oort cloud distance) . Should aliens and/or their gadgets withstand decelerations of $2g$, the overall deceleration time would take about half a year, and it should start at the closer distance of 7,600 AU = 0.12 lt-yr from Earth.

Let us now go back to the relativistic decelerated speed $v(t)$ given by (11.50) and consider the radial distance $r(t)$ covered by the spacecraft during the deceleration phase:

$$\frac{dr(t)}{dt} = v(t) = c\sqrt{1 - \left(\frac{t^{2H-1}}{T^{2H-1}}\right)^2}; \tag{11.103}$$

that is

$$R_H(t) = \int_0^{T_H} r(t)\,dt = \int_0^{T_H} c\sqrt{1 - \left(\frac{t^{2H-1}}{T^{2H-1}}\right)^2}\,dt. \tag{11.104}$$

Unfortunately, this integral cannot be computed in a closed form, and we are thus prevented from fully extending our investigation to any value of H larger than $\frac{3}{4}$. We shall thus confine ourselves to the two values $H = \frac{3}{4}$ and $H = 1$, for which one finds

$$R_{\frac{3}{4}}(T) = \int_0^{T_H} r(t)\,dt = \int_0^{T_H} c\sqrt{1 - \frac{t}{T}}\,dt = \tfrac{2}{3}cT_{\frac{3}{4}} = 0.66cT_{\frac{3}{4}} \tag{11.105}$$

and

$$R_1(T) = \int_0^{T_H} r(t)\,dt = \int_0^{T_H} c\sqrt{1 - \left(\frac{t}{T}\right)^2}\,dt = \frac{\pi}{4} cT_1 = 0.78cT_1, \quad (11.106)$$

respectively.

Next we are going to focus only on (11.105) because this is the case where the deceleration of the alien spacecraft is "smoothest" (i.e., less sudden).

The total mean energy emitted by the alien spacecraft in the form of electromagnetic waves (= signals + noise) during the time $T_{\frac{3}{4}}$ is given by (11.90) with $H = \frac{3}{4}$; that is

$$K_1 = E\{\varepsilon\} = \frac{T_{\frac{3}{4}}^2}{2H(2H+1)} = \frac{4}{15} T_{\frac{3}{4}}^2 = 0.266 T_{\frac{3}{4}}^2. \quad (11.107)$$

The variance of the total energy is given by (11.91) again with $H = \frac{3}{4}$

$$K_2 = \sigma_\varepsilon^2 = \frac{T_{\frac{3}{4}}^4}{2H^2(2H+1)(4H+1)} = \frac{4}{45} T_{\frac{3}{4}}^4 = 0.088 T_{\frac{3}{4}}^4. \quad (11.108)$$

Thus, the total mean energy of the electromagnetic waves emitted by the approaching alien spacecraft lies within the range

$$E\{\varepsilon\} \pm \sigma_\varepsilon = \frac{4}{15} T_{\frac{3}{4}}^2 \pm \frac{2}{3\sqrt{5}} T_{\frac{3}{4}}^2 = (0.266 \pm 0.298) T_{\frac{3}{4}}^2. \quad (11.109)$$

This is the "energy bandwidth" upon which any detector of electromagnetic radiation emitted by the alien spacecraft must be built.

The topics discussed in this section were first presented by the author in October 1994 at the *International Astronautical Congress*, held in Jerusalem [10].

11.7 REFERENCES

[1] P. G. Bergmann, *Introduction to the Theory of Relativity*, Prentice-Hall, Englewood Cliffs, NJ, 1942.

[2] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 213–217.

[3] L. Arnold, *Stochastic Differential Equations*, Wiley, New York, 1973.

[4] R. Ash, *Information Theory*, Interscience, New York, 1965.

[5] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1953.

[6] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1966.

[7] M. Abramowitz, and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1970.

[8] N. N. Lebedev, *Special Functions and Their Applications*, Dover, New York, 1972.

[9] R. H. Cameron, and W. R. Martin, "The Wiener Measure of Hilbert Neighborhoods in the Space of Real Continuous Functions," *Journal of Mathematics and Physics of the M.I.T.*, **23** (1944), 195–201.

[10] C. Maccone, "Mathematical Algorithm to Detect Alien Spacecraft Approaching the Earth," paper IAA-94-IAA.4.1.656 presented at the *International Astronautical Congress held in Jerusalem, Israel, October 11, 1994*.

12

KLT of radio signals from relativistic spaceships in hyperbolic motion

12.1 INTRODUCTION

A spaceship, traveling at a constant acceleration g in its own reference frame, exemplifies the relativistic interstellar flight. If a Gaussian noise (Brownian motion) is emitted in units of the spaceship's proper time, it undergoes a time rescaling when measured in units of the coordinate time. This noise is studied in this chapter in terms of its KL expansion. All topics discussed in this chapter were first published by the author between 1988 and 1990 [1, 2].

12.2 HYPERBOLIC MOTION

A classical topic in special relativity is the so-called *hyperbolic motion*, first considered by Minkowski in 1908 [3], and discussed in most textbooks (see [4, p. 41].[1] Spaceflight did not exist in the time of Minkowski, so he believed that his formulas about the hyperbolic motion could only be applied to the physics of elementary particles then known to exist, such as electrons. Here, however, we shall give the topic of hyperbolic motion a space-travel cut, in view of the applications to telecommunications that will be made in the rest of this book.

Imagine a spacecraft traveling faster and faster with respect to its own reference frame, so that the crew experience a constant acceleration that, for their maximum comfort, we assume numerically equal to $g = 9.8\ \text{m/s}^2$. The longitudinal force (see [5, p. 205]) is

$$f_{\|} = \left[1 - \frac{v^2(t)}{c^2}\right]^{-\frac{3}{2}} m \frac{dv(t)}{dt} \tag{12.1}$$

[1] The adjective "hyperbolic" refers to the fact that the $x(t)$ curve in the (x, t) plane is a hyperbola—given by Equation (13.24)—and that hyperbolic functions are used in the analysis.

and so we must find the unknown $v(t)$ in the differential equation

$$\left[1 - \frac{v^2(t)}{c^2}\right]^{-\frac{3}{2}} m \frac{dv(t)}{dt} = mg. \tag{12.2}$$

Separating the variables, and setting $v(t) = c \sin \Omega(t)$, one easily finds

$$\Omega(t) = \arctan\left(\frac{g}{c} t\right), \tag{12.3}$$

whence

$$v(t) = c \sin\left[\arctan\left(\frac{g}{c} t\right)\right] \tag{12.4}$$

but

$$\sin[\arctan x] = \sqrt{\frac{x^2}{1 + x^2}} \tag{12.5}$$

so that the velocity $v(t)$ in (11.16) is given by

$$v(t) = \frac{gt}{\sqrt{1 + \left(\frac{g}{c} t\right)^2}}. \tag{12.6}$$

Note that as $t \to \infty$, (12.6) gives $v(t) \to c$, as one would expect. The function $f(t)$ for the hyperbolic motion is then found from (11.16) and (12.6)

$$f(t) = \frac{1}{\left[1 + \left(\frac{g}{c} t\right)^2\right]^{\frac{1}{4}}}. \tag{12.7}$$

Unfortunately, it is quite difficult to handle this function. For instance, its integral

$$\int \frac{dx}{[1 + x^2]^{\frac{1}{4}}} \tag{12.8}$$

can be shown to be expressed by hypergeometric functions inasmuch as it is a binomial integral, but not of an elementary type. Thus, we will not attempt to study (12.7) directly, but shall consider its asymptotic expansion in Section 12.4.

A few more results, however, can still be derived from (12.7). In fact, one has (see [6, p. 86])

$$\begin{aligned}\tau(t) = \int_0^t f^2(s)\, ds = \int_0^t \frac{ds}{\sqrt{1 + \left(\frac{g}{c} s\right)^2}} = \frac{c}{g} \operatorname{arcsinh}\left(\frac{g}{c} t\right) \\ = \frac{c}{g} \ln\left[\frac{g}{c} t + \sqrt{1 + \left(\frac{g}{c} t\right)^2}\right].\end{aligned} \tag{12.9}$$

Thus, the time-rescaled Brownian motion corresponding to the hyperbolic motion of special relativity is

$$X(t) = B(\tau) = B\left(\frac{c}{g}\operatorname{arcsinh}\left(\frac{g}{c}t\right)\right)$$
$$= B\left(\frac{c}{g}\ln\left[\frac{g}{c}t + \sqrt{1 + \left(\frac{g}{c}t\right)^2}\right]\right). \tag{12.10}$$

We shall simply refer to it as the hyperbolic motion.

12.3 TOTAL ENERGY OF SIGNALS FROM RELATIVISTIC SPACESHIPS IN HYPERBOLIC MOTION

In this section we shall show that it is possible (by virtue of the formulas derived in Appendix F) to compute both the mean total energy and total energy variance of the signals emitted by relativistic spaceships in hyperbolic motion.

Let us start with the mean total energy (F.60). This, by substituting (12.9), takes the form of the definite integral

$$E\{\varepsilon\} = \int_0^T dt \int_0^t f^2(s)\, ds = \frac{c}{g}\int_0^T \operatorname{arcsinh}\left(\frac{g}{c}t\right) dt$$
$$= \frac{c^2}{g^2}\left[x\operatorname{arcsinh}(x) - \sqrt{1+x^2}\right]_0^{\frac{g}{c}T} \tag{12.11}$$

where we make use of the substitution $(gt)/c = x$ and of [6, p. 88, entry 4.6.43]. Thus, the mean total energy of the hyperbolic motion is

$$E\{\varepsilon\} = \frac{c^2}{g^2}\left[\frac{gT}{c}\operatorname{arcsinh}\left(\frac{gT}{c}\right) - \sqrt{1 + \left(\frac{gT}{c}\right)^2} + 1\right]. \tag{12.12}$$

It is also possible to derive a closed-form expression for the total energy variance starting from (F.62) and (12.9), but the calculations are more involved. To this end, let us first note that

$$\int \operatorname{arcsinh}^2(s)\, ds = s\operatorname{arcsinh}^2(s) - 2\sqrt{1+s^2}\operatorname{arcsinh}(s) + 2s + C \tag{12.13}$$

This result can be used to prove the more complicated expression

$$\int x\operatorname{arcsinh}^2(x)\, dx = \tfrac{1}{2}x^2\operatorname{arcsinh}^2(x) - x\sqrt{1+x^2}\operatorname{arcsinh}(x)$$
$$+ \frac{x^2}{2} + \int \sqrt{1+x^2}\operatorname{arcsinh}(x)\, dx + C. \tag{12.14}$$

This leads us to compute a further integral

$$\int \sqrt{1+x^2}\,\operatorname{arcsinh}(x)\,dx = \tfrac{1}{4}[(2\sqrt{1+x^2}\,\operatorname{arcsinh}(x) - x)x + \operatorname{arcsinh}^2(x)] + C. \tag{12.15}$$

These preliminary results enable us to tackle σ_ε defined in (F.62) using (12.9)

$$\sigma_\varepsilon^2 = 4\int_0^T dt \int_0^t du \left[\int_0^u f^2(s)\,ds\right]^2 = 4\left(\frac{c}{g}\right)^2 \int_0^T dt \int_0^t du\, \operatorname{arcsinh}^2\left(\frac{g}{c}u\right). \tag{12.16}$$

Now (12.13) and the substitution $((g/c)u = s)$ change this into

$$\begin{aligned}\sigma_\varepsilon^2 &= 4\left(\frac{c}{g}\right)^3 \int_0^T dt \left[s\,\operatorname{arcsinh}^2(s) - 2\sqrt{1+s^2}\,\operatorname{arcsinh}(s) + 2s\right]_0^{\left(\frac{g}{c}t\right)} \\ &= 4\left(\frac{c}{g}\right)^3 \left[\int_0^T \left(\frac{g}{c}t\right)\operatorname{arcsinh}^2\left(\frac{g}{c}t\right)\,dt - 2\int_0^T \sqrt{1+\left(\frac{g}{c}t\right)^2}\operatorname{arcsinh}\left(\frac{g}{c}t\right)\,dt + 2\int_0^T \frac{g}{c}t\,dt\right].\end{aligned}$$

The further substitution $(g/c)t = x$ and (12.14) yield

$$\begin{aligned}\sigma_\varepsilon^2 = 4\left(\frac{c}{g}\right)^4 \Bigg[&\frac{1}{2}\left(\frac{g}{c}T\right)^2 \operatorname{arcsinh}^2\left(\frac{g}{c}T\right) - \left(\frac{g}{c}T\right)\sqrt{1+\left(\frac{g}{c}T\right)^2}\operatorname{arcsinh}\left(\frac{g}{c}T\right) + \frac{3}{2}\left(\frac{g}{c}T\right)^2 \\ &- \int_0^{\left(\frac{g}{c}T\right)} \sqrt{1+x^2}\,\operatorname{arcsinh}(x)\,dx\Bigg]\end{aligned}$$

hence the integral (12.15) finally yields

$$\begin{aligned}\sigma_\varepsilon^2 = \left(\frac{c}{g}\right)^4 \Bigg[&2\left(\frac{g}{c}T\right)^2 \operatorname{arcsinh}^2\left(\frac{g}{c}T\right) - 4\left(\frac{g}{c}T\right)\sqrt{1+\left(\frac{g}{c}T\right)^2}\operatorname{arcsinh}\left(\frac{g}{c}T\right) \\ &+6\left(\frac{g}{c}T\right)^2 - 2\left(\frac{g}{c}T\right)\sqrt{1+\left(\frac{g}{c}T\right)^2}\operatorname{arcsinh}\left(\frac{g}{c}T\right) + \left(\frac{g}{c}T\right)^2 - \operatorname{arcsinh}\left(\frac{g}{c}T\right)\Bigg].\end{aligned}$$

Rearranging, the total energy variance for the hyperbolic motion is obtained

$$\begin{aligned}\sigma_\varepsilon^2 = \left(\frac{c}{g}\right)^4 \Bigg\{&\left[2\left(\frac{g}{c}T\right)^2 - 1\right]\operatorname{arcsinh}^2\left(\frac{g}{c}T\right) \\ &-6\left(\frac{g}{c}T\right)\sqrt{1+\left(\frac{g}{c}T\right)^2}\operatorname{arcsinh}\left(\frac{g}{c}T\right) + 7\left(\frac{g}{c}T\right)^2\Bigg\}.\end{aligned} \tag{12.17}$$

12.4 KLT FOR SIGNALS EMITTED IN ASYMPTOTIC HYPERBOLIC MOTION

The obvious asymptotic formula

$$\lim_{x\to\infty} \sqrt{1+x^2} = \lim_{x\to\infty} x$$

and its consequence

$$\lim_{x\to\infty} \ln[x + \sqrt{1+x^2}] = \lim_{x\to\infty} \ln[2x] \tag{12.18}$$

form the starting point to investigate the asymptotic hyperbolic motion. In fact, from (12.10), we see that, when $t \to \infty$, $X(t)$ approaches

$$B\left(\frac{c}{g}\ln\left(2\frac{g}{c}t\right)\right). \tag{12.19}$$

This we shall call the asymptotic hyperbolic motion and shall study it thoroughly.

By comparing (12.19) against (F.40), we immediately find

$$\int_0^t f^2(s)\,ds = \frac{c}{g}\ln\left(2\frac{g}{c}t\right). \tag{12.20}$$

Then, differentiating and taking the square root, we are led to

$$f(t) = \sqrt{\frac{c}{g}}\frac{1}{\sqrt{t}}. \tag{12.21}$$

This is the $f(t)$ function for the hyperbolic motion.

Integrating (12.21), one then gets

$$\int_0^t f(s)\,ds = 2\sqrt{\frac{c}{g}}\sqrt{t}. \tag{12.22}$$

By virtue of (12.21) and (12.22) the $\chi(t)$ function defined by (11.10) reads

$$\chi(t) = \sqrt{f(t)\int_0^t f(s)\,ds} = \sqrt{2\frac{c}{g}} \tag{12.23}$$

a constant. This circumstance is vital in order to develop the asymptotic hyperbolic case, inasmuch as it simplifies things greatly. In fact, from

$$\chi'(t) = 0 \tag{12.24}$$

and from (11.9), it can be seen at once that $\nu(t)$ vanishes identically

$$\nu(t) = 0 \tag{12.25}$$

(i.e., the order of the Bessel functions is zero). Thus, the KL expansion is given by functions of the form

$$J_0\left(\gamma_n \frac{\int_0^t f(s)\,ds}{\int_0^T f(s)\,ds}\right) = J_0\left(\gamma_n\frac{\sqrt{t}}{\sqrt{T}}\right). \tag{12.26}$$

Our next task is to find the meaning of the constants γ_n, formally given as the real positive zeros of (11.11). Letting $\chi'(t) = 0$ and $\nu'(t) = 0$, and getting rid of all multiplicative factors, one easily sees that (11.11) simplifies to

$$J_0'(\gamma_n) = 0. \tag{12.27}$$

Thus, the γ_n are the positive zeros, arranged in ascending order of magnitude, of the derivative of $J_0(x)$. In other words, they are the abscissas of the maxima and minima of $J_0(x)$, which are known to follow each other alternately. However, a different interpretation of the γ_n follows from the Bessel function property (see [7, p. 12, entry (55) (set $\nu = 0$)]

$$J'_\nu(x) = \frac{\nu}{x} J_\nu(x) - J_{\nu+1}(x). \tag{12.28}$$

In fact, (12.27) now becomes equivalent to

$$J_1(\gamma_n) = 0 \tag{12.29}$$

and one may also say that the γ_n are the real positive zeros of $J_1(x)$. The first 40 among them are listed in [8, p. 748], and one finds, for instance,

$$\gamma_1 = 3.8317060 \quad \gamma_2 = 7.0155867 \quad \gamma_{40} = 126.4461387. \tag{12.30}$$

No explicit formula yielding these zeros exactly is known. However, it is possible to get an approximated expression by setting $\nu = 1$ into the asymptotic formula for $J_\nu(x)$ (see [9, p. 134])

$$\lim_{x\to\infty} J_\nu(x) = \lim_{x\to\infty} \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \tag{12.31}$$

from which

$$\cos\left(\gamma_n - \frac{3\pi}{4}\right) \approx 0 \tag{12.32}$$

or

$$\gamma_n - \frac{3\pi}{4} \approx n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots). \tag{12.33}$$

Thus,

$$\gamma_n \approx n\pi + \frac{\pi}{4}. \tag{12.34}$$

We may see how good this approximation is by setting $n = 1, 2, \ldots, 40$

$$\gamma_1 \approx 3.9269908 \quad \gamma_2 \approx 7.0685835 \quad \gamma_{40} \approx 126.4491 \tag{12.35}$$

and checking these results against (12.30). Of course, the agreement improves with increasing n. As for the eigenvalues λ_n, they are related to the γ_n by (11.13)

$$\lambda_n = \frac{4cT}{g} \frac{1}{(\gamma_n)^2} \tag{12.36}$$

and are also variances of the independent Gaussian random variables Z_n.

Finally, we turn to the normalization constants N_n that are obtained from (11.12) after inserting (12.22) and (12.25). The resulting condition for N_n is

$$1 = N_n^2 \frac{4cT}{g} \int_0^1 x[J_0(\gamma_n x)]^2 \, dx. \tag{12.37}$$

This integral of (12.37) is calculated within the framework of the Dini expansion in series of Bessel functions (see [7, p. 71]), and one finds

$$1 = N_n^2 \frac{4cT}{g} \left\{ \frac{1}{2} \left[J_0'^2(\gamma_n) + \left(1 - \frac{0}{\gamma_n^2}\right) J_0^2(\gamma_n) \right] \right\}$$
$$= N_n^2 \frac{2cT}{g} [J_0'^2(\gamma_n) + J_0^2(\gamma_n)] = N_n^2 \frac{2cT}{g} J_0^2(\gamma_n) \tag{12.38}$$

where (12.27) was used in the last step. Solving with respect to N_n requires the introduction of the modulus of $J_0(\gamma_n)$, and one has

$$N_n = \frac{\sqrt{g}}{\sqrt{2cT} |J_0(\gamma_n)|}. \tag{12.39}$$

This is the exact expression of the normalization constants.

For an approximated expression for N_n, we substitute the Bessel function in its asymptotic form (12.31) with γ_n given in (12.34):

$$|J_0(\gamma_n)| \approx \left| \sqrt{\frac{2}{\pi \gamma_n}} \cos\left(\gamma_n - \frac{\pi}{4}\right) \right| = \sqrt{\frac{2}{\pi \gamma_n}} |\cos(n\pi)| = \sqrt{\frac{2}{\pi \gamma_n}}. \tag{12.40}$$

Then, from (12.39) and (12.40) we get the approximated N_n:

$$N_n \approx \frac{\pi}{2} \sqrt{\frac{g}{cT}} \sqrt{n + \tfrac{1}{4}}. \tag{12.41}$$

All the results obtained in this section may now be summarized by writing the exact KL expansion

$$B\left(\frac{c}{g} \ln\left(2 \frac{g}{c} t\right)\right) = \sum_{n=1}^{\infty} Z_n \frac{\sqrt{c}}{\sqrt{g}} \cdot \frac{1}{\sqrt{2}\sqrt{T} |J_0(\gamma_n)|} J_0\left(\gamma_n \frac{\sqrt{t}}{\sqrt{T}}\right) \tag{12.42}$$

and the approximated expansion—found by virtue of (12.31) and (12.41)

$$B\left(\frac{c}{g} \ln\left(2 \frac{g}{c} t\right)\right) = \sum_{n=1}^{\infty} Z_n \frac{\sqrt{c}}{\sqrt{g}} \cdot \frac{1}{\sqrt{2} T^{\frac{1}{4}} t^{\frac{1}{4}}} \cos\left(\gamma_n \frac{\sqrt{t}}{\sqrt{T}} - \frac{\pi}{4}\right). \tag{12.43}$$

The physical range of validity of (12.42) and (12.43) is provided by the relativistic condition (11.7). Since, from (12.21)

$$f^4(t) = \frac{c^2}{g^2} \cdot \frac{1}{t^2}, \tag{12.44}$$

(11.7) yields the velocity of the asymptotic hyperbolic motion

$$v(t) = c \sqrt{1 - \frac{c^2}{g^2} \cdot \frac{1}{t^2}}. \tag{12.45}$$

In order to have a non-negative radicand, the inequality

$$\frac{c^2}{g^2} \cdot \frac{1}{t^2} \le 1 \tag{12.46}$$

must hold, meaning

$$t \geq \frac{c}{g} = 3.0612245 \cdot 10^7 \text{ s} \approx 0.96996974 \text{ years} \approx 1 \text{ year}. \tag{12.47}$$

Thus, the asymptotic approximation to the hyperbolic motion holds only after about 1 year of travel. Since any trip to even the nearest stars will certainly last longer than that, this approximation may be reagarded as physically acceptable.

12.5 CHECKING THE KLT OF ASYMPTOTIC HYPERBOLIC MOTION BY MATLAB SIMULATIONS

Just look at Figure 12.1.

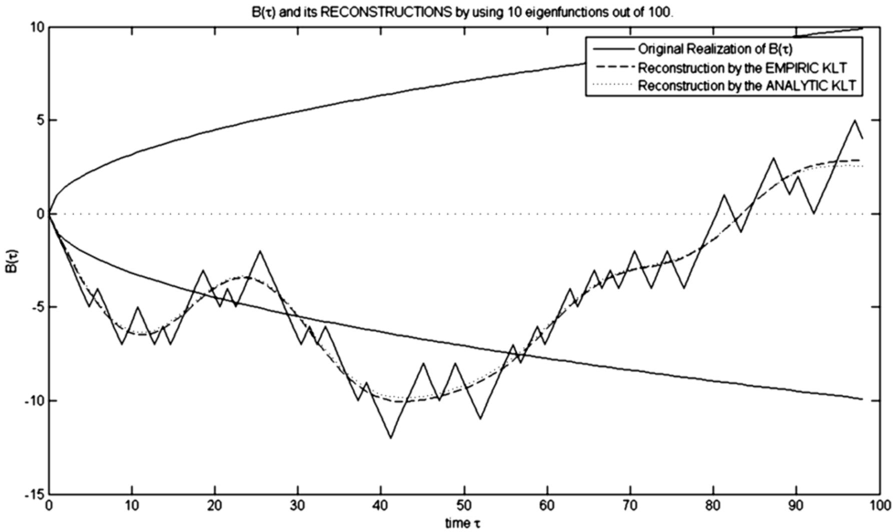


Figure 12.1. The time-rescaled Brownian motion $X(t)$ of (12.43) vs. time t simulated as a random walk over 100 time instants. This $X(t)$ represents the "noisy signal" received on Earth (whence the use of the coordinate time $t =$ Earth time) from a relativistic spaceship moving away from the Earth in an asymptotic hyperbolic motion, as in the science fiction novel *Tau Zero*. Next to the "bumpy curve" of $X(t)$, two more "smooth curves" are shown that *interpolate at best* the bumpy $X(t)$. These two curves are the KLT reconstruction of $X(t)$ *by using the first ten eigenfunctions only*. It is important to note that the two smooth curves are *different* in this case because the KLT expansion (12.43) is *approximated*. Actually, it is an approximated KLT expansion because the asymptotic expansion of the Bessel functions (12.31) was used. So, the two curves are different from each other, but both still interpolate $X(t)$ at best. Note that, were we taking into account the full set of 100 KLT eigenfuctions—rather than just 10—then the *empirical* reconstruction would overlap $X(t)$ *exactly*, but the *analytic* reconstruction would not because of the use of the asymptotic expansion (12.31) of the Bessel functions.

12.6 SIGNAL TOTAL ENERGY AS A STOCHASTIC PROCESS OF *T*

Formulas (F.60) and (12.20) enable us to obtain the total energy mean value

$$E\{\varepsilon_{Asy}\} = \int_0^T dt \int_0^t f^2(s)\, ds = \frac{c}{g} \int_0^T \ln\left(\frac{2g}{c} t\right) dt. \tag{12.48}$$

The substitution $x = (2g/c)t$ then results in

$$E\{\varepsilon_{Asy}\} = \frac{1}{2}\left(\frac{c}{g}\right)^2 [x(\ln x - 1)]_0^{(\frac{2g}{c})T} = \frac{cT}{g}\left[\ln\left(\frac{2g}{c} T\right) - 1\right].$$

Thus, the asymptotic mean total energy reads

$$E\{\varepsilon_{Asy}\} = \frac{cT}{g}\left[\ln\left(\frac{2g}{c} T\right) - 1\right]. \tag{12.49}$$

Note that the same asymptotic result is obtained from the exact expression (12.12) upon substituting arcsinh by log, and disregarding all the +1 that disappear for large T.

Next let us turn to the asymptotic total energy variance by resorting to

$$\int \ln^2 x\, dx = x \ln^2 x - 2x \ln x + 2x + C \tag{12.50}$$

$$\int x \ln^2 x\, dx = \frac{x^2}{2} \ln^2 x - \frac{x^2}{2} \ln x + \frac{x^2}{4} + C \tag{12.51}$$

$$\int x \ln x\, dx = \frac{x^2}{2} \ln x - \frac{x^2}{4} + C. \tag{12.52}$$

In fact, inserting (12.20) into the expression for σ_ε^2 in (F.62), one finds

$$\begin{aligned} \sigma^2_{\varepsilon_{Asy}} &= 4 \int_0^T dt \int_0^t du \left[\int_0^u f^2(s)\, ds\right]^2 \\ &= 4\left(\frac{c}{g}\right)^2 \int_0^T dt \int_0^t \ln^2\left(\frac{2g}{c} u\right) du \end{aligned} \tag{12.53}$$

whence the substitution $[(2g)/c]u = x$ and the integral in (12.50) yield

$$\sigma^2_{\varepsilon_{Asy}} = 2\left(\frac{c}{g}\right)^3 \int_0^T dt \left[x \ln^2 x - 2x \ln x + 2x\right]_0^{\frac{2g}{c}t}.$$

The further substitution $[(2g/c)]t = x$ now leads to the couple of integrals (12.51) and (12.52)

$$
\begin{aligned}
\sigma^2_{\varepsilon_{Asy}} &= \left(\frac{c}{g}\right)^4 \left[\int_0^{\frac{2g}{c}T} x \ln^2 x\, dx - 2\int_0^{\frac{2g}{c}T} x \ln x\, dx + 2\int_0^{\frac{2g}{c}T} x\, dx\right] \\
&= \left(\frac{c}{g}\right)^4 \left[\frac{x^2}{2}\ln^2 x - \frac{x^2}{2}\ln x - x^2 \ln x + \frac{x^2}{4} + \frac{x^2}{2} + x^2\right]_0^{\frac{2g}{c}T} \\
&= \left(\frac{c}{g}\right)^4 \frac{1}{4}\left[x^2(2\ln^2 x - 6\ln x + 7)\right]_0^{\frac{2g}{c}T} \\
&= \left(\frac{cT}{g}\right)^2 \left[2\ln^2\left(\frac{2g}{c}T\right) - 6\ln\left(\frac{2g}{c}T\right) + 7\right].
\end{aligned}
$$

Thus, the asymptotic total energy variance reads

$$\sigma^2_{\varepsilon_{Asy}} = \left(\frac{cT}{g}\right)^2 \left[2\ln^2\left(\frac{2g}{c}T\right) - 6\ln\left(\frac{2g}{c}T\right) + 7\right]. \tag{12.54}$$

Note that just as (12.49) is the asymptotic version of (12.12), so (12.54) is the asymptotic form of (12.17), and could have been found by substituting arcsinh by log, and forgetting all the additive +1 that are dwarfed for large T.

The square root of (12.74) is the asymptotic total energy standard deviation

$$\sigma_{\varepsilon_{Asy}} = \pm\frac{cT}{g}\sqrt{2\ln^2\left(\frac{2g}{c}T\right) - 6\ln\left(\frac{2g}{c}T\right) + 7}. \tag{12.55}$$

Setting

$$\frac{gT}{c} = x \tag{12.56}$$

we see that the radicand of (12.55) is the quadratic in $\xi \equiv \ln x$

$$2\xi^2 - 6\xi + 7 > 0. \tag{12.57}$$

This is positive for any ξ because $\Delta = -20 < 0$.

Let us regard the noise asymptotic total energy as a stochastic process of T. The process behavior in time is characterized by its mean value curve (12.49) and by the upper and lower (mean value ± standard deviation) curves given by

$$E\{\varepsilon_{Asy}\} \pm \sigma_{\varepsilon_{Asy}}. \tag{12.58}$$

The first column of Table 12.1 shows the numerical values of the independent variable x defined by (12.56) ranging from 0 to 20. In units of time, T ranges from 0 to 20 years since

$$\frac{c}{g} \approx 3.0612245 \cdot 10^7 \text{ s} \approx 0.96699947 \text{ years} \approx 1 \text{ year}. \tag{12.59}$$

Table 12.1. Noise asymptotic total energy

$$x = \frac{gT}{c}$$

$$M = x(\ln(2x) - 1)$$

$$V = x^2(2\ln^2(2x) - 6\ln(2x) + 7)$$

x	M	$M - \sqrt{V}$	$M + \sqrt{V}$
0	0	0	0
1	−0.30685	−2.25673	1.643024
2	0.772588	−2.40600	3.951178
3	2.375278	−2.52698	7.277545
4	4.317766	−2.80572	11.44125
5	6.512925	−3.21883	16.24468
6	8.909439	−3.73346	21.55234
7	11.47340	−4.32692	27.27327
8	14.18070	−4.98419	33.34561
9	17.01334	−5.69497	39.72166
10	19.95732	−6.45182	46.36646
11	23.00146	−7.24919	53.25213
12	26.13664	−8.08277	60.35607
13	29.35525	−8.94912	67.65963
14	32.65086	−9.84541	75.14714
15	36.01796	−18.7693	82.80522
16	39.45177	−11.7188	90.62235
17	42.94812	−12.6922	98.58846
18	46.50334	−13.6880	106.6947
19	50.11413	−14.7049	114.9332
20	53.77758	−15.7418	123.2970

The second column gives the numerical values of the asymptotic mean value (12.49) of the noise total energy apart from a factor $(c/g)^2$. The third and fourth columns, respectively, show the values of the lower (minus sign) and upper (plus sign) curves (12.58), again apart from a factor $(c/g)^2$.

One may check the above asymptotic total energy results against the corresponding exact results derived at the end of Section 12.3. Table 12.2 shows the same items as Table 12.1, but is calculated by using the exact total energy variance (12.17). We see that the agreement is not as good for very small values of T, while it increases for increasing T, and the dispersion of the total energy around its mean value increases roughly by the same amount as the total energy itself.

The conclusion to this section is that the KL eigenfunction expansion has been derived for the noise emitted by a spaceship traveling at a constantly accelerated relativistic motion. Though the mathematical difficulties forced us to confine ourselves to the asymptotic theory for values of time larger than 1 year, the study of the noise total energy (where both asymptotic and exact results can be obtained) shows that the errors of the asymptotic version are not very large.

12.7 INSTANTANEOUS NOISE ENERGY FOR ASYMPTOTIC HYPERBOLIC MOTION: PREPARATORY CALCULATIONS

In Appendix I, as well as in [2], the process $Y(t)$ defined by

$$Y(t) = X^2(t) - E\{X^2(t)\} \tag{12.59}$$

was considered. According to (I.35), the KL eigenfunction expansion of that process reads

$$Y(t) = \sum_{n=1}^{\infty} \tilde{Z}_n \tilde{N}_n \sqrt{\tilde{f}(t) \int_0^t \tilde{f}(s)\, ds} \cdot J_{\tilde{\nu}(t)}\left(\tilde{\gamma}_n \frac{\int_0^t \tilde{f}(s)\, ds}{\int_0^T \tilde{f}(s)\, ds} \right), \tag{12.60}$$

where the function $\tilde{f}(t)$ is defined in terms of $f(t)$ by (I.24). That is,

$$\tilde{f}(t) \equiv 2f(t)\sqrt{\int_0^t f^2(z)\, dz}. \tag{12.61}$$

This section is devoted to finding the KL expansion of the zero-mean square process $Y(t)$, in the asymptotic hyperbolic case, and its physical meaning for relativistic interstellar flight will be examined in the coming section. In this section we just pave the mathematical way to the coming section by performing the necessary calculations.

Table 12.2. Noise exact total energy

$$x = \frac{gT}{c}$$

$$M = x\ln(x + \sqrt{1+x^2}) - \sqrt{1+x^2} + 1$$

$$V = (2x^2 - 1)[\ln^2(x + \sqrt{1+x^2}) - 6x\sqrt{1+x^2}\ln(x + \sqrt{1+x^2}) + 7x^2]$$

0	0	0	0
1	0.467160	−0.07884	1.013160
2	1.651202	−0.31139	3.613797
3	3.293061	−0.67013	7.256257
4	5.255744	−1.12713	11.63862
5	7.463172	−1.66295	16.58929
6	9.867916	−2.26419	22.00002
7	12.43777	−2.92124	27.79679
8	15.14952	−3.62692	33.92596
9	17.98561	−4.37568	40.34690
10	20.93235	−5.16310	47.02780
11	23.97876	−5.98558	53.94311
12	27.11583	−6.84016	61.07182
13	30.33603	−7.72432	68.39640
14	33.633301	−8.63592	75.90195
15	37.00130	−9.57310	83.57571
16	40.43615	−18.5342	91.40657
17	43.93342	−11.5179	99.38482
18	47.48945	−12.5229	107.5018
19	51.10098	−13.5481	115.7500

According to (12.61), we must first obtain the function $\tilde{f}(t)$, which follows at once from (12.20) and (12.21)

$$\tilde{f}(t) = \frac{2c}{g}\sqrt{\frac{\ln\left(\frac{g}{c}t\right)}{t}}. \tag{12.62}$$

We now proceed to construct the complicated expression (4.26), or, alternatively, (3.50) with $f(t)$ substituted by $\tilde{f}(t)$, to find the time-dependent order $\tilde{\nu}(t)$. But a glance at (I.26) and (12.62) shows that considerable analytical difficulties are involved. For instance, evaluation of the integral appearing in (G.50) with $f(t)$ substituted by $\tilde{f}(t)$, namely

$$\int_0^t \tilde{f}(s)\,ds = \int_0^t \frac{2c}{g}\sqrt{\frac{\ln\left(2\frac{g}{c}s\right)}{s}}\,ds \tag{12.63}$$

does not seem to be feasible in terms of elementary transcendental functions.

Nevertheless, these difficulties may be overcome by keeping in mind that we are seeking the asymptotic version of (12.62) for large values of time. Therefore, one is led to consider the limit

$$\lim_{t\to\infty}\tilde{f}(t) = \lim_{t\to\infty}\frac{2c}{g}\sqrt{\frac{\ln\left(2\frac{g}{c}t\right)}{t}} = \frac{2c}{g}\sqrt{\lim_{t\to\infty}\frac{\ln\left(2\frac{g}{c}t\right)}{t}} = \frac{\infty}{\infty} \tag{12.64}$$

where the indefinite form forces us to resort to L'Hospital's rule, and yields

$$\lim_{t\to\infty}\tilde{f}(t) = \frac{2c}{g}\sqrt{\lim_{t\to\infty}\frac{\left(2\frac{g}{c}\right)(1)}{\left(2\frac{g}{c}t\right)(1)}} = \frac{2c}{g}\lim_{t\to\infty}\frac{1}{\sqrt{t}} = \lim_{t\to\infty}\frac{2c}{g}\frac{1}{\sqrt{t}}.$$

Concluding the calculation at the last limit, and checking this against the initial limit, we thus obtain the following "ultimate" asymptotic version of (12.62), which from now on we shall regard as the asyptotic replacement to (12.62) for large values of time t

$$\tilde{f}(t) = \frac{2c}{g}\frac{1}{\sqrt{t}}. \tag{12.65}$$

This formula is simple enough for us to perform the remaining calculations involved with the KL expansion (12.60).

Beginning with the computation of the Bessel function order (G.50) with $f(t)$ substituted by $\tilde{f}(t)$, (12.65) yields at once

$$\ln\tilde{f}(t) = \ln\left(\frac{2c}{g}\right) - \tfrac{1}{2}\ln t \tag{12.66}$$

whence

$$\frac{d\ln\tilde{f}(t)}{dt} = -\frac{1}{2}\frac{1}{t} \tag{12.67}$$

$$\frac{d^2\ln\tilde{f}(t)}{dt^2} = +\frac{1}{2}\frac{1}{t^2} \tag{12.68}$$

and

$$\int_0^t \tilde{f}(s)\,ds = \frac{2c}{g}\int_0^t \frac{1}{\sqrt{s}}\,ds = \frac{4c}{g}\sqrt{t}. \tag{12.69}$$

Therefore, (G.50), with $f(t)$ substituted by $\tilde{f}(t)$, yields

$$\begin{aligned}\tilde{\nu}(t) &= \sqrt{\frac{1}{4} + \left[\frac{\frac{4c}{g}\sqrt{t}}{\frac{2c}{g}\frac{1}{\sqrt{t}}}\right]^2 \left\{\frac{3}{4}\left(-\frac{1}{2}\frac{1}{t}\right)^2 - \frac{1}{2}\left(\frac{1}{2}\frac{1}{t^2}\right)\right\}} \\ &= \sqrt{\frac{1}{4} + [2t]^2\left\{\frac{3}{16}\frac{1}{t^2} - \frac{1}{4}\frac{1}{t^2}\right\}} = \sqrt{\frac{1}{4} + 4t^2\cdot\frac{1}{t^2}\left\{\frac{3}{16} - \frac{1}{4}\right\}}. \end{aligned} \tag{12.70}$$

The time variable is thus seen to disappear from the last formula, leaving

$$\tilde{\nu}(t) = \sqrt{\frac{1}{4} + 4\left\{\frac{3-4}{16}\right\}} = \sqrt{\frac{1}{4} + \frac{-1}{4}} = \sqrt{0} = 0.$$

That is, the order of the Bessel function vanishes identically

$$\tilde{\nu}(t) = 0 \tag{12.71}$$

and this circumstance helps to simplify further calculations considerably. Intuitively speaking, (12.71) is quite a reasonable result. In fact, on the one hand, the corresponding Bessel function order in the KL expansion of the $X(t)$ process vanished too

$$\nu(t) = 0, \tag{12.72}$$

which is Equation (12.25), or eq. (68) in [1]. On the other hand, (12.71) truly mirrors the asymptotic character of the KL expansion under consideration, since the Bessel function of order zero is the only Bessel function of the first kind to have its initial value equal to one rather than zero, pointing out the non-validity of this theory for values of time near to the origin.

Let us now proceed to finding the function $\tilde{\chi}(t)$ defined by (I.25). By virtue of (12.65) and (12.69), it follows that

$$\tilde{\chi}(t) = \sqrt{\frac{2c}{g}\frac{1}{\sqrt{t}}\cdot\frac{4c}{g}\sqrt{t}} = 2\sqrt{2}\frac{c}{g}. \tag{12.73}$$

Once again, the time variable cancels out from the last formula, yielding a constant rather than a time function. An immediate consequence of (12.73) is, of course,

$$\tilde{\chi}'(t) = 0 \tag{12.74}$$

which helps to simplify further calculations also.

Reverting now to the KL expansion of (12.60), we see that the Bessel function must have the form

$$J_0\left(\tilde{\gamma}_n \frac{\int_0^t \tilde{f}(s)\,ds}{\int_0^T \tilde{f}(s)\,ds}\right) = J_0\left(\tilde{\gamma}_n \frac{\sqrt{t}}{\sqrt{T}}\right). \tag{12.75}$$

Our next task is to find the meaning of the constants $\tilde{\gamma}_n$, given by (I.27). As $\tilde{\chi}'(t) = 0$ and $\tilde{\nu}'(t) = 0$, and getting rid of all multiplicative factors, one easily sees that (I.27) yields

$$J_0'(\tilde{\gamma}_n) = 0. \tag{12.76}$$

Thus, the $\tilde{\gamma}_n$ are the positive zeros, arranged in ascending order of magnitude, of the derivative of $J_0(x)$. In other words, they are the abscissas of the maxima and minima of $J_0(x)$, that are known to follow each other alternately. However, a different interpretation of the $\tilde{\gamma}_n$ follows from (see [7, p. 12, entry 55 ($\nu = 0$ must be set)]

$$J_\nu'(x) = \frac{\nu}{x} J_\nu(x) - J_{\nu+1}(x). \tag{12.77}$$

Using (12.77), (12.76) now becomes equivalent to

$$J_1(\tilde{\gamma}_n) = 0 \tag{12.78}$$

and one may also say that the $\tilde{\gamma}_n$ are the real positive zeros of $J_1(x)$ The first 40 among them are listed in [8, p. 748], and one finds, for instance:

$$\tilde{\gamma}_1 = 3.8317060 \quad \tilde{\gamma}_2 = 7.0155867 \quad \tilde{\gamma}_{40} = 126.4461387. \tag{12.79}$$

No explicit formula yielding these zeros exactly is known. However, it is possible to get an approximated expression for them on setting $\nu = 1$ into the asymptotic formula for $J_\nu(x)$ (see [9, p. 134])

$$\lim_{x\to\infty} J_\nu(x) = \lim_{x\to\infty} \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right) \tag{12.80}$$

getting

$$\cos\left(\tilde{\gamma}_n - \frac{3\pi}{4}\right) \approx 0 \tag{12.81}$$

whence

$$\tilde{\gamma}_n - \frac{3\pi}{4} \approx n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots) \tag{12.82}$$

and finally

$$\tilde{\gamma}_n \approx n\pi + \frac{\pi}{4}. \tag{12.83}$$

We can see how good this approximation is by setting $n = 1, 2, \ldots, 40$,

$$\tilde{\gamma}_1 \approx 3.9269908 \quad \tilde{\gamma}_2 \approx 7.0685835 \quad \tilde{\gamma}_{40} \approx 126.4491 \tag{12.84}$$

and checking these results against (12.79): the agreement improves with increasing n.

As for the eigenvalues $\tilde{\lambda}_n$, they are related to the $\tilde{\gamma}_n$ by (I.29), and, by virtue of (12.69), take the form

$$\tilde{\lambda}_n = \frac{16c^2 T}{g^2} \frac{1}{(\tilde{\gamma}_n)^2}. \tag{12.85}$$

Finally, we turn to the normalization constants $\tilde{N}_n$ that can be discovered from (I.28) by inserting (12.69) and (12.71). Therefore

$$1 = \tilde{N}_n^2 \frac{16c^2 T}{g^2} \int_0^1 x[J_0(\tilde{\gamma}_n x)]^2 \, dx. \tag{12.86}$$

This integral is evaluated within the framework of the Dini expansion in the series of Bessel functions [5, p. 71], and one finds

$$\begin{aligned} 1 &= \tilde{N}_n^2 \frac{16c^2 T}{g^2} \left\{ \frac{1}{2} \left[J_0'^2(\tilde{\gamma}_n) + \left(1 - \frac{0}{\tilde{\gamma}_n^2}\right) J_0^2(\tilde{\gamma}_n) \right] \right\} \\ &= \tilde{N}_n^2 \frac{8c^2 T}{g^2} [J_0'^2(\tilde{\gamma}_n) + J_0^2(\tilde{\gamma}_n)] = \tilde{N}_n^2 \frac{8c^2 T}{g^2} J_0^2(\tilde{\gamma}_n) \end{aligned} \tag{12.87}$$

where (12.76) was used in the last step. Solving for $\tilde{N}_n$ requires the introduction of the modulus of $J_0(\tilde{\gamma}_n)$, and one has

$$\tilde{N}_n = \frac{g}{2\sqrt{2}c\sqrt{T}|J_0(\tilde{\gamma}_n)|}. \tag{12.88}$$

This is the exact expression of the normalization constants. We can, however, derive an approximated expression for them on substituting the Bessel function by virtue of (12.80) and (12.83)

$$|J_0(\tilde{\gamma}_n)| \approx \left| \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} \cos\left(\tilde{\gamma}_n - \frac{\pi}{4}\right) \right| = \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} |\cos(n\pi)| = \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}. \tag{12.89}$$

Thus, from (12.88), by virtue of (12.89) and (12.83), we get the approximated $\tilde{N}_n$:

$$\tilde{N}_n \approx \frac{\pi}{4} \frac{g}{c\sqrt{T}} \sqrt{n + \tfrac{1}{4}} \tag{12.90}$$

which completes our set of preliminary calculations.

12.8 KL EXPANSION FOR THE INSTANTANEOUS ENERGY OF THE NOISE EMITTED BY A RELATIVISTIC SPACESHIP

When dealing with a noise represented by a stochastic process $X(t)$, an important distinction is between its instantaneous energy, given by the square process

$$X^2(t) \tag{12.91}$$

and the total energy, given by the stochastic integral of the instantanous energy (12.91) over the finite time span, $0 \le t \le T$ during which the noise is observed:

$$I = \int_0^T X^2(s)\, ds. \tag{12.92}$$

This section is devoted to finding the KL expansion of the process (12.91), whereas both mean value and variance of the random variable (12.92) have already been obtained in Section 12.3, as well as in section 5 of [1]. A related paper, [10], may also be consulted.

Let us then consider the mean value of (12.91), given by (F.59); that is,

$$E\{X^2(t)\} = \int_0^t f^2(s)\, ds, \tag{12.93}$$

where E denotes mean value operator, or ensemble average. By virtue of (12.20), (12.93) is changed into

$$E\{X^2(t)\} = \frac{c}{g}\ln\left(2\frac{g}{c}t\right). \tag{12.94}$$

Thus, the zero-mean square process $Y(t)$, defined by (12.59), takes the form

$$Y(t) = X^2(t) - \frac{c}{g}\ln\left(2\frac{g}{c}t\right) \tag{12.95}$$

whence

$$X^2(t) = \frac{c}{g}\ln\left(2\frac{g}{c}t\right) + Y(t). \tag{12.96}$$

Let us now consider the KL expansion of the $Y(t)$ process. By sustituting into (12.60) the normalization constants (12.90), the $\tilde{\chi}(t)$ function (12.73), and the Bessel function (12.75), we come up with

$$Y(t) = \sum_{n=1}^{\infty} \tilde{Z}_n \frac{g}{2\sqrt{2}c\sqrt{T}|J_0(\tilde{\gamma}_n)|} 2\sqrt{2}\frac{c}{g} J_0\left(\tilde{\gamma}_n \frac{\sqrt{t}}{\sqrt{T}}\right)$$

from which both c and g disappear, yielding

$$Y(t) = \sum_{n=1}^{\infty} \tilde{Z}_n \frac{1}{\sqrt{T}|J_0(\tilde{\gamma}_n)|} J_0\left(\tilde{\gamma}_n \frac{\sqrt{t}}{\sqrt{T}}\right). \tag{12.97}$$

Thus, by virtue of (12.96) and (12.97), we conclude that the exact KL expansion of the instantaneous energy $X^2(t)$ reads

$$X^2(t) = \frac{c}{g}\ln\left(2\frac{g}{c}t\right) + \sum_{n=1}^{\infty} \tilde{Z}_n \frac{1}{\sqrt{T}|J_0(\tilde{\gamma}_n)|} \frac{2\sqrt{2}c}{g} J_0\left(\tilde{\gamma}_n \frac{\sqrt{t}}{\sqrt{T}}\right). \tag{12.98}$$

From this exact expansion we may also derive an approximated one by resorting to the usual asymptotic formula (12.80) for both the Bessel functions appearing in (12.98). The result is

$$X^2(t) \approx \frac{c}{g}\ln\left(2\frac{g}{c}t\right) + \sum_{n=1}^{\infty} \tilde{Z}_n \frac{1}{T^{\frac{1}{4}}t^{\frac{1}{4}}} \frac{2\sqrt{2}c}{g} \cos\left(\tilde{\gamma}_n \frac{\sqrt{t}}{\sqrt{T}} - \frac{\pi}{4}\right), \tag{12.99}$$

which, after substituting the $\tilde{\gamma}_n$ by the approximated version (12.83), takes the final form

$$X^2(t) \approx \frac{c}{g}\ln\left(2\frac{g}{c}t\right) + \sum_{n=1}^{\infty} \tilde{Z}_n \frac{1}{T^{\frac{1}{4}}t^{\frac{1}{4}}} \frac{2\sqrt{2}c}{g} \cos\left(\left(n\pi + \frac{\pi}{4}\right) \frac{\sqrt{t}}{\sqrt{T}} - \frac{\pi}{4}\right). \tag{12.100}$$

This is the approximated (i.e., asymptotic) KL expansion of the noise instantaneous energy for large values of time. The computational advantage of (12.100) over (12.98) is that the Bessel functions have been substituted by a cosine.

12.9 CONCLUSION

A surprising property of both the instantaneous energy KL expansions (12.98) and (12.100) is revealed by checking them, respectively, against the corresponding KL expansions (12.42) and (12.43) of the noise process $X(t)$. In fact, on the one hand, one should note that the $\tilde{\gamma}_n$ (12.78) of the $Y(t)$ process are just the same as the $\tilde{\gamma}_n$ of the $X(t)$ process, given by (12.29), inasmuch as both are the real positive zeros of $J_1(t)$. Moreover, a glance shows that (12.98) has just the same eigenfunctions as (12.42), and (12.100) as (12.43). Therefore, we reach the unexpected conclusion that, when dealing with the noise emitted by a relativistic spaceship in asymptotic hyperbolic motion, the best orthonormal basis in the Hilbert space (i.e., the basis spanned by the eigenfunctions) is the same for both the noise and its own zero-mean instantaneous energy. Alternatively, if we prefer to give up the zero-mean restriction, we may say that the noise and its own instantaneous energy share parallel optimal reference frames, or bases, in the Hilbert space. This unusual feature should bear consequences in the design of a correct signal analysis procedure to filter out the noise received on Earth from a relativistically moving spaceship in asymptotic hyperbolic motion.

12.10 REFERENCES

[1] C. Maccone, "Relativistic Interstellar Flight and Gaussian Noise," *Acta Astronautica*, **17** (1988), 1019–1027.

[2] C. Maccone, "Relativistic Interstellar Flight and Instantaneous Noise Energy," *Acta Astronautica*, **21** (1990), 155–159.
[3] H. Minkowski, *Raum und Zeit*, English translation in *The Principle of Relativity*, Dover, New York, 1952, in particular pp. 85–86.
[4] W. Rindler, *Special Relativity*, Oliver & Boyd, Edinburgh, 1960.
[5] H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1950.
[6] M. Abramowitz, and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1970.
[7] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1953.
[8] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1966.
[9] N. N. Lebedev, *Special Functions and Their Applications*, Dover, New York, 1972.
[10] G. M. Anderson, and D. T. Greenwood, "Relativistic Flight with a Constant Thrust Rocket," *Astronautica Acta*, **16** (1970), 153–158.

13

KLT of radio signals from relativistic spaceships in arbitrary motion

13.1 INTRODUCTION

In three papers [1–3] the present author applied the concept of time-rescaled Brownian motion to aspects of relativistic interstellar flight ranging from communication theory to genetics. The content of the present chapter was fully published in paper form in [11]. In particular, the Gaussian noise (Brownian motion) $X(t)$, emitted by a relativistic spaceship traveling at a *constant* acceleration g in its own reference frame, was shown to be—see Equation (12.10) or [1, eq. (53)]

$$X(t) = B\left(\frac{c}{g}\operatorname{arcsinh}\left(\frac{g}{c}t\right)\right) = B\left(\frac{c}{g}\ln\left[\frac{g}{c}t + \sqrt{1 + \left(\frac{g}{c}t\right)^2}\right]\right), \tag{13.1}$$

where c is the speed of light, $B(t)$ denotes standard Brownian motion with mean zero, variance t, and initial condition $B(0) = 0$, and time ranges within the finite interval $0 \leq t \leq T$.

An approximated (i.e., asymptotic) version of (13.1) for large values of time follows by ignoring the additive $+1$ under the root sign that is dwarfed by the other terms, and reads

$$B\left(\frac{c}{g}\ln\left(2\frac{g}{c}t\right)\right). \tag{13.2}$$

For this stochastic process, it was proved in Equation (10.42) that the KL eigenfunction expansion is given by

$$B\left(\frac{c}{g}\ln\left(2\frac{g}{c}t\right)\right) = \sum_{n=1}^{\infty} Z_n \frac{1}{\sqrt{T}|J_0(\gamma_n)|} J_0\left(\gamma_n \frac{\sqrt{t}}{\sqrt{T}}\right), \tag{13.3}$$

where

(1) The constants γ_n $(n = 1, 2, \ldots)$, appearing in the argument of the Bessel function of order 0, $J_0(\ldots)$, are the (infinite) real positive zeros of the Bessel function of order 1; that is

$$J_1(\gamma_n) = 0. \tag{13.4}$$

(2) The eigenvalues λ_n are expressed in terms of the constants γ_n by the formula

$$\lambda_n = \frac{4cT}{g} \frac{1}{(\gamma_n)^2}.$$

(3) The Z_n are a set of orthogonal Gaussian random variables with mean zero and variance λ_n; that is

$$E\{Z_m Z_n\} = \lambda_n \delta_{mn}. \tag{13.5}$$

By resorting to the asymptotic expansion of the Bessel function of the first kind for large values of its argument—see Chapter 12, Equation (12.80)—it is also possible to derive an approximated KL eigenfunction expansion reading

$$B\left(\frac{c}{g}\ln\left(2\frac{g}{c}t\right)\right) \approx \sum_{n=1}^{\infty} Z_n \frac{1}{T^{\frac{1}{4}}t^{\frac{1}{4}}} \cos\left(\gamma_n \frac{\sqrt{t}}{\sqrt{T}} - \frac{\pi}{4}\right). \tag{13.6}$$

Here again, the meaning of the quantities appearing in (13.6) is the same as for those appearing in (13.3), with the only exception that the γ_n are now replaced by the approximated formula

$$\gamma_n \approx n\pi + \frac{\pi}{4}. \tag{13.7}$$

All the above results apply if, and only if, the spaceship acceleration in its own reference frame (proper acceleration) equals a *constant*, here denoted by g. Indeed, customary terminology such as "relativistic rocket", "relativistic interstellar flight", and the like, is almost always understood in this restricted constant g sense. However, for the sake of completeness it is desirable to extend the above results of Equations (13.1)–(13.7) to the general case where the constant g is replaced by an arbitrary time function

$$gA(t). \tag{13.8}$$

Here the time function $A(t)$ has the dimensions of a purely numerical factor, and represents the non-constant spaceship acceleration measured in units of g with respect to the spaceship reference frame. Clearly, the constant g case can now be regarded as the particular case for which the $A(t)$ function is given by

$$A(t) = 1. \tag{13.9}$$

The present chapter is devoted to generalization of the mentioned results (13.1)–(13.7) to the non-constant g case, and to exploring how many of these equations can be cast in a closed analytical form without resorting to numerical techniques for solving them.

13.2 ARBITRARY SPACESHIP ACCELERATION

13.2.1 Relativistic interstellar flight with an arbitrary spaceship acceleration profile

Analogous to the constant g case in (10.2), the differential equation of motion for a relativistic spaceship moving with arbitrary acceleration with respect to its own reference frame reads

$$\left[1 - \frac{v^2(t)}{c^2}\right]^{-\frac{3}{2}} \frac{dv(t)}{dt} = gA(t) \tag{13.10}$$

with the initial condition

$$v(0) = 0. \tag{13.11}$$

The left-hand side of (13.10) is the so-called "longitudinal force" of special relativity seen before as (12.1), and the right-hand side is just (13.8). The variables in (13.10) may be separated to achieve the analytical expression of the unknown velocity $v(t)$ by setting

$$v(t) = c \sin \Omega(t). \tag{13.12}$$

In fact, integrating both sides, one gets

$$c \tan \Omega(t) = g \int_0^t A(s)\, ds \tag{13.13}$$

whence an inversion yields

$$\Omega(t) = \arctan\left(\frac{g}{c}\int_0^t A(s)\, ds\right). \tag{13.14}$$

Replacing this into (13.12), we get for the velocity

$$v(t) = c \sin\left(\arctan\left(\frac{g}{c}\int_0^t A(s)\, ds\right)\right). \tag{13.15}$$

But

$$\sin(\arctan(x)) = \frac{x}{\sqrt{1+x^2}} \tag{13.16}$$

so that, in conclusion, the spaceship velocity is given by

$$v(t) = \frac{g\displaystyle\int_0^t A(s)\, ds}{\sqrt{1 + \left(\dfrac{g}{c}\displaystyle\int_0^t A(s)\, ds\right)^2}}. \tag{13.17}$$

In the constant g case, this reduces to (12.6), namely

$$v(t) = \frac{gt}{\sqrt{1 + \left(\dfrac{g}{c}t\right)^2}}. \tag{13.18}$$

The distance $x(t)$ covered by the spaceship up to time t and measured by Earth standards is defined by

$$x(t) = \int_0^t v(s)\, ds \tag{13.19}$$

with the consequential initial condition

$$x(0) = 0. \tag{13.20}$$

To perform the integration (13.19) for the function (13.17) in a closed form, we note that one has

$$\frac{d}{dt}\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\, ds\right)^2} = \frac{\frac{g}{c}\int_0^t A(s)\, ds}{\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\, ds\right)^2}} \cdot \frac{g}{c} A(t). \tag{13.21}$$

This enables us to rewrite (13.17) in the form

$$v(t) = \frac{c^2}{g}\frac{1}{A(t)}\frac{d}{dt}\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\, ds\right)^2} \tag{13.22}$$

whence integration (by parts on the right-hand side) yields the required distance:

$$x(t) = \frac{c^2}{g}\left[\frac{\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\, ds\right)^2} - 1}{A(t)} + \int_0^t \frac{\frac{dA(s)}{ds}}{A^2(s)}\sqrt{1+\left(\frac{g}{c}\int_0^s A(w)\, dw\right)^2}\, ds\right] \tag{13.23}$$

In the constant g case, this reduces to (see, for example, [5, p. 199, eq. (13)])

$$x(t) = \frac{c^2}{g}\left[\sqrt{1+\left(\frac{g}{c}t\right)^2} - 1\right]. \tag{13.24}$$

Finally, the proper time τ (i.e., the time measured aboard the spaceship) is given by

$$\tau(t) = \int_0^t \sqrt{1-\frac{v^2(s)}{c^2}}\, ds. \tag{13.25}$$

On replacing (13.17), some rearranging yields

$$\tau(t) = \int_0^t \frac{1}{\sqrt{1+\left(\frac{g}{c}\int_0^s A(w)\, dw\right)^2}}\, ds. \tag{13.26}$$

In the constant g case, both integrals in (13.26) can be performed analytically, whence Equation (12.9) is found again

$$\tau(t) = \int_0^t \frac{ds}{\sqrt{1 + \left(\frac{g}{c}s\right)^2}} = \frac{c}{g}\operatorname{arcsinh}\left(\frac{g}{c}t\right) = \frac{c}{g}\ln\left[\frac{g}{c}t + \sqrt{1 + \left(\frac{g}{c}t\right)^2}\right]. \quad (13.27)$$

13.2.2 KL expansion of the Gaussian noise emitted by a spaceship having an arbitrary acceleration profile

We now turn to the problem of finding the KL expansion for the Brownian motion (Gaussian noise) emitted when the spaceship moves according to the arbitrary acceleration law $gA(t)$. From a purely mathematical point of view, the formulas solving this problem were obtained by the author in [6]. However, that paper only dealt with analytical developments and not with their application to special relativity, not to mention starflight. It thus appears appropriate to recast the content of [6] here so as to make it fit the new developments required by introduction of the generic acceleration function $A(t)$.

The important point expressed in Section F.8 is that the time-dilation effect of special relativity induces a time-rescaling in the Brownian motion argument that may be expressed by the formula

$$B(\tau) = B\left(\int_0^t f^2(s)\, ds\right), \quad (13.28)$$

where the function $f(t)$ is called the "time-rescaling" function (see [6, p. 213, eq. (1.1)]. From this and the definition (13.25) of proper time, one gets

$$\int_0^t f^2(s)\, ds = \int_0^t \sqrt{1 - \frac{v^2(s)}{c^2}}\, ds \quad (13.29)$$

whence, differentiating and taking the square root, one finds

$$f(t) = \left[1 - \frac{v^2(t)}{c^2}\right]^{\frac{1}{4}}. \quad (13.30)$$

This is the relationship between the Brownian motion time-rescaling function, $f(t)$, and the spaceship velocity, $v(t)$. Inversion of (13.30) immediately leads to

$$v(t) = c\sqrt{1 - f^4(t)}. \quad (13.31)$$

We next develop the relationship between the time-rescaling function, $f(t)$, and the acceleration, $gA(t)$. Inserting (13.17) into (13.30), with a few reductions, yields

$$f(t) = \frac{1}{\left[1 + \left(\frac{g}{c}\int_0^t A(s)\, ds\right)^2\right]^{\frac{1}{4}}} = \left[1 + \left(\frac{g}{c}\int_0^t A(s)\, ds\right)^2\right]^{-\frac{1}{4}}. \quad (13.32)$$

This formula will be extensively used in the rest of the present chapter in either of the above forms. To invert it, one first solves for the integral

$$\int_0^t A(s)\,ds = \frac{c}{g}\sqrt{\frac{1}{f^4(t)} - 1} \tag{13.33}$$

and then differentiates with respect to t, getting

$$A(t) = -\frac{2c}{g}\frac{\frac{df(t)}{dt}}{f^3(t)\sqrt{1-f^4(t)}}. \tag{13.34}$$

We are now ready to state the main theorem of the present section.

Theorem 1 *The KL expansion for the Gaussian noise emitted by a spaceship moving at an arbitrary proper acceleration $gA(t)$ is given by*

$$X(t) \equiv B(\tau) = \left[1 + \left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2\right]^{-\frac{1}{8}} \cdot \sqrt{\int_0^t \frac{ds}{\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{\frac{1}{4}}}} \cdot \sum_{n=1}^{\infty} Z_n N_n J_{\nu(t)}\left(\gamma_n \frac{\int_0^t \frac{ds}{\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{\frac{1}{4}}}}{\int_0^T \frac{ds}{\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{\frac{1}{4}}}}\right). \tag{13.35}$$

Here

(1) The Bessel function of the first kind appearing in (13.35) has the time-dependent order

$$\nu(t) = \sqrt{-\chi^3(t)\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2} \cdot \sqrt{\frac{d}{dt}\left[\chi'(t)\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2}\right]}} \quad \ldots(13.36)$$

where the auxiliary function $\chi(t)$ has been defined

$$\chi(t) = \left[1 + \left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2\right]^{-\frac{1}{8}} \cdot \sqrt{\int_0^t \frac{ds}{\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{\frac{1}{4}}}}. \tag{13.37}$$

(2) The real positive zeros, arranged in ascending order of magnitude, of the equation

$$\chi'(T)\cdot J_{\nu(T)}(\gamma_n)+\chi(T)\cdot$$
$$\cdot\left\{\frac{\left[1+\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2\right]^{-\frac{1}{4}}\cdot\gamma_n}{\int_0^T\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{4}}ds}J'_{\nu(T)}(\gamma_n)+\frac{J_{\nu(T)}(\gamma_n)}{\nu}\nu'(T)\right\}=0 \quad (13.38)$$

define the constants γ_n.

(3) The normalization condition fulfilled by the eigenfunctions reads

$$1=N_n^2\left[\int_0^T\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{4}}ds\right]^2\cdot\int_0^1 xJ^2_{\nu((x))}(\gamma_n x)\,dx \quad (13.39)$$

and defines the normalization constants N_n. In (13.39) the new transformed order $\nu((x))$ is obtained from the order $\nu(t)$ of (13.36) via the transformation

$$\int_0^t\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{4}}ds=x\int_0^T\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{4}}ds. \quad (13.40)$$

(4) The eigenvalues are given by

$$\lambda_n=\left[\int_0^T\left[1+\left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{4}}ds\right]^2\frac{1}{(\gamma_n)^2}. \quad (13.41)$$

(5) Finally, the Gaussian random variables Z_n are orthogonal with mean zero and variance λ_n, as in (13.5).

The proof of Theorem 1 will not be repeated here, for it is just the same as the proof we give in Appendix G (see [6, p. 214, theorem 1.1]):[1] one just needs replace $f(t)$ by the arbitrary function $A(t)$ by virtue of their relationship given by (13.32).

13.2.3 Total noise energy

One of the main advantages of the KL expansion over other expansion types is that it allows analytical computation of the stochastic integral

$$\varepsilon=\int_0^T X^2(s)\,ds. \quad (13.42)$$

Physically, this integral represents the total noise energy over the finite time span $0\le t\le T$ during which the noise is observed on Earth. Mathematically, the expression in (13.42) is a random variable whose cumulants K_n, for $n=1,2,\ldots,$ are

[1] First published in 1984.

obtained in Appendix F, Equation (F.56)—see also [1, eqs. (21)–(30)]—and read, when rewritten in terms of the $A(t)$ function

$$K_n = 2^{n-1}(n-1)!\left[\int_0^T \left[1 + \left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{4}} ds\right]^{2n} \sum_{m=1}^{\infty} \frac{1}{(\gamma_m)^{2n}}. \tag{13.43}$$

The mean value of the total noise energy may be directly expressed in terms of the $A(t)$ function by the procedure proven in [1, p. 1021]. The result is eq. (34) of [1], which, translated and generalized by virtue of (13.32), now takes the form

$$E\{\varepsilon\} = \int_0^T dt \int_0^t \frac{1}{\sqrt{1 + \left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2}}\,ds. \tag{13.44}$$

In the constant g case, this reduces to Equation (12.12) (or [1, eq. (55)]), namely

$$E\{\varepsilon\} = \frac{c^2}{g^2}\left[\frac{gT}{c}\operatorname{arcsinh}\left(\frac{gT}{c}\right) - \sqrt{1 + \left(\frac{gT}{c}\right)^2} + 1\right]. \tag{13.45}$$

Total noise energy variance may also be expressed in a similar fashion, as described in Appendix F, Equation (F.62) (or [1, eqs. (56)–(60)]). When generalized and recast in terms of $A(t)$ by virtue of (13.32), this variance reads

$$\sigma_\varepsilon^2 = 4\int_0^T dt \int_0^t du \left[\int_0^u \frac{1}{\sqrt{1 + \left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2}}\,ds\right]^2 \tag{13.46}$$

and reduces to (12.17) (or [1, eq. (60)]) in the constant g case; that is

$$\sigma_\varepsilon^2 = \left(\frac{c}{g}\right)^4 \left\{\left[2\left(\frac{g}{c}T\right)^2 - 1\right]\operatorname{arcsinh}^2\left(\frac{g}{c}T\right) - 6\left(\frac{g}{c}T\right)\sqrt{1 + \left(\frac{g}{c}T\right)^2}\operatorname{arcsinh}\left(\frac{g}{c}T\right) + 7\left(\frac{g}{c}T\right)^2\right\}. \tag{13.47}$$

13.2.4 KL expansion of noise instantaneous energy

The instantaneous energy of noise is the square process

$$B^2(\tau) = X^2(t) \tag{13.48}$$

and we are now going to determine its KL expansion in terms of the arbitrary acceleration $A(t)$. In a strictly mathematical sense, the problem of finding the KL expansion of (13.48) was solved by the author in [7], with no reference to special relativity or starflight. The application of these results to the constant g case of starflight was found in [2], which will now be generalized to encompass the general $A(t)$ acceleration case.

Consider the process $Y(t)$ defined by

$$Y(t) = X^2(t) - E\{X^2(t)\}. \tag{13.49}$$

It is natural to call this *the zero-mean square process of time-rescaled Brownian motion* $X(t)$, for (13.49) is just the square of $X(t)$ centered around the latter's mean value.

Define now the new function

$$\tilde{f}(t) = 2f(t)\sqrt{\int_0^t f^2(s)\,ds}. \tag{13.50}$$

Recast in terms of $A(t)$ by virtue of (13.32), this becomes

$$\tilde{f}(t) = 2\left[1 + \left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2\right]^{-\frac{1}{4}} \cdot \sqrt{\int_0^t \left[1 + \left(\frac{g}{c}\int_0^s A(w)\,dw\right)^2\right]^{-\frac{1}{2}} ds}. \tag{13.51}$$

Then, Equation (13.35) (or [7, eq. (3.24)][2]) is the proof of the following:

Theorem 2 *The KL expansion of the zero-mean instantaneous energy of the Gaussian noise emitted by a spaceship moving at proper arbitrary acceleration* $A(t)$ *is given by*

$$X^2(t) - E\{X^2(t)\} = Y(t) = \sqrt{\tilde{f}(t)\int_0^t \tilde{f}(s)\,ds} \cdot \sum_{n=1}^{\infty} \tilde{Z}_n \tilde{N}_n J_{\tilde{\nu}(t)}\left(\tilde{\gamma}_n \frac{\int_0^t \tilde{f}(s)\,ds}{\int_0^T \tilde{f}(s)\,ds}\right). \tag{13.52}$$

Here

(1) The Bessel function of the first kind appearing in (13.52) has the time-dependent order

$$\tilde{\nu}(t) = \sqrt{-\frac{\tilde{\chi}^3(t)}{\tilde{f}^2(t)} \cdot \frac{d}{dt}\left[\frac{\tilde{\chi}'(t)}{\tilde{f}^2(t)}\right]} \tag{13.53}$$

where the auxiliary $\tilde{\chi}(t)$ function is given by

$$\tilde{\chi}(t) = \sqrt{\tilde{f}(t)\int_0^t \tilde{f}(s)\,ds}. \tag{13.54}$$

Alternatively, it is possible to express the order by virtue of the single formula

$$\tilde{\nu}(t) = \sqrt{\frac{1}{4} + \left[\frac{\int_0^t \tilde{f}(s)\,ds}{\tilde{f}(t)}\right]^2 \cdot \left\{\frac{3}{4}\left[\frac{d\ln\tilde{f}(t)}{dt}\right]^2 - \frac{1}{2}\frac{d^2\ln\tilde{f}(t)}{dt^2}\right\}}. \tag{13.55}$$

proved by the author in the appendix to [8] and in (G.50).

[2] First published in 1988.

(2) The real positive zeros, arranged in ascending order of magnitude, of the equation

$$\tilde{\chi}'(T)\cdot J_{\tilde{\nu}(T)}(\tilde{\gamma}_n)+\tilde{\chi}(T)\cdot\left[\frac{\tilde{f}(T)\cdot\tilde{\gamma}_n}{\int_0^T \tilde{f}(s)\,ds}J'_{\tilde{\nu}(T)}(\tilde{\gamma}_n)+\frac{\partial J_{\tilde{\nu}(T)}(\tilde{\gamma}_n)}{\partial\tilde{\nu}}\tilde{\nu}'(T)\right]=0 \quad (13.56)$$

define the constants $\tilde{\gamma}_n$.

(3) The normalization condition fulfilled by the eigenfunctions reads

$$1=\tilde{N}_n^2\left[\int_0^T \tilde{f}(s)\,ds\right]^2\int_0^1 xJ^2_{\tilde{\nu}((x))}(\tilde{\gamma}_n x)\,dx \quad (13.57)$$

and defines the normalization constants $\tilde{N}_n$. In (13.57) the new transformed order $\tilde{\nu}((x))$ is obtained from the order $\tilde{\nu}(t)$ of either (13.53) or (13.55) via the transformation

$$\int_0^t \tilde{f}(s)\,ds = x\int_0^T \tilde{f}(s)\,ds. \quad (13.58)$$

(4) The eigenvalues are given by

$$\tilde{\lambda}_n=\left[\int_0^T \tilde{f}(s)\,ds\right]^2\frac{1}{(\tilde{\gamma}_n)^2}. \quad (13.59)$$

(5) Finally, the gamma-type probability density

$$f_{\tilde{Z}_n}(z)=\frac{1}{\sqrt{\pi}(2\tilde{\lambda}_n)^{\frac{1}{4}}}e^{-\frac{z+\sqrt{\frac{\tilde{\lambda}_n}{2}}}{\sqrt{2\tilde{\lambda}_n}}}\cdot\left[z+\sqrt{\frac{\tilde{\lambda}_n}{2}}\right]^{-\frac{1}{2}}U\left(z+\sqrt{\frac{\tilde{\lambda}_n}{2}}\right), \quad (13.60)$$

where $U(x)$ is the unit step function, is the probabilistic law according to which the random variables $\tilde{Z}_n$ are distributed.

13.3 ASYMPTOTIC ARBITRARY SPACESHIP ACCELERATION

13.3.1 Asymptotic motion with arbitrary acceleration

The apparently complicated mathematical nature of most results derived so far, in particular the two KL expansions of Theorems 1 and 2, may seem to give little hope for their closed-form application to interesting cases of starflight. However, this is not the case. In fact, if we confine ourselves to the *asymptotic* for $t \to \infty$ motion at an arbitrary acceleration, all the forgoing results get easier. They will be further simplified in the power-like acceleration case that we are going to present in Section 13.4. So, let us now examine under what circumstances the asymptotic approximation is physically acceptable.

The primary idea behind performing the asymptotic approximation is to replace the exact radical

$$\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2} \tag{13.61}$$

by what is left of it when the +1 term is dwarfed by the other term; that is

$$\sqrt{1+\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2} \approx \sqrt{\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2} = \frac{g}{c}\left|\int_0^t A(s)\,ds\right|. \tag{13.62}$$

In other words, we suppose that

$$\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2 \gg 1 \tag{13.63}$$

whence we find the *asymptotic condition*

$$\left|\int_0^t A(s)\,ds\right| > \frac{c}{g} \approx 3.061\cdot 10^7 \text{ seconds} \approx 0.9699 \text{ years} \approx 1 \text{ year}. \tag{13.64}$$

In the constant g case, (13.64) reduces to

$$t > \frac{c}{g} \approx 1 \text{ year} \tag{13.65}$$

which is (12.47) (or [1, eq. (90)]) and tells us that the asymptotic approximation holds only after about one year of travel. Since the left-hand side of (13.64) should not numerically differ too much from that of (13.65) for whatever "reasonable" $A(t)$ function we may adopt, and since any trip to even the nearest stars will certainly take longer than one year, we regard the asymptoticity condition (13.64) as physically acceptable.

In the remainder of this chapter we shall denote all asymptotic formulas derived from the corresponding exact formulas by the "*Asy*" suffix.

The first formula that we must asymptotically simplify with the aid of (13.62) is (13.28), for this is the argument for the time-rescaled Brownian motion $B(\tau)$ and leads to the asymptotic time-rescaling function "*Asy*". Thus, (13.28), (13.32), and (13.62) yield

$$B(\tau_{Asy}) \equiv B\left(\int_0^t \frac{1}{\frac{g}{c}\left|\int_0^s A(w)\,dw\right|}\,ds\right) = B\left(\int_0^t f^2_{Asy}(s)\,ds\right) \tag{13.66}$$

whence we infer

$$f^2_{Asy}(t) = \frac{1}{\frac{g}{c}\left|\int_0^t A(s)\,ds\right|} \tag{13.67}$$

and, finally, taking the square root, the asymptotic time-rescaling function

$$f_{Asy}(t) = \frac{1}{\sqrt{\frac{g}{c}}\sqrt{\left|\int_0^t A(s)\,ds\right|}} \tag{13.68}$$

is obtained.

In the constant g case, this reduces to (12.21) (or [1, eq. (64)]), namely

$$f_{Asy}(t) = \sqrt{\frac{c}{g}}\frac{1}{\sqrt{t}}. \tag{13.69}$$

The spaceship asymptotic velocity is obtained from (13.68) and (13.31), and reads

$$v_{Asy}(t) = c\sqrt{1 - f^4_{Asy}(t)} = c\sqrt{1 - \frac{1}{\left(\frac{g}{c}\int_0^t A(s)\,ds\right)^2}}. \tag{13.70}$$

In the constant g case, this reduces to (12.45) (or [1, eq. (88)]); that is

$$v_{Asy}(t) = c\sqrt{1 - \frac{c^2}{g^2}\cdot\frac{1}{t^2}}. \tag{13.71}$$

It should be noted that the reality of the radicand in (13.70) again yields the same asymptoticity condition (13.64) as did (13.63).

13.3.2 Asymptotic KL expansion for noise

The asymptotic version of Theorem 1 is obtained on rewriting formulas (13.35)–(13.41) by the aid of the asymptotic radical (13.62). One then gets

Theorem 3 *The asymptotic KL expansion for the Gaussian noise emitted by a spaceship moving at a proper acceleration $gA(t)$ is given by*

$$X_{Asy}(t) \equiv B_{Asy}(\tau) = \sqrt{\frac{c}{g}}\left[\left|\int_0^t A(s)\,ds\right|\right]^{-\frac{1}{4}} \cdot \sqrt{\int_0^t \frac{ds}{\sqrt{\left|\int_0^s A(w)\,dw\right|}}}\cdot$$
$$\cdot\sum_{n=1}^{\infty} Z_{n_{Asy}} N_{n_{Asy}} J_{\nu(t)_{Asy}}\left(\gamma_{n_{Asy}}\frac{\int_0^t \frac{ds}{\sqrt{\left|\int_0^s A(w)\,dw\right|}}}{\int_0^T \frac{ds}{\sqrt{\left|\int_0^s A(w)\,dw\right|}}}\right). \tag{13.72}$$

Here

(1) The Bessel function of the first kind appearing in (13.72) has the time-dependent order

$$\nu_{Asy}(t) = \frac{g}{c}\sqrt{-\chi^3_{Asy}(t)\left|\int_0^t A(s)\,ds\right|}\cdot\sqrt{\frac{d}{dt}\left[\chi'_{Asy}(t)\left|\int_0^t A(s)\,ds\right|\right]}. \quad (13.73)$$

where the auxiliary function $\chi_{Asy}(t)$ has been defined

$$\chi_{Asy}(t) = \sqrt{\frac{c}{g}}\left[\left|\int_0^t A(s)\,ds\right|\right]^{-\frac{1}{4}}\cdot\sqrt{\int_0^t \frac{ds}{\sqrt{\left|\int_0^s A(w)\,dw\right|}}}. \quad (13.74)$$

(2) The real positive zeros, arranged in ascending order of magnitude, of the equation

$$\chi'_{Asy}(T)\cdot J_{\nu_{Asy}(T)}(\gamma_{n_{Asy}}) + \chi_{Asy}(T)\cdot$$
$$\cdot\left\{\frac{\left[\left|\int_0^t A(s)\,ds\right|\right]^{-\frac{1}{2}}\gamma_{n_{Asy}}}{\int_0^T\left[\left|\int_0^s A(w)\,dw\right|\right]^{-\frac{1}{2}}ds}J'_{\nu_{Asy}(T)}(\gamma_{n_{Asy}}) + \frac{\partial J_{\nu_{Asy}(T)}(\gamma_{n_{Asy}})}{\partial\nu_{Asy}}\nu'_{Asy}(T)\right\} = 0. \quad (13.75)$$

define the constants $\gamma_{n_{Asy}}$.

(3) The normalization condition fulfilled by the eigenfunctions reads

$$1 = N^2_{n_{Asy}}\frac{c}{g}\left[\int_0^T\left[\left|\int_0^s A(w)\,dw\right|\right]^{-\frac{1}{2}}ds\right]^2\cdot\int_0^1 xJ^2_{\nu_{Asy}((x))}(\gamma_{n_{Asy}}x)\,dx \quad (13.76)$$

and defines the normalization constants $N_{n_{Asy}}$. In (13.76) the new transformed order $\nu_{Asy}((x))$ is obtained from the order $\nu_{Asy}(t)$ of (13.73) via the transformation

$$\int_0^t\left[\left|\int_0^s A(w)\,dw\right|\right]^{-\frac{1}{2}}ds = x\int_0^T\left[\left|\int_0^s A(w)\,dw\right|\right]^{-\frac{1}{2}}ds. \quad (13.77)$$

(4) The eigenvalues are given by

$$\lambda_{n_{Asy}} = \frac{c}{g}\left[\int_0^T\left[\left|\int_0^s A(w)\,dw\right|\right]^{-\frac{1}{2}}ds\right]^2\frac{1}{(\gamma_{n_{Asy}})^2}. \quad (13.78)$$

(5) Finally, the Gaussian random variables $Z_{n_{Asy}}$ are orthogonal with mean zero and variance $\lambda_{n_{Asy}}$.

13.3.3 Asymptotic total noise energy

In the asymptotic version, characterized by (13.62), the asymptotic total noise energy cumulants (13.43) take the form

$$K_{n_{Asy}} = 2^{n-1}(n-1)!\left(\frac{c}{g}\right)^n \cdot \left[\int_0^T \left[\left|\int_0^s A(w)\,dw\right|\right]^{-\frac{1}{2}} ds\right]^{2n} \sum_{m=1}^{\infty} \frac{1}{(\gamma_{m_{Asy}})^{2n}}. \quad (13.79)$$

The asymptotic total noise energy derived from (13.44) and (13.62) reads

$$E\{\varepsilon_{Asy}\} = \frac{c}{g}\int_0^T dt \int_0^t \frac{1}{\left|\int_0^s A(w)\,dw\right|}\,ds. \quad (13.80)$$

Finally, asymptotic total noise energy variance follows from (13.46) and (13.62), yielding

$$\sigma^2_{\varepsilon_{Asy}} = 4\left(\frac{c}{g}\right)^2 \int_0^T dt \int_0^t du \left[\int_0^u \frac{1}{\left|\int_0^s A(w)\,dw\right|}\,ds\right]^2. \quad (13.81)$$

13.3.4 Asymptotic KL expansion for noise instantaneous energy

Using (13.62) once again, the content of Theorem 2 is turned into the asymptotic KL expansion for noise instantaneous energy. Let us start by defining the new function, corresponding to (13.50) in the asymptotic limit,

$$\tilde{f}_{Asy}(t) = 2f_{Asy}(t)\sqrt{\int_0^t f^2_{Asy}(s)\,ds}. \quad (13.82)$$

Expressed in terms of $A(t)$ by virtue of (13.68), this becomes

$$\tilde{f}_{Asy}(t) = 2\frac{c}{g}\left[\left|\int_0^t A(s)\,ds\right|\right]^{-\frac{1}{2}} \cdot \sqrt{\int_0^t \frac{1}{\left|\int_0^s A(w)\,dw\right|}\,ds}. \quad (13.83)$$

Theorem 2 then immediately yields the following:

Theorem 4 *The asymptotic KL expansion of the zero-mean instantaneous energy of the Gaussian noise emitted by a spaceship moving at a proper acceleration $gA(t)$ is given by*

$$X^2_{Asy}(t) - E\{X^2_{Asy}(t)\} = Y_{Asy}(t)$$

$$= \sqrt{\tilde{f}_{Asy}(t)\int_0^t \tilde{f}_{Asy}(s)\,ds} \cdot \sum_{n=1}^{\infty} \tilde{Z}_{n_{Asy}} \tilde{N}_{n_{Asy}} J_{\tilde{\nu}_{Asy}(t)}\left(\tilde{\gamma}_{n_{Asy}} \frac{\int_0^t \tilde{f}_{Asy}(s)\,ds}{\int_0^T \tilde{f}_{Asy}(s)\,ds}\right). \quad (13.84)$$

Here

(1) The Bessel function of the first kind appearing in (13.84) has the time-dependent order

$$\tilde{\nu}_{Asy}(t) = \sqrt{-\frac{\tilde{\chi}^3_{Asy}(t)}{\tilde{f}^2_{Asy}(t)} \cdot \frac{d}{dt}\left[\frac{\tilde{\chi}'_{Asy}(t)}{\tilde{f}^2_{Asy}(t)}\right]} \tag{13.85}$$

where the auxiliary function is defined by

$$\tilde{\chi}_{Asy}(t) = \sqrt{\tilde{f}_{Asy}(t) \int_0^t \tilde{f}_{Asy}(s)\, ds}. \tag{13.86}$$

Alternatively, it is possible to express the order by virtue of the single formula

$$\tilde{\nu}_{Asy}(t) = \sqrt{\frac{1}{4} + \left[\frac{\int_0^t \tilde{f}_{Asy}(s)\, ds}{\tilde{f}_{Asy}(t)}\right]^2 \cdot \left\{\frac{3}{4}\left[\frac{d \ln \tilde{f}_{Asy}(t)}{dt}\right]^2 - \frac{1}{2}\frac{d^2 \ln \tilde{f}_{Asy}(t)}{dt^2}\right\}} \tag{13.87}$$

proved in Appendix G, Equation (G.50).

(2) The real positive zeros, arranged in ascending order of magnitude, of the equation

$$\tilde{\chi}'_{Asy}(T) \cdot J_{\tilde{\nu}_{Asy}(T)}(\tilde{\gamma}_{n_{Asy}}) + \tilde{\chi}_{Asy}(T) \cdot \left[\frac{\tilde{f}_{Asy}(T) \cdot \tilde{\gamma}_{n_{Asy}}}{\int_0^T \tilde{f}_{Asy}(s)\, ds} J'_{\tilde{\nu}_{Asy}(T)}(\tilde{\gamma}_{n_{Asy}}) + \frac{J_{\tilde{\nu}_{Asy}(T)}(\tilde{\gamma}_{n_{Asy}})}{\tilde{\nu}_{Asy}} \tilde{\nu}'_{Asy}(T)\right] = 0 \tag{13.88}$$

define the constants $\tilde{\gamma}_{n_{Asy}}$.

(3) The normalization condition fulfilled by the eigenfunctions reads

$$1 = \tilde{N}^2_{n_{Asy}} \left[\int_0^T \tilde{f}_{Asy}(s)\, ds\right]^2 \cdot \int_0^1 x J^2_{\tilde{\nu}_{Asy}((x))}(\tilde{\gamma}_{n_{Asy}} x)\, dx \tag{13.89}$$

and defines the normalization constants $\tilde{N}_{n_{Asy}}$. In (13.89) the new transformed order $\tilde{\nu}_{Asy}((x))$ is obtained from the order $\tilde{\nu}_{Asy}(x)$ of either (13.85) or (13.87) via the transformation

$$\int_0^1 \tilde{f}_{Asy}(s)\, ds = x \int_0^T \tilde{f}_{Asy}(s)\, ds. \tag{13.90}$$

(4) The eigenvalues are given by

$$\tilde{\lambda}_{n_{Asy}} = \left[\int_0^T \tilde{f}_{Asy}(s)\, ds\right]^2 \frac{1}{(\tilde{\gamma}_{n_{Asy}})^2}. \tag{13.91}$$

(5) Finally, the gamma-type probability density

$$f_{\tilde{Z}_{n_{Asy}}}(z) = \frac{1}{\sqrt{\pi}(2\tilde{\lambda}_{n_{Asy}})^{\frac{1}{4}}} e^{-\frac{z+\sqrt{\frac{\tilde{\lambda}_{n_{Asy}}}{2}}}{\sqrt{2\tilde{\lambda}_{n_{Asy}}}}} \cdot \left[z + \sqrt{\frac{\tilde{\lambda}_{n_{Asy}}}{2}} \right]^{-\frac{1}{2}} U\left(z + \sqrt{\frac{\tilde{\lambda}_{n_{Asy}}}{2}} \right), \quad (13.92)$$

where $U(x)$ is the unit step function, is the probability law obeyed by the random variables $\tilde{Z}_{n_{Asy}}$.

13.4 POWER-LIKE ASYMPTOTIC SPACESHIP ACCELERATION

13.4.1 Asymptotic motion with power-like acceleration

There exists a particular case of the spaceship acceleration $A(t)$ for which all the asymptotic formulas previously worked out, including the KL expansions of Theorems 3 and 4, can be analytically developed in a closed form. This is the case where the acceleration changes in time like some real power α of time; that is

$$A(t) = t^{\alpha}. \quad (13.93)$$

The exponent α is subjected to the limitations

$$-1 < \alpha < 1 \quad (13.94)$$

that we shall determine in the present section.

In what follows, we shall denote all the formulas derived from the power-like hypothesis (13.93) by a "$|P$" suffix. Moreover, since we are going to apply the power-like hypothesis (13.93) to asymptotic formulas only, in practice the new suffix will be "$Asy|P$".

Note also that the asymptotic constant g case, already studied in detail in [1, 2], corresponds to the particular case

$$\alpha = 0 \quad (13.95)$$

of the asymptotic power-like acceleration theory developed in the present and following sections. All the results of [1, 2] are thus considerably generalized in the present chapter.

Turning next to the general formulas for the asymptotic power-like acceleration case, consider first the asymptotic time-rescaling function $f_{Asy|P}(t)$ defined by (13.68). By virtue of (13.93) one gets

$$f_{Asy|P}(t) = \frac{1}{\sqrt{\frac{g}{c}}\sqrt{\left|\int_0^t s^{\alpha}\, ds\right|}} = \frac{1}{\sqrt{\frac{g}{c}\frac{\sqrt{t^{1+\alpha}}}{\sqrt{1+\alpha}}}}. \quad (13.96)$$

That is,

$$f_{Asy|P}(t) = \frac{\sqrt{c}\sqrt{1+\alpha}}{\sqrt{g}}\, t^{-\frac{1+\alpha}{2}}. \tag{13.97}$$

Clearly, the integration in (13.96) converges, and $\sqrt{1+\alpha}$ is not complex if, and only if

$$1+\alpha > 0. \tag{13.98}$$

On the other hand, the presence in Theorem 4 of many integrals of the type

$$\int_0^t f_{Asy|P}(s)\, ds = \frac{\sqrt{c}\sqrt{1+\alpha}}{\sqrt{g}} \int_0^t s^{-\frac{1+\alpha}{2}} ds = \frac{2\sqrt{c}\sqrt{1+\alpha}}{\sqrt{g}(1-\alpha)}\, t^{\frac{1-\alpha}{2}} \tag{13.99}$$

clearly requires

$$1-\alpha > 0. \tag{13.100}$$

Thus, the limitations (13.98) and (13.100) yield (13.94).

The asymptotic spaceship velocity follows from (13.70) and (13.97), and reads

$$v_{Asy|P}(t) = c\sqrt{1 - \frac{c^2(1+\alpha)^2}{g^2} \cdot \frac{1}{t^{2(1+\alpha)}}}. \tag{13.101}$$

Thus, to avoid complex variables, the inequality

$$\frac{c^2(1+\alpha)^2}{g^2} \cdot \frac{1}{t^{2(1+\alpha)}} \le 1 \tag{13.102}$$

must apply, whence we infer the *power-like asymptotic condition*, generalizing (13.65)

$$t \ge \left[\frac{c(1+\alpha)}{g}\right]^{\frac{1}{1+\alpha}} \approx [1 \text{ year} \cdot (1+\alpha)]^{\frac{1}{1+\alpha}}. \tag{13.103}$$

In other words, in order that the asymptotic power-like acceleration case be physically acceptable, the time elapsed since departure must be larger than the $1/(1+\alpha)$ power of $(1+\alpha)$ years.

13.4.2 Power-like asymptotic KL expansion for noise

The power-like acceleration (13.93) has the remarkable advantage over other acceleration types that its KL expansions stated in Theorems 3 and 4 can be fully obtained in the closed form. As a matter of fact, most of the purely mathematical features of the content of the present section appeared in [9], with no reference to either physics or engineering applications. In [8, pp. 333–339], the material of the present section first appeared cast in the language of special relativity, but it was oriented to particle physics rather than to starflight: the terminology "decelerated motion" appearing there refers to deceleration with respect to the Earth, not with respect to a moving point. We believe that recasting and completing the mentioned material of [8, 9] into the framework of the *proper* acceleration $A(t)$, as required by starflight, will be of help to scientists and engineers wishing to refer to a single book chapter rather than to a widely scattered literature.

Theorem 5 *The asymptotic KL expansion for the Gaussian noise emitted by a spaceship moving at a proper power-like acceleration is given by*

$$X_{Asy|P}(t) \equiv B_{Asy|P}(\tau)$$
$$= \frac{\sqrt{1-\alpha}\,t^{-\frac{\alpha}{2}}}{T^{\frac{1-\alpha}{2}}} \cdot \sum_{n=1}^{\infty} Z_{n_{Asy|P}} \frac{1}{|J_{\nu_{Asy|P}}(\gamma_{n_{Asy|P}})|} J_{\nu_{Asy|P}}\left(\gamma_{n_{Asy|P}} \frac{t^{\frac{1-\alpha}{2}}}{T^{\frac{1-\alpha}{2}}}\right). \tag{13.104}$$

Here

(1) The Bessel function of the first kind appearing in (13.104) has the constant order

$$\nu_{Asy|P} = \frac{|\alpha|}{1-\alpha}. \tag{13.105}$$

The auxiliary function defined by (13.74) now reads

$$\chi_{Asy|P}(t) = \sqrt{\frac{c}{g}}\sqrt{2}\frac{\sqrt{1+\alpha}}{\sqrt{1-\alpha}}t^{-\frac{\alpha}{2}}. \tag{13.106}$$

(2) The constants $\gamma_{n_{Asy|P}}$ are the real positive zeros, arranged in ascending order of magnitude, of the Bessel function of the first kind of order $\nu + 1$. That is,

$$J_{\nu_{Asy|P}+1}(\gamma_{n_{Asy|P}}) = 0. \tag{13.107}$$

(3) The normalization constants for the eigenfunctions are given by

$$N_{n_{Asy|P}} = \sqrt{\frac{g}{c}} \frac{1-\alpha}{\sqrt{2}\sqrt{1+\alpha}T^{\frac{1-\alpha}{2}}|J_{\nu_{Asy|P}}(\gamma_{n_{Asy|P}})|}. \tag{13.108}$$

(4) The eigenvalues are given by

$$\lambda_{n_{Asy|P}} = \frac{c}{g}\frac{4T^{1-\alpha}(1+\alpha)}{(1-\alpha)^2} \cdot \frac{1}{\gamma^2_{n_{Asy|P}}}. \tag{13.109}$$

(5) Finally, the Gaussian random variables $Z_{n_{Asy|P}}$ are orthogonal with mean zero and variance $\lambda_{n_{Asy|P}}$, as in (13.5).

We shall just highlight the proof of the foregoing theorem, omitting all lengthy calculations. The $\chi_{Asy|P}(t)$ function of (13.106) immediately follows from (13.74) and (13.93). Then, a vital simplification occurs: the time disappears identically from (13.73), yielding the constant order (13.105). This, in turn, simplifies the writing of (13.75), which, after several reductions, takes the final form

$$-\frac{|\alpha|}{1-\alpha}J_{\nu_{Asy|P}}(\gamma_{n_{Asy|P}}) + \gamma_{n_{Asy|P}}J'_{\nu_{Asy|P}}(\gamma_{n_{Asy|P}}) = 0. \tag{13.110}$$

Let us now recall the Bessel function property (for a proof, see [10, p. 12, entry (55)])

$$zJ'_\nu(z) - \nu J_\nu(z) = -zJ_{\nu+1}(z). \tag{13.111}$$

By virtue of (13.105), the left-hand sides of (13.110) and (13.111) coincide, yielding

$$-\gamma_{n_{Asy|P}} J_{\nu_{Asy|P}+1}(\gamma_{n_{Asy|P}}) = 0 \tag{13.112}$$

whence, as $\gamma_{n_{Asy|P}} \neq 0$, (13.107) is obtained.

Let us next turn to the normalization constants (13.108). In this regard, we note that a formula arising from the Dini expansion in a series of Bessel functions (see [10, p. 71, entry (49)]), that is

$$\int_0^1 x J_\nu^2(\gamma_n x)\, dx = \frac{1}{2\gamma_n^2}\{\gamma_n^2 J_\nu'^2(\gamma_n) + (\gamma_n^2 - \nu^2) J_\nu^2(\gamma_n)\} \tag{13.113}$$

may be simplified, by virtue of (13.107) and (13.112), to yield

$$\int_0^1 x J^2_{\nu_{Asy|P}}(\gamma_{n_{Asy|P}} x)\, dx = \frac{J^2_{\nu_{Asy|P}}(\gamma_{n_{Asy|P}})}{2}. \tag{13.114}$$

This and (13.76), where the order transformation (13.77) is not required because the order (13.105) is a constant, lead then to (13.108).

Finally, the eigenvalues (13.109) are obtained at once from (13.78) and (13.93). In the constant g case, use of (13.95) shows that formulas (13.104)–(13.109) reduce to the corresponding formulas appearing in section 4 of [1].

13.4.3 Approximated power-like asymptotic KL expansion for noise

Theorem 5 is perhaps still too complicated for engineering purposes inasmuch as the Bessel functions may not by easy to evaluate numerically. It seems thus useful to derive the *approximated* version of Theorem 5 that follows on replacing the Bessel function of the first kind by means of its asymptotic expansion for large values of its argument

$$\lim_{x\to\infty} J_\nu(x) = \lim_{x\to\infty} \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right). \tag{13.115}$$

In the following we shall denote all the formulas derived for this approximated case by a further "$|A$" suffix. Thus, as we are considering the asymptotic power-like approximated case, in practice the whole suffix will be "$Asy|P|A$". When applied to (13.107), (13.115) yields

$$\cos\left(\gamma_{n_{Asy|P|A}} - \frac{(\nu_{Asy|P}+1)\pi}{2} - \frac{\pi}{4}\right) \approx 0 \tag{13.116}$$

whence

$$\gamma_{n_{Asy|P|A}} - \frac{(\nu_{Asy|P}+1)\pi}{2} - \frac{\pi}{4} \approx n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots). \tag{13.117}$$

Thus, invoking (13.105), we get the following approximated expression for the constants $\gamma_{n_{Asy|P|A}}$

$$\gamma_{n_{Asy|P|A}} \approx n\pi + \frac{\pi}{4} + \frac{\nu_{Asy|P}\pi}{2} = \pi\left(n + \frac{1}{4} + \frac{|\alpha|}{2(1-\alpha)}\right). \tag{13.118}$$

We now set out to determine the approximated normalization constants $N_{n_{Asy|P|A}}$. First, the Bessel function appearing in (13.108) has to be replaced by its asymptotic version (13.115), yielding

$$|J_{\nu_{Asy|P}}(\gamma_{n_{Asy|P|A}})| \approx \left|\sqrt{\frac{2}{\pi\gamma_{n_{Asy|P|A}}}}\right| \cdot \left|\cos\left(\gamma_{n_{Asy|P|A}} - \frac{\nu_{Asy|P}\pi}{2} - \frac{\pi}{4}\right)\right|$$
$$= \sqrt{\frac{2}{\pi\gamma_{n_{Asy|P|A}}}}|\cos(n\pi)| = \sqrt{\frac{2}{\pi\gamma_{n_{Asy|P|A}}}}. \tag{13.119}$$

Then, (13.108) and (13.119) yield the approximated normalization constants $N_{n_{Asy|P|A}}$

$$N_{n_{Asy|P|A}} \approx \frac{\pi}{2}\sqrt{\frac{g}{cT^{1-\alpha}}}\frac{1-\alpha}{\sqrt{1+\alpha}} \cdot \sqrt{n + \frac{1}{4} + \frac{|\alpha|}{2(1-\alpha)}}. \tag{13.120}$$

As for the eigenvalues, one merely has to insert (13.118) into (13.109) to get

$$\lambda_{n_{Asy|P|A}} \approx \frac{4cT^{1-\alpha}(1+\alpha)}{g(1-\alpha)^2} \cdot \frac{1}{\pi^2\left(n + \frac{1}{4} + \frac{|\alpha|}{2(1-\alpha)}\right)^2}. \tag{13.121}$$

These also are the variances of Gaussian zero-mean orthogonal Gaussian random variables $Z_{n_{Asy|P|A}}$.

In conclusion, we have proven the following:

Theorem 6 *The approximated asymptotic KL expansion for the Gaussian noise emitted by a spaceship moving at a proper power-like acceleration is given by*

$$X_{Asy|P|A}(t) \equiv B_{Asy|P|A}(\tau)$$
$$\approx \frac{\sqrt{1-\alpha}}{T^{\frac{1-\alpha}{4}}t^{\frac{1+\alpha}{4}}} \cdot \sum_{n=1}^{\infty} Z_{n_{Asy|P|A}} \cos\left(\gamma_{n_{Asy|P|A}} \frac{t^{\frac{1-\alpha}{2}}}{T^{\frac{1-\alpha}{2}}} - \frac{\pi}{4} - \frac{|\alpha|}{1-\alpha} \cdot \frac{\pi}{2}\right). \tag{13.122}$$

In the constant g case, this theorem reduces to (12.43) (or [1, eq. (86)]).

13.4.4 Power-like asymptotic total noise energy

The results of Section 13.3.3 on asymptotic total noise energy will now be applied to the power-like acceleration (13.93). However, the required convergence of all integrals that we have to use will force us to restrict the range of validity of the exponent

$$-1 < \alpha < 0. \tag{13.123}$$

Note that the upper inequality implies that the constant g case, characterized by (13.95), cannot be derived as a particular case of the theory developed within the present section. For instance, consider the integral

$$\int_0^t f^2_{Asy|P}(s)\,ds = \frac{c(1+\alpha)}{g}\left[\frac{s^{-\alpha}}{-\alpha}\right]_0^t = \frac{c(1+\alpha)}{g}\frac{t^{-\alpha}}{-\alpha}. \tag{13.124}$$

Clearly, the last step is possible only if α is negative, whence the upper limitation in (13.123).

The integral in (13.124) is required for computation of the mean total noise energy (13.80). In fact, invoking (13.93), one obtains

$$E\{\varepsilon_{Asy|P}\} = \int_0^T dt \int_0^t f^2_{Asy|P}(s)\,ds = \frac{c}{g}\frac{1+\alpha}{(-\alpha)(1-\alpha)}T^{1-\alpha}. \tag{13.125}$$

The power-like asymptotic total noise energy variance, derived from (13.81) by virtue of (13.93) and (13.124), reads

$$\sigma^2_{\varepsilon_{Asy|P}} = 4\int_0^T dt\int_0^t du\left[\int_0^u f^2_{Asy|P}(s)\,ds\right]^2 = 2\frac{c^2}{g^2}\frac{(1+\alpha)^2}{\alpha^2(\alpha-1)^2(\alpha-2)(2\alpha-3)}T^{2(2-\alpha)}. \tag{13.126}$$

whence, by taking the square root, we get the power-like asymptotic total noise energy standard deviation

$$\sigma_{\varepsilon_{Asy|P}} = \pm\sqrt{2}\frac{c}{g}\cdot\frac{1+\alpha}{\alpha(\alpha-1)\sqrt{(\alpha-2)(2\alpha-3)}}T^{2-\alpha}. \tag{13.127}$$

Finally, the power-like asymptotic total noise energy cumulants (13.79) take the form

$$K_{n_{Asy|P}} = 2^{3n-1}(n-1)!\left(\frac{c}{g}\right)^n\frac{(1+\alpha)^n}{(1-\alpha)^{2n}}T^{n(1-\alpha)}\cdot\sum_{m=1}^{\infty}\frac{1}{(\gamma_{m_{Asy|P}})^{2n}}. \tag{13.128}$$

This formula holds for the full α range (13.94), rather than for the limited range (13.123), for (13.124) was not required for its derivation.

13.4.5 Power-like asymptotic KL expansion for noise instantaneous energy

Just as Theorem 5 of Section 13.4.2 was the power-like subcase of Theorem 3 of Section 13.3.2, so it is natural to seek a Theorem 7 to be the power-like subcase of Theorem 4 of Section 13.3.4. In the present section we shall prove that this guess is true. However, as in the previous section, we shall have to restrict the range of validity of the exponent α to

$$-1 < \alpha < 0. \tag{13.129}$$

Once again the upper inequality implies that the constant g case, characterized by (13.95), cannot be derived as a particular case of the theory now to be established.

To prove the upper inequality in (13.129), rewrite the definition (13.82) in terms of the new $\tilde{f}_{Asy|P}(t)$ function

$$\tilde{f}_{Asy|P}(t) = 2f_{Asy|P}(t)\sqrt{\int_0^t f^2_{Asy|P}(s)\, ds}. \tag{13.130}$$

Invoking (13.97), it is observed that the convergence of the integral just requires the upper inequality in (13.129):

$$\int_0^t f^2_{Asy|P}(s)\, ds = \frac{c(1+\alpha)}{g}\left[\frac{s^{-\alpha}}{-\alpha}\right]_0^t = \frac{c(1+\alpha)}{g}\frac{t^{-\alpha}}{-\alpha}. \tag{13.131}$$

Then, inserting (13.97) and (13.131) into (13.130), we get the explicit expression

$$\tilde{f}_{Asy|P}(t) = \frac{c}{g}\frac{2(1+\alpha)}{\sqrt{-\alpha}}t^{-\frac{2\alpha+1}{2}}. \tag{13.132}$$

Finally, the integral of (13.132) reads

$$\int_0^t \tilde{f}_{Asy|P}(s)\, ds = \frac{c}{g}\cdot\frac{4(1+\alpha)}{\sqrt{-\alpha}(1-2\alpha)}t^{\frac{1-2\alpha}{2}}. \tag{13.133}$$

We are now ready to state the following theorem:

Theorem 7 *The power-like asymptotic KL expansion of the zero-mean instantaneous energy of the Gaussian noise emitted by a spaceship moving at a proper acceleration $gA(t)$ is given by*

$$\begin{aligned} Y_{Asy|P}(t) &\equiv X^2_{Asy|P}(\tau) - E\{X^2_{Asy|P}(\tau)\} \\ &= \sqrt{1-2\alpha}\,\frac{t^{-\alpha}}{T^{\frac{1}{2}-\alpha}}\cdot\sum_{n=1}^{\infty}\tilde{Z}_{n_{Asy|P}}\tilde{N}_{n_{Asy|P}}J_{\tilde{\nu}_{Asy|P}}\left(\tilde{\gamma}_{n_{Asy|P}}\frac{t^{\frac{1}{2}-\alpha}}{T^{\frac{1}{2}-\alpha}}\right). \end{aligned} \tag{13.134}$$

Here

(1) The Bessel function of the first kind appearing in (13.134) has the constant order

$$\tilde{\nu}_{Asy|P} = \frac{|2\alpha|}{|2\alpha - 1|}. \tag{13.135}$$

The auxiliary function $\tilde{\chi}_{Asy|P}(t)$ defined by (13.86), by the aid of (13.132) and (13.133), takes the form

$$\tilde{\chi}_{Asy|P}(t) = \frac{c}{g}\cdot\frac{2\sqrt{2}(1+\alpha)}{\sqrt{-\alpha}\sqrt{1-2\alpha}}t^{-\alpha}. \tag{13.136}$$

(2) The constants $\tilde{\gamma}_{n_{Asy|P}}$ are the real positive zeros, arranged in ascending order of magnitude, of the Bessel function of the first kind of order $(\tilde{\nu}_{n_{Asy|P}} - 1)$; that is

$$J_{\tilde{\nu}_{n_{Asy|P}}-1}(\tilde{\gamma}_{n_{Asy|P}}) = 0. \tag{13.137}$$

(3) The normalization constants for the eigenfunctions are given by

$$\tilde{N}_{n_{Asy|P}} = \frac{g}{c} \frac{\sqrt{2}\sqrt{-\alpha}(1-2\alpha)T^{\alpha-\frac{1}{2}}}{4(1+\alpha)|J_{\tilde{\nu}_{n_{Asy|P}}}(\tilde{\gamma}_{n_{Asy|P}})|}. \tag{13.138}$$

(4) The eigenvalues are given by

$$\tilde{\lambda}_{n_{Asy|P}} = \frac{c^2}{g^2} \frac{16(1+\alpha)^2 T^{1-2\alpha}}{(-\alpha)(1-2\alpha)^2} \cdot \frac{1}{(\tilde{\gamma}_{n_{Asy|P}})^2}. \tag{13.139}$$

(5) Finally, the probability density followed by the orthogonal random variables $\tilde{Z}_{n_{Asy|P}}$ is the gamma-type:

$$f_{\tilde{Z}_{n_{Asy|P}}}(z) = \frac{1}{\sqrt{\pi}(2\tilde{\lambda}_{n_{Asy|P}})^{\frac{1}{4}}} e^{-\frac{z+\sqrt{\frac{\tilde{\lambda}_{n_{Asy|P}}}{2}}}{\sqrt{2\tilde{\lambda}_{n_{Asy|P}}}}} \cdot \left[z + \sqrt{\frac{\tilde{\lambda}_{n_{Asy|P}}}{2}}\right]^{-\frac{1}{2}} U\left(z + \sqrt{\frac{\tilde{\lambda}_{n_{Asy|P}}}{2}}\right) \tag{13.140}$$

where $U(x)$ is the unit step function.

We now sketch the proof of Theorem 7. The $\tilde{\chi}_{Asy|P}(t)$ function of (13.136) immediately follows from the definition (13.86) by invoking (13.132) and (13.133). Then, just as happened for Theorem 5, time disappears *identically* from (13.85) or (13.87) yielding the constant order (13.135). Whether this simplification (which is essential in order to simplify all subsequent analytical developments) occurs just "by chance" or has any deeper meaning is unknown to this author at the present time. The reader, however, may wish to speculate on the following unpublished result: the analytical technique used in [6, 7] to obtain the KL expansions of $B(\tau)$ and $B^2(\tau)$, respectively, *cannot* be extended to $B^3(\tau)$ and $B^4(\tau)$. Thus, it would appear that the mentioned simplifications, occurring in the $B(\tau)$ and $B^2(\tau)$ cases, are indeed a "lucky circumstance".

Continuing the proof of our Theorem 7, we see that the constancy of $\tilde{\nu}_{Asy|P}$ simplifies the writing of (13.88), and, after several reductions, it takes the form

$$-\frac{2\alpha}{2\alpha-1} J_{\tilde{\nu}_{Asy|P}}(\tilde{\gamma}_{n_{Asy|P}}) + \tilde{\gamma}_{n_{Asy|P}} J'_{\tilde{\nu}_{Asy|P}}(\tilde{\gamma}_{n_{Asy|P}}) = 0. \tag{13.141}$$

Then the Bessel function property (see [10, p. 11, entry (19.54)])[3]

$$zJ'_{\tilde{\nu}}(z) + \tilde{\nu}J_{\tilde{\nu}}(z) = zJ_{\tilde{\nu}-1}(z) \tag{13.142}$$

and (13.141) yield

$$-\tilde{\gamma}_{n_{Asy|P}} J_{\tilde{\nu}_{Asy|P}}(\tilde{\gamma}_{n_{Asy|P}}) = 0 \tag{13.143}$$

from which

$$\tilde{\gamma}_{n_{Asy|P}} \neq 0 \tag{13.144}$$

and (13.137) is obtained.

[3] Note that this is not the same as (13.111).

The normalization constants (13.138) come next. Just as for the proof of (13.108), the Dini expansion formula for the Bessel functions (13.113) may be invoked to perform the integration in (13.89), with the result

$$\int_0^1 x J^2_{\tilde{\nu}_{Asy|P}}(\tilde{\gamma}_{n_{Asy|P}} x)\, dx = \frac{J^2_{\tilde{\nu}_{Asy|P}}(\tilde{\gamma}_{n_{Asy|P}} x)}{2}. \tag{13.145}$$

Equation (13.89) then leads to (13.138) by invoking (13.133) and (13.145). The eigenvalues (13.139) are immediately obtained from (13.91) and (13.133). Finally, the probability density (13.140) is the same as (13.60), apart from the replacement of $\tilde{\lambda}_{n_{Asy}}$ by $\tilde{\lambda}_{n_{Asy|P}}$.

13.4.6 Approximated power-like asymptotic KL expansion for noise instantaneous energy

Just as Theorem 6 is the asymptotic approximation of Theorem 5, so in the present section we are going to derive a Theorem 8 that is the asymptotic approximation to Theorem 7. Again, the leading idea is to replace all Bessel functions by their simpler asymptotic form (13.115), which has the advantage of expressing them in terms of a cosine.

As in Section 13.4.3, all formulas derived for the asymptotic approximation will be denoted by a further "$|A$" suffix. Thus, the whole suffix will now be "$Asy|P|A$".

When applied to (13.137), (13.115) yields

$$\cos\left(\tilde{\gamma}_{n_{Asy|P|A}} - \frac{(\tilde{\nu}_{Asy|P} - 1)\pi}{2} - \frac{\pi}{4}\right) \approx 0 \tag{13.146}$$

whence

$$\tilde{\gamma}_{n_{Asy|P|A}} - \frac{(\tilde{\nu}_{Asy|P} - 1)\pi}{2} - \frac{\pi}{4} \approx n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots) \tag{13.147}$$

Invoking (13.135), we now get the following approximated expression for the $\tilde{\gamma}_{n_{Asy|P|A}}$

$$\tilde{\gamma}_{n_{Asy|P|A}} \approx (n-1)\pi + \frac{\pi}{4} + \frac{\tilde{\nu}_{Asy|P}\pi}{2} = \pi\left(n - \frac{3}{4} + \frac{|\alpha|}{|2\alpha - 1|}\right). \tag{13.148}$$

The approximated normalization constants $\tilde{N}_{n_{Asy|P|A}}$ come next. However, the Bessel function appearing in (13.138) must first be replaced by its asymptotic version (13.115), yielding

$$\begin{aligned} |J_{\tilde{\nu}_{Asy|P}}(\tilde{\gamma}_{n_{Asy|P|A}})| &\approx \left|\sqrt{\frac{2}{\pi\tilde{\gamma}_{n_{Asy|P|A}}}}\right| \left|\cos\left(\tilde{\gamma}_{n_{Asy|P}} - \frac{\tilde{\nu}_{Asy|P}\pi}{2} - \frac{\pi}{4}\right)\right| \\ &= \sqrt{\frac{2}{\pi\tilde{\gamma}_{n_{Asy|P|A}}}} |\cos((n-1)\pi)| = \sqrt{\frac{2}{\pi\tilde{\gamma}_{n_{Asy|P|A}}}}. \end{aligned} \tag{13.149}$$

Then, (13.138) and (13.149) yield the approximated normalization constants $\tilde{N}_{n_{Asy|P|A}}$

$$\tilde{N}_{n_{Asy|P|A}} \approx \frac{g}{c} \frac{\sqrt{-\alpha}(1-2\alpha)T^{\alpha-\frac{1}{2}}}{4(1+\alpha)} \pi \cdot \sqrt{n - \frac{3}{4} + \frac{|\alpha|}{|2\alpha - 1|}}. \tag{13.150}$$

As for the eigenvalues, one merely has to substitute (13.148) into (13.139) to get

$$\tilde{\lambda}_{n_{Asy|P|A}} \approx \frac{c^2}{g^2} \frac{16(1+\alpha)^2 T^{1-2\alpha}}{\pi^2(-\alpha)(1-2\alpha)^2} \cdot \frac{1}{\left(n - \frac{3}{4} + \frac{|\alpha|}{|2\alpha - 1|}\right)^2}. \tag{13.151}$$

These are also the variances of Gaussian zero-mean orthogonal random variables $\tilde{Z}_{n_{Asy|P|A}}$ whose probability density is the same as (13.140).

In conclusion, we have proved the following:

Theorem 8 *The approximated asymptotic KL expansion of the instantaneous energy of the Gaussian noise emitted by a spaceship moving at a power-like acceleration is given by*

$$Y^2_{Asy|P|A}(t) \equiv X^2_{Asy|P|A}(\tau) - E\{X^2_{Asy|P|A}(\tau)\}$$
$$\approx \sqrt{1-2\alpha} \frac{t^{-\frac{\alpha}{2}-\frac{1}{4}}}{T^{\frac{1}{4}-\frac{\alpha}{2}}} \cdot \sum_{n=1}^{\infty} Z_{n_{Asy|P|A}} \cos\left(\gamma_{Asy|P|A} \frac{t^{\frac{1}{2}-\alpha}}{T^{\frac{1}{2}-\alpha}} - \frac{\pi}{4} - \frac{|2\alpha|}{|2\alpha - 1|} \cdot \frac{\pi}{2}\right). \tag{13.152}$$

13.5 CONCLUSION

Eight theorems have been proved concerning application of the KL eigenfunction expansion to radio communications between the Earth and a relativistic spaceship. First, we obtained the KL expansion for the noise emitted by the spaceship moving at an *arbitrary* acceleration with respect to its own reference frame. The results of the previous Chapter 12, where this acceleration was supposed to remain constant ($= g$), have thus been fully generalized. Yet the KL expansion for the general case is not much more complicated than for constant g: in all cases, in fact, the eigenfunctions are Bessel functions of the first kind, and the eigenvalues are the zeros of such Bessel functions. The noise total energy cumulants, mean value, and variance were obtained as byproducts of this KL expansion.

A second KL expansion, already studied in Chapter 12 for constant g, was also fully generalized to arbitrary acceleration: this is the instantaneous energy (i.e., the zero-mean square noise) eigenfunction expansion. Again, the eigenfunctions are Bessel functions of the first kind, and the eigenvalues are their zeros.

All these results undergo remarkable simplifications when one considers the *asymptotic* acceleration of the spaceship about one year after departure from Earth. In particular, if the accelaration behaves like a power of time, the formulas are simple enough to allow Bessel functions to be replaced by suitable sinusoids. The way is thus paved for computer numeric simulations, good for engineering design.

The KL expansion is *optimal* among all possible transforms inasmuch as the eigenfunction set is, by definition, the best orthonormal basis in the Hilbert space. Finding this basis only by means of numerical techniques may be time-consuming, even on fast computers, for one has to determine both the eigenvalues and eigenvectors of the large symmetric matrix representing noise autocorrelation. However, we have solved this problem by providing explicit—if only asymptotically approximated—eigenfunctions and eigenvalues directly expressed in terms of the acceleration of the emitting source. The first step has thus been taken in designing the optimal signal analysis procedure to filter out the noise received on Earth from a relativistically moving spaceship.

13.6 REFERENCES

[1] C. Maccone, "Relativistic Interstellar Flight and Gaussian Noise," *Acta Astronautica*, **17**(9) (1988), 1019–1027.

[2] C. Maccone, "Relativistic Interstellar Flight and Instantaneous Noise Energy," *Acta Astronautica*, **21**(3) (1990), 155–159.

[3] C. Maccone, "Relativistic Interstellar Flight and Genetics," *Journal of the British Interplanetary Society*, **43** (1990), 569–572.

[4] E. Mallove, and G. Matloff, *The Starflight Handbook: A Pioneer's Guide to Interstellar Travel*, Wiley, New York, 1989.

[5] D. Goldsmith, *The Quest for Extraterrestrial Life: A Book of Readings*, University Science Books, Mill Valley, CA, 1980.

[6] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 213–219.

[7] C. Maccone, "The Karhunen–Loève Expansion of the Zero-Mean Square Process of a Time-Rescaled Gaussian Process," *Bollettino dell'Unione Matematica Italiana*, Series 7, **2-A** (1988), 221–229.

[8] C. Maccone, "Special Relativity and the Karhunen–Loève Expansion of Brownian Motion," *Il Nuovo Cimento*, Series B, **100** (1987), 329–341.

[9] C. Maccone, "The Time-Rescaled Brownian Motion $B(t^{2H})$," *Bollettino dell'Unione Matematica Italiana*, Series 6, **4-C** (1985), 363–378.

[10] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1953.

[11] C. Maccone, "Relativistic Interstellar Flight Communication Theorems," *Acta Astronautica*, **26**(2) (1992), 105–117.

14

Genetics aboard relativistic spaceships

14.1 INTRODUCTION

This chapter was born out of the need to merge two topics apparently unrelated thus far, namely:

(i) the theory of relativistic interstellar flight; and
(ii) the stochastic processes of genetics.

Their unification is achieved by virtue of the notion of time-rescaled Brownian motion that embodies both time rescaling, typical of relativity, and Brownian motion, typical of the stochastic processes of genetics. Though the mathematics involved is not difficult, to set out the calculations in detail would require too much space. Thus, the main lines of thought only have been highlighted.

The rationale behind the chapter is that the huge time differences between the crew of a relativistic spacecraft and people on Earth are not as "unbelievable" as they seem, inasmuch as the biological laws of genetics are mathematically proved to apply equally well to both. It is true that, whenever statistical mechanics and thermodynamics are involved, transformations from a reference frame at rest to a moving reference frame are controversial at the present time. However, we shall not take thermodynamics into account, nor statistical mechanics as it is understood in the theory of gases. Our proof only considers a special-relativistic transformation of coordinates and time affecting a stochastic process representing the number of living members in a finite population traveling aboard a relativistic spacecraft.

All topics discussed in this chapter were first published by the author in 1990 in [18].

14.2 DIFFUSION PARTIAL DIFFERENTIAL EQUATION FOR *X(t)*

This section deals with diffusion phenomena and their relationship to time-rescaled Brownian motion. We first consider the partial differential equation for a Brownian diffusion that is rescaled in time, and show that its solution coincides with the time-rescaled Gaussian density (F.42). Second, we derive the probability density function for the first-passage time of the $X(t)$ process at any positive value $X = a$. This topic thus provides a further example of application of the mathematical theory developed throughout this book.

Consider the partial differential equation

$$\frac{\partial p(x,t)}{\partial t} + \eta(t)\frac{\partial p(x,t)}{\partial x} = \frac{\sigma^2(t)}{2}\frac{\partial^2 p(x,t)}{\partial t^2} \tag{14.1}$$

with the initial condition

$$p(x,0) = \delta(x). \tag{14.2}$$

In (14.1) we assume $p(x,t)$ to be a probability density in x, and the physical meaning of (14.2) is that the particle randomly moves forward and backward along the x-axis, departing from the origin (with probability one) at the initial instant $t = 0$. We may then introduce the corresponding characteristic function (i.e., the Fourier transform), that depends on both "independent variable" ζ and "parameter" t

$$\Phi(\zeta,t) = \int_{-\infty}^{\infty} e^{i\zeta x} p(x,t)\, dx. \tag{14.3}$$

The first step solving the partial differential equation (14.1), subjected to the initial condition (14.2), is re-writing (14.2) in terms of the characteristic function (14.3), rather than in terms of the probability density function $p(x,t)$. To this end, we merely have to set $t = 0$ into (14.3) and substitute (14.2) into (14.3)

$$\Phi(\zeta,0) = \int_{-\infty}^{\infty} e^{i\zeta x} p(x,0)\, dx = \int_{-\infty}^{\infty} e^{i\zeta x}\delta(x)\, dx = [e^{i\zeta x}]_{x=0} = e^0 = 1; \tag{14.4}$$

that is

$$\Phi(\zeta,0) = 1. \tag{14.5}$$

We now want to find a new differential equation in the unknown function $\Phi(\zeta,t)$ by differentiating (14.3) with respect to t followed by a substitution of the right-hand side of (14.1):

$$\frac{\partial \Phi(\zeta,t)}{\partial t} = \int_{-\infty}^{\infty} e^{i\zeta x} \frac{\partial p(x,t)}{\partial t}\,dx$$
$$= \int_{-\infty}^{\infty} e^{i\zeta x}\left[-\eta(t)\frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2(t)}{2}\frac{\partial^2 p(x,t)}{\partial t^2}\right] dx$$
$$= -\eta(t)\int_{-\infty}^{\infty} e^{i\zeta x}\frac{\partial p(x,t)}{\partial x}\,dx + \frac{\sigma^2(t)}{2}\int_{-\infty}^{\infty} e^{i\zeta x}\frac{\partial^2 p(x,t)}{\partial t^2}\,dx. \qquad (14.6)$$

We can now integrate both integrals by parts and note that all the integrated parts vanish at $\pm\infty$ because of the complex exponential. The result is

$$\frac{\partial \Phi(\zeta,t)}{\partial t} = i\zeta\eta(t)\Phi(\zeta,t) + \frac{\sigma^2(t)}{2}(-i\zeta)\int_{-\infty}^{\infty} e^{i\zeta x}\frac{\partial p(x,t)}{\partial x}\,dx$$

and one more integration by parts finally yields

$$\frac{\partial \Phi}{\partial t} = i\zeta\eta(t)\Phi(\zeta,t) - \frac{\sigma^2(t)}{2}\zeta^2\Phi(\zeta,t) = \left[i\zeta\eta(t) - \frac{\sigma^2(t)}{2}\zeta^2\right]\Phi(\zeta,t).$$

In conclusion, we have found the new differential equation for the unknown characteristic function $\Phi(\zeta,t)$

$$\frac{\partial \Phi(\zeta,t)}{\partial t} = \left[i\zeta\eta(t) - \frac{\sigma^2(t)}{2}\zeta^2\right]\Phi(\zeta,t) \qquad (14.7)$$

This is an ordinary differential equation that may be solved by separating the variables

$$\frac{d\Phi(\zeta,t)}{\Phi(\zeta,t)} = \left[i\zeta\eta(t) - \frac{\sigma^2(t)}{2}\zeta^2\right] dt. \qquad (14.8)$$

Hence, an integration between the limits 0 and t yields

$$\ln\Phi(\zeta,t) - \ln\Phi(\zeta,0) = i\zeta\int_0^t \eta(s)\,ds - \frac{\zeta^2}{2}\int_0^t \sigma^2(s)\,ds. \qquad (14.9)$$

If we invoke the initial condition (14.5), we get

$$\ln\Phi(\zeta,0) = \ln 1 = 0 \qquad (14.10)$$

and this changes (14.9) into

$$\Phi(\zeta,t) = e^{i\zeta\int_0^t \eta(s)\,ds - \frac{\zeta^2}{2}\int_0^t \sigma^2(s)\,ds}. \qquad (14.11)$$

Having thus found the characteristic function $\Phi(\zeta,t)$, we must now inverse-Fourier-transform in order to get the probability density function $p(x,t)$. Fortunately, this is no problem in the case of (14.11), for (14.11) is just the inverse Fourier of the Gaussian probability density function

$$p(x,t) = \frac{1}{\sqrt{2\pi \int_0^t \sigma^2(s)\, ds}} e^{-\frac{(x - \int_0^t \eta(s)\, ds)^2}{2\int_0^t \sigma^2(s)\, ds}}. \tag{14.12}$$

This is thus the full solution to the partial differential equation (14.1) subjected to the initial condition (14.2).

Let me now check (14.12) against the time-rescaled Gaussian probability density (F.42) of the process $X(t)$. They are seen to coincide if we set

$$\begin{cases} \int_0^t \eta(s)\, ds = 0, \\ \int_0^t \sigma^2(s)\, ds = \int_0^t f^2(s)\, ds. \end{cases} \tag{14.13}$$

By differentiating both equations with respect to time, these equations become

$$\begin{cases} \eta(t) \equiv 0, \\ \sigma(t) = f(t). \end{cases} \tag{14.14}$$

The function $\sigma(t) \equiv f(t)$ is sometimes called *infinitesimal variance*, especially in the applications of stochastic processes to genetics.

Finally, by substituting (14.14) into (14.1), we see that the diffusion partial differential equation for the time-rescaled Gaussian process $X(t)$ is given by

$$\frac{\partial f_{X(t)}(x)}{\partial t} = \frac{f^2(t)}{2} \cdot \frac{\partial^2 f_{X(t)}(x)}{\partial x^2} \tag{14.15}$$

with the initial condition

$$f_{X(0)}(x) = \delta(x). \tag{14.16}$$

In particular, since the standard Brownian motion is the special case $f(t) \equiv 1$ we infer from (14.15) that the diffusion partial differential equation for $B(t)$ is

$$\frac{\partial f_{B(t)}(x)}{\partial t} = \frac{1}{2} \cdot \frac{\partial^2 f_{B(t)}(x)}{\partial x^2} \tag{14.17}$$

again with the initial condition

$$f_{B(0)}(x) = \delta(x). \tag{14.18}$$

14.3 FIRST-PASSAGE TIME FOR *X(t)*

A particle moves back and forth along the x-axis according to the process $X(t)$. The time required for the particle to reach the fixed value $X = a$ for the first time is a random variable, called the *first-passage time*, and denoted by T_a. We shall now find its probability density by extending it to the process $X(t)$ the method used by Feller [1, p. 174] to determine the first-passage time probability density of standard Brownian motion.

The probability that the process $X(t)$ remains smaller than the threshold a is given by

$$P\{X(t) < a\} = \int_{-\infty}^{a} f_{X(t)}(x)\,dx = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma_{X(t)}} e^{-\frac{x^2}{2\sigma^2_{X(t)}}}\,dx. \tag{14.19}$$

The substitution

$$y^2 = \frac{x^2}{\sigma^2_{X(t)}}$$

yields then

$$P\{x < a\} = \int_{-\infty}^{\frac{a}{\sigma_{X(t)}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\,dy = N\left(\frac{a}{\sigma_{X(t)}}\right),$$

where the N notation for the normal distribution was adopted. On the other hand, we can follow up Feller's proof by considering the probability

$$F_{T_a}(t) = P\{T_a < t\} = 2P\{X(t) > a\} = 2\left[1 - N\left(\frac{a}{\sigma_{X(t)}}\right)\right], \tag{14.20}$$

which is the first-passage time distribution function, as $X(T_a) = a$.

The derivative of (14.20) with respect to t is the required probability density of the first-passage time

$$f_{T_a}(t) = \frac{dF_{T_a}(t)}{dt} = -2\frac{dN\left(\frac{a}{\sigma_{X(t)}}\right)}{d\left(\frac{a}{\sigma_{X(t)}}\right)} \cdot \frac{d\left(\frac{a}{\sigma_{X(t)}}\right)}{d\sigma_{X(t)}} \cdot \frac{d\sigma_{X(t)}}{dt};$$

that is

$$f_{T_a}(t) = \frac{1}{\sigma_{X(t)}} \cdot \frac{d\sigma_{X(t)}}{dt} \cdot \frac{\sqrt{2}|a|}{\sqrt{\pi}} e^{-\frac{a^2}{2\sigma^2_{X(t)}}} \tag{14.21}$$

where we had to introduce the absolute value of a because (14.20) is a probability density.

An interesting result, first proved by Borovkov [2], is also found at once by equalizing the two densities (in x and t, respectively) (14.21) and (G.8)

$$f_{T_a}(t) = \frac{2|a|}{\sigma_{X(t)}} \cdot \frac{d\sigma_{X(t)}}{dt} \cdot f_{X(t)}(a). \tag{14.22}$$

This completes the investigation of the first-passage time probability density for the $X(t)$ process.

14.4 RELATIVISTIC INTERSTELLAR FLIGHT

Imagine a spacecraft accelerating with respect to its own reference frame, in such a way that the crew experience constant acceleration which, for their comfort, we assume numerically equal to

$$g = 9.8 \text{ m s}^{-2}. \tag{14.23}$$

Then let

t = time on Earth since departure (coordinate time)
τ = time on the spacecraft since departure (proper time)
$v(t)$ = spacecraft velocity with respect to the Earth
m = spacecraft (rest) mass.

Special relativity leads to the longitudinal force [5]

$$f_{\parallel} = \left[1 - \frac{v^2(t)}{c^2}\right]^{-\frac{3}{2}} m \frac{dv(t)}{dt}. \tag{14.24}$$

So we have to solve the differential equation in the unknown velocity

$$\left[1 - \frac{v^2(t)}{c^2}\right]^{-\frac{3}{2}} m \frac{dv(t)}{dt} = mg. \tag{14.25}$$

Let us set

$$v(t) = c \sin \Omega(t) \tag{14.26}$$

By separating the variables in (14.25) and performing elementary integration, one gets

$$\Omega(t) = \arctan\left(\frac{g}{c} t\right), \tag{14.27}$$

whence

$$v(t) = c \sin\left(\arctan\left(\frac{g}{c} t\right)\right). \tag{14.28}$$

But

$$\sin(\arctan(x)) = \frac{x}{\sqrt{1 + x^2}}. \tag{14.29}$$

So that the spacecraft velocity is given by

$$v(t) = \frac{gt}{\sqrt{1 + \left(\frac{g}{c} t\right)^2}}. \tag{14.30}$$

From special relativity it is also known that the spacecraft proper time is given by

$$\tau(t) = \int_0^t \sqrt{1 - \frac{v^2(s)}{c^2}}\, ds. \tag{14.31}$$

By substituting (14.30) into (14.31), and performing elementary integration, one derives the spacecraft proper time as

$$\tau(t) = \frac{c}{g}\operatorname{arcsinh}\left(\frac{g}{c}t\right) = \frac{c}{g}\ln\left[\frac{g}{c}t + \sqrt{1 + \left(\frac{g}{c}t\right)^2}\right]. \tag{14.32}$$

The physical meaning of (14.32) is that the time on board the moving spacecraft elapses much more slowly than on Earth. For instance, by turning the spacecraft at mid-way, and then decelerating constantly, the center of our Galaxy might be reached in just 21 years of proper time, although it lies 30,000 lt-yr from us.

14.5 TIME-RESCALED BROWNIAN MOTION

Let us consider ordinary Brownian motion (or the Wiener–Lévy process) with mean zero and variance t

$$B(t) = \text{Brownian motion}, \tag{14.33}$$

$$B(0) = 0. \tag{14.34}$$

Its (first order) probability density function is the Gaussian

$$f_{B(t)}(x) = \frac{1}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{x^2}{2t}}, \tag{14.35}$$

with initial condition

$$f_{B(0)}(x) = \delta(x) = \text{Dirac delta function}. \tag{14.36}$$

A white-noise integral is a stochastic process defined by

$$X(t) = \int_0^t f(s)\, dB(s) \tag{14.37}$$

with

$$f(t) = \text{time-rescaling function} \tag{14.38}$$

which may be any time function, continuous, non-negative, and with a non-negative derivative. Clearly, ordinary Brownian motion is the particular case of a white-noise integral when the time-rescaling function equals 1. Also, the initial condition for (14.37) is

$$X(0) = 0 \tag{14.39}$$

and it can be proved [3] that (14.37) is a time-rescaled Brownian motion, namely

$$X(t) = B\left(\int_0^t f^2(s)\, ds\right). \tag{14.40}$$

In other words, (14.40) is just a Brownian motion having a new time variable defined by

$$\tau(t) = \int_0^t f^2(s)\, ds. \tag{14.41}$$

For future reference, we prefer to rewrite (14.41) in the slightly different form that follows by differentiating with respect to the coordinate time and then taking the reciprocal of both sides

$$\frac{dt}{d\tau} = \frac{1}{f^2(t)}. \tag{14.42}$$

From (14.40) it may be inferred that (14.37) is a Gaussian process with zero-mean variance

$$\sigma^2_{X(t)} = \int_0^t f^2(s)\, ds, \tag{14.43}$$

and (first-order) probability density given by the time-rescaled Gaussian

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi}\sqrt{\int_0^t f^2(s)\, ds}}\, e^{-\frac{x^2}{2\int_0^t f^2(s)\, ds}}. \tag{14.44}$$

An important topic related to Brownian motion is the diffusion partial differential equation (also called the *Kolmogorov forward equation, or Fokker–Planck equation*) fulfilled by the Gaussian density (14.35)

$$\frac{\partial f_{B(t)}(x)}{\partial t} = \frac{1}{2}\frac{\partial^2 f_{B(t)}(x)}{\partial x^2}. \tag{14.45}$$

This may be proved either by direct differentiation of (14.35) or by Fourier transform techniques, as in Section 14.2. Correspondingly, the diffusion partial differential equation fulfilled by the time-rescaled Gaussian density (14.44) may be proved to be

$$\frac{\partial f_{X(t)}(x)}{\partial t} = \frac{f^2(t)}{2}\frac{\partial^2 f_{X(t)}(x)}{\partial x^2}, \tag{14.46}$$

with the initial condition

$$f_{X(0)}(x) = \delta(x). \tag{14.47}$$

Having thus paved the mathematical way, the genetics may now be considered.

14.6 GENETICS

Biology is a scientific discipline not yet recast in mathematical terms. Nevertheless, a few areas of biology may be understood mathematically. Genetics is one of them. Mendel's laws, published in 1866, were ignored by the scientific community until 1900, and only after 1908, when the Hardy–Weinberg law was formulated, did the field of population genetics take a precise mathematical form. Probability played a central role in this development, but it should be remembered that, within the wide field of probability, the theory of stochastic processes did not shape up neatly until about 1930. Subsequently, R. A. Fisher, S. Wright, and others concluded that the partial differential equation

$$\frac{\partial\phi(p,x;t)}{\partial t} = \frac{1}{4N}\frac{\partial^2}{\partial x^2}[x(1-x)\phi(p,x;t)] \tag{14.48}$$

represents the diffusion of a single gene belonging to one of two genotypes [8].

Let us now skip the definition of the terms "allele" used by biologists. A certain gene (e.g., the gene accounting for one's eye color) may have several alternative forms (i.e., one's eyes may be brown, blue, green, etc.). Each of these alternative forms of the same gene is said to be an "allele" of that gene. Now, denote a certain gene by A—namely, let A be one allele of the biallelic (A and a) population under consideration. Then the meaning of the other quantities appearing in (14.48) is

$t =$ time measured with one generation as the unit; (14.49)

$N =$ population size (i.e., there are N diploid individuals)—"diploid" means that each member of the population has *two* copies of each gene, because he/she got one from his/her father and one from his/her mother. This number N stays constant over generations; (14.50)

$\phi(p,x;t) =$ probability that the frequency of allele A in the population becomes x in the tth generation, given that it is p at $t = 0$ (14.51)

In other words, one has

$$0 \le x \le 1 \tag{14.52}$$

with the initial condition

$$\phi(p,x;0) = p. \tag{14.53}$$

The exact solution to (14.48), fulfilling the initial condition (14.53), was obtained by M. Kimura in 1955 [9] and is expressed as follows:

$$\begin{aligned}\phi(p,x;t) &= \sum_{i=1}^{\infty} p(1-p)i(i+1)(2i+1)\\ &= F(1-i,i+2;2;p)\cdot F(1-i,i+2;2;x)\cdot e^{-\frac{i(i+1)t}{4N}}\end{aligned} \tag{14.54}$$

where $F(a,b;c;x)$ is the hypergeometric function.

The intuitive meaning of (14.48) is that gene fluctuations are a Brownian motion with two absorbing barriers at

$$x = 0 \quad \text{and} \quad x = 1. \tag{14.55}$$

These barriers are mathematically represented by the term

$$x\cdot(1-x) \tag{14.56}$$

on the right-hand side of (14.48). In words, the percentage (or the relative frequency) of individuals in the population having gene A changes randomly from generation to generation, just like a Brownian motion. However, sooner or later, the time will come when either no individual in the population has the gene A (i.e., the barrier at zero is reached), or every individual has the gene A (i.e., the barrier at one is reached). The situation will then no longer change, and all the population will stay homozygous for the allele a or A, respectively. Thus, (14.48) is the mathematical formulation of the law of tendency to homozygosity (see [10, 11, and 12].[1]

14.7 RELATIVISTIC GENETICS

It is now possible to lay the foundations of relativistic genetics. To this end, let us identify the proper time of special relativity, given by (14.31), with the new time variable of time-rescaled Brownian motion, given by (14.41)

$$\int_0^t \sqrt{1 - \frac{v^2(s)}{c^2}}\, ds = \int_0^t f^2(s)\, ds. \tag{14.57}$$

Differentiating both sides with respect to time gives the relationship between the velocity of the moving spacecraft and the time-rescaling function

$$\sqrt{1 - \frac{v^2(t)}{c^2}} = f^2(t). \tag{14.58}$$

This is valid for any kind of special-relativistic motion. Reference to relativistic interstellar flight, as defined by (14.30), and (14.58), yield the time-rescaling function

$$f^2(t) = \frac{1}{\sqrt{1 + \left(\frac{g}{c}t\right)^2}}. \tag{14.59}$$

The next matter is the diffusion partial differential equation (14.48) for the single biallelic gene. Since the gene fluctuations, defined by (14.48), are just a Brownian motion with two absorbing barriers, we must substitute the probability density

$$f_{X(t)}(x) = \phi(p, x; t) \tag{14.60}$$

into (14.46) to let the unknown functions of (14.46) and (14.48) coincide. These are actually the same thing because nothing more than time rescaling occurs when transferring from the Earth reference frame to the moving spacecraft reference frame. Rearranging puts (14.46) in the form

$$\frac{\partial \phi(p, x; t)}{\partial t} \frac{1}{f^2(t)} = \frac{1}{2} \frac{\partial^2 \phi(p, x; t)}{\partial x^2}. \tag{14.61}$$

[1] In the last reference [12], however, the Kolmogorov backward equation is used, rather than the forward one, as here.

To let (14.61) represent the gene fluctuations, the two-barriers term (14.56) (divided by twice the population size) must be inserted inside its right-hand side[2]

$$\frac{\partial \phi(p,x;t)}{\partial t}\frac{1}{f^2(t)} = \frac{1}{4N}\frac{\partial^2}{\partial x^2}[x(1-x)\phi(p,x;t)]. \tag{14.62}$$

Then, using (14.42) and the chain rule produces

$$\frac{\partial \phi}{\partial t}\frac{1}{f^2(t)} = \frac{\partial \phi}{\partial t}\frac{dt}{d\tau} = \frac{\partial \phi}{\partial \tau} \tag{14.63}$$

and (14.62) takes the form

$$\frac{\partial \phi(p,x;\tau)}{\partial \tau} = \frac{1}{4N}\frac{\partial^2}{\partial x^2}[x(1-x)\phi(p,x;\tau)] \tag{14.64}$$

(14.64) is the equivalent of the diffusion equation (14.48) in the moving spacecraft reference frame.

In conclusion, we have shown how gene fluctuations obey just the same mathematical laws on both the Earth and the relativistic spacecraft, once the time-rescaling required by special relativity has been taken into account.

14.8 A GLANCE AHEAD

The exact expression of time-rescaled Brownian motion (14.40) is difficult to handle because of the complicated expression of proper time for relativistic interstellar flight. It is possible, however, to substitute (14.32) by the approximated version that one gets for large values of the time. In fact, the 1 under the square root is then dwarfed by the other term, and one may write

$$\frac{c}{g}\ln\left[\frac{g}{c}t + \sqrt{1+\left(\frac{g}{c}t\right)^2}\right] \approx \frac{c}{g}\ln\left[2\frac{g}{c}t\right]. \tag{14.65}$$

Thus, we call asymptotic Brownian motion the stochastic process

$$B\left(\frac{c}{g}\ln\left[2\frac{g}{c}t\right]\right). \tag{14.66}$$

In Chapter 12 of this book (and, earlier, in [13]) the author showed that the above approximation starts being physically acceptable about one year after the spacecraft departure from Earth, and that the exact Karhunen–Loève eigenfunction expansion of the Gaussian process (14.66) reads

$$B\left(\frac{c}{g}\ln\left[2\frac{g}{c}t\right]\right) = \sum_{n=1}^{\infty} Z_n \frac{1}{\sqrt{T}|J_0(\gamma_n)|} J_0\left(\gamma_n \frac{\sqrt{t}}{\sqrt{T}}\right) \tag{14.67}$$

[2] For a mathematical justification of this, see, for instance, [8].

where $\gamma_n = n$th real positive zero of the Bessel function $J_1(x)$;
$Z_n =$ Gaussian random variable with zero mean and variance $\dfrac{4cT}{g}\dfrac{1}{\gamma_n^2}$;

and that (14.67) is valid for $0 \le t \le T$. Finally, by resorting to the well-known asymptotic expansion of the Bessel function $J_0(\ldots)$ for large values of its argument, (14.67) may be put in the form

$$B\left(\frac{c}{g}\ln\left[2\frac{g}{c}t\right]\right) \approx \sum_{n=1}^{\infty} Z_n \frac{1}{\sqrt{\sqrt{T}\sqrt{t}}}\cos\left(\gamma_n \frac{\sqrt{t}}{\sqrt{T}} - \frac{\pi}{4}\right). \tag{14.68}$$

Whether this result has genetical significance is unclear.

In a series of papers [14–18], this author also gave the Karhunen–Loève eigenfunction expansion of the general time-rescaled Brownian motion (14.40), but the calculations are then made difficult by Bessel functions whose order changes in time. Once again, whether these results have genetical significance still is an open question. This leads one to wonder whether the KL eigenfunction expansion, so useful in communication theory, might become just as useful in genetics.

14.9 REFERENCES

[1] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd Edition, Wiley, New York, 1971.

[2] A. A. Borovkov, "On the First Passage Time for One Class of Processes with Independent Increments," *Theory of Probability and Applications*, **X**(2) (1965), 331–334.

[3] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 213–220.

[4] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

[5] W. Rindler, *Special Relativity*, Oliver & Boyd, Edinburgh, 1960, p. 41.

[6] L. Arnold, *Stochastic Differential Equations*, Wiley, New York, 1973, pp. 84–87.

[7] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1965, p. 552.

[8] M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, 1983, pp. 36–40.

[9] M. Kimura, "Solution of a process of random genetic drift with a continuous model," *Proc. Natl. Acad. Sci. U.S.A.*, **41** (1955), 144–150.

[10] P. A. P. Moran, *The Statistical Processes of Evolutionary Theory*, Clarendon Press, Oxford, 1962, pp. 75–78.

[11] M. Kimura and T. Ohta, *Theoretical Aspects of Population Genetics*, Princeton University Press, Princeton, NJ, 1971, pp. 167–196.

[12] T. Maruyama, "Stochastic Problems in Population Genetics," *Lecture Notes in Biomathematics*, No. 17, pp. 24–28, Springer-Verlag, Berlin, 1977.

[13] C. Maccone, "Relativistic Interstellar Flight and Gaussian Noise," *Acta Astronautica*, **17**(9) (1988), 1019–1027.

[14] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 213–220.

[15] C. Maccone, "Sur un mouvement brownien dont le temps est retardé exponentiellement," *Publications de l'Institut de Statistique de l'Université de Paris*, **30** (1985), 61–72.
[16] C. Maccone, "Special Relativity and the Karhunen–Loève Expansion of Brownian Motion," *Il Nuovo Cimento*, Series B, **100** (1987), 329–342.
[17] C. Maccone, "The Karhunen–Loève Expansion of the Zero-Mean Square Process of a Time-Rescaled Gaussian Process," *Bollettino dell'Unione Matematica Italiana*, Series 7, **2-A** (1988), 221–229.
[18] C. Maccone, "Relativistic Interstellar Flight and Genetics," *Journal of the British Interplanetary Society*, **43** (1990), 569–572.

Appendix A

Engineering tradeoffs for the "FOCAL" spacecraft antenna

The antenna is clearly the most important part of the FOCAL spacecraft, for it acts as both the receiver of the radio waves focused by the Sun and also as the radio link with the Earth. Thus, optimally sizing the antenna dish will be the primary task of the designers of the FOCAL spacecraft. To help them with some preliminary considerations, a series of diagrams are shown plotting the gain vs. the antenna size, measured in meters, for a variety of sizes, gradually ranging from 0 meters up to 1,000 meters.

The six plots (Figures A.1 through A.6) show the antenna's and Sun's share of the total gain of the combined system Sun + spacecraft antenna. These plots clearly show the advantage of the combined system Sun + spacecraft antenna over the antenna system alone, which is why the FOCAL is being designed.

Antennas up to 10 meters in diameter are already within today's technological realm. For instance, the first large space antennas orbiting the Earth in the form of radio-astronomy satellites, that is, the Japanese *VSOP/HALCA* and Russian–International *Radioastron* spacecraft, have antennas 8 meters and 10 meters in diameter, respectively. The FOCAL space mission, however, would require a larger antenna (actually an antenna as large as possible) in order to insure reliable and fast communications with the Earth from the huge distances of 550 AU to 1,000 AU. Therefore, it does not seem inappropriate to include in our gallery of "gain vs. antenna diameter" plots of antennas of diameter up to about 100 meters.

Large antennas for use in space can hardly be of the "rigid" type, like the VSOP/HALCA one. They would rather have to be of the "inflatable" type that was prototyped around 1985 for the intended NASA/ESA *QUASAT* radioastronomy satellite, never constructed due to lack of funds (see [1]).

Finally, one more revolutionary idea pops up in the picture: that of a solar sail being also a very large space antenna. This idea was nurtured at the first conferences about the FOCAL space missions and it is not entirely "crazy" to think that one day a 1 km diameter sail could also be a wonderful 1 km antenna.

The following sequence of six graphs is intended to help engineers to realize what are the antenna gains in the game in order to design the FOCAL spacecraft antenna. Each of the six graphs shows two sets of curves: The set of four lower curves represents the spacecraft antenna gain (in dB) as a function of the dish increasing diameter from 0 m up to 1,000 m. The set of four upper curves represents the overall gain (i.e., Sun gain + spacecraft antenna gain) as a function of the increasing spacecraft dish diameter from 0 m up to 1,000 m. Of course, the Sun gain is the same for all graphs and marks the difference in height (in dB) between any one of the lower four curves and the corresponding upper curve. Let us now clarify which curve corresponds to which observing frequency: the thick two curves at the top of each of the two sets correspond to the frequency of the peak in the cosmic microwave backround (CMB), the primordial "fossil" radiation that filled the universe about 300,000 years after the Big Bang, at the time when radiation and matter separated from each other. In Chapter 9 of this book we show that this frequency equals 160.378 GHz (see Eq. (9.3)), and it is also stated that CMB radiation would be the ideal target radiation for FOCAL to measure because of its uniform distribution all over the celestial sphere. Thus, CMB peak measurement does not imply any precision tracking of the spacecraft at distances greater than 550 AU. In conclusion, the thick top curve represents the overall gain of the (Sun + spacecraft) antenna, while the lower thick curve shows the gain of the spacecraft antenna alone. The two second curves from the top (dash–dash curves) correspond to the observing frequency of 22 GHz. This is the water maser frequency, and one of the most important astrophysical spectral lines in the radio part ot the spectrum. The two third curves from the top (dot–dot curves) correspond to the observing frequency of 1.420 GHz, the famous "spin transition of the electron with respect to the proton" in neutral hydrogen. This is regarded as the most important frequency for SETI research as well. Finally, the two lowest curves in the two sets (dash–dot curves) correspond to the frequency of 327 MHz, just to give an example of a frequency in the megahertz range. Note that the (Sun + spacecraft) gain associated with this 327 MHz frequency is lower than the gain associated with the spacecraft alone when observing at the CMB peak of 160.378 GHz.

REFERENCE

[1] A. Hawkyard, and A. Anselmi, "QUASAT Industrial Phase A Study," Executive Summary, *Aeritalia GSS Report QS-RP-A1-0004*, 1988.

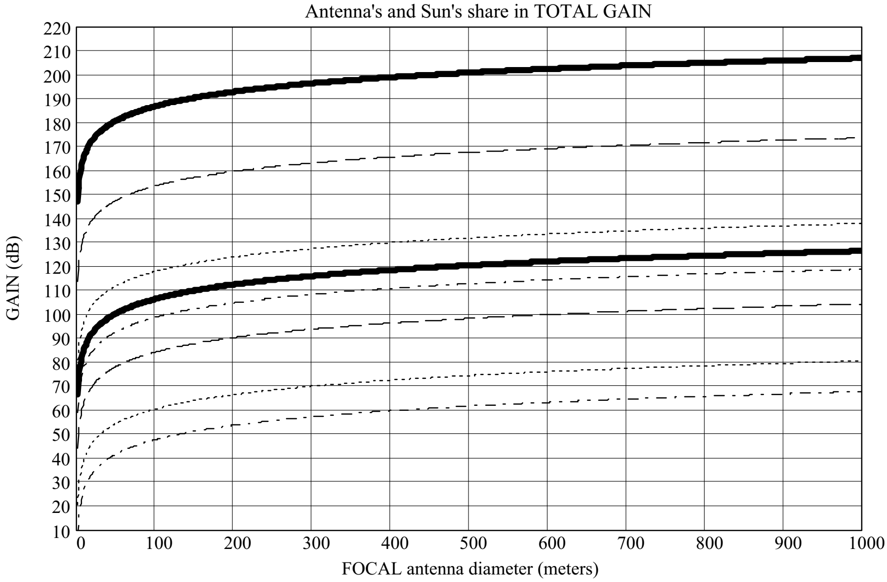


Figure A.1. This is the "overall picture" of the two sets of curves (Sun + spacecraft gain and spacecraft antenna gain alone) for an increasing spacecraft antenna diameter raising from 0 m up to 1,000 m. In the following five graphs, we "chop" this graph into five different increasing ranges of the FOCAL spacecraft antenna to let FOCAL design engineers get a feeling for the tradeoffs involved in the game.

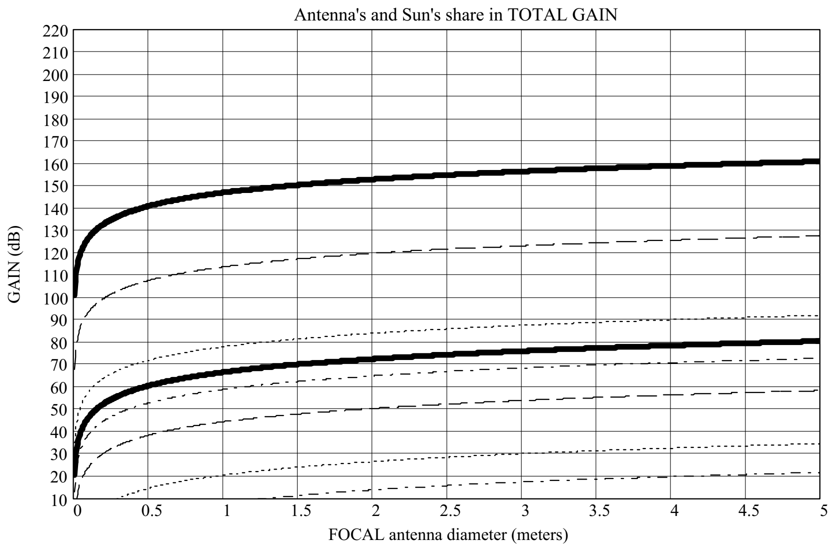


Figure A.2. This is the "first chop" of Figure A.1: the FOCAL spacecraft antenna is supposed to be small, ranging from 0 m to 5 m in diameter.

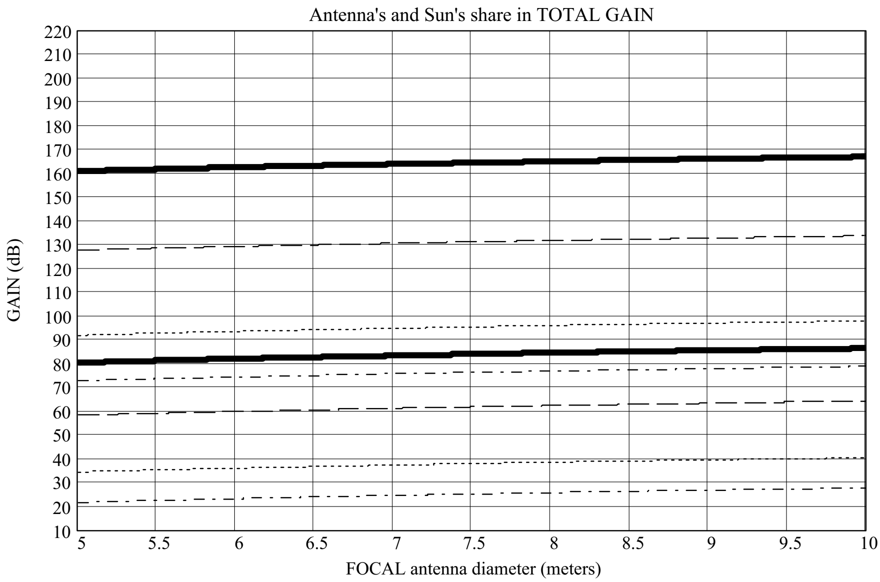


Figure A.3. This is the "second chop" of Figure A.1: the FOCAL antenna ranges from 5 m to 10 m in diameter. This is probably the size of the real FOCAL antenna when it will finally be designed and launched if the antenna is to be rigid.

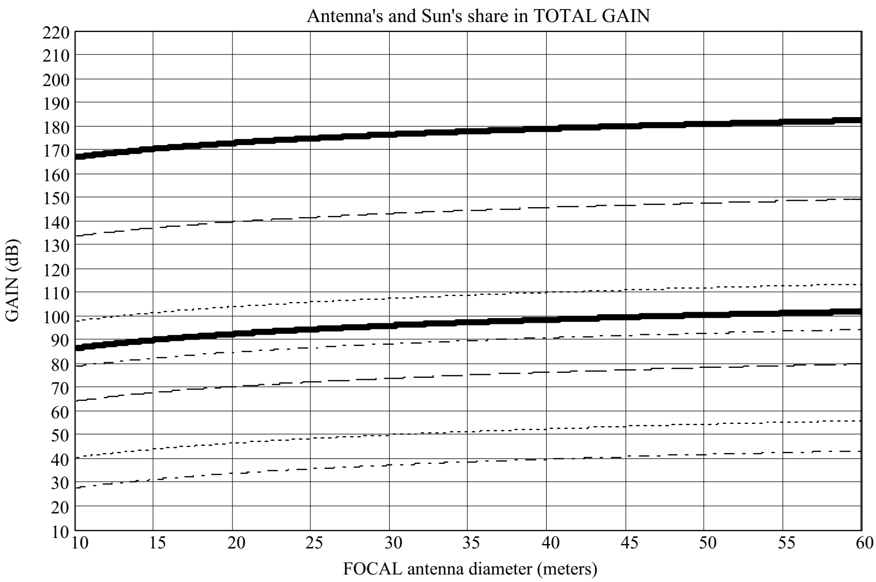


Figure A.4. Third chop of Figure A.1: the FOCAL antenna ranges from 10 m to 60 m in diameter. This is probably the size of the real FOCAL antenna in case it is going to be an inflatable antenna, or a solar sail that could somehow be used as an antenna as well.

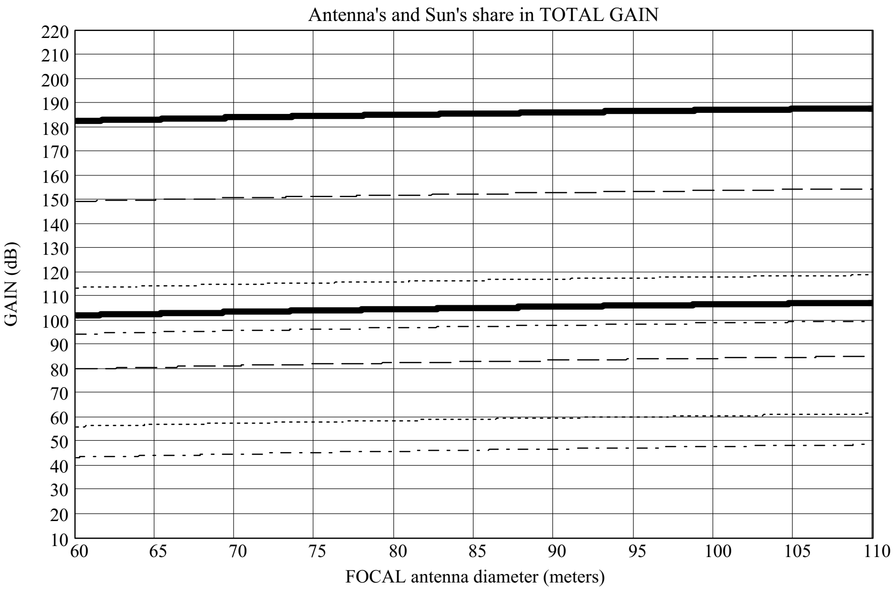


Figure A.5. Fourth chop of Figure A.1: antenna diameter ranging from 60 m up to 110 m.

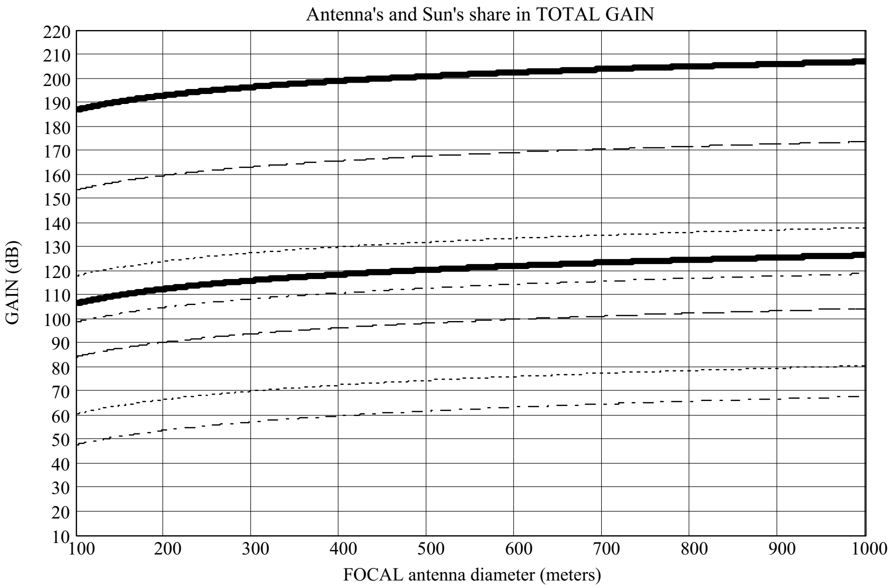


Figure A.6. Last chop of Figure A.1: antenna diameter ranging from 100 m to 1,000 m. We can hardly envisage how such an antenna could be constructed. At best, it would probably be a solar sail, to be somehow transformed into an antenna when the spacecraft is so far away from the Sun for the solar sail to be of use.

Appendix B

"FOCAL" Sun flyby characteristics

The numerical tables in this appendix give the parameters of the Sun flyby according to *four* different values of approach speed of the FOCAL spacecraft. These four different values of the approach speed are justified on the basis of the following four astronautical arguments, respectively:

(1) 11.185 km s^{-1}—this is the escape velocity from the Earth and so, assuming that the launch from Earth is orthogonal to the Earth trajectory around the Sun, it is the *minimal* speed by which the FOCAL spacecraft can possibly approach the Sun (Tables B.1 and B.2);
(2) 30 km s^{-1}—this is a spacecraft speed that could easily be achieved if the spacecraft was going to make some planetary flyby within the solar system (e.g., a Venus flyby) before approaching the Sun for the final flyby (Tables B.3 and B.4);
(3) 50 km s^{-1}—this is the spacecraft speed by which the Sun would be approached if the spacecraft was going to use the "Jupiter–Sun–Jupiter–Sun flyby", which was described in Chapter 4 (Tables B.5 and B.6);
(4) 75 km s^{-1}—this is a speed that could conceivably be achieved by virtue of several flybys within the solar system possibly improved by the use of thrusters (Tables B.7 and B.8).

Note: if the minimum spacecraft distance from the Sun surface turns out to be negative, the relevant Sun flyby cannot obviously take place, as the spacecraft would fall into the Sun! A safe rule of thumb for a "realistic" minimal spacecraft distance from the Sun surface seems to be about the distance of Mercury (i.e., 0.38 AU or about 80 Sun radii).

Table B.1. Sun flyby characteristics for the nearest 25 stellar systems for the Sun approach speed $v_\infty = 11.185$ km/s. Since this Sun approach speed is also the FOCAL spacecraft exit speed out of the solar system, at this speed it would take 233 years for FOCAL to reach 550 AU, and 423 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
1	0.553	119.010	0.548	118.010	57.637	0.00019
2	0.489	105.154	0.484	104.154	61.177	0.00020
3	0.218	46.877	0.213	45.877	91.157	0.00030
4	0.000	0.005	−0.005	−0.995	11,955.31	0.03988
5	0.210	45.139	0.205	44.139	92.874	0.00031
6	0.414	89.016	0.409	88.016	66.313	0.00022
7	0.182	39.087	0.177	38.087	99.188	0.00033
8	0.000	0.053	−0.004	−0.947	2,888.710	0.00964
9	0.494	106.284	0.489	105.284	61.069	0.00020
10	0.193	41.610	0.189	40.610	96.181	0.00032
11	0.000	0.011	−0.005	−0.989	5,174.544	0.01726
12	0.007	1.498	0.002	0.498	498.209	0.00166
13	0.796	171.327	0.792	170.327	48.577	0.00016
14	0.487	104.721	0.482	103.721	61.299	0.00020
15	0.069	14.865	0.064	13.865	159.949	0.00053
16	2.278	490.110	2.273	489.110	30.094	0.00010
17	0.410	88.137	0.405	87.137	66.879	0.00022
18	0.204	43.928	0.200	42.928	93.650	0.00031
19	0.164	35.314	0.159	34.314	104.274	0.00035
20	0.128	27.625	0.124	26.625	117.708	0.00039
21	0.129	27.823	0.125	26.823	117.293	0.00039
22	0.000	0.052	−0.004	−0.948	2,907.923	0.00970
23	1.430	307.577	1.425	306.577	36.923	0.00012
24	1.067	229.485	1.062	228.485	42.343	0.00014
25	0.164	35.346	0.160	34.346	104.228	0.00035

Table B.2. Sun flyby characteristics for the next nearest 25 stellar systems for the Sun approach speed $v_\infty = 11.185$ km/s. At this speed it would take 233 years for FOCAL to reach 550 AU, and 423 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
26	0.028	6.095	0.024	5.095	252.205	0.00084
27	0.000	0.004	−0.005	−0.996	13,327.44	0.4446
28	0.037	7.932	0.032	6.932	220.969	0.00074
29	0.010	2.104	0.005	1.104	421.248	0.00141
30	0.310	66.749	0.306	65.749	76.287	0.00025
31	0.353	75.939	0.348	74.939	71.920	0.00024
32	0.131	28.131	0.126	27.131	116.658	0.00039
33	0.000	0.047	−0.004	−0.953	3,078.777	0.01027
34	0.285	61.303	0.280	60.303	79.527	0.00027
35	0.113	24.317	0.108	23.317	125.365	0.00042
36	2.701	581.181	2.697	580.181	27.988	0.00009
37	0.956	205.780	0.952	204.780	44.443	0.00015
38	0.008	1.824	0.004	0.824	452.126	0.00151
39	0.200	42.959	0.195	41.959	94.684	0.00032
40	0.160	34.410	0.155	33.410	106.225	0.00035
41	0.021	4.584	0.017	3.584	291.057	0.00097
42	0.240	51.605	0.235	50.605	86.938	0.00029
43	0.187	40.200	0.182	39.200	98.350	0.00033
44	1.254	269.797	1.249	268.797	39.288	0.00013
45	0.641	137.889	0.636	136.889	53.873	0.00018
46	0.373	80.273	0.368	79.273	69.996	0.00023
47	0.332	71.396	0.327	70.396	74.123	0.00025
48	0.001	0.305	−0.003	−0.695	1,085.764	0.00362
49	0.062	13.407	0.058	12.407	168.342	0.00056
50	0.003	0.650	−0.002	−0.350	782.594	0.00261

Table B.3. Sun flyby characteristics for the nearest 25 stellar systems for the Sun approach speed $v_\infty = 30$ km/s. At this speed, it would take 86 years for FOCAL to reach 550 AU, and 158 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed *c*
1	0.077	16.543	0.0.72	15.543	154.593	0.00052
2	0.068	14.617	0.063	13.617	164.088	0.00055
3	0.030	6.516	0.026	5.516	244.498	0.00082
4	0.000	0.001	−0.005	−0.999	32,066.12	0.10696
5	0.029	6.275	0.025	5.275	249.103	0.00083
6	0.058	12.374	0.053	11.374	177.862	0.00059
7	0.025	5.433	0.021	4.433	266.039	0.00089
8	0.000	0.007	−0.005	−0.993	7,747.993	0.02584
9	0.069	14.774	0.064	13.774	163.796	0.00055
10	0.027	5.784	0.022	4.784	257.973	0.00086
11	0.000	0.001	−0.005	−0.999	13,878.97	0.04630
12	0.001	0.208	−0.004	−0.792	1,336.278	0.00446
13	0.111	23.815	0.106	22.815	130.291	0.00043
14	0.068	14.557	0.063	13.557	164.415	0.00055
15	0.010	2.066	0.005	1.006	429.009	0.00143
16	0.317	68.128	0.312	67.128	80.717	0.00027
17	0.057	12.251	0.052	11.251	179.380	0.00060
18	0.028	6.106	0.024	5.106	251.185	0.00084
19	0.023	4.909	0.018	3.909	279.680	0.00093
20	0.018	3.840	0.013	2.840	315.712	0.00105
21	0.018	3.868	0.013	2.868	314.598	0.00105
22	0.000	0.007	−0.005	−0.993	7,799.526	0.02602
23	0.199	42.755	0.194	41.755	99.032	0.00033
24	0.148	31.899	0.144	30.899	113.571	0.00038
25	0.023	4.913	0.018	3.913	279.557	0.00093

Table B.4. Sun flyby characteristics for the nearest next 25 stellar systems for the Sun approach speed $v_\infty = 30$ km/s. At this speed, it would take 86 years for FOCAL to reach 550 AU, and 158 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
26	0.004	0.847	−0.001	−0.153	676.455	0.00226
27	0.000	0.001	−0.005	−0.999	35,746.39	0.11924
28	0.005	1.103	0.000	0.103	592.676	0.00198
29	0.001	0.293	−0.003	−0.707	1,129.857	0.00377
30	0.043	9.278	0.038	8.278	204.615	0.00068
31	0.049	10.556	0.044	9.556	192.901	0.00064
32	0.018	3.910	0.014	2.910	312.895	0.00104
33	0.000	0.007	−0.005	−0.993	8,257.784	0.02754
34	0.040	8.521	0.035	7.521	213.305	0.00071
35	0.016	3.380	0.011	2.380	336.250	0.00112
36	0.375	80.787	0.371	79.787	75.067	0.00025
37	0.133	28.604	0.128	27.604	119.203	0.00040
38	0.001	0.253	−0.003	−0.747	1,212.676	0.00405
39	0.028	5.971	0.023	4.971	253.957	0.00085
40	0.022	4.783	0.018	3.783	284.912	0.00095
41	0.003	0.637	−0.002	−0.363	780.663	0.00260
42	0.033	7.173	0.029	6.173	233.181	0.00078
43	0.026	5.588	0.021	4.588	263.790	0.00088
44	0.174	37.503	0.170	36.503	105.377	0.00035
45	0.089	19.167	0.084	18.167	144.496	0.00048
46	0.052	11.158	0.047	10.158	187.741	0.00063
47	0.046	9.924	0.041	8.924	198.810	0.00066
48	0.000	0.042	−0.004	−0.958	2,912.196	0.00971
49	0.009	1.864	0.004	0.864	451.521	0.00151
50	0.000	0.090	−0.004	−0.910	2,099.045	0.00700

Table B.5. Sun flyby characteristics for the nearest 25 stellar systems for the Sun approach speed $v_\infty = 50$ km/s. At this speed, it would take 52 years for FOCAL to reach 550 AU, and 94 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
1	0.028	5.955	0.023	4.955	257.655	0.00086
2	0.024	5.262	0.020	4.262	273.480	0.00091
3	0.011	2.346	0.006	1.346	407.497	0.00136
4	0.000	0.000	−0.005	−1.000	53,443.53	0.17827
5	0.010	2.259	0.006	1.259	415.172	0.00138
6	0.021	4.455	0.016	3.455	296.437	0.00099
7	0.009	1.956	0.004	0.956	443.398	0.00148
8	0.000	0.003	−0.005	−0.997	12,913.32	0.04307
9	0.025	5.319	0.020	4.319	272.994	0.00091
10	0.010	2.082	0.005	1.082	429.955	0.00143
11	0.000	0.001	−0.005	−0.999	23,131.62	0.07716
12	0.000	0.075	−0.004	−0.925	2,227.130	0.00743
13	0.040	8.574	0.035	7.574	217.152	0.00072
14	0.024	5.240	0.020	4.240	274.025	0.00091
15	0.003	0.744	−0.001	−0.256	715.015	0.00239
16	0.114	24.526	0.109	23.526	134.529	0.00045
17	0.020	4.411	0.016	3.411	298.967	0.00100
18	0.010	2.198	0.006	1.198	418.641	0.00140
19	0.008	1.767	0.004	0.767	466.134	0.00155
20	0.006	1.382	0.002	0.382	526.186	0.00176
21	0.006	1.392	0.002	0.392	524.330	0.00175
22	0.000	0.003	−0.005	−0.997	12,999.21	0.04336
23	0.072	15.392	0.067	14.392	165.054	0.00055
24	0.053	11.484	0.049	10.484	189.284	0.00063
25	0.008	1.769	0.004	0.769	465.929	0.00155

Table B.6. Sun flyby characteristics for the next 25 stellar systems for the Sun approach speed $v_\infty = 50$ km/s. At this speed, it would take 52 years for FOCAL to reach 550 AU, and 94 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
26	0.001	0.305	−0.003	−0.695	1127.425	0.00376
27	0.000	0.000	−0.005	−1.000	59,577.33	0.19873
28	0.002	0.397	−0.003	−0.603	987.793	0.00329
29	0.000	0.105	−0.004	−0.895	1,883.095	0.00628
30	0.016	3.340	0.011	2.340	341.025	0.00114
31	0.018	3.800	0.013	2.800	321.501	0.00107
32	0.007	1.408	0.002	0.408	521.492	0.00174
33	0.000	0.002	−0.005	−0.998	13,762.97	0.04591
34	0.014	3.068	0.010	2.068	355.508	0.00119
35	0.006	1.217	0.001	0.217	560.417	0.00187
36	0.135	29.083	0.131	28.083	125.112	0.00042
37	0.048	10.298	0.043	9.298	198.672	0.00066
38	0.000	0.091	−0.004	−0.909	2,021.127	0.00674
39	0.010	2.150	0.005	1.150	423.262	0.00141
40	0.008	1.722	0.003	0.722	474.854	0.00158
41	0.001	0.229	−0.004	−0.771	1,301.105	0.00434
42	0.012	2.582	0.007	1.582	388.635	0.00130
43	0.009	2.012	0.005	1.012	439.650	0.00147
44	0.063	13.501	0.058	12.501	175.628	0.00059
45	0.032	6.900	0.027	5.900	240.827	0.00080
46	0.019	4.017	0.014	3.017	312.902	0.00104
47	0.017	3.573	0.012	2.573	331.349	0.00111
48	0.000	0.015	−0.005	−0.985	4,853.660	0.01619
49	0.003	0.671	−0.002	−0.329	752.535	0.00251
50	0.000	0.033	−0.004	−0.967	3,498.408	0.01167

Table B.7. Sun flyby characteristics for the nearest 25 stellar systems for the Sun approach speed $v_\infty = 75$ km/s. At this speed, it would take 34 years for FOCAL to reach 550 AU, and 63 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
1	0.012	2.647	0.008	1.647	386.482	0.00129
2	0.011	2.339	0.006	1.339	410.220	0.00137
3	0.005	1.043	0.000	0.043	611.246	0.00204
4	0.000	0.000	−0.005	−1.000	80,165.30	0.26740
5	0.005	1.004	0.000	0.004	622.758	0.00208
6	0.009	1.980	0.005	0.980	444.656	0.00148
7	0.004	0.869	−0.001	−0.131	665.097	0.00222
8	0.000	0.001	−0.005	−0.999	19,369.98	0.06461
9	0.011	2.364	0.006	1.364	409.491	0.00137
10	0.004	0.925	0.000	−0.075	644.933	0.00215
11	0.000	0.000	−0.005	−1.000	34,697.43	0.11574
12	0.000	0.033	−0.004	−0.967	3340.696	0.01114
13	0.018	3.810	0.013	2.810	325.727	0.00109
14	0.011	2.329	0.006	1.329	411.038	0.00137
15	0.002	0.331	−0.003	−0.669	1,072.522	0.00358
16	0.051	10.900	0.046	9.900	201.793	0.00067
17	0.009	1.960	0.004	0.960	448.451	0.00150
18	0.005	0.977	0.000	−0.023	627.962	0.00209
19	0.004	0.785	−0.001	−0.215	699.201	0.00233
20	0.003	0.614	−0.002	−0.386	789.279	0.00263
21	0.003	0.619	−0.002	−0.381	786.495	0.00262
22	0.000	0.001	−0.005	−0.999	19,498.81	0.06504
23	0.032	6.841	0.027	5.841	247.581	0.00083
24	0.024	5.104	0.019	4.104	283.927	0.00095
25	0.004	0.786	−0.001	−0.214	698.893	0.00233

Table B.8. Sun flyby characteristics for the next 25 stellar systems for the Sun approach speed $v_\infty = 75$ km/s. At this speed, it would take 34 years for FOCAL to reach 550 AU, and 63 years to reach 1,000 AU.

Stellar system #	Perihelion distance (AU)	Perihelion distance (Sun radii)	Minimum distance from Sun surface (AU)	Minimum distance from Sun surface (Sun radii)	Perihelion speed (km/s)	Perihelion speed w.r.t. light speed c
26	0.001	0.136	−0.004	−0.864	1,691.138	0.00564
27	0.000	0.000	−0.005	−1.000	89,365.99	0.29809
28	0.001	0.176	−0.004	−0.824	1481.690	0.00494
29	0.000	0.047	−0.004	−0.953	2,824.643	0.00942
30	0.007	1.485	0.002	0.485	511.537	0.00171
31	0.008	1.689	0.003	0.689	482.251	0.00161
32	0.003	0.626	−0.002	−0.374	782.238	0.00261
33	0.000	0.001	−0.005	−0.999	20,644.46	0.06886
34	0.006	1.363	0.002	0.363	533.262	0.00178
35	0.003	0.541	−0.002	−0.459	840.625	0.00280
36	0.060	12.926	0.055	1.926	187.668	0.00063
37	0.021	4.577	0.017	3.577	298.009	0.00099
38	0.000	0.041	−0.004	−0.959	3,031.690	0.01011
39	0.004	0.955	0.000	−0.045	634.892	0.00212
40	0.004	0.765	−0.001	−0.235	712.281	0.00238
41	0.000	0.102	−0.004	−0.898	1,951.657	0.00651
42	0.005	1.148	0.001	0.148	582.953	0.00194
43	0.004	0.894	0.000	−0.106	659.476	0.00220
44	0.028	6.000	0.023	5.000	263.441	0.00088
45	0.014	3.067	0.010	2.067	361.240	0.00120
46	0.008	1.785	0.004	0.785	469.354	0.00157
47	0.007	1.588	0.003	0.588	497.024	0.00166
48	0.000	0.007	−0.005	−0.993	7,280.491	0.02429
49	0.001	0.298	−0.003	−0.702	1,128.802	0.00377
50	0.000	0.014	−0.005	−0.986	5,247.612	0.01750

Appendix C

Mission to the solar gravitational focus by solar sailing

G. Vulpetti[1]

The aim of this appendix is to show that near-term technology together with a recent discovery of astrodynamics would allow designing a mission to the closest *solar gravitational focus* (SGF, at 550 AU from the Sun) with acceptable flight time. In the past, other examples of fast trajectory by rocket to the SGF distance have been published in the specialized literature (e.g., [1]). Before discussing a specific example mission and for easing its understanding, we shall introduce a minimal set of concepts and equations for solar-sailing trajectory computation, recommending the interested reader the indicated references for an extended study, in particular [2]. Here, we limit ourselves to heliocentric trajectories. (Details on the geocentric trajectory equations for solar sail spacecraft can be found in [2, 3].)

C.1 SOME CONCEPTS OF SOLAR-SAIL DYNAMICS

Solar gravity and radiation pressure drive the motion of a sail-based spacecraft or *sailcraft*. Solar radiation pressure is not a mere perturbation to the motion of such a spacecraft, but can bring about continuous thrust acceleration comparable with the gravitational one. Formally, the dynamical equations can be simple. However, even in the case of a planar sail, equations become complicated when one introduces the actual features of the source of light, its photon spectrum and the interaction of photons with the sail materials. In addition, since we have to deal with a space vehicle, the sailcraft mass breakdown (namely, a mass model of its systems and subsystems) couples with its trajectory. One should realize two basic points: (a) in general, the force field actually acting on sailcraft is not conservative (because the incidence angle of the photon beam impinging on the sail can be greater than zero), (b) the sailcraft motion can be controlled by steering the sail axis (of symmetry).

[1] Retired Scientist, Telespazio SpA, Rome, Italy. Email: *giovanni_vulpetti@fastwebnet.it*

For our purposes, we need two frames of reference to describe the sailcraft motion: the *heliocentric inertial frame* (HIF) and *the extended heliocentric orbital frame* (EHOF). HIF could be identified with the mean-ecliptic and equinox at J2000. EHOF is the generalization of the usual heliocentric orbital frame by including trajectory branches separated by a finite number of points where the sailcraft's orbital angular momentum per unit mass (**H**) vanishes. The strict definition and properties of EHOF can be found in [4]. The axes EHOF are given by the columns of the matrix (**rh** × **rh**), where **r** denotes the direction of the sailcraft position vector (say, **R** in the HIF) and **h** is either the **H** direction for a direct trajectory arc or the −**H** direction for a retrograde trajectory arc or their (common) limit when $\mathbf{H} = \mathbf{0}$.

The key point for writing the general sailcraft motion equation is to introduce the time-dependent vector function named *the lightness vector* [5, 6], denoted by **L** as follows:

$$\mathbf{L} \equiv \begin{bmatrix} \lambda_r \\ \lambda_t \\ \lambda_n \end{bmatrix} \qquad \lambda \equiv |\mathbf{L}| \tag{C.1}$$

L is defined in EHOF. Its components (also called the *radial*, the *transversal*, and the *normal* lightness numbers) represent the components of the solar pressure–induced vector acceleration in units of the *local* gravitational acceleration or μ/R^2, where μ denotes the solar gravitational constant. Thus, the classical dynamics equations of sailcraft motion can be written as

$$\begin{aligned} \frac{d}{dt}\mathbf{R} &= \mathbf{V} \\ \frac{d}{dt}\mathbf{V} &= \frac{\mu}{R^2}[-(1-\lambda_r)\,\mathbf{r} + \lambda_t\mathbf{h}\times\mathbf{r} + \lambda_n\mathbf{h}] \end{aligned} \tag{C.2}$$

(Here, for simplicity, Equations (C.2) do not contain the mass rate equation; actually, if a sailcraft is controlled by small attitude rockets then $dM/dt < 0$, where M is the vehicle mass.) One then realizes that sailcraft trajectory can be analyzed in terms of the **L** vector only, even though an actual control shall operate on the sail orientation. This is an important point that contributed to the discovery of high-speed trajectory families as recently as the 1990s. In order to highlight the different roles of the **L** components, we report the main equations for orbital sailcraft energy E, angular momentum and their time rates ([4–6] contain full discussions of such equations and their significant consequences for sailcraft dynamics):

$$\begin{aligned} E &= \tfrac{1}{2}V^2 - (1-\lambda_r)\,\frac{\mu}{R}, & \frac{d}{dt}E &= \frac{H}{R^2}\frac{d}{dt}H \\ \mathbf{H}\times\frac{d}{dt}\mathbf{H} &= H\lambda_n\frac{\mu}{R}\mathbf{r}, & \frac{d}{dt}\mathbf{H} &= \frac{\mu}{R}(\lambda_t\mathbf{h} - \lambda_n\mathbf{h}\times\mathbf{r}) \\ \frac{d}{dt}H &= \lambda_t\frac{\mu}{R}, & \mathbf{H} &= H\mathbf{h} \end{aligned} \tag{C.3}$$

The main quantity in Equations (C.3) is the invariant H—namely, the projection of the angular momentum onto the z-axis of the heliocentric orbital frame. Its derivative depends on the *transversal* lightness number, drives the E change, and determines the history of $\mathbf{H}$. Note that the normal lightness number governs the bending of $\mathbf{H}$; an important consequence is analyzed in [4].

In general, $\mathbf{L}$ is a complicated function of the sailcraft mass (M) on sail area (S) ratio (or the spacecraft *sail loading*, usually denoted by σ), the thermo-optical properties of the sail materials, the sail axis control angles, the spacecraft velocity, and the characteristics of the source of light [2, 3]. As usually conceived, a practical sail consists of a multi-layer film: the *reflective* layer, the *emissive* layer, and the *substrate* the other two layers are deposited on. The reflective layer is always facing the Sun in a heliocentric trajectory, whereas the emissive layer allows keeping the sail temperature sufficiently low. (They are known also as the front side and the back side of the sail, respectively.) Although specular reflection is the dominant effect for photon–sail momentum exchange, other non-negligible effects have to be taken into account; an appropriate discussion can be found, for example, in [2] and [5, the appendix]. If the substrate (the heaviest component of a sailcraft) is removed,[2] then one has an all-metal sail capable of achieving very high speeds [4–6]. This is the configuration we shall consider for a fast mission to SGF.

Neglecting the Sun's finite size and limb-darkening effects[3] and retaining the linear terms in the sailcraft velocity, one can arrive at the following expression for the lightness vector:

$$\mathbf{L} = \lambda_0 \cos\alpha\cos\delta \left\{ \left[(2r\cos\alpha\cos\delta + \chi_f d + \kappa a)(1-2\beta_x) - 2r\sin\alpha\cos\delta\beta_y\right]\mathbf{n} \right.$$
$$\left. +(a+d)\begin{bmatrix} 1-2\beta_x \\ -\beta_y \\ 0 \end{bmatrix} \right\}$$

where

$$\mathbf{n} \equiv \begin{bmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{bmatrix}, \quad \beta \equiv \frac{V}{C}\begin{bmatrix} \cos\varphi \\ \sin\varphi \\ 0 \end{bmatrix}, \quad \varphi \equiv \widehat{\mathbf{RV}}, \quad V \equiv |\mathbf{V}|$$
$$\lambda_0 \equiv \tfrac{1}{2}\sigma_c \frac{S}{M}, \quad \sigma_c \equiv 2\frac{W_{1AU}}{g_{1AU}C}; \quad 0.001539 \text{ kg m}^{-2},$$
$$W_{1AU} = 1{,}368 \text{ W m}^{-2}, \quad g_{1AU} = 0.00593 \text{ m s}^{-2} \tag{C.4}$$

[2] The first experiments on plastic substrate removal, simulating a process in space, were performed at ENEA (Rome, Italy) in 1998 [10].

[3] References [2, 3] contain formulas for dealing with such sail irradiance reduction effects.

In Equations (C.4), α and δ denote the azimuth and elevation of the sail axis **n** (oriented backward with respect to the reflective sail side) in EHOF. The set (r, d, a) denote the specularly reflected, the diffused reflected, and the absorbed fractions of the solar incident flux by the sail materials, respectively. Their exact meaning, numeric handling, and relationship to orbit determination can be found extensively in [6, 7]. χ_f denotes the surface coefficient[4] of the front side, whereas κ is a function of the sail temperature [5, 6]. The vector β (resolved in EHOF) accounts for the aberration effect, which is a linear one in the sailcraft velocity and, therefore, is not negligible for a high-speed flight. The quantity σ_c represents the so-called *critical* density. One has the following important relationship

$$\sigma = \sigma_c \tau / \lambda \tag{C.5}$$

where τ is the thrust efficiency [5]. Note that g_{1AU} is the solar gravitational acceleration at 1 AU, whereas W_{1AU} denotes the solar constant; the value given above is compliant with another reference[5] [8].

Equations (C.4) can be called the *connection equations* since they represent the link between the direct control variables and parameters and the lightness numbers that enter the motion equations.

A sailcraft is not an easy-to-build space vehicle, at least not at the present time. Its overall system has to be carefully tested in a number of real missions in the Earth–Moon system before attempting to deliver a scientific payload in the solar system and beyond. Here, we are forced to assume that such tests have been successful in order to simulate a fast flight beyond the heliosphere. The theory behind such a mission is based on that developed in [4–6]. The related trajectory classes, named the H-reversal mode for solar sailing, are all characterized by reversal of the orbital angular momentum of the sailcraft. This is the only way to achieve the absolute maximum of the vehicle energy. In terms of speed, this means that the sailcraft is able to achieve a cruise value considerably higher than the Earth's heliocentric orbital speed.[6] A quite important issue for an all-metal sail spacecraft is that it requires no additional propulsion systems for both escaping the Earth–Moon system and flying-by the Sun. Therefore, a mission to a distant target could be carried out without *planetary* launch windows.

C.2 EXAMPLE SAILCRAFT FOR SGF MISSION

Although we do not enter the details of a sailcraft model for an SGF mission, nevertheless we mention just a few points about its main systems. The sail system would consist of a film of aluminum and chromium (130 nm and 10 nm thick,

[4] It takes on 2/3 for a Lambertian surface.

[5] This quantity has to be measured accurately for a sailing mission, but this could be performed when an experimental sailcraft flight starts.

[6] This value amounts to 2π AU/yr. For comparison, the fastest space vehicle made so far (namely, Voyager 1) has a cruise speed of 3.5 AU/yr. 1 AU/yr = 4.7405 km/s.

respectively) deployed and kept open all the flight by a technique [9] studied by the Aurora Collaboration (AC), which also has been comparing some methods of in-orbit plastic removal [10]. For this study, AC has selected an attitude control system based on chemical cold-gas microengines developed by the European Space Agency. Power system selection focused on the Pu^{238}-based general purpose heat system (successfully used in Voyager, Galileo, Ulysses, etc.) plus a thermo-photo-voltaic converter [11]. This system has high reliability and efficiency. Considering the enormous target distance from the Earth, AC's choice has fallen on a laser system based on Nd:YAG [12]. The SGF mission would need a minimum of 100 kg of scientific instruments (from technology of the 1990s) that are meant to work up to 750 AU, for which a bit rate (with coding) of 200 baud is considered. Computer simulations tell us that, using current or near-term technology, the sailcraft in the heliocentric field can have a mass as low as 300 kg (including contingency) and a circular sail area of 0.24 km^2. Thus, the vehicle sail loading takes on 1.25 g/m^2—namely, about 80% of the critical density—whereas the mass of the complete sail system amounts to 46% of the sailcraft mass.

The above goal of 1.25 g/m^2 allows applying the H-reversal mode theory to achieve a very high cruise speed, or better a *mean radial* speed very close to the cruise one. This entails that the time to reach the perihelion is considerably short (200 days in our example mission) compared with 4 years to 5 years of the conventional technique by a Jupiter flyby.

C.3 TRAJECTORY PROFILE FOR SGF MISSION

In this specific example, the sailcraft points at the so-called Galactic anti-center direction. The importance of this target is detailed elsewhere in this book. The spherical coordinates of such a direction in the above HIF are approximately: longitude = 86.83°, latitude = 5.537°. The minimum operational distance for payload is 550 AU. These three numbers represent the target of our example mission or the SGF target, for short.

Reasoning in terms of lightness numbers, the flight design we are going to discuss is the solution to the following problem:

> What is the history of the 3-dimensional **L**, relative to an H-reversal sailing mode, which minimizes the flight time to the SGF-target, subject to the constraint that the sail temperature peak does not exceed 60%[7] of the aluminum melting point?

Obviously, this temperature constraint determines a lower limit on the reachable perihelion.[8] We have used the nonlinear dynamics approach—in particular, the

[7] This value is chosen to preserve the mechanical properties of the aluminum film and, at the same time, to allow sailcraft to reach a low perihelion.

[8] Additional constraints may come from requirements of attitude time rate when the sailcraft approaches the perihelion.

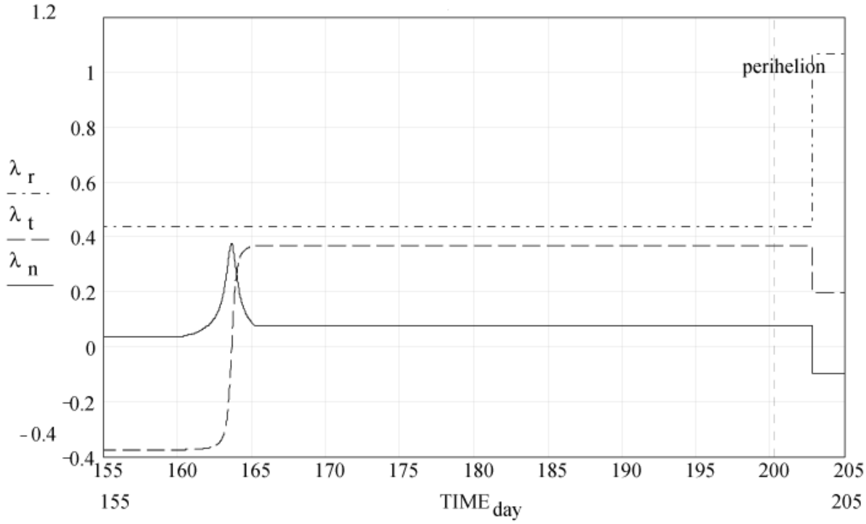


Figure C.1. Lightness vector.

Levenberg–Marquardt–Morrison method—for minimizing the (Euclidean) norm of a complicated penalty vector function such as that stemming from the H-reversal mode theory. We shall show *nominal* time behaviors of meaningful quantities zoomed on the interval where they vary appreciably. Outside such an interval, quantities are either constant or asymptotically constant. This avoids displaying compressed plots.

The optimal flight time of the current problem results in 23.46 years, the mean radial speed taking on 23.55 AU/yr whereas the cruise speed achieves 24.01 AU/yr (or 113.82 km/s)—very high values indeed. Figure C.1 shows the time profile of the *lightness* numbers (defined in Section C.1) which induce this dynamical output.

In general, a low sail-loading value entails a high value of the lightness number, as in this case, compared with the typical values (0.01–0.1) related to sail-craft with some polymer-supported sail.[9] Note that the *transversal* lightness number crosses zero while the *normal* one achieves its maximum. After the perihelion, there is an attitude maneuver that turns the *radial* lightness number into a value greater than 1.[10]

In terms of sail axis angles, the above control can be read in Figures C.2 and C.3, where both profiles in EHOF (or HOF) and HIF are shown. Note that the histories of azimuth and elevation in HOF are particularly simple.

[9] Such sails are considerably easier to make and could be appropriate for multiple interplanetary transfers.
[10] This is allowed by a subcritical value of the sail loading and high thrust efficiency, according to Equation (C.5).

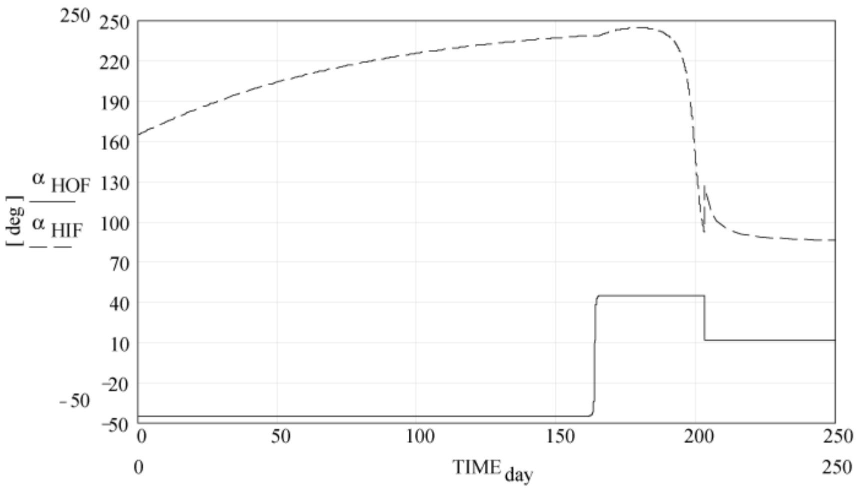


Figure C.2. Sail axis azimuth and longitude.

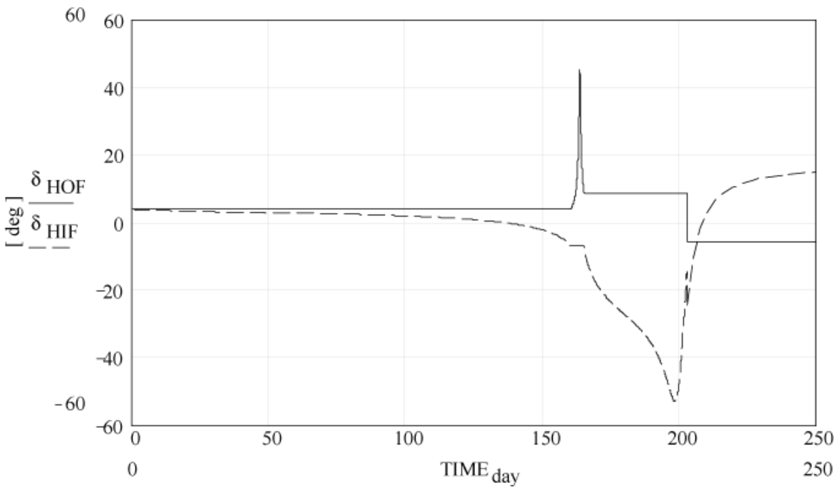


Figure C.3. Sail axis elevation and latitude.

It is meaningful that the angle of photon incidence onto the sail, shown in Figure C.4, remains practically constant down to the post-perihelion attitude maneuver. This means that the sail temperature changes only with solar distance. In contrast, when the temperature decreases sufficiently after its peak, an attitude maneuver can be performed to further increase the subsequent cruise speed.

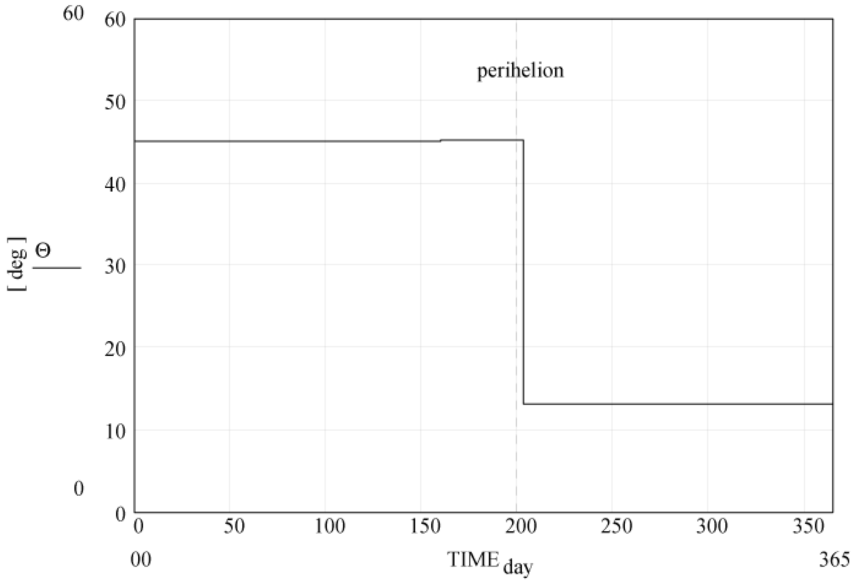


Figure C.4. Incidence angle.

Figures C.5 and C.6 show that the trajectory, induced by the above control Earth longitude at sailcraft injection[11] into the solar branch, has been optimized. Injection date is on October 23, every year. The computation of the related window is outside the scope of this appendix. However, preliminary results indicate an injection window of 2 weeks. Figures C.5 and C.6 highlight special events in the unusual trajectory shape characteristic of the H-reversal mode families.

Together with Figure C.7, which displays the vehicle speed, and Figure C.8, which displays energy and invariant H, one can note the general policy underlying a high-speed trajectory. At injection, the sailcraft begins by decelerating and passes through a point of minimum speed. Then, the along-track component of the *total* vector acceleration becomes positive and keeps on increasing. At the minimum value of H, also the sailcraft energy achieves its minimum, while the vector **H** crosses the ecliptic plane and then reverses.[12] As a consequence, the spacecraft can draw close to the Sun on acceleration. Figures C.7 and C.8 show that *both* energy and speed never

[11] Sailcraft of such a low sail loading could take about 30–40 days, depending on the allowed control history, to escape the Earth–Moon system.

[12] At the reversal point, even angular momentum achieves a non-zero minimum magnitude in this trajectory family [4], in contrast to the other two for which $\mathbf{H}_{rev} = 0$ [5–6].

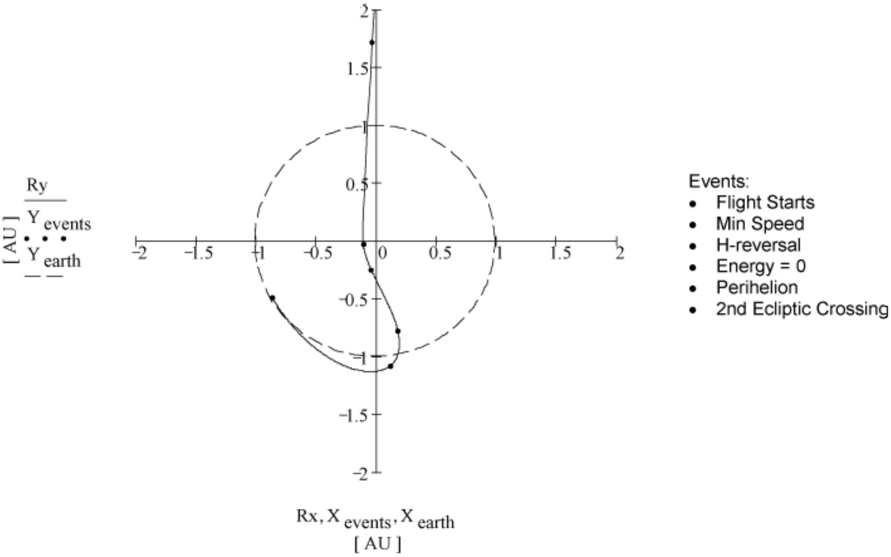


Figure C.5. In-ecliptic trajectory.

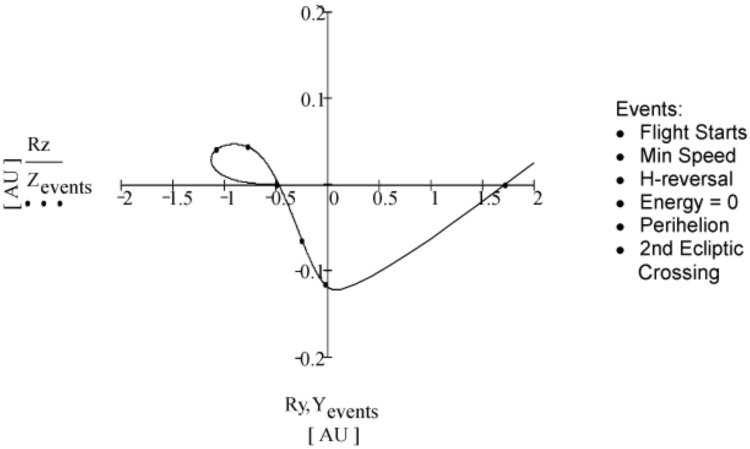


Figure C.6. Out-of-ecliptic trajectory.

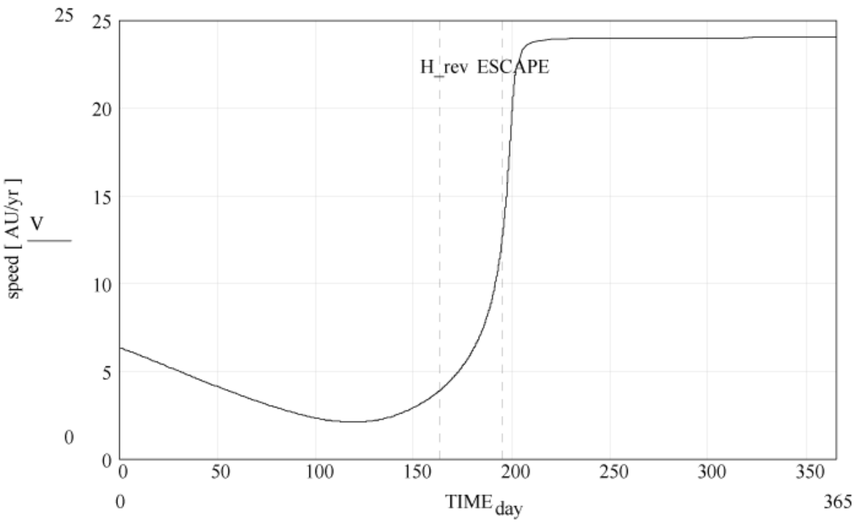


Figure C.7. Sailcraft speed.

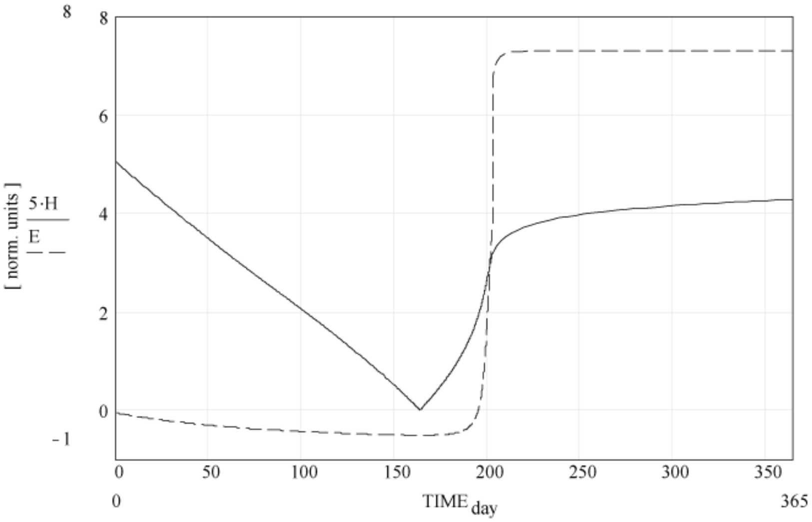


Figure C.8. Energy, H-function.

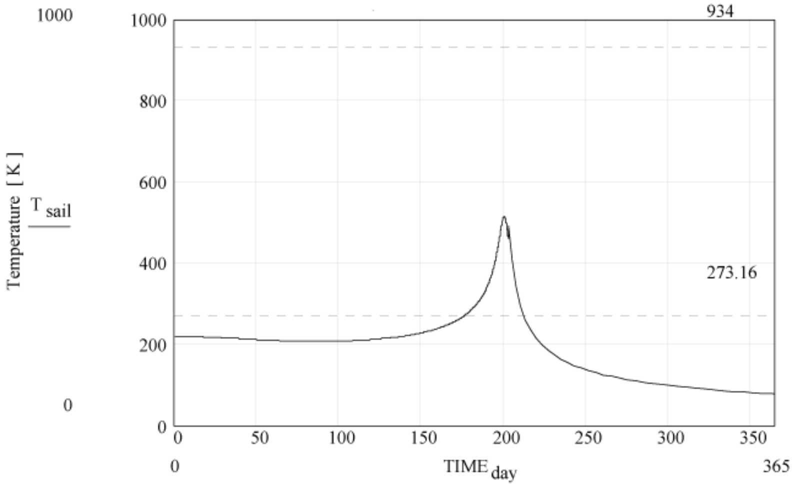


Figure C.9. Sail temperature.

cease increasing.[13] In practice, 250 days after injection one could consider speed as the cruise speed, equal to 24.01 AU/yr in this case. In Figure C.7, the vertical line on the right indicates where the sailcraft energy achieves the zero value—namely, the escape speed.

Figure C.9 shows the sail temperature behavior. Two horizontal lines indicate 0°C and the aluminum melting point, respectively. Some important information can be read. Except for 36 days in 23.5 years, the sail temperature is below 273 K. This is caused by (1) the sail absorptance ranges from 0.072 to 0.075, and (2) the sail always being tilted at sufficiently large angles near the Sun, as shown in Figure C.4. The peak temperature (514 K) is achieved at the perihelion (0.1655 AU) tagged at 200.315 days after injection. This maximum satisfies the problem constraint. Finally, in Figure C.10 we plotted the main contributions to the radiation–pressure thrust around the perihelion and the final attitude maneuver.

As expected, the dominant term is due to photon specular reflection on the sail. However, neglecting the other contributions would entail error in the trajectory computation where sensitivity is the highest. Error propagation in the sailcraft's final direction to SGF would result in the target being missed.

Sailcraft navigation and guidance are key areas to be deeply studied. As a point of fact, although solar sailing represents a true propulsion mode continuously acting

[13] No sail attitude controls other than the H-reversal mode are able to reach such high cruise speeds. In particular, if the maximum lightness number $\lambda(\alpha = \delta = 0)$ is lower than unity, then the sailcraft cruise speed is lower than the maximum value the vehicle achieves shortly after perihelion [4–6]. If λ is higher than 1—as in this example mission—then an appropriate attitude maneuver causes the *radial* lightness number $\lambda_r > 1$. As a result, speed continues to increase (even at large distances from the Sun).

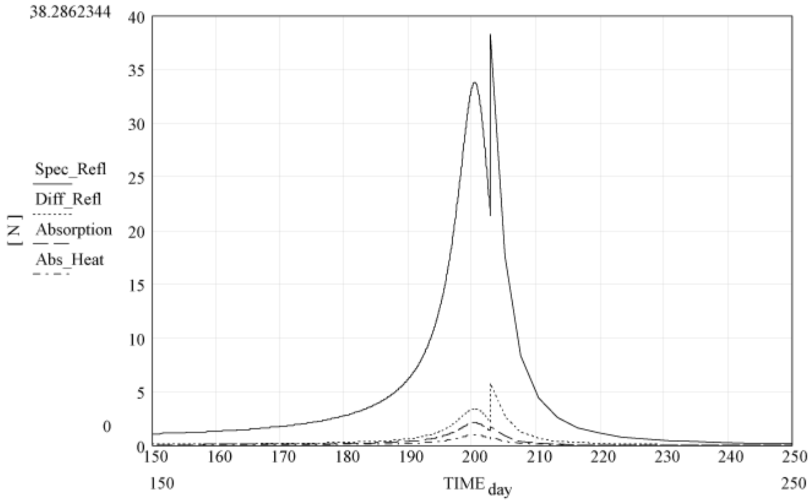


Figure C.10. Main thrust components.

for all the flight, sufficiently far from the Sun one does not expect to be able to correct previous errors (due to many present-day uncertainties) by the rapidly decreasing thrust level.

C.4 CONCLUSIONS

The evolution of spacecraft system technology *and* the advancement of solar sail dynamics induce us to deem it possible—in the near term—to start solar sailing as a significant part of a new era in space exploration. Not only does it seem possible to send frequent and low-cost scientific probes in the Earth–Moon system and beyond, but also it appears just as feasible to make a low-mass sailcraft to reach the heliopause in just a few years and the solar gravitational lens region in a couple of decades. Such a quality jump is fully compliant with the current views of modern spaceflight.

C.5 REFERENCES

[1] S. A. Carpenter, M. E. Deveny, and G. Vulpetti, "Assessing New Materials, Technologies, and Implementing Strategies for Space Fusion Propulsion as Applied to an Extra-Solar-System Mission," *11th Space Nuclear Power Systems Symposium*, held in Albuquerque, NM, January 1994.

[2] C. R. McInnes, *Solar Sailing: Technology*, *Dynamics, and Mission Applications*, Springer-Praxis Series in Space Science and Technology, May 1999.

[3] G. Vulpetti, "The Aurora Project: Flight Design of a Technology Demonstration Mission," *Proceedings of the First IAA Symposium on Realistic Near-Term Advanced Scientific Space Missions*, held in Turin, Italy, June 25–27, 1996.
[4] G. Vulpetti, "General 3D H-Reversal Trajectories for High-Speed Sailcraft," *Acta Astronautica*, **44**(1) (1999), 67–73.
[5] G. Vulpetti, "Sailcraft at High Speed by Orbital Angular Momentum Reversal," *Acta Astronautica*, **40**(10) (1997), 733–758.
[6] G. Vulpetti, "3D High-Speed Escape Heliocentric Trajectories by All-Metallic-Sail Low-Mass Sailcraft," *Acta Astronautica*, **39**(1–4) (1996), 161–170.
[7] G. Vulpetti, and S. Scaglione, "The Aurora Project: Estimation of the Optical Sail Parameters," special issue of *Acta Astronautica*, **44**(2–4) (1999), 123–132.
[8] J. L. Wright, *Space Sailing*, Gordon & Breach, Amsterdam, 1993.
[9] G. Genta, and E. Brusa, "The Aurora Project: A New Sail Layout," special issue of *Acta Astronautica*, **44**(2–4) (1999), 141–146.
[10] S. Scaglione, and G. Vulpetti, "The Aurora Project: Removal of Plastic Substrate to Obtain an All-Metal Solar Sail," special issue of *Acta Astronautica*, **44**(2–4) (1999), 147–150.
[11] A. Shock, C. Or, and V. Kumar, "Design and Integration of Small RTPV with New Millennium Spacecraft for Outer Solar System," *Acta Astronautica*, **41**(12) (1997).
[12] J. R. Lesh, C. J. Ruggier, and R. J. Cesarone, "Space Communications Technologies for Interstellar Missions," *Planetary Society Conference on Practical Robotic Interstellar Flight*, held in New York, August 29–September 1, 1994.

Appendix D

"FOCAL" radio interferometry by a tethered system

D.1 A TETHERED SYSTEM TO GET MAGNIFIED RADIO PICTURES OF THE GALACTIC CENTER FROM 550 AU

The goal of this appendix is to put forward for the first time the notion of a tethered system for the FOCAL spacecraft. In fact, we are going to show that the length of this tether system does not need to be very long: actually, just a couple of kilometers or so is sufficient, and this is a good result because a 2 km tether is certainly technologically feasible. It is important to point out that the tether could possibly be replaced by a *truss*. This would of course increase system stability. To build a 2 km long truss in space, however, is a very difficult engineering task. We thus prefer to speak about a tethered system rather than a truss system.

We start by facing the problem of Corona Plasma fluctuations with the relevant disturbances that affect radio waves passing through the Corona itself, as described in Chapter 8. Finding a solution to this problem is vital for the success of the FOCAL space mission. In this appendix we claim that the best way to solve the Corona problem is by doing *interferometry* in between *two* antennas of the FOCAL spacecraft. Thus, the FOCAL spacecraft, rather than having just one antenna (inflatable and, say, 12 m in diameter), must have *two identical* antennas in the new configuration proposed here. This doubles the sensitivity of the system, and introduces the new and fruitful idea of a tether tying each of them to the main cylindrical body of the FOCAL spacecraft, as depicted in Figure D.1.

The tethered FOCAL system we wish to propose is described as follows:

(1) The whole spacecraft moves away from the Sun along a rectilinear, purely radial trajectory.
(2) When the distance from the Sun is, say, 400 AU to 500 AU, all "engines" (solar sails? nuclear-electric? antimatter?) are turned off, so we can assume that, at least beyond 500 AU, the Sun speed of the whole system is *uniform*.

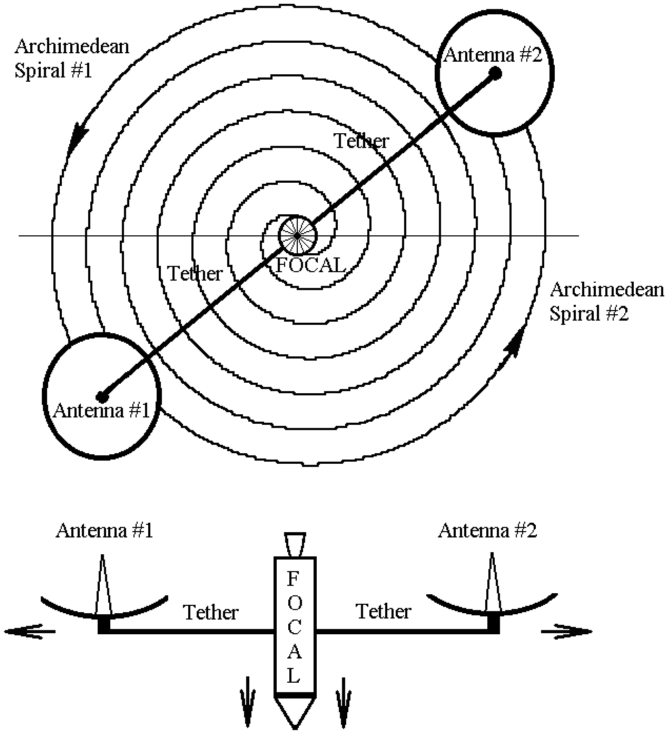


Figure D.1. A tethered system, letting two antennas describe Archimedean spirals around the FOCAL spacecraft. The whole system moves at uniform speed away from the Sun along a purely radial trajectory.

(3) Uniform speed means no acceleration. So, one can start deploying the tether. The body of the FOCAL spacecraft is supposed to be cylindrical and kept in rotation at a suitable angular speed (i.e., FOCAL is supposed to be spin-stabilized). On two opposite sides of the cylinder, the two packed, inflatable antennas are put out of the spacecraft. And each antenna is tied to the spacecraft by a tether kept taut because of the angular rotation of the whole system.
(4) The two antennas are inflated at the same time when they have reached the minimal safety distance from the spacecraft.
(5) The two antennas are oriented and pointed toward the Sun. Notice that this does not mean that the two antenna axes are parallel to each other. In practice, a huge isosceles triangle is created in space, having as its basis the distance in between the antennas and as its apex the center of the Sun (at a distance greater than 550 AU).
(6) Slowly, equal lengths of both tethers are deployed on each side of FOCAL. Because of the uniform angular rotation of the whole system, this means that the endpoints of the tether (i.e., the center of each antenna) are made to describe

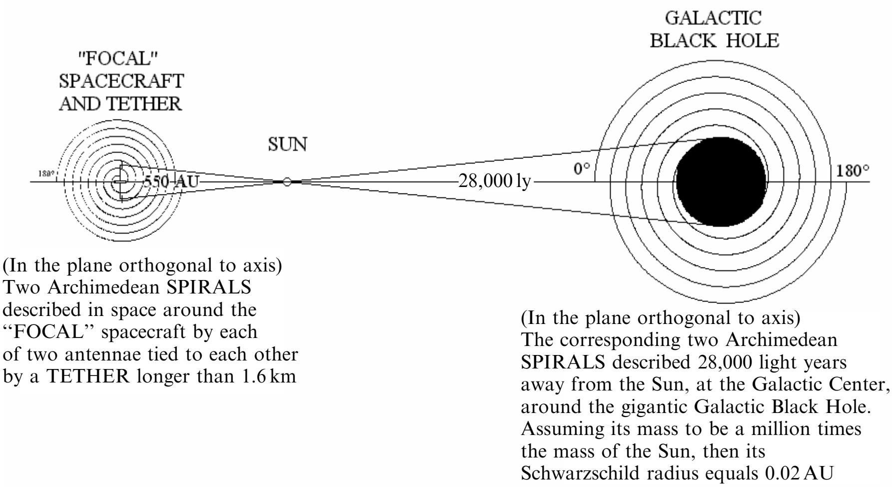


Figure D.2. Similar triangles relating the FOCAL tether length, the FOCAL spacecraft distance from the Sun, the size of the Galactic black hole and its distance from the Sun.

an Archimedean spiral (i.e., a spiral with polar equation $r = \text{const} \cdot \theta$) around the axis of the FOCAL cylindrical spacecraft. And, in turn, this fact actually means much more: since each antenna is pointing to the Sun, then ...

(7) On the other side of the Sun, at the distance of the Galactic center (i.e., about 28,000 lt-yr away) two "huge" Archimedean spirals are correspondingly being described around the Galactic center. Just at the center, a "huge" black hole is suspected to exist, as depicted in Figure D.2. We call this gigantic black hole in this appendix the "Galactic black hole", and provisionally assign to it the estimated mass of a million times that of the Sun. Consequently, the Schwarzschild radius of the Galactic black hole is a million times larger than the Sun Schwarzschild radius (i.e., it equals $\sim 2.95 \times 10^9$ km $\sim$ 0.01976 AU). The linearity between mass and Schwarzschild radius appears in Equation (1.7).

(8) We are now able to estimate the minimal tether length necessary to include the whole of the Galactic black hole within the area encompassed by the Archimedean spirals. Figure D.2 clearly shows two "similar" isosceles triangles: (i) the "small" one, between the tethered FOCAL system and the Sun, and (ii) the "large" one, between the Sun and the Galactic black hole. These two similar triangles yield immediately:

$$\frac{\text{minimal tether length}}{550\,\text{AU}} = \frac{2r_{\textit{Schwarzschild of Galactic black hole}}}{28{,}000\ \text{light years}} \tag{D.1}$$

Solving for the tether length and inserting the relevant numerical values, one finally gets

$$\text{minimum tether length} = 1.8\,\text{km} \tag{D.2}$$

Since the actual tether length must be higher than this, we reach the conclusion that a tether about 2 km long would certainly allow us to see not just the Galactic black hole, but also a host of astrophysical phenomena taking place around it, like the "swallowing" of stars, etc. by the Galactic black hole. We remind the reader that Table 1.6 gives the spatial resolution for any object at the distance of the Galactic center, and so it also gives the spatial resolution for the Galactic black hole.

In conclusion, it is believed that the 21st and following centuries are likely to see a host of FOCAL space missions, each one devoted to a different astrophysical target and thus launched along a different direction out of the solar system. But the guess is made here that all of them will use a tethered system as described in this appendix in order to avoid, by virtue of interferometry, all the problems caused by random fluctuations occurring within the solar Corona. The tethered system could be replaced by a long truss to make the system rigid, but we shall not discuss this truss here.

D.2 REFERENCES

[1] A. Einstein, "Lens-like Action of a Star by the Deviation of Light in the Gravitational Field," *Science*, **84** (1936), 506–507.

[2] S. Liebes, Jr., "Gravitational Lenses," *Physical Review*, **133** (1964), B835–B844.

[3] H. C. Ohanian, "On the Focusing of Gravitational Radiation," *International Journal of Theoretical Physics*, **9** (1974), 425–437.

[4] E. Herlt, and H. Stephani, "Diffraction of a Plane Electromagnetic Wave at a Schwarzschild Black Hole," *International Journal of Theoretical Physics*, **12** (1975), 81–93.

[5] E. Herlt, and H. Stephani, "Wave Optics of the Spherical Gravitational Lens. Part 1: Diffraction of a Plane Electromagnetic Wave by a Large Star," *International Journal of Theoretical Physics*, **15** (1976), 45–65.

[6] V. Eshleman, "Gravitational Lens of the Sun: Its Potential for Observations and Communications over Interstellar Distances," *Science*, **205** (1979), 1133–1135.

[7] F. Drake, "Stars as Gravitational Lenses," *Proceedings of the Bioastronomy International Conference*, held in Balatonfüred, Hungary, June 22–27, 1987, G. Marx (editor), pp. 391–394.

[8] N. Cohen, "The Pro's and Con's of Gravitational Lenses in CETI," *Proceedings of the Bioastronomy International Conference*, held in Balatonfüred, Hungary, June 22–27, 1987, G. Marx (editor), p. 395.

[9] F. Drake, and D. Sobel, *Is Anyone Out There?*, Delacorte Press, New York, 1992, see in particular pp. 230–234.

[10] N. Cohen, *Gravity's Lens*, Wiley Science Editions, New York, 1988.

[11] C. Maccone, "Space Missions outside the Solar System to Exploit the Gravitational Lens of the Sun," in *Proceedings of the International Conference on Space Missions and Astrodynamics*, held in Turin, Italy, June 18, 1992, C. Maccone (editor), *Journal of the British Interplanetary Society*, **47** (1994), 45–52.

[12] C. Maccone, "FOCAL, A New space Mission to 550 AU to Exploit the Gravitational Lens of the Sun," a proposal for an M3 Space Mission submitted to the European Space Agency (ESA) on May 20, 1993, on behalf of an international team of scientists and

engineers. Later (October 1993) re-considered by ESA within the "Horizon 2000 Plus" space missions plan.

[13] J. Heidmann, and C. Maccone, "AstroSail and FOCAL: Two extraSolar System missions to the Sun's gravitational focuses," *Acta Astronautica*, **35** (1994), 409–410, 1994.

[14] C. Maccone, "The SETISAIL Project," in *Progress in the Search for Extraterrestrial Life: Proceedings of the 1993 Bioastronomy Symposium*, held at the University of California at Santa Cruz, August 16–20, 1993, G. Seth Shostak (editor), *Astronomical Society of the Pacific Conference Series*, **74** (1995), 407–417.

Appendix E

Interstellar propulsion by Sunlensing

E.1 INTRODUCTION

An entirely new way of exploiting the gravitational lens of the Sun arose in the mind of this author and of Prof. Gregory L. Matloff of New York University (NYU) in the year 2000: this is "interstellar propulsion by Sunlensing"—namely, the *propulsion for interstellar flight* that the focusing power of the Sun would insure on an interstellar spacecraft made up by a suitable microwave sail like Bob Forward's "StarWisp".

To put it in easy terms, just imagine that you have a bright star (called "the source" hereafter), the Sun, and a spacecraft to be pushed away from the Sun in the direction exactly opposite to the source on the celestial sphere. Then, the Sun's gain (i.e., the Sun's geometrical focusing) would be so strong that one might imagine the spacecraft on the opposite side being pushed away at higher and higher speeds—namely, at a (roughly) uniformly accelerated motion, so that even relativistic speeds could be achieved in a relatively short time. The only important point is that a perfect alignment between source, Sun, and spacecraft (the three "Ss") be kept at all times. This could be achieved by putting aboard the spacecraft a sort of PLL (Phase Locked Loop device) capable of steering the sail so as to keep it perfectly aligned with the Sun and the source at all times.

It was in this form that the notion of interstellar propulsion by Sunlensing was put forth publicly for the first time by this author at the *STAIF-2001 Conference* (STAIF is an acronym for "Space Technology and Applications International Forum") held in Albuquerque, New Mexico, February 12–14, 2001, and published in its Proceedings (see [1]).

In this chapter we investigate this concept by adding Gregory Matloff's suggestion of using a space-based power station as the source and by utilizing a similar equation to (E.1). It is clear, however, that a much deeper study of this new form of interstellar propulsion by Sunlensing must be done as soon as possible.

E.2 HIGHLIGHTS ON RESEARCH AREAS IN INTERSTELLAR PROPULSION BY SUNLENSING

This author's belief is that the problems listed in this section are very relevant to interstellar propulsion by Sunlensing, but they have not yet reached the stage of maturity required to be cast in mathematical form.

Let us list a few of them:

(1) The alignment between the star acting as the "propeller", the center of the Sun, and the sail must be extremely tight. Thus, probes propelled by Sunlensing must head toward "nothing"—namely, toward radial directions from the Sun that are of no interest. Nevertheless, one could just let the probe fly along one such "forced" radial trajectory until it reached the closest point to the target star. Afterward, more traditional propulsion systems could be used for the final approach to the target star. Optimal "two-leg" trajectories of this kind have already been proposed by Giovanni Vulpetti in a series of *JBIS* papers for the exploration of the nearby stars by relativistic probes. However, he was not thinking of propulsion by Sunlensing, but rather of nuclear propulsion; so his trajectories must be re-computed.

(2) According to a suggestion first put forward publicly by Gregory Matloff at the *STAIF-2001 Conference*, one could imagine putting a microwave source on one side of the Sun and let the probe be pushed to interstellar distances just in the opposite direction by virtue of Sunlensing. This of course solves the exit direction problem just mentioned at (1) since we can place the microwave generator where we want around the Sun, and thus we can select at will the sail exit direction. Figure E.1, drawn for and presented in this book for the first time, depicts this situation.

We would like to remark, however, that in this case one must use a different formula for the minimal focal distance than (1.8). In fact, in this case the source no longer is placed at infinity and formula (1.10) must now be used, where the source distance is the power station distance and the spacecraft distance is the solar sail distance. Thus, the correct formula yielding the distance of the solar sail from the Sun as a function of the distance of the power station from the Sun (on the opposite side) is:

$$d_{solar\,sail} = \frac{r_{Sun}^2}{\dfrac{4GM_{Sun}}{c^2} - \dfrac{r_{Sun}^2}{d_{power\,station}}} \tag{E.1}$$

This formula yields an infinite distance for the solar sail when the denominator on the right-hand side approaches zero. In this case, it is immediately seen that the power station distance has just the value (1.9) of 550 AU (for the naked Sun) and so we infer that: *the solar power station must be initially located at distances from the Sun much greater than 550 AU. Only later will the solar power station be forced to get closer and closer to the Sun, but always at distances higher than 550 AU, so as to gradually push the solar sail farther and farther out.*

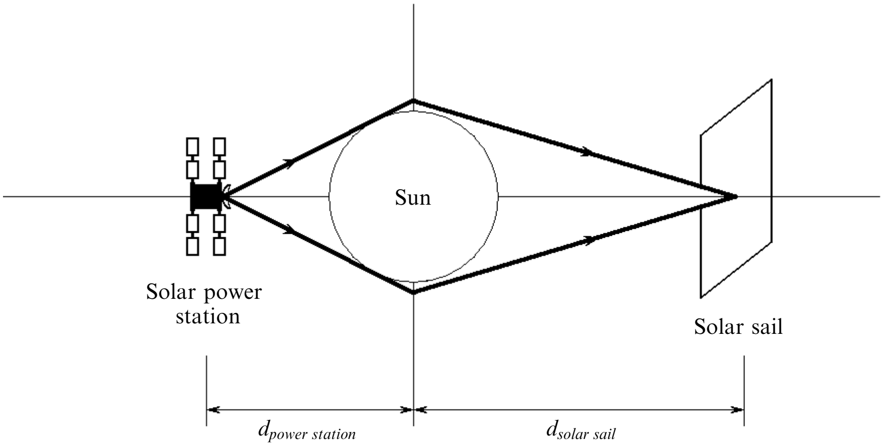


Figure E.1. A solar power station pushing a solar sail (more properly a microwave sail like Bob Forward's "StarWisp") away from the Sun in exactly the opposite direction. The two distances are related by (E.1).

Put another way, use of (E.1) instead of (1.8) changes the old picture of just "reaching 550 AU" into a new one that is essentially a tradeoff between the distance of the microwave generator and the sail distance. This author's suggestion is that it could be more convenient to initially send the microwave generator farther out than 550 AU, and then let it approach 550 AU "from outside" to let the sail be gradually pushed to high relativistic speeds.

(3) The final problem concerns the Cosmic Microwave Background (CMB). In Chapter 9 and earlier (in [2]), this author studied how to utilize NASA's Interstellar Probe (ISP) to find out "what happens" to the CMB at a distance of 763 AU where (9.22) predicts the ISP would be hit by the CMB focused by Sunlensing. We would just like to add here that our feeling is that although CMB focusing by the Sun may well exist, it would have a very tiny effect. In fact, we think that the amount of focused CMB in any direction around the Sun simply equals the amount of CMB radiation shielded by the Sun's disk in that direction, but we do not propose to write down the integrals now.

E.3 AN EXAMPLE: LIGHT FROM SIRIUS, NAKED SUN GRAVITY LENS, AND RELEVANT SOLAR SAIL SIZE

To give just one example, recall from Section 1.3 that the gain of the gravitational lens of the Sun can be proved to be given by (1.15); that is

$$G_{Sun} = 4\pi^2 \frac{r_{Schwarzschild}}{\lambda} = \frac{8\pi^2 G M_{Sun}}{c^2} \cdot \frac{1}{\lambda} = \frac{8\pi^2 G M_{Sun}}{c^3} \cdot \nu \qquad \text{(E.2)}$$

Applying this formula to spacecraft *propulsion* means

(1) considering the light from a bright star (e.g., suppose the source is Sirius) that reaches the Sun;
(2) computing where the Sun's gravitational lens focuses such a light beam from Sirius (presumably, much farther out than 550 AU, especially if one takes the Corona into account, as outlined in Chapter 8);
(3) placing a spacecraft like a solar sail at the minimal distance where Sirius's light is focused by the Sun and then let it be pushed away from the solar system as a result of this light pressure on the sail. The goal is of course to compute how fast the sail will move (presumably with a constant acceleration).

In order to perform this calculation, only formula (E.2) matters now, since it yields the numerical values of the Sun's gain according to the frequencies emitted by the source. Of course, Sirius and all stars emit, especially in visible light ($\lambda = 400$ nm through 700 nm; i.e., $\nu = 5.5 \times 10^5$ GHz through 4.3×10^5 GHz), but, for the sake of completeness, in Table E.1 we give the Sun gain for radio as well as for visible frequencies—(the definition of dB is $N\,\text{dB} = 10\log_{10}(N) = 10\ln(N)/\ln(10)$. Please note that these are the gain values for the "naked" Sun (i.e., the Sun as if it had no flames). The flames—namely, the Corona—have important (negative) consequences on the Sun's focusing, as briefly described in the coming section. These coronal effects, however, fade out with increasing spacecraft distance from the nearest focus.

Table E.1 shows that for visible light the on-axis gain ranges in between 112 dB (for red light) and 114 dB (for violet light), so, on the average, one might say that for visible light the Sun gain equals about 113 dB.

We would now like to point out that this 113 dB Sun gain for visible light is actually too optimistic. In fact, (E.2) is simply the maximum of the more complicated formula yielding the antenna patterns of the Sun lens as

$$G_{Sun}(\lambda, \rho, z) = 4\pi^2 \frac{r_g}{\lambda} \cdot J_0^2\left(\frac{2\pi\rho}{\lambda}\sqrt{\frac{2r_g}{z}}\right). \tag{E.3}$$

Table E.1. On-axis *gain* of the gravitational lens of the naked Sun for seven important frequencies.

Line	Neutral H	OH	H_2O	Ka band	CMB peak	**Visible red**	**Visible violet**
Frequency ν **(GHz)**	1.420	1.6	22	32	160	$\mathbf{4.3 \times 10^5}$	$\mathbf{5.5 \times 10^5}$
Wavelength λ **(cm)**	21	18	1.35	0.937	0.106	**700 nm**	**400 nm**
Naked Sun gain (dB)	57.4	57.9	69.3	71.46	80.40	112.22	114.65

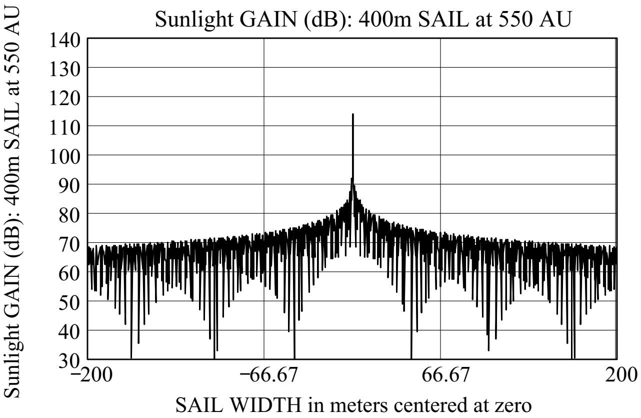


Figure E.2. (Naked) Sun gain for a 400-meter solar sail at 550 AU from the Sun.

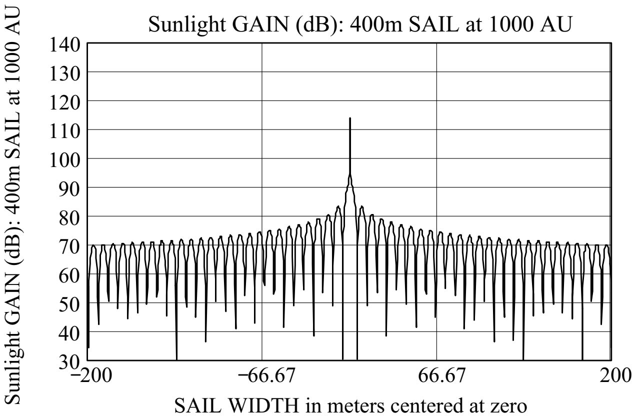


Figure E.3. (Naked) Sun gain for a 400-meter solar sail at 1,000 AU from the Sun.

This formula (Andersson and Turyshev [3]; West [4]) yields the naked Sun gain at distance ρ from the focal axis, and at spacecraft distance z from the Sun, $J_0(x)$ being the Bessel function of order zero and argument x. Since $J_0(0) = 1$, (E.3) reduces to (E.2) for $\rho \to 0$, but notice also that it does the same for $z \to \infty$. Having accepted (E.3), the surprise comes when one plots it for different ranges of the off-axis distance ρ, as shown in Figures E.2 through E.5.

In Figures E.2 and E.3 the off-axis distance ranges in between −200 m and 200 m. This is just the case for a solar sail 400-meter large, like the one currently under consideration by NASA for the Interstellar Probe (ISP) to be launched toward the direction of the incoming interstellar gas (see Mewaldt and Liewer [5] or the ISP website *http://interstellar.jpl.nasa.gov/*).

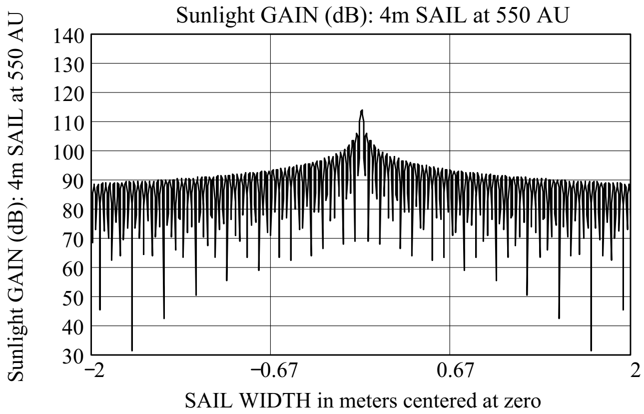


Figure E.4. (Naked) Sun gain for a 4-meter solar sail at 550 AU from the Sun. Rather than a sail, it could be just a solid antenna, as planned by NASA for the so-called NASA InterStellar Probe (ISP) described in Chapter 9. From this diagram the overall solar pressure on this rigid antenna could be computed, leading a nearly-uniformly accelerated motion away from the Sun beyond 550 AU.

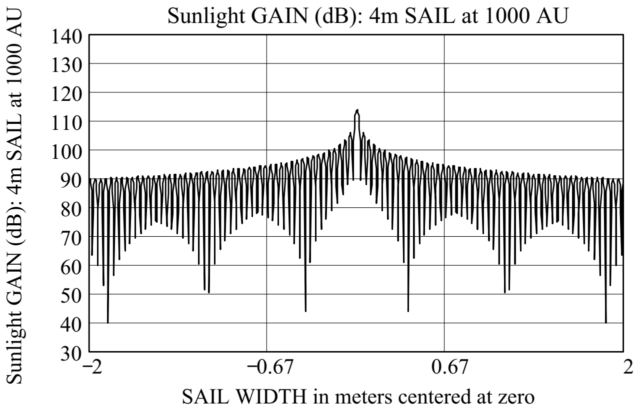


Figure E.5. Same as above, but with the spacecraft now at 1,000 AU from the Sun.

In Figure E.3 the off-axis distance ranges in between -2 m and 2 m (i.e., one considers a 4-m solar sail only).

The antenna patterns in Figures E.2 through E.5 clearly show that

(1) The on-axis peak gain for visible light is ~ 113 dB, but the peak goes off to much smaller values even if one keeps at very short distances (meters or less) from the axis. This means *that a very robust control mechanism must keep the sail center* ***on axis*** *at all times*.

(2) The actual *average* Sun gain on the sail is in fact much lower than the promised 113 dB. In practice, *the average Sun gain decreases for increasing sail sizes*. For instance, a glance at Figures E.2 and E.3 shows that for the 400-m ISP solar sail the average Sun gain is about 75 dB, but for the much smaller 4-m sail in Figures E.4 and E.5 the gain is about 95 dB.

E.4 CONCLUSIONS

Exploiting the gravitational lens of the Sun to achieve propulsion of suitable solar sails is a new topic that deserves much further study.

This author thinks that even relativistic speeds could possibly be achieved by these sails were they constantly pushed into an accelerated interstellar motion by the radiation focused on them by the gravitational lens of the Sun.

Other civilizations in the Galaxy—maybe much more advanced than ours—have probably already exploited the gravitational lens of their mother star not only for telecommunications and SETI, but for interstellar propulsion as well.

E.5 REFERENCES

[1] C. Maccone, "Interstellar Propulsion by SunLensing," a paper presented at the *STAIF-2001 Conference*, held in Albuquerque, NM, February 11–14, 2001. STAIF is an acronym for "Space Technology and Applications International Forum." See pp. 606–611 of the relevant Proceedings, published in CD-ROM form by Mohamed El-Genk (editor), University of New Mexico (all papers were collected in advance in the fall of 2000).

[2] C. Maccone, "Cosmic Microwave Background and Gravitational Lens of the Sun," Academy Transactions Note published in *Acta Astronautica*, Vol. 46, pp. 605–614, 2000.

[3] B. G. Andersson, and Slava G. Turyshev, "550 AU Mission: Science Rationale, Goals and Requirements," unpublished JPL internal report dated April 10, 1998.

[4] J. L. West, "Design Issues for a Mission to Exploit the Gravitational Lens of the Sun at 550 AU," *Acta Astronautica*, Vol. 44, pp. 99–107, 1999.

[5] R. Mewaldt, and P. Liewer, *Interstellar Probe*, NASA's Interstellar Probe Science and Technology Definition Team Booklet, published by JPL in October 1999.

Appendix F

Brownian motion and its time rescaling

F.1 INTRODUCTION

Let us now change the topic of the discussion, and consider *Brownian motion*. Since Brownian motion has been investigated by physicists and mathematicians for about a century, a number of aspects of both theoretical and practical interest have been brought to light, and a large book would thus be required to cover them. In this book we shall confine ourselves to a very particular feature that we will call "time rescaling".

By time rescaling we mean making a change in the Brownian motion time variable in such a way that the new time variable does not elapse uniformly. The new resulting Brownian motion is mathematically called *time-inhomogeneous Brownian motion*, and will be called time-rescaled Brownian motion in the present book. Thus, our time rescaling is a suitable nonlinear transformation, which may be adapted to represent a host of scientific phenomena in the fields of mathematics and physics as well as in those of economics and biology.

For instance, it is evident that the time-rescaling transformation is of interest in the theory of relativity, inasmuch as two time variables, "coordinate time" and "proper time", exist in relativity, and Brownian motion may then be a function of either of them. In addition to restricting our presentation to time-rescaled Brownian motion, we want to restrict it further to those aspects of Brownian motion that are related to the KLT.

Now, the basic result proved in this book (Maccone First KLT Theorem, proved in Appendix G) shows that the KL eigenfunctions for time-rescaled Brownian motion are Bessel functions of the first kind having nonconstant order. This appears to be an original contribution, not only with regard to time-rescaled Brownian motion, but also to the theory of Bessel functions of the first kind. In fact, several properties of Bessel functions that are well-known to hold good for constant order are naturally extended here to a general, time-dependent order.

F.2 BROWNIAN MOTION ESSENTIALS

Brownian motion—or, better, *standard Brownian motion*—is the easiest and most important non-stationary Gaussian process. We shall denote it by $B(t)$, and define it as the stochastic process with probability density function (of the first order) equal to the well-known Gaussian

$$f_{B(t)}(x) = \frac{1}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{x^2}{2t}}. \tag{F.1}$$

From this definition, it immediately follows that the mean value of Brownian motion is identically zero; that is, it equals zero for all values of time

$$E\{B(t)\} = \int_{-\infty}^{\infty} x f_{B(t)}(x)\, dx = 0 \tag{F.2}$$

and taking $t \to 0$ in (F.1), it follows that $B(t)$ fulfills the initial condition

$$B(0) = 0 \tag{F.3}$$

In other words, at $t = 0$, $B(t)$ becomes deterministic with precise value $B(0) = 0$. Another way of stating this is to say that as $t \to 0$ the Gaussian density (F.1) approaches the Dirac delta function $\delta(x)$.

Sometimes, Brownian motion is called the *Wiener process*, or the *Wiener–Lévy process*, according to the aspects of the topic that the authors desire to stress. We are not going to prove here the mathematical properties of Brownian motion. The interested reader may find, for instance, a nice presentation of them in [2, pp. 292–293]. Instead, we just highlight the main results, as well as those special features that will be used in the remainder of the present book.

The relevant variance and standard deviation are immediately seen to be given, respectively, by

$$\sigma^2_{B(t)} = t \tag{F.4}$$

$$\sigma_{B(t)} = \pm\sqrt{t}. \tag{F.5}$$

If we plot $B(t)$ against t, the resulting graph is a continuous curve, randomly moving above and below the time axis, in such a way that the standard deviation curve (F.5) is a parabola having its vertex at the origin and its axis coinciding with the time axis. Figure F.1 illustrates this.

An interesting property of Brownian motion which we use in this book is its self-similarity to the order 1/2, expressed by the formula

$$B(ct) = \sqrt{c}B(t) \quad (c > 0) \tag{F.6}$$

To understand why this is true, just replace $t \to ct$ into the Gaussian density (F.3), and then rearrange as follows:

$$f_{B(ct)}(x) = \frac{1}{\sqrt{c}} \frac{1}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{1}{2t}\left(\frac{x}{\sqrt{c}}\right)^2} = f_{\sqrt{c}B(t)}(x). \tag{F.7}$$

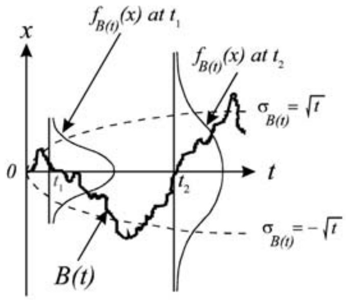


Figure F.1. Illustration of Brownian motion $B(t)$ vs. t. The dotted lines are the upper and lower branches of the standard deviation parabola, having the equation $\pm\sqrt{t}$.

To complete this overview of Brownian motion properties, we state without proof (e.g., [2, p. 293]) that the autocorrelation of Brownian motion is given by

$$E\{B(t_1)B(t_2)\} = \begin{cases} t_1 & \text{for } t_1 < t_2 \\ t_2 & \text{for } t_1 > t_2 \end{cases}. \tag{F.8}$$

This circumstance may be re-phrased by introducing a new symbol, called the minimum (= smallest of) and denoted $\wedge$, so that (F.8) takes the form

$$E\{B(t_1)B(t_2)\} = t_1 \wedge t_2. \tag{F.9}$$

Let us now express the minimum in terms of the unit step function, defined by

$$U(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t > 0 \end{cases}. \tag{F.10}$$

The unit step function is clearly discontinuous at the origin, and its value there is not defined by (F.10).[1] Moreover, the derivative of the unit step function (in the sense of the theory of distributions) is the Dirac delta function familiar to physicists and engineers alike (see [3, pp. 255–282])

$$\frac{dU(t)}{dt} = \delta(t). \tag{F.11}$$

Let us now go back to the minimum $t_1 \wedge t_2$. An easy but essential result about the minimum $t_1 \wedge t_2$ is that it can be re-written in terms of the unit step function as

[1] Actually, within the mathematically rigorous context of the *theory of distributions* it can be shown that the value of (F.10) at the origin may be any value between zero and one, for this point is a "topological" one. However, we shall not elaborate further topics because it is irrelevant to the current purpose.

follows:

$$t_1 \wedge t_2 = t_1 U(t_2 - t_1) + t_2 U(t_1 - t_2). \tag{F.12}$$

A generalization of this is

$$F(t_1 \wedge t_2) = F(t_1)\, U(t_2 - t_1) + F(t_2)\, U(t_1 - t_2) \tag{F.13}$$

where $F(\ldots)$ is *any* function of the minimum $t_1 \wedge t_2$. The full power of (F.13) in helping one to get rid of a number of apparent difficulties related to minimum will show up in Appendix I (particularly its equations).

F.3 KLT OF BROWNIAN MOTION

We are now ready to compute the KL expansion by solving the integral equation (10.18) for standard Brownian motion. This exercise yields a fundamental insight into the mathematical methods that will later be developed to find new results.

In practice, the integral equation (10.18) must be solved with the autocorrelation (F.9) of Brownian motion. However, dealing with the minimum $t_1 \wedge t_2$ as if it was a continuous function of both its arguments may not be easy, so that we replace it by the equivalent expression (F.12) in terms of unit step functions. The integral equation (10.18) then becomes

$$\begin{aligned} \lambda_n \phi_n(t_1) &= \int_0^T [t_1 U(t_2 - t_1) + t_2 U(t_1 - t_2)] \phi_n(t_2)\, dt_2 \\ &= t_1 \int_{t_1}^T \phi_n(t_2)\, dt_2 + \int_0^{t_1} t_2 \phi_n(t_2)\, dt_2. \end{aligned} \tag{F.14}$$

If $t_1 = 0$, the right-hand side of the last expression vanishes, and we get the initial condition for the eigenfunctions

$$\phi_n(0) = 0. \tag{F.15}$$

Differentiating both sides of (F.14) with respect to t_1 reduces the integral equation to a differential equation. This procedure actually amounts to realizing that the integral equation (10.18) is not a *Fredholm-type* for the particular autocorrelation (F.9) of Brownian motion, but is actually a *Volterra-type*. We thus get

$$\lambda_n \frac{d\phi_n(t_1)}{dt_1} = \int_{t_1}^T \phi_n(t_2)\, dt_2 - t_1 \phi_n(t_1) + t_1 \phi_n(t_1). \tag{F.16}$$

The last two terms cancel each other (and similar cancellations will prove to be vital in future calculations), so that, by setting $t_1 = T$, (F.16) yields the end point condition for the eigenfunctions

$$\left.\frac{d\phi_n(t)}{dt}\right|_{t=T} \equiv \phi_n'(T) = 0. \tag{F.17}$$

The presence of a remaining integral in (F.16) suggests performing a further differentiation with respect to t_1 changing (F.16) into the differential equation

$$\frac{d^2\phi_n(t)}{dt^2} + \frac{1}{\lambda_n}\phi_n(t) = 0. \tag{F.18}$$

This is the harmonic oscillator differential equation, whose general integral is a linear combination of a sine and a cosine

$$\phi_n(t) = A\sin\left(\frac{t}{\sqrt{\lambda_n}}\right) + B\cos\left(\frac{t}{\sqrt{\lambda_n}}\right) \tag{F.19}$$

where A and B are integration constants.

Determining the integration constants and λ_n from the two boundary conditions (F.15) and (F.17) and the normalization condition (10.4), it follows that

$$\phi_n(t) = \sqrt{\frac{2}{T}}\sin\left(\frac{\pi(2n-1)}{2T}t\right) \tag{F.20}$$

$$\lambda_n = \frac{4T^2}{\pi^2(2n-1)^2} \quad (n = 1, 2, \ldots). \tag{F.21}$$

These are just the KL expansion coefficients for standard Brownian motion. We may thus formally write the KL eigenfunction expansion of standard Brownian motion by substituting (F.20) into (10.2)

$$B(t) = \sqrt{\frac{2}{T}}\sum_{n=1}^{\infty} Z_n \sin\left(\frac{\pi(2n-1)}{2T}t\right). \tag{F.22}$$

Our purpose is to generalize this basic result (F.22) to forms of Brownian motion for which time does not elapse uniformly. This leads to many applications of interest in physics and in relativistic spaceflight.

F.4 WHITE NOISE AS THE DERIVATIVE OF BROWNIAN MOTION WITH RESPECT TO TIME

White noise $W(t)$ is an important notion with which physicists and engineers alike are acquainted. It is usually introduced as the one stochastic process whose power spectrum is a constant—in practice over a finite range of frequencies. That is, for which

$$P(\omega) \equiv \int_{-\infty}^{\infty} e^{i\omega t} W^2(t)\, dt$$

is a constant.

A stochastic process is said to be *stationary* if its autocorrelation is a function of the time difference $(t_2 - t_1)$ *alone*, rather than being a function of t_1 and t_2 independently. Then, the *Wiener–Khinchin theorem* guarantees that the power spectrum and the autocorrelation of the stationary processes are the Fourier transforms of each other. Thus

$$E\{W(t_1)W(t_2)\} = \int_{-\infty}^{\infty} P(\omega)\, e^{-i\omega(t_2-t_1)} d\omega$$
$$= const \int_{-\infty}^{\infty} e^{-i\omega(t_2-t_1)} d\omega = const\, \delta(t_2 - t_1)$$

and, apart from the constant, the autocorrelation of white noise is the delta function.

In this section we will show that the derivative of Brownian motion $B(t)$ is white noise in the sense that if we define $\dot{B}(t)$ such that

$$E\{\dot{B}(t_1)\dot{B}(t_2)\} \equiv \frac{\partial^2 E\{B(t_1)B(t_2)\}}{\partial t_1\, \partial t_2} \tag{F.23}$$

then

$$E\{\dot{B}(t_1)\dot{B}(t_2)\} = \delta(t_1 - t_2) \tag{F.24}$$

just like *white noise*.

To do this we first recall from (F.9) that

$$E\{B(t_1)B(t_2)\} = t_1 \wedge t_2.$$

We now form the left member as in (F.23) and use the expression from (F.12) to get

$$E\{\dot{B}(t_1)\dot{B}(t_2)\} \equiv \frac{\partial^2 E\{B(t_1)B(t_2)\}}{\partial t_1\, \partial t_2} = \frac{\partial^2 (t_1 \wedge t_2)}{\partial t_1\, \partial t_2}$$
$$= \frac{\partial^2}{\partial t_1\, \partial t_2}[t_1 U(t_2 - t_1) + t_2 U(t_1 - t_2)]$$

using the fact that $dU(x)/dx = \delta(x)$

$$= \frac{\partial}{\partial t_2}[U(t_2 - t_1) - t_1\delta(t_2 - t_1) + t_2\delta(t_1 - t_2)]$$

using the fact that $x\delta(x) = 0$

$$= \frac{\partial}{\partial t_2}[U(t_2 - t_1) + (t_2 - t_1)\,\delta(t_2 - t_1)] = \frac{\partial}{\partial t_2} U(t_2 - t_1)$$
$$= \delta(t_2 - t_1) \tag{F.25}$$

as required. Thus (F.24) is proved. The interpretation of white noise as the derivative of Brownian motion will be used in Section F.6 to prove basic results about the time rescaling of Brownian motion.

F.5 INTRODUCTION TO TIME RESCALING

The rest of this Appendix explores what happens to Brownian motion if one lets the time elapse according to an arbitrary law, rather than uniformly. We will call this *time-rescaling* the Brownian motion. Others call it time-inhomogeneous Brownian motion, among other things.

The results of the present chapter are not original. They have been known for years, but have seldom been exploited. We exploited them to the advantage of physics in Chapters 11 through 13 by establishing their relationship to the theory of relativity. Further scientific applications of these results may occur in mathematical genetics, and we dealt with some of them in Chapter 14.

The idea underpinning our work is that the Gaussian character of Brownian motion is not altered if time is arbitrarily rescaled. That is, all the Gaussian properties remain the same for the rescaled process.

F.6 THE WHITE NOISE INTEGRAL AND ITS AUTOCORRELATION

The term *white noise integral* refers to any stochastic process having the form

$$X(t) \equiv \int_0^t f(s)\, dB(s) \tag{F.26}$$

where $B(t)$ is standard Brownian motion (with zero mean and variance t) and for the time being the function $f(t)$ may be any arbitrary function continuous over the real positive axis. Since the integral in (F.26) is over a stochastic process $B(t)$, and hence difficult to define, it makes more sense to use the alternate definition

$$X(t) \equiv \int_0^t f(s)\, \dot{B}(s)\, ds \tag{F.27}$$

and $\dot{B}(t)$ is the white noise by virtue of Section F.5.

A few words about notation: $f(t)$ shall henceforth be the (almost) arbitrary scaling function of time, while $f_X(x)$ with a subscript will always refer to the probability density of the stochastic process $X(t)$—specifically the time-rescaled Gaussian probability density in this book. While in Chapter 10 $X(t)$ referred to an arbitrary stochastic process, in this section and in the rest of the Appendix it will refer to the white noise integral (F.26).

The stochastic process $X(t)$ defined by the white noise integral (F.26) is a Gaussian process. This should be obvious from the fact that the linear integral operator in (F.26) acts only on time and not on the statistical nature of the process $dB(s) = \dot{B}(s)\, ds$, which is Gaussian. However, a more sophisticated argument to realize that this is indeed the case proceeds as follows:

(1) When one has two (independent) Gaussian random variables X_1 and X_2 (with respective means m_1 and m_2 and variances σ_1^2 and σ_2^2), the random variable *sum*

$X_1 + X_2$ is again Gaussian (with mean $m_1 + m_2$ and variance $\sigma_1^2 + \sigma_2^2$)—for a proof of this well-known fact, see, for instance, [2, p. 250].

(2) Clearly, the same argument can be extended to the sum of any finite number of Gaussian random variables, which is thus one more Gaussian random variable.

(3) Finally, the integral is the limit of a sum, so that a white noise integral like (F.1) also is a Gaussian random variable (i.e., the stochastic process $X(t)$ is Gaussian).

Now from (F.26) the initial condition for the $X(t)$ process is immediately seen to be $X(0) = 0$. By taking the mean value of both sides of (F.27) and interchanging the mean value operator with the integral sign, it is also evident from (F.27) that

$$E\{X(t)\} = E\left\{\int_0^t f(s)\dot{B}(s)\,ds\right\} = \int_0^t f(s)\,E\{\dot{B}(s)\}\,ds$$
$$= \int_0^t f(s)\frac{d}{ds}[E\{B(s)\}]\,ds = 0 \tag{F.28}$$

since Brownian motion satisfies the standard condition $E\{B(t)\} = 0$ for all t.

Let us now compute the autocorrelation of the process (F.27):

$$E\{X(t_1)X(t_2)\} = E\left\{\int_0^{t_1} f(s)\,\dot{B}(s)\,ds \int_0^{t_2} f(t)\,\dot{B}(t)\right\}$$
$$= \int_0^{t_1} ds\,f(s)\int_0^{t_2} f(t)\,E\{\dot{B}(s)\dot{B}(t)\}\,dt$$

by virtue of (F.24)

$$= \int_0^{t_1} ds\,f(s)\int_0^{t_2} f(t)\,\delta(t-s)\,dt.$$

Now the inner integral differs from zero only if the singularity of the delta function lies between zero and t_2—that is, only if one has $0 \le s \le t_2$. By resorting to the unit step function, we can rewrite this in the form

$$E\{X(t_1)X(t_2)\} = \int_0^{t_1} ds\,f(s)[f(s)U(t_2 - s)] = \int_0^{t_1\wedge t_2} f^2(s)\,ds \tag{F.29}$$

where the notion of minimum (= smallest of) t_1 and t_2 was used in the last step. In conclusion, autocorrelation of the white-noise integral (F.26) or (F.27) reads

$$E\{X(t_1)X(t_2)\} = \int_0^{t_1\wedge t_2} f^2(s)\,ds. \tag{F.30}$$

The variance of $X(t)$ can now be found at once by noticing that

$$\sigma^2_{X(t)} = E\{X^2(t)\} - E^2\{X(t)\} = E\{X(t)X(t)\}. \tag{F.31}$$

Then, one merely has to set $t_1 = t_2 = t$ into the autocorrelation (F.30), and use the obvious formula $t \wedge t = t$ to get

$$\sigma^2_{X(t)} = \int_0^t f^2(s)\, ds. \tag{F.32}$$

The square root of this is the standard deviation

$$\sigma_{X(t)} = \pm\sqrt{\int_0^t f^2(s)\, ds}. \tag{F.33}$$

Equation (F.33) defines two curves on the $(t, X(t))$-plane, named standard deviation curves, that lie above and below the (zero) mean value axis, and are quite helpful for qualitative understanding of the behavior of the $X(t)$ process in time.

F.7 TIME RESCALING AND GAUSSIAN PROPERTIES OF $X(t)$

In this section we are going to prove formally that the Gaussian process $X(t)$ defined by (F.27) is just a time-rescaled version of Brownian motion. To this end, consider the following two processes:

(1) The process $X(t)$ defined by (F.27).
(2) The Brownian motion $B(t)$, where the ordinary time variable t is replaced by a rescaled time variable expressed by

$$\int_0^t f^2(s)\, ds. \tag{F.34}$$

That is, we want to consider time-rescaled Brownian motion

$$B\left(\int_0^t f^2(s)\, ds\right). \tag{F.35}$$

Evidently, both processes (F.1) and (F.9) fulfill the same initial condition

$$X(0) = 0, \quad B(0) = 0 \tag{F.36}$$

and also have the same mean value (zero):

$$E\{X(t)\} = 0, \quad E\left\{B\left(\int_0^t f^2(s)\, ds\right)\right\} = 0. \tag{F.37}$$

The crucial point lies in processes (F.27) and (F.35) also having the same autocorrelation (i.e., the relationship holds)

$$E\{X(t_1)X(t_2)\} = E\left\{B\left(\int_0^{t_1} f^2(s)\, ds\right)B\left(\int_0^{t_2} f^2(s)\, ds\right)\right\}. \tag{F.38}$$

To prove this, one first notices that the right-hand side of (F.38)—that is, the autocorrelation of the time-rescaled Brownian motion $B(\ldots)$—by virtue of (F.9) is the minimum

$$E\left\{B\left(\int_0^{t_1} f^2(s)\,ds\right)B\left(\int_0^{t^2} f^2(s)\,ds\right)\right\} = \left(\int_0^{t_1} f^2(s)\,ds\right) \wedge \left(\int_0^{t_2} f^2(s)\,ds\right). \qquad \text{(F.39)}$$

Second, that this minimum just equals the integral

$$\int_0^{t_1 \wedge t_2} f^2(s)\,ds \quad \text{or} \quad E\{X(t_1)X(t_2)\}$$

which, in turn, is the autocorrelation (F.30), completing the proof of (F.38).

To sum up, we have shown that the Gaussian processes (F.27) and (F.35) have the same initial condition, the same mean value, and the same autocorrelation. Because of its Gaussian nature, time-rescaled Brownian motion has all its higher moments fully determined by its first two moments (see, e.g., [2, section 9.3]). Moreover, the same fact must be true for the process $X(t)$ as well, and, since *all* the moments of the two processes are identical, we conclude that the two processes coincide completely

$$X(t) \equiv B\left(\int_0^t f^2(s)\,ds\right). \qquad \text{(F.40)}$$

This basic result reveals the nature of time-inhomogeneous, or time-rescaled, Brownian motion: all the action of the $f(t)$ function consists in altering the time behavior of the original Brownian motion, $B(t)$, but not in changing its Gaussian character. In fact, the probability density of $X(t)$ is

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi}\sigma_{X(t)}} e^{-\frac{x^2}{2\sigma^2_{X(t)}}} \qquad \text{(F.41)}$$

expressing a Gaussian whose variance changes in time according to (F.32). When rewritten explicitely in terms of the $f(t)$ function, (F.41) reads

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi}\sqrt{\int_0^t f^2(s)\,ds}} e^{-\frac{x^2}{2\int_0^t f^2(s)\,ds}}. \qquad \text{(F.42)}$$

Also, one may notice that, in agreement with the initial condition $X(0) = 0$, both (F.41) and (F.42) tend to the delta function, $\delta(x)$, for $t \to 0$. Finally, it follows immediately that the Gaussian (F.42) reduces to the familiar Gaussian density of Brownian motion in the case $f(t) = 1$:

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{x^2}{2t}}. \qquad \text{(F.43)}$$

As for the higher moments of the Gaussian process $X(t)$, they can be found at once from the corresponding higher moments of the Gaussian density (F.43) simply by replacing t by virtue of the variance (F.32). The proofs of the Gaussian-higher-moments formulas can be found, for instance, in [2, section 5.4]. Thus, the higher moments of $X(t)$ read, for even and odd values of n, respectively

$$E\{X^n(t)\} = \begin{cases} 1 \cdot 3 \cdot \ldots \cdot (n-1) \cdot \left[\int_0^t f^2(s)\,ds\right]^{n/2}, & \text{even } n \\ 0, & \text{odd } n \end{cases}. \tag{F.44}$$

F.8 ORTHOGONAL INCREMENTS FOR NONOVERLAPPING TIME INTERVALS

Consider four consecutive instants t_1, t_2, t_3, t_4 on the positive time axis, starting from left to right. The two time intervals, ranging from t_1 to t_2, and from t_3 to t_4, respectively, can be arranged in any one of three ways: completely disjoint intervals, contiguous intervals (i.e., with $t_2 = t_3$), and overlapping intervals.

Let us then consider the corresponding two increments taken by $X(t)$ over the two time intervals

$$[X(t_2) - X(t_1)] \tag{F.45}$$

and

$$[X(t_4) - X(t_3)]. \tag{F.46}$$

By virtue of (F.30), we may write the mean value of their product as

$$E\{[X(t_2) - X(t_1)][X(t_4) - X(t_3)]\}$$
$$= \int_0^{t_2 \wedge t_4} f^2(s)\,ds - \int_0^{t_2 \wedge t_3} f^2(s)\,ds - \int_0^{t_1 \wedge t_4} f^2(s)\,ds + \int_0^{t_1 \wedge t_3} f^2(s)\,ds.$$

Owing to the relative positions of the four points t_1, t_2, t_3, t_4, the four integrals in the last formula may or may not cancel against each other. Therefore, in the three cases stated above, one gets, respectively,

$$E\{[X(t_2) - X(t_1)][X(t_4) - X(t_3)]\} = \begin{cases} 0, & \text{for completely disjoint intervals} \\ 0, & \text{for continuous intervals} \\ \int_{t_3}^{t_4} f^2(s)\,ds, & \text{for overlapping intervals.} \end{cases}$$

We summarize these results by saying that the increments of the process $X(t)$ are orthogonal for non-overlapping time intervals only.

F.9 AN APPLICATION OF THE KLT: FINDING THE TOTAL ENERGY OF $X(t)$

One of the most important applications of the KL expansion (10.2) is the calculation of the stochastic integral

$$\varepsilon \equiv \int_0^T X^2(t)\,dt \tag{F.47}$$

representing the total energy of the time-rescaled Brownian motion $X(t)$. In fact, ε is a random variable that, by virtue of the orthonormality property of the eigenfunctions $\phi_n(t)$

$$\int_0^T \phi_m(t)\phi_n(t)\,dt = \delta_{mn} \tag{F.48}$$

is immediately seen to be given by the series

$$\varepsilon = \int_0^T \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} Z_m Z_n \phi_m(t)\phi_n(t)\,dt = \sum_{n=1}^{\infty} Z_n^2. \tag{F.49}$$

This series expansion for the total energy ε can be further investigated by means of Fourier transforms, which are called *characteristic functions* in probability theory. Let us set $i = \sqrt{-1}$ and define the characteristic function $\Phi_{X(t)}(\zeta)$ of the stochastic process $X(t)$ with two alternative notations (see [4, p. 153]):

(1) the traditional Fourier transform integral, here applied to the probability density $f_{X(t)}(x)$ of the stochastic process $X(t)$; and
(2) the probabilistic notation exploiting the notion of mean value

$$\Phi_{X(t)}(\zeta) \equiv \int_{-\infty}^{\infty} e^{i\zeta x} f_{X(t)}(x)\,dx = E\{e^{i\zeta x}\}.$$

The use of characteristic functions (i.e., Fourier transforms) in probability theory simplifies things greatly, and was initiated by the French mathematician Paul Lévy in the 1920s. For instance, consider the simple integral

$$\int_0^{\infty} e^{i\zeta x}\, e^{-\alpha x}\,dx = \frac{1}{\alpha - i\zeta} \qquad \alpha > 0.$$

From this it follows that the exponential probability density

$$f(x) = \alpha\, e^{-\alpha x} U(x)$$

has the characteristic function

$$\Phi(\zeta) = \frac{\alpha}{\alpha - i\zeta}.$$

By differentiating the above integral n times with respect to ζ, one gets

$$\int_0^{\infty} e^{i\zeta x}\, e^{-\alpha x} x^n\,dx = \frac{n!}{(\alpha - i\zeta)^{n+1}}.$$

This prompts us to take a "bold" step: replacement of the positive integer n by the *real* variable ν (it actually is better to replace n by $\nu - 1$). To prove that this is correct would take be a step too far, so we will skip the proof here; however, the reader should be aware that the extension of the factorial from discrete to continuum values is given by definition of Euler's gamma function $\Gamma(\nu)$ (first conceived by Euler around 1744)

$$n! \to (\nu - 1)! = \Gamma(\nu) \equiv \int_0^\infty e^{-t} t^{\nu-1} dt.$$

With this extension to the continuum, the integral above becomes

$$\int_0^\infty e^{i\zeta x} \, e^{-\alpha x} x^{\nu-1} \, dx = \frac{\Gamma(\nu)}{(\alpha - i\zeta)^\nu}.$$

The so-called "gamma-type" probability density is defined by

$$f_{X,\text{gamma}}(x) = \frac{\alpha^\nu}{\Gamma(\nu)} e^{-\alpha x} x^{\nu-1} U(x).$$

The characteristic function of the gamma density is immediately seen from the above integral to be

$$\Phi_{\text{gamma}}(\zeta) = \int_0^\infty e^{i\zeta x} \frac{\alpha^\nu}{\Gamma(\nu)} e^{-\alpha x} x^{\nu-1} U(x) \, dx = \frac{\alpha^\nu}{(\alpha - i\zeta)^\nu} = \frac{1}{\left(1 - i\dfrac{\zeta}{\alpha}\right)^\nu}.$$

Let us now go back to the random variables Z_n^2. From the KL expansion of $X(t)$ we already know that the Z_n are Gaussian with mean zero and variance equal to the eigenvalues λ_n. On the other hand, a famous theorem in the theory of probability—for the proof, see, for instance, [4, p. 130]—states that if X is a Gaussian random variable with mean value zero and standard deviation σ, the square of this random variable obeys a distribution of the gamma type given by

$$f_{X^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x}{2\sigma^2}} \frac{1}{\sqrt{x}} U(x).$$

Checking this against the general gamma-type density previously studied, it is seen that the density of the square random variables Z_n^2 is a particular gamma-type density (sometimes also called a χ^2-type density) having

$$\begin{cases} \sigma = \sqrt{\lambda_n} \\ \nu = \dfrac{1}{2} \\ \alpha = \dfrac{1}{2\sigma^2} \end{cases}$$

and, in conclusion, we see that the characteristic function of each Z_n^2 is given by

$$\Phi_{Z_n^2}(\zeta) = \frac{1}{\left(1 - i\dfrac{\zeta}{\alpha}\right)^\nu} = \frac{1}{(1 - 2i\lambda_n\zeta)^{1/2}} = (1 - 2i\lambda_n\zeta)^{-1/2}. \tag{F.50}$$

Another useful property of the Fourier transforms that we need is the convolution theorem, which states that the Fourier transform of a convolution is the product of Fourier transforms. (This theorem was first proved by the French mathematician Duhamel around 1833.) In probability theory, convolution of the two densities of the independent random variables X and Y is very important, because it is just the density of the random variable sum, $X + Y$; that is,

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-t) f_Y(t)\, dt.$$

This result, plus the convolution theorem, clearly imply that for the corresponding characteristic functions of the *independent* random variables X and Y one has

$$\Phi_{X+Y}(\zeta) = \Phi_X(\zeta)\, \Phi_Y(\zeta).$$

Let us apply this to the random variables Z_n that we know already to be independent because they are orthogonal and Gaussian. From the series (F.49) it just follows that the energy characteristic function is found as the infinite product

$$\Phi_\varepsilon(\zeta) = \prod_{n=1}^{\infty} \Phi_{Z_n^2}(\zeta) = \prod_{n=1}^{\infty} (1 - 2i\lambda_n \zeta)^{-1/2}. \tag{F.51}$$

In order to deal with series rather than with products, it is convenient to introduce the so-called second characteristic function $\Psi_\varepsilon(\zeta)$, simply defined as the natural logarithm of the (first) characteristic function

$$\Psi_\varepsilon(\zeta) = \ln \Phi_\varepsilon(\zeta). \tag{F.52}$$

Applying this definition to the infinite product (F.51), the latter is changed into the infinite series

$$\Psi_\varepsilon(\zeta) = -\frac{1}{2} \sum_{n=1}^{\infty} \ln(1 - 2i\lambda_n \zeta). \tag{F.53}$$

Differentiating n times, we get

$$\Psi_\varepsilon^{(n)}(\zeta) = \frac{1}{2} (2i)^n\, (n-1)! \sum_{n=1}^{\infty} \frac{(\lambda_m)^n}{(1 - 2i\lambda_m \zeta)^n} \tag{F.54}$$

We will now introduce the statistical quantity called the *cumulants*, denoted K_n, of a random variable X. Once again, this may be a notion that not all readers may be familiar with, because it is mainly used by applied statisticians. Therefore, we do not regard it a waste of time to briefly describe how cumulants get the intuitive justification that later leads to their formal mathematical definition in (F.55).

Consider the familiar Maclaurin expansion of the exponential

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

and use this to expand the complex exponential appearing in the definition of the characteristic function of a generic random variable X:

$$\Phi_X(\zeta) = \int_{-\infty}^{\infty} e^{i\zeta x} f_X(x)\,dx$$
$$= \int_{-\infty}^{\infty} \sum_{k=1}^{\infty} \frac{(i\zeta x)^k}{k!} f_X(x)\,dx = \sum_{k=1}^{\infty} \frac{(i\zeta)^k}{k!} \int_{-\infty}^{\infty} x^k f_X(x)\,dx = \sum_{k=1}^{\infty} \frac{i^k m_k}{k!} \zeta^k$$

where we used the definition of the absolute kth moment; that is,

$$m_k = \int_{-\infty}^{\infty} x^k f_X(x)\,dx.$$

The above Maclaurin expansion for the characteristic function is evidently related to the absolute kth moment by the formula

$$\Phi_X^{(k)}(0) = i^k m_k$$

meaning that, if the characteristic function of a certain probability density is known, all the absolute moments can be computed by differentiating the characteristic function and then setting the independent variable to zero

$$m_k = \frac{\Phi_X^{(k)}(0)}{i^k}.$$

Precisely the same definition as this, applied to the second characteristic function rather than to the first, yields all the cumulants K_n—whence the moments may be found (see [5, p. 27]). That is,

$$K_n = \frac{\Psi_\varepsilon^{(n)}(0)}{i^n}. \tag{F.55}$$

Besides this formal definition, however, there is much more that one might say about cumulants. For instance, it is easy to prove that the mean value of the random variable ε is given by

$$E\{\varepsilon\} = K_1 = \Psi'_\varepsilon(0)$$

and the variance is given by

$$\sigma_\varepsilon^2 = K_2 = -\Psi''_\varepsilon(0).$$

Reverting now to our problem of determining the total energy distribution ε of $X(t)$, by setting $\zeta = 0$ into (F.54) and making use of (G.48) (to be proven in Appendix G) we get

$$K_n = 2^{n-1}(n-1)! \left[\int_0^T f(s)\,ds \right]^{2n} \sum_{m=1}^{\infty} \frac{1}{(\gamma_m)^{2n}} \tag{F.56}$$

yielding all the cumulants of the energy distribution of $X(t)$.

The mean energy of ε is a special case $n = 1$ of (F.56)

$$E\{\varepsilon\} = K_1 = \left[\int_0^T f(s)\,ds \right]^2 \sum_{m=1}^{\infty} \frac{1}{(\gamma_m)^2}. \tag{F.57}$$

However, one can also write the following expression for $E\{\varepsilon\}$:

$$E\{\varepsilon\} = E\left\{\int_0^T X^2(t)\,dt\right\} = \int_0^T E\{X^2(t)\}dt. \tag{F.58}$$

On the other hand, (F.30) with $t_1 = t_2 = t$ leads at once to

$$E\{X^2(t)\} = \int_0^t f^2(s)\,ds. \tag{F.59}$$

Thus, (F.58) yields for the mean energy of $X(t)$

$$E\{\varepsilon\} = \int_0^T dt \int_0^t f^2(s)\,ds. \tag{F.60}$$

Of course, (F.57) and (F.60) are equivalent. Therefore, one finds

$$\sum_{m=1}^{\infty} \frac{1}{(\gamma_m)^2} = \frac{\displaystyle\int_0^T dt \int_0^t f^2(s)\,ds}{\left[\displaystyle\int_0^T f(s)\,ds\right]^2}. \tag{F.61}$$

We conclude that, although no explicit expression for the γ_n is known, it is possible to sum a series like (F.61) involving them. The explanation of this apparent paradox must lie deep in the theory of Bessel functions, and is unknown to mathematicians at this time.

Having found an explicit formula for the mean energy in terms of $f(t)$ such as (F.60), it is natural to seek a similar expression yielding energy variance. We shall now prove that it reads

$$\sigma_\varepsilon^2 = 4\int_0^T dt \int_0^t dv \left[\int_0^v f^2(s)\,ds\right]^2. \tag{F.62}$$

Though this result is simple, its proof is not so. The starting point is, of course, the definition of variance as

$$\sigma_\varepsilon^2 = E\{\varepsilon^2\} - E^2\{\varepsilon\}. \tag{F.63}$$

The second term is known by (F.60). The first term is calculated to be

$$E\{\varepsilon^2\} = E\left\{\int_0^T X^2(t)\,dt \int_0^T X^2(s)\,ds\right\} = \int_0^T dt \int_0^T ds\, E\{X^2(t)X^2(s)\} \tag{F.64}$$

In order to rewrite the integrand, note that $X(t)$ and $X(s)$ are Gaussian with zero mean. Therefore, one is allowed to apply the following property

$$\begin{aligned} E\{X^2(t)X^2(s)\} &= E\{X^2(t)\}E\{X^2(s)\} + 2E^2\{X(t)X(s)\} \\ &= \int_0^t f^2(x)\,dx \int_0^s f^2(y)\,dy + 2\left[\int_0^{t\wedge s} f^2(z)\,dz\right]^2 \end{aligned} \tag{F.65}$$

where (F.30) and (F.32) were used in the last step. Thus, (F.64) yields

$$\begin{aligned}
E\{\varepsilon^2\} &= \int_0^T dt \int_0^T ds \left\{ \int_0^t f^2(x)\,dx \int_0^s f^2(y)\,dy + 2\left[\int_0^{t\wedge s} f^2(z)\,dz\right]^2 \right\} \\
&= \int_0^T dt \int_0^t f^2(x)\,dx \int_0^T ds \int_0^s f^2(y)\,dy + 2\int_0^T dt \int_0^T ds \left[\int_0^{t\wedge s} f^2(z)\,dz\right]^2 \\
&= E\{\varepsilon\}E\{\varepsilon\} + 2\int_0^T dt \int_0^T ds \left[\int_0^{t\wedge s} f^2(z)\,dz\right]^2
\end{aligned} \tag{F.66}$$

where (F.60) was used twice to rewrite the first term. A remarkable simplification now occurs. On inserting (F.66) into (F.63): the two terms $E^2\{\varepsilon\} - E^2\{\varepsilon\}$ cancel, and the energy variance becomes

$$\sigma_\varepsilon^2 = 2\int_0^T dt \int_0^T ds \left[\int_0^{t\wedge s} f^2(x)\,dx\right]^2. \tag{F.67}$$

The next difficulty lies in handling the minimum, $t \wedge s$, which is greatly simplified by the use of the unit step function:

$$U(t) = \begin{cases} 1 & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}. \tag{F.68}$$

In fact, the minimum may be rewritten as a sum of two terms, resulting in

$$\begin{aligned}
\sigma_\varepsilon^2 &= 2\int_0^T dt \int_0^T ds \left\{ \left[\int_0^t f^2(z)\,dz\right]^2 U(s-t) + \left[\int_0^s f^2(z)\,dz\right]^2 U(t-s) \right\} \\
&= 2\int_0^T dt \left[\int_0^t f^2(z)\,dz\right]^2 \int_0^T ds U(s-t) + 2\int_0^T ds \left[\int_0^s f^2(z)\,dz\right]^2 \int_0^T dt U(t-s).
\end{aligned} \tag{F.69}$$

The two terms differ only in the variables t and s, so

$$\sigma_\varepsilon^2 = 4\int_0^T dt \left[\int_0^t f^2(z)\,dz\right]^2 \int_t^T ds = 4\int_0^T (T-t)\left[\int_0^t f^2(z)\,dz\right]^2 dt.$$

Finally, we use integration by parts to get:

$$\begin{aligned}
\sigma_\varepsilon^2 &= 4\left[(T-t)\int_0^t dv \left[\int_0^v f^2(z)\,dz\right]^2\right]_{t=0}^{t=T} + 4\int_0^T dt \int_0^t dv \left[\int_0^v f^2(z)\,dz\right]^2 \\
&= 4\int_0^T dt \int_0^t dv \left[\int_0^v f^2(z)\,dz\right]^2
\end{aligned}$$

which is recognized to be the same as (F.62), the desired result.

F.10 REFERENCES

[1] R. Ash, *Information Theory*, Interscience, Wiley, New York, 1965.
[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.
[3] A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, New York, 1962.
[4] A. Papoulis, *Signal Analysis*, McGraw-Hill, New York, 1977.
[5] E. Lukacs, *Characteristic Functions*, Griffin, London, 1970.

Appendix G

Maccone First KLT Theorem: KLT of all time-rescaled Brownian motions

G.1 INTRODUCTION

In Section F.4 the problem of finding the KL expansion of standard Brownian motion was solved completely. That was possible because the differential equation for the eigenfunctions was just a simple harmonic oscillator equation, whose solution is trivial.

In the present chapter we face the problem of finding the KL expansion for the time-rescaled Brownian motion defined in Appendix F. This problem is rather involved from the analytical point of view, but we are going to show that it can be solved completely by resorting to an unusual type of Bessel function of the first kind whose order is not constant in time.[1]

Some mathematicians might thus be tempted to explore these new special functions more in depth. Physicists might do so for their applications, some of which will be brought to light in the forthcoming chapters.

G.2 SELF-ADJOINT FORM OF A SECOND-ORDER DIFFERENTIAL EQUATION

When we present the KL eigenfunction expansion of the $X(t)$ process in Section G.3, the calculations will be lengthy. Therefore, it appears convenient to isolate within the present section the content and proof of a lemma that will later be used in Section G.3. This lemma deals with the general self-adjoint form of a linear differential equation of the second order, and it will help putting a certain differential equation in Section G.3 into its own self-adjoint form.

[1] The original results appearing in the present chapter were first published by the author in 1984 in [1].

Consider the most generic linear homogeneous ordinary differential equation of the second order with non-constant coefficients

$$A(t)\,y''(t) + B(t)y'(t) + C(t)y(t) = 0. \tag{G.1}$$

Since the $A(t)$ coefficient cannot vanish identically, we can divide the entire equation by it, illustrating that it actually has two independent coefficients rather than three:

$$y''(t) + \frac{B(t)}{A(t)}y'(t) + \frac{C(t)}{A(t)}y(t) = 0. \tag{G.2}$$

We are now going to prove that any differential equation of the form (G.2) may be put into its own self-adjoint form that reads

$$\frac{d}{dt}[P(t)y'(t)] + R(t)y(t) = 0 \tag{G.3}$$

which is the well-known *Sturm–Liouville* form of the equation. In other words, we must prove that the new coefficients $P(t)$ and $R(t)$ may uniquely be expressed in terms of the old coefficients $A(t)$, $B(t)$, and $C(t)$.

To this end, the derivative in (G.3) is expanded to give

$$P(t)y''(t) + P'(t)y'(t) + R(t)y(t) = 0 \tag{G.4}$$

and the equation divided by $P(t)$ (which cannot vanish identically) to give

$$y''(t) + \frac{P'(t)}{P(t)}y'(t) + \frac{R(t)}{P(t)}y(t) = 0. \tag{G.5}$$

By comparing (G.2) and (G.5), the following pair of simultaneous equations is found:

$$\begin{cases} \dfrac{B(t)}{A(t)} = \dfrac{P'(t)}{P(t)} = \dfrac{d}{dt}[\ln P(t)] & \text{(G.6)} \\ \dfrac{C(t)}{A(t)} = \dfrac{R(t)}{P(t)}. & \text{(G.7)} \end{cases}$$

Now (G.6) may be integrated at once, yielding the solution

$$\begin{cases} P(t) = e^{\int \frac{B(t)}{A(t)}\,dt} & \text{(G.8)} \\ R(t) = P(t)\dfrac{C(t)}{A(t)} = \dfrac{C(t)}{A(t)}e^{\int \frac{B(t)}{A(t)}\,dt} & \text{(G.9)} \end{cases}$$

These are the required formulas yielding $P(t)$ and $R(t)$ in terms of $A(t)$, $B(t)$, and $C(t)$, and the theorem is thus proved.

Let us now put this theorem to work by reducing the following differential equation to its own self-adjoint form:

$$\frac{\chi(t)}{f^2(t)}y_n''(t) + \left\{\frac{\chi'(t)}{f^2(t)} + \frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right]\right\}y_n'(t) + \left\{\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] + \frac{\chi(t)}{\lambda_n}\right\}y_n(t) = 0. \tag{G.10}$$

Evidently, (G.10) coincides with (G.1) if we set

$$A(t) = \frac{\chi(t)}{f^2(t)} \tag{G.11}$$

$$B(t) = \frac{\chi'(t)}{f^2(t)} + \frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right] \tag{G.12}$$

$$C(t) = \frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] + \frac{\chi(t)}{\lambda_n} \tag{G.13}$$

and if the unknown functions $y_n(t)$ with $y(t)$ can be identified.

We now want to find the new pair of coefficients $P(t)$ and $Q(t)$ defined by (G.8) and (G.9), respectively. Dividing (G.12) by (G.11), we get

$$\frac{B(t)}{A(t)} = \frac{\frac{\chi'(t)}{f^2(t)} + \frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right]}{\frac{\chi(t)}{f^2(t)}}. \tag{G.14a}$$

The key to further steps lies in noticing that both terms are logarithmic derivatives. Hence

$$\begin{aligned}\frac{B(t)}{A(t)} &= \frac{\chi'(t)}{\chi(t)} + \frac{\frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right]}{\frac{\chi(t)}{f^2(t)}} \\ &= \frac{d}{dt}[\ln\chi(t)] + \frac{d}{dt}\left[\ln\left[\frac{\chi(t)}{f^2(t)}\right]\right] \\ &= \frac{d}{dt}\left[\ln\chi(t) + \ln\left[\frac{\chi(t)}{f^2(t)}\right]\right] \\ &= \frac{d}{dt}\left[\ln\left[\chi(t)\frac{\chi(t)}{f^2(t)}\right]\right] = \frac{d}{dt}\left[\ln\left[\frac{\chi^2(t)}{f^2(t)}\right]\right].\end{aligned} \tag{G.14b}$$

By inserting (G.14b) into (G.8), we can now find the coefficient $P(t)$:

$$P(t) = e^{\int\frac{B(t)}{A(t)}dt} = e^{\int\frac{d}{dt}\left[\ln\left[\frac{\chi^2(t)}{f^2(t)}\right]\right]dt} = e^{\ln\left[\frac{\chi^2(t)}{f^2(t)}\right]} = \frac{\chi^2(t)}{f^2(t)}. \tag{G.15}$$

The coefficient $R(t)$ is also found by virtue of (G.9):

$$R(t) = \frac{C(t)}{A(t)}P(t) = \chi(t)\left\{\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] + \frac{\chi(t)}{\lambda_n}\right\}. \tag{G.16}$$

In conclusion, the self-adjoint version of the differential equation (G.10) reads

$$\frac{d}{dt}\left[\frac{\chi^2(t)}{f^2(t)}y_n'(t)\right] + \chi(t)\left\{\frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right] + \frac{\chi(t)}{\lambda_n}\right\}y_n(t) = 0. \tag{G.17}$$

G.3 EXACT SOLUTION OF THE INTEGRAL EQUATION FOR KLT EIGENFUNCTIONS OF ALL BROWNIAN MOTIONS OF WHICH THE TIME IS NOT ELAPSING UNIFORMLY

In the present section we completely solve the problem of finding the KLT of any time-rescaled Brownian motion (i.e., of any Gaussian process, in practice).

Our proof develops in three subsequent steps that may be summarized as follows:

(1) The KLT integral equation (10.18), with the time-rescaled Brownian motion autocorrelation (F.30), turns out to be a Volterra-type (and not Fredholm-type) integral equation. In other words, it can be transformed into a linear differential equation of the second order in the unknown KLT eigenfunctions jointly with two boundary conditions on the eigenfunctions that we will soon discover: an initial condition and a final condition.
(2) Back in 1984 this author discovered (and published) the fact that this differential equation can actually be reduced to the standard Bessel differential equation by virtue of two analytical transformations: (i) suitably changing the unknown function and (ii) suitably changing the independent (time) variable. The latter time transformation is actually a general (non-linear) time-rescaling paving the way to the applications of our results in the theory of relativistic telecommunications.
(3) The general form of the KLT eigenfunctions for all the time-rescaled Brownian motions (i.e., Gaussian processes) is thus the product of a time-rescaled Bessel function of the first kind $J_{\nu(t)}(t)$ (where the order ν may itself depend on time, in some cases) multiplied by another time function that represents one more rescaling in time. As for the KLT eigenvalues, we will show that they are essentially the zeros of certain linear combinations of the $J_{\nu(t)}(t)$ and their derivatives.

Let us start from the KLT integral equation (10.18) with the autocorrelation (F.4)

$$\int_0^T \left[\int_0^{t_1 \wedge t_2} f^2(s)\, ds \right] \phi_n(t_2)\, dt_2 = \lambda_n \phi_n(t_1). \tag{G.18}$$

A first consequence of (G.18) is easily found by setting $t_1 = 0$. In fact, because of the minimum $t_1 \wedge t_2 = 0 \wedge t_2 = 0$, the entire left-hand side of (G.18) vanishes, and one is left with

$$\phi_n(0) = 0. \tag{G.19}$$

This is the initial condition fulfilled by the eigenfunctions.

Let us now proceed toward the full solution to (G.18). A typical feature of the Volterra-type integral equations is that any such equation may be changed into a differential equation with two boundary conditions. For instance, in the case of (G.18), we can temporarily set

$$m(t_1) \equiv t_1 \wedge t_2 \tag{G.20}$$

to denote the minimum $t_1 \wedge t_2$ as a function of t_1. Then we can use the Leibniz theorem for the differentiantion of an integral to get

$$\frac{\partial}{\partial t_1}\int_0^{m(t_1)} f^2(s)\,ds = \int_0^{m(t_1)} \frac{\partial f^2(s)}{\partial t_1}\,ds + f^2(t_1)\frac{dm(t_1)}{dt_1} - f^2(0)\frac{d0}{dt_1}$$
$$= f^2(t_1)\frac{dm(t_1)}{dt_1} = f^2(t_1)\,U(t_2 - t_1) \tag{G.21}$$

where use of (F.36) was made in the last step. It follows that if both sides of (G.18) are differentiated with respect to t_1, (G.20) yields

$$\int_0^T \phi_n(t_2) f^2(t_1)\,U(t_2 - t_1)\,dt_2 = \lambda_n \phi_n'(t_1). \tag{G.22}$$

That is,

$$f^2(t_1)\int_{t_1}^T \phi_n(t_2)\,dt_2 = \lambda_n \phi_n'(t_1). \tag{G.23}$$

A glance at (G.23) shows that setting $t_1 = T$ results in

$$\phi_n'(T) = 0. \tag{G.24}$$

This is the final condition fulfilled by the eigenfunctions.

Let us now rewrite (G.23) in the form

$$\int_T^{t_1} \phi_n(t_2)\,dt_2 = -\lambda_n \frac{\phi_n'(t_1)}{f^2(t_1)}.$$

In order to let the integral disappear, we must differentiate both sides of this with respect to t_1. That finally yields, with some rearranging,

$$\frac{d}{dt}\left[\frac{1}{f^2(t)}\phi_n'(t)\right] + \frac{1}{\lambda_n}\phi_n(t) = 0. \tag{G.25}$$

This is the differential equation fulfilled by the eigenfunctions. It is already cast into its own self-adjoint form.

Having thus changed the integral equation (G.18) into the differential equation (G.25), we must now solve the latter subject to the boundary conditions (G.19) and (G.24). Now a great result comes. We discovered that it is possible to reduce (G.25) to the standard Bessel differential equation (see, e.g., [2, p. 4])

$$\frac{d}{dx}\left[x\frac{dy(x)}{dx}\right] + \left[x - \frac{\nu^2}{x}\right]y(x) = 0 \tag{G.26}$$

on replacing the eigenfunctions by means of a product of two unknown functions like

$$\phi_n(t) = \chi(t) y_n(t). \tag{G.27}$$

Then differentiating (G.27), and inserting it into (G.25), the latter is turned into

$$\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}y_n(t) + \frac{\chi(t)}{f^2(t)}y_n'(t)\right] + \frac{\chi(t)}{\lambda_n}y_n(t) = 0. \tag{G.28}$$

Performing differentiations with some rearranging of the terms, (G.28) becomes

$$\frac{\chi(t)}{f^2(t)} y_n''(t) + \left\{ \frac{\chi'(t)}{f^2(t)} + \frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right] \right\} y_n'(t) + \left\{ \frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] + \frac{\chi(t)}{\lambda_n} \right\} y_n(t) = 0.$$

This differential equation is just (G.10), and, as we already know from Section G.2, it may be cast into its own self-adjoint form given by (G.17)

$$\frac{d}{dt}\left[\frac{\chi^2(t)}{f^2(t)} y_n'(t)\right] + \chi(t)\left\{ \frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] + \frac{\chi(t)}{\lambda_n} \right\} y_n(t) = 0. \qquad \text{(G.29)}$$

By so doing, we have performed one of the two allowed substitutions in any differential equation like (G.28), i.e., changing the unknown function. The only remaining change allowed by the theory is a change of the independent variable (i.e., time). Changing time means time rescaling. Thus, let us make this time rescaling $x = \psi(t)$ on the Bessel differential equation (G.26), turning it into

$$\frac{d}{dt}\left[\frac{\psi(t)}{\psi'(t)} y'(t)\right] + [\psi^2(t) - \nu^2]\frac{\psi'(t)}{\psi(t)} y(t) = 0. \qquad \text{(G.30)}$$

Now, the differential equations (G.29) and (G.30) must coincide in order to yield the solution $y_n(t) =$ Bessel function of $\psi_n(t)$ By equating the coefficients of each term, we get a pair of simultaneous equations:

$$\begin{cases} \dfrac{\chi^2(t)}{f^2(t)} = \dfrac{\psi_n(t)}{\psi_n'(t)} & \text{(G.31)} \\ \chi(t)\left\{ \dfrac{d}{dt}\left[\dfrac{\chi'(t)}{f^2(t)}\right] + \dfrac{\chi(t)}{\lambda_n} \right\} = [\psi_n^2(t) - \nu^2]\dfrac{\psi_n'(t)}{\psi_n(t)}. & \text{(G.32)} \end{cases}$$

Our next task is the full solution of this pair of simultaneous equations, in terms of the only known function $f(t)$ Let us start by inspecting (G.31). Its left-hand side does not depend on the subscript (i.e., variable) n. We thus infer that the same thing must happen to the right-hand side of (G.31). Hence, we must have the following functional dependence:

$$\psi_n(t) = l_n\, \theta(t). \qquad \text{(G.33)}$$

In other words, a new constant l_n has been introduced as well as a new time-dependent function $\theta(t)$ to perform the separation of the variables t and n. Let us now rewrite the simultaneous equations (G.31) and (G.32) by aid of (G.33) to get the new pair of simultaneous equations

$$\begin{cases} \dfrac{\chi^2(t)}{f^2(t)} = \dfrac{\theta(t)}{\theta'(t)} & \text{(G.34)} \\ \chi(t)\left\{ \dfrac{d}{dt}\left[\dfrac{\chi'(t)}{f^2(t)}\right] + \dfrac{\chi(t)}{\lambda_n} \right\} = l_n^2\theta(t)\,\theta'(t) - \nu^2\dfrac{\theta'(t)}{\theta(t)}. & \text{(G.35)} \end{cases}$$

Once again, only two terms depend on n in (G.35), causing (G.35) to split into three more simultaneous equations. Adding these to the previous equation (G.34), we now get a set of four simpler simultaneous equations

$$\begin{cases} \dfrac{\chi^2(t)}{f^2(t)} = \dfrac{\theta(t)}{\theta'(t)} & \text{(G.36)} \\ \chi(t)\dfrac{d}{dt}\left[\dfrac{\chi'(t)}{f^2(t)}\right] = -\nu^2\dfrac{\theta'(t)}{\theta(t)} & \text{(G.37)} \\ \chi^2(t) = \theta(t)\theta'(t) & \text{(G.38)} \\ \dfrac{1}{\lambda_n} = l_n^2. & \text{(G.39)} \end{cases}$$

To solve this set, we eliminate $\chi^2(t)$ between (G.36) and (G.38), finding an easy differential equation in the unknown function $\theta(t)$

$$\theta(t)\,\theta'(t) = f^2(t)\frac{\theta(t)}{\theta'(t)}.$$

This differential equation may be solved at once by separation of variables, yielding the solution

$$\theta(t) = \int_0^t f(s)\,ds \tag{G.40}$$

where the plus sign must be taken in front of all square roots because the rescaled time is always positive. Equations (G.33), (G.39), and (G.40) then yield the $\psi_n(t)$ function

$$\psi_n(t) = \frac{1}{\sqrt{\lambda_n}}\int_0^t f(s)\,ds \tag{G.41}$$

while Equations (G.38) and (G.40) yield the expression

$$\chi(t) = \sqrt{f(t)\int_0^t f(s)\,ds} \tag{G.42}$$

for the $\chi(t)$ function. Finally, Equations (G.37) and (G.36) yield, for the order ν of the Bessel functions, the new time function

$$\nu(t) = \sqrt{-\frac{\chi^3(t)}{f^2(t)}\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right]}. \tag{G.43}$$

Only after several attempts was this author able to further transform this expression into an easier one. That will be presented in Section G.4.

So far, we have been discussing Bessel functions without actually determining the precise kind that applies to the particular problem we are facing. We now do so by noting that in the standard Brownian motion case, described by the functions

$$\begin{cases} f(t) \equiv 1 \\ \chi(t) = \sqrt{t} \\ \nu(t) = +\dfrac{1}{2} \end{cases}$$

one gets just the classical sine eigenfunctions (F.20) if, and only if, Bessel functions of the first kind are chosen. Therefore, only Bessel functions of the first kind fit our problem, and only they will be retained among all the possible kinds of Bessel functions, making the solution to the differential equation (G.26) read

$$y_n(t) = J_{\nu(t)}(\psi_n(t)) = J_{\nu(t)}\left(\frac{1}{\sqrt{\lambda_n}}\int_0^t f(s)\,ds\right).$$

Next, we turn to the orthogonality property for the eigenfunctions $\phi_n(t)$

$$\begin{aligned} 0 &= \int_0^T \phi_m(t)\phi_n(t)\,dt = \int_0^T N_m\chi(t)y_m(t)\cdot N_n\chi(t)y_n(t)\,dt \\ &= N_m N_n \int_0^T dt\, f(t) \int_0^T ds\, f(s) J_{\nu(t)}\left(\frac{1}{\sqrt{\lambda_m}}\int_0^t f(s)\,ds\right) J_{\nu(t)}\left(\frac{1}{\sqrt{\lambda_n}}\int_0^t f(s)\,ds\right) \end{aligned} \tag{G.44}$$

holding for $m \neq n$. Note that the new, unknown normalization constants N_m had to be introduced here. We now wish to prove that (G.44) is just a disguised form of the orthogonality condition (holding for $m \neq n$)

$$\int_0^1 x\, J_\nu(g_m x)\, J_\nu(g_n x)\, dx = 0 \quad \text{for } m \neq n$$

which is known to be fulfilled by Bessel functions of the first kind $J_\nu(x)$. This is called the *Dini orthogonality condition* because it can be used to expand an arbitrary function over the set of Bessel functions $J_\nu(x)$, and this series expansion is called the "Dini series" because it was proved for the first time by the Italian mathematician Ulisse Dini of Pisa in 1877. For a thorough description of these topics in the theory of Bessel functions, the reader may wish to consult [2, p. 70, entry (48)].

We now substitute a new variable x for t in (G.44) by virtue of the equation

$$x\int_0^T f(s)\,ds = \int_0^t f(s)\,ds. \tag{G.45}$$

It is immediately seen that for $t \to 0$ then $x \to 0$, and for $t \to T$ then $x \to 1$. Differentianting (G.45), we get

$$dx\int_0^T f(s)\,ds = f(t)\,dt.$$

Hence

$$dt = \frac{\int_0^T f(s)\,ds}{f(t)}\,dx.$$

When this is replaced into (G.44), a "miracle" occurs (i.e., the time-rescaling function $f(t)$ is canceled out from the integrand). This is a hint that we are on the right track to reach the Dini orthogonality condition. In fact, by virtue of the new definition

$$\gamma_n = \frac{1}{\sqrt{\lambda_n}}\int_0^T f(s)\,ds \tag{G.46}$$

the orthogonality expressed by (G.44) becomes

$$\int_0^1 x\, J_{\nu((x))}(\gamma_m x)\, J_{\nu((x))}(\gamma_n x)\, dx = 0.$$

This is, indeed, the desired Dini orthogonality condition rewritten in the notation required by our needs.

Also, from the integral equation (G.18) it appears that the eigenfunctions $\phi_n(t)$ may be multiplied by the arbitrary constant called the normalization constant and denoted by N_n. The numerical values of these normalization constants may be established in a fully arbitrary fashion. However, it is customary to determine them so that the eigenfunctions have "unit length" in the Hilbert space; that is,

$$1 = \int_0^T \phi_n^2(t)\, dt = N_n^2 \left[\int_0^T f(s)\, ds\right]^2 \int_0^1 x\, [J_{\nu((x))}(\gamma_n x)]^2\, dx \tag{G.47}$$

where the new Bessel function of order $\nu((x))$ is given by (G.43) with the substitution (G.45). The normalization condition (G.47) is the case $m = n$ of the orthogonality condition (G.44), and the right-hand side of (G.47) is a definite integral that must be computed, either analytically or numerically, in order to determine the normalization constants N_n.

To complete the solution to the integral equation (G.18), we still need to find the eigenvalues λ_n. From (G.46) it follows that:

$$\lambda_n = \left[\int_0^T f(s)\, ds\right]^2 \frac{1}{(\gamma_n)^2}. \tag{G.48}$$

This formula establishes a one-to-one relationship between the eigenvalues λ_n and the unknown constants γ_n. Finding the λ_n thus means finding the γ_n, but how can that be done? A clue comes from consideration of the standard Brownian motion case developed in Section F.3. There Equation (F.22) virtually determines the λ_n by virtue of the final condition (F.17), which is the same as (G.24). Since the initial condition (G.19) is identically fulfilled by any Bessel function of the first kind, we are forced to resort to the final condition (G.24). This final condition plus the expression (G.27) for the eigenfunctions $\phi_n(t)$ yield, after one differentiation

$$\chi'(T) J_{\nu(T)}(\gamma_n) + \chi(T) \left[\frac{f(T)\gamma_n}{\displaystyle\int_0^T f(s)\, ds} J'_{\nu(T)}(\gamma_n) + \frac{\partial J_{\nu(T)}(\gamma_n)}{\partial \nu} \nu'(T)\right] = 0. \tag{G.49}$$

This is a linear combination of the Bessel functions and their partial derivatives, whose zeros are the required γ_n. Finding an analytical expression for the zeros of (G.49) is, in general, impossible, and one must thus resort to a numerical solution of (G.49). However, there may exist some particular cases of the $f(t)$ function for which the zeros of (G.49) can be found, albeit in an approximated form. In Appendices H, I, and J we will give examples of how that can be done.

G.4 A SIMPLER FORMULA FOR BESSEL FUNCTION ORDER

As mentioned before, the expression (G.43) for the order of Bessel functions

$$\nu(t) = \sqrt{-\frac{\chi^3(t)}{f^2(t)}\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right]}$$

is cumbersome because of the expression of (G.42)

$$\chi(t) = \sqrt{f(t)\int_0^t f(s)\,ds}$$

in terms of the only known function $f(t)$.

It would thus be desirable to have a single formula yielding the order $\nu(t)$ directly in terms of the $f(t)$ function, without invoking the $\chi(t)$ function. The present section is devoted to proving that such a formula reads

$$\nu(t) = \sqrt{\frac{1}{4} + \left[\frac{\int_0^t f(s)\,ds}{f(t)}\right]^2 \left\{\frac{3}{4}\left[\frac{d\ln f(t)}{dt}\right]^2 - \frac{1}{2}\frac{d^2\ln f(t)}{dt^2}\right\}}. \tag{G.50}$$

Since the proof is rather lengthy, we set, for convenience

$$f \equiv f(t) \tag{G.51}$$

$$f' \equiv \frac{df(t)}{dt} \tag{G.52}$$

$$f'' \equiv \frac{d^2f(t)}{dt^2} \tag{G.53}$$

$$\int \equiv \int_0^t f(s)\,ds. \tag{G.54}$$

Then, from (G.42) it follows that:

$$\ln\chi(t) = \frac{1}{2}\ln f + \frac{1}{2}\ln\int \tag{G.55}$$

$$\frac{d\ln\chi(t)}{dt} = \frac{1}{2}\frac{f'}{f} + \frac{1}{2}\frac{f}{\int} = \frac{f'\int + f^2}{2f\int} \tag{G.56}$$

$$\frac{d^2\ln\chi(t)}{dt^2} = \frac{f''f\int^2 - f'^2\int^2 + f'f^2\int - f^4}{2f^2\int^2} \tag{G.57}$$

The order $\nu(t)$ of (G.43), by virtue of (G.56) and (G.57), now reads

$$\begin{aligned}
\nu^2(t) &= -\frac{\chi^3(t)}{f^2(t)}\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right] \\
&= -\frac{\chi^3(t)}{f^2(t)}\frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\frac{d\ln\chi(t)}{dt}\right] \\
&= -\frac{\chi^3(t)}{f^2(t)}\left\{\frac{d}{dt}\left[\frac{\chi(t)}{f^2(t)}\right]\frac{d\ln\chi(t)}{dt}+\frac{\chi(t)}{f^2(t)}\frac{d^2\ln\chi(t)}{dt^2}\right]\right\} \\
&= -\frac{1}{2}\frac{d}{dt}\left[\left(\frac{\chi(t)}{f^2(t)}\right)^2\right]\frac{f'\int+f^2}{2}-\frac{f''f\int^2-f'^2\int^2+f'f^2\int-f^4}{2f^4} \\
&= -\frac{1}{2}\frac{d}{dt}\left[\frac{\int}{f^3}\right]\frac{f'\int+f^2}{2}-\frac{f''f\int^2-f'^2\int^2+f'f^2\int-f^4}{2f^4} \\
&= -\frac{1}{4f^4}\left[-f^4-5f'^2\int^2+2f''f\int^2\right] \\
&= \frac{1}{4}+\frac{3}{4}\left[\frac{f'\int}{f^2}\right]^2+\frac{1}{2}\left[\frac{f'}{f}\right]^2\left[\frac{\int}{f}\right]^2-\frac{1}{2}\frac{f''}{f}\left[\frac{\int}{f}\right]^2 \\
&= \frac{1}{4}+\frac{3}{4}\left[\frac{f'\int}{f^2}\right]^2-\frac{1}{2}\left[\frac{\int}{f}\right]^2\left[\frac{f''f-f'^2}{f^2}\right] \\
&= \frac{1}{4}+\frac{3}{4}\left[\frac{\int}{f}\right]^2\left[\frac{f'}{f}\right]^2-\frac{1}{2}\left[\frac{\int}{f}\right]^2\frac{d^2\ln f(t)}{dt^2} \\
&= \frac{1}{4}+\left[\frac{\int}{f}\right]^2\left\{\frac{3}{4}\left[\frac{d\ln f(t)}{dt}\right]^2-\frac{1}{2}\frac{d^2\ln f(t)}{dt^2}\right\}
\end{aligned} \tag{G.58}$$

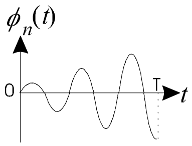
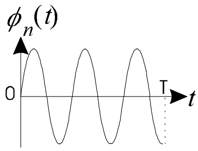
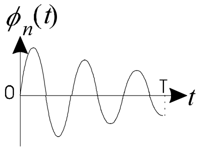
and (G.50) is proved.

G.5 STABILITY CRITERION FOR EIGENFUNCTIONS

An amazing feature of eigenfunctions $\phi_n(t)$ is that their behavior in time may be predicted even without knowing their actual analytical expression, as we prove in the present section. The starting point is the *Sonine–Pólya theorem* (see [2, p. 205]), which states that, if in the differential equation

$$\frac{d}{dt}[K(x)y'(x)]+\phi(x)y(x)=0 \tag{G.59}$$

Table G.1. Stability criterion for eigenfunctions $\phi_n(t)$.

Sign of the logarithmic derivative of $f(t)$	Shape of the KL eigenfunctions $\phi_n(t)$	Description when T is finite	Description when T is infinite
Positive		Divergent	Asymptotic unstable
Zero		Sine/cosine type	Simply stable
Negative		Convergent	Asymptotic stable

$K(x)$ and $\phi(x)$ are positive and continuously differentiable, and if $K(x)\phi(x)$ is monotonic, then the successive (relative) maxima of $|y(x)|$ form an increasing or decreasing sequence according as $K(x)\phi(x)$ is decreasing or increasing.

We are going to apply this theorem to the differential equation (G.25), which reads

$$\frac{d}{dt}\left[\frac{1}{f^2(t)}\phi_n'(t)\right] + \frac{1}{\lambda_n}\phi_n(t) = 0.$$

By checking this against (G.59), one evidently finds the correspondences

$$\begin{cases} x = t \\ y(x) = \phi_n(t) \\ K(x) = \dfrac{1}{f^2(t)} > 0 \\ \phi(x) = \dfrac{1}{\lambda_n} = \left[\dfrac{\gamma_n}{\displaystyle\int_0^T f(s)\,ds}\right]^2 = \text{constant} > 0 \end{cases} \tag{G.60}$$

where (G.48) was used in the last formula of (G.60). We thus see that the requirements of the Sonine–Pólya theorem are fulfilled only if $f^2(t)$ is assumed to be monotonic, which will indeed be the case in all physical applications we are going to consider in the book.

Now, it is important to know whether $K(x)\phi(x)$ is decreasing or increasing—namely, whether its derivative is negative or positive. With easy steps from (G.60), it follows that

$$\frac{d}{dt}\left[\frac{1}{f^2(t)}\frac{1}{\lambda_n}\right] = \frac{1}{\lambda_n}\frac{-2f(t)f'(t)}{f^4(t)}$$
$$= -\frac{2}{\lambda_n f^2(t)}\frac{f'(t)}{f(t)} = \begin{pmatrix}\text{negative}\\ \text{quantity}\end{pmatrix}\frac{d\ln f(t)}{dt}. \tag{G.61}$$

Combining this with the Sonine–Pólya theorem statement, we get the scheme shown in Table G.1. In Chapter 11 we extended this stability criterion to the theory of relativity: see Table 11.1.

G.6 REFERENCES

[1] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 212–219.

[2] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1952.

Appendix H

KLT of the $B(t^{2H})$ time-rescaled Brownian motion

H.1 INTRODUCTION

The topics considered in the present chapter are twofold: on the one hand, they can be regarded as a particular application of the results obtained in Appendices F–G to a case that allows analytic calculations to be easily carried through to completion; on the other hand, new light is shed on the theory of certain H-self-similar stochastic processes, in the wake of the celebrated results obtained by Benoit B. Mandelbrot in his theory of fractals.

H.2 THE TIME-RESCALED BROWNIAN MOTION $B(t^{2H})$

Consider the process $B_{PH}(t)$—where subscript PH refers to the "power H"—defined by the white noise integral

$$B_{PH}(t) \equiv \int_0^t \frac{s^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)}\, dB(s). \tag{H.1}$$

By setting

$$f(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \tag{H.2}$$

the definite integral of the square of this function from 0 to T is given by

$$\int_0^t f^2(s)\, ds = \int_0^t \frac{s^{2H-1}}{\Gamma^2\left(H+\frac{1}{2}\right)}\, ds = \frac{t^{2H}}{2H\,\Gamma^2\left(H+\frac{1}{2}\right)}.$$

Equation (F.40) shows that (H.1) is the same as time-rescaled Brownian motion

$$B_{PH}(t) \equiv B\left(\frac{t^{2H}}{2H\,\Gamma^2\left(H+\frac{1}{2}\right)}\right). \tag{H.3}$$

This formula, by virtue of the self-similarity to the order 1/2 expressed by (F.6), namely

$$B(ct) = \sqrt{c}B(t), \quad c \geq 0 \tag{H.4}$$

becomes

$$B_{PH}(t) \equiv \frac{1}{\sqrt{2H}\,\Gamma\left(H+\frac{1}{2}\right)} B(t^{2H}), \quad H \geq 0. \tag{H.5}$$

We thus see that $B_{PH}(t)$ is essentially a new Brownian motion whose time variable does not elapse uniformly. Rather, it is accelerated t^{2H}-like for $H > 1/2$, and decelerated t^{2H}-like for $H < 1/2$.

The application of (H.4) to (H.5) immediately yields the important self-similarity property

$$B_{PH}(ct) = c^H B_{PH}(t), \quad c \geq 0 \tag{H.6}$$

which is called the *self-similarity to the order* H, or H*-self-similarity of* $B_{PH}(ct)$—again, the subscript PH ("power H") in our notation reminds us of this.

Let us now consider two more processes:

$$B_{LH}(t) \equiv \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} dB(s) \tag{H.7}$$

and

$$B_{WH}(t) \equiv \int_{-\infty}^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} dB(s), \tag{H.8}$$

both of which are H-self-similar (as can easily be proved):

$$B_{LH}(ct) = c^H B_{LH}(t), \tag{H.9}$$

$$B_{WH}(ct) = c^H B_{WH}(t). \tag{H.10}$$

The process $B_{LH}(t)$ is the Riemann–Liouville fractional integral to the order $H - 1/2$ of Brownian motion: it is an integral for $H > 1/2$ and a derivative for $H < 1/2$ (see 1, p. 115], [2, Vol. 2, p. 181], and [3, the whole book][1]). The definition (H.7) was given in 1953 by the French mathematician Paul Lévy [4, p. 357], who confined himself to finding the process variance without further investigations. The process

[1] However, only fractional integrals of functions—and not of stochastic processes—are considered in these works.

$B_{WH}(t)$ is the Weyl fractional integral to the order $H - 1/2$ of Brownian motion, and was first considered in 1940 by Kolmogorov [5]. In 1965 Mandelbrot used its self-similarity to account for a hydrological law discovered in 1949 by Hurst [6]. Ever since, Mandelbrot and co-workers were mainly responsible for developing the computer applications of the process $B_{WH}(t)$ (notably in [7]). An excellent list of references to this and related topics, updated to 1982, can be found in Mandelbrot's book about fractals [8]. It should also be noted that a detailed analytical study of both processes (H.7) and (H.8) offers considerable difficulties. It is possible to show that (H.1) may be regarded as the first-order approximation to both (H.7) and (H.8). As a result of this we believe our results are interesting, particularly with regard to the energy distribution of the process $B_{PH}(t)$, the study of which can be carried on with a considerable amount of details (Section H.4) by exploiting the KLT of $B_{PH}(t)$ that we derive in the next section.

H.3 KL EXPANSION OF $B_{PH}(t)$

We are now going to derive the KL expansion for $B_{PH}(t)$ as in (H.1). Evidently, we are dealing with the special case of $f(t)$ given by

$$f(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \tag{H.11}$$

Standard Brownian motion corresponds to the special case $H = 1/2$ of all results to follow, because at the denominator of (H.11) one then uses the formula $\Gamma(1) = 1$.

Let us first consider autocorrelation and variance. From (H.11) and (F.30) integration easily yields the required autocorrelation of $B_{PH}(t)$

$$E\{B_{PH}(t_1)B_{PH}(t_2)\} = \int_0^{t_1 \wedge t_2} f^2(s)\, ds = \frac{(t_1 \wedge t_2)^{2H}}{2H\,\Gamma^2\left(H+\frac{1}{2}\right)}. \tag{H.12}$$

Substituting $t_1 = t_2 = t$ into (H.12) immediately gives the variance

$$\sigma^2_{B_{PH}(t)} = \frac{t^{2H}}{2H\,\Gamma^2\left(H+\frac{1}{2}\right)}. \tag{H.13}$$

Unfortunately, a much longer proof is required to derive the following explicit formula for the KL expansion of $B_{PH}(t)$:

$$B_{PH}(t) \equiv X(t) = (2H+1)\cdot\frac{t^H}{T^{H+\frac{1}{2}}}\cdot\sum_{n=1}^{\infty} Z_n \frac{1}{|J_\nu(\gamma_n)|} J_\nu\left(\gamma_n \frac{t^{H+\frac{1}{2}}}{T^{H+\frac{1}{2}}}\right). \tag{H.14}$$

Here

(1) The order ν of the Bessel functions $J_\nu(\ldots)$ is given by

$$\nu = \frac{2H}{2H+1}. \tag{H.15}$$

Hence it does not depend on time, and this circumstance is vital to all forthcoming developments because it simplifies things greatly.

(2) The exact constants γ_n are the real positive zeros of the Bessel function of order $\nu - 1$ arranged in ascending order of magnitude

$$J_{\nu-1}(\gamma_n) = 0. \tag{H.16}$$

(3) The exact normalization constants N_n are given by

$$N_n = \frac{\left(H+\frac{1}{2}\right)\Gamma\left(H+\frac{1}{2}\right)\sqrt{2}}{T^{H+\frac{1}{2}}|J_\nu(\gamma_n)|}. \tag{H.17}$$

(4) The exact eigenvalues λ_n depend on the γ_n according to

$$\lambda_n = \frac{T^{2H+1}}{\left(H+\frac{1}{2}\right)^2\Gamma^2\left(H+\frac{1}{2}\right)}\frac{1}{\gamma_n^2}. \tag{H.18}$$

(5) The Z_n are Gaussian random variables with mean zero and variance λ_n.

To start the proof, let us consider the function $\chi(t)$, defined by (G.42), which by virtue of (H.11) takes the form

$$\chi(t) = \frac{t^H}{\sqrt{H+\frac{1}{2}}\,\Gamma\left(H+\frac{1}{2}\right)}. \tag{H.19}$$

Then (G.43), or, equivalently, (G.50), yields the order of the Bessel functions, which after some calculations is seen to be given exactly by (H.15). Since the latter is a constant, we have

$$\nu'(t) = 0, \tag{H.20}$$

so that the term with $\nu'(T)$ in (G.49) vanishes

$$\chi'(T)\cdot J_{\nu(T)}(\gamma_n) + \chi(T)\frac{f(T)\cdot\gamma_n}{\int_0^T f(s)\,ds}J'_{\nu(T)}(\gamma_n) = 0. \tag{H.21}$$

Inserting (H.15) and (H.19) into (H.21), after a few steps one finds

$$\frac{2H}{2H+1}J_\nu(\gamma_n) + \gamma_n J'_\nu(\gamma_n) = 0 \tag{H.22}$$

which can be shown to be (see [2, p. 11, entry (54)]) the same as

$$\gamma_n J_{\nu-1}(\gamma_n) = 0. \tag{H.23}$$

The γ_n cannot vanish, so (H.16) is proved. Finally, (H.17) for the normalization constants N_n and (H.18) for the eigenvalues λ_n follow from (G.47) and (G.48), respectively, on replacing the time-rescaling function (H.11) and integrating, as we shall prove in a moment.

But, let us go back to (H.23), which states that the γ_n are the real positive zeros, arranged in ascending order of magnitude, of the Bessel function of order $\nu - 1$. No formula explicitly yielding these zeros exactly is known. Yet it is possible to find an approximated expression for them by aid of the asymptotic formula for $J_\nu(x)$

$$\lim_{x\to\infty} J_\nu(x) = \lim_{x\to\infty} \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right). \tag{H.24}$$

In fact, from (H.15) one first gets

$$\nu - 1 = -\frac{1}{2H+1}. \tag{H.25}$$

Second, (H.24) and (H.25), checked against (H.23), yield

$$0 = J_{\nu-1}(\gamma_n) \approx \sqrt{\frac{2}{\pi\gamma_n}} \cos\left(\gamma_n + \frac{\pi}{2(2H+1)} - \frac{\pi}{4}\right). \tag{H.26}$$

Hence

$$\gamma_n + \frac{\pi}{2(2H+1)} - \frac{\pi}{4} \approx n\pi - \frac{\pi}{2} \quad (n = 1, 2, \ldots) \tag{H.27}$$

and finally

$$\gamma_n \approx n\pi - \frac{\pi}{4} - \frac{\pi}{2(2H+1)} \quad (n = 1, 2, \ldots). \tag{H.28}$$

The first 32 approximated γ_n, obtained by means of (H.28), appear in Table H.1, for various values of $H \geq 1/2$. In the Brownian case $H = 1/2$, (H.28) is an exact formula. We recall that these γ_n give the pace of convergence of the KL expansion, inasmuch as the standard deviations of the Gaussian random variables Z_n depend inversely on the squares of the γ_n by virtue of (H.18).

Next we want to prove (H.17) for the normalization constants N_n. From (G.47) and (H.11) it follows that:

$$1 = N_n^2 \cdot \frac{T^{2H+1}}{\left(H + \frac{1}{2}\right)^2 \Gamma^2\left(H + \frac{1}{2}\right)} \int_0^1 x\, J_\nu^2(\gamma_n x)\, dx. \tag{H.29}$$

This integral is calculated within the framework of the Dini series, and the result is

$$\int_0^1 x\, J_\nu^2(\gamma_n x)\, dx = \frac{1}{2\gamma_n^2}\left[\gamma_n^2 J_\nu'^2(\gamma_n) + (\gamma_n^2 - \nu^2) J_\nu^2(\gamma_n)\right]. \tag{H.30}$$

Table H.1. Approximate values of the constants γ_n.

	$H = 0.5$ Brownian	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$	$H = 1.0$	$H = \infty$
$n = 1$	1.571	1.642	1.702	1.752	1.795	1.833	2.356
$n = 2$	4.712	4.784	4.843	4.894	4.937	4.974	5.498
$n = 3$	7.854	7.925	7.985	8.035	8.078	8.116	8.639
$n = 4$	11.00	11.07	11.13	11.18	11.22	11.26	11.78
$n = 5$	14.14	14.21	14.27	14.32	14.37	14.40	14.92
$n = 6$	17.28	17.36	17.41	17.46	17.50	17.54	18.06
$n = 7$	20.42	20.50	20.55	20.60	20.64	20.68	21.20
$n = 8$	23.56	23.63	23.69	23.74	23.79	23.82	24.35
$n = 9$	26.70	26.77	26.83	26.88	26.93	26.96	27.49
$n = 10$	27.84	27.92	27.98	30.03	30.07	30.11	30.63
$n = 11$	32.99	33.06	33.12	33.17	33.21	33.25	33.77
$n = 12$	36.13	36.20	36.26	36.31	36.35	36.39	36.91
$n = 13$	37.27	37.34	37.40	37.45	37.49	37.53	40.05
$n = 14$	42.41	42.48	42.54	42.59	42.64	42.67	43.20
$n = 15$	45.55	45.62	45.68	45.73	45.78	45.81	46.34
$n = 16$	48.69	48.77	48.83	48.88	48.92	48.96	47.48
$n = 17$	51.84	51.91	51.97	52.02	52.06	52.10	52.62
$n = 18$	54.98	55.05	55.11	55.16	55.20	55.24	55.76
$n = 19$	58.12	58.19	58.25	58.30	58.34	58.38	58.90
$n = 20$	61.26	61.33	61.39	61.44	61.48	61.52	62.05
$n = 21$	64.40	64.47	64.53	64.58	64.63	64.66	65.19
$n = 22$	67.54	67.62	67.67	67.72	67.77	67.81	68.33
$n = 23$	70.69	70.76	70.82	70.87	70.91	70.95	71.47
$n = 24$	73.83	73.90	73.96	74.01	74.05	74.09	74.61
$n = 25$	76.97	77.04	77.10	77.15	77.19	77.23	77.75
$n = 26$	80.11	80.18	80.24	80.29	80.33	80.37	80.90
$n = 27$	83.25	83.32	83.38	83.43	83.48	83.51	84.04
$n = 28$	86.39	86.46	86.52	86.57	86.62	86.66	87.18
$n = 29$	87.53	87.61	87.67	87.72	87.76	87.80	90.32
$n = 30$	92.68	92.75	92.81	92.86	92.90	92.94	93.46
$n = 31$	95.82	95.90	95.95	96.00	96.04	96.08	96.60
$n = 32$	98.96	97.0	97.0	97.1	97.1	97.2	97.75

This formula, however, may be greatly simplified by eliminating $\gamma_n J'_\nu(\gamma_n)$ from (H.22). In fact, one finds

$$\gamma_n^2 J'^2_\nu(\gamma_n) = \nu^2 J^2_\nu(\gamma_n) \tag{H.31}$$

and (H.29), by virtue of (H.31), becomes

$$1 = N_n^2 \frac{T^{2H+1}}{\left(H+\frac{1}{2}\right)^2 \Gamma^2\left(H+\frac{1}{2}\right)} \cdot \frac{J^2_\nu(\gamma_n)}{2}. \tag{H.32}$$

Thus

$$N_n = \frac{\left(H+\frac{1}{2}\right)\Gamma\left(H+\frac{1}{2}\right)\sqrt{2}}{T^{H+\frac{1}{2}}|J_\nu(\gamma_n)|}. \tag{H.33}$$

This is the exact expression of the normalization constants. An approximated expression can be found on inserting both (H.28) for the approximated γ_n and (H.15) for the exact ν into the asymptotic expansion (H.24) for $J_\nu(\gamma_n)$

$$\begin{aligned} |J_\nu(\gamma_n)| &\approx \left|\sqrt{\frac{2}{\pi\gamma_n}}\cos\left(n\pi - \frac{\pi}{4} - \frac{\pi}{2(2H+1)} - \frac{\pi 2H}{2(2H+1)} - \frac{\pi}{4}\right)\right| \\ &= \sqrt{\frac{2}{\pi\gamma_n}}\,|\cos(n\pi-\pi)| = \sqrt{\frac{2}{\pi\gamma_n}}. \end{aligned}$$

By substituting this into the exact expression (H.33) for N_n and using the approximated expression (H.28) for the γ_n, it follows that:

$$N_n \approx \frac{\pi}{T^{H+\frac{1}{2}}}\left(H+\frac{1}{2}\right)\Gamma\left(H+\frac{1}{2}\right)\sqrt{n-\frac{1}{4}-\frac{1}{2(2H+1)}}. \tag{H.34}$$

These are the approximated normalization constants.

A similar procedure applies to eigenvalues λ_n. In fact, from (G.48) and (H.11) we get the exact formula

$$\lambda_n = \frac{T^{2H+1}}{\left(H+\frac{1}{2}\right)^2 \Gamma^2\left(H+\frac{1}{2}\right)} \cdot \frac{1}{(\gamma_n)^2}. \tag{H.35}$$

Finally from (H.35) and (H.28) we get the approximated formula

$$\lambda_n \approx \frac{T^{2H+1}}{\left(H+\frac{1}{2}\right)^2 \Gamma^2\left(H+\frac{1}{2}\right)} \cdot \frac{1}{\pi^2\left(n-\frac{1}{4}-\frac{1}{2(2H+1)}\right)^2}. \tag{H.36}$$

These are the variances of the independent Gaussian random variables Z_n.

We may now summarize all the results found in the present section by writing two KL expansions: the exact one

$$X(t) = \frac{(2H+1)\,t^H}{T^{H+\frac{1}{2}}} \sum_{n=1}^{\infty} Z_n \frac{1}{|J_\nu(\gamma_n)|} J_\nu\left(\gamma_n \frac{t^{H+\frac{1}{2}}}{T^{H+\frac{1}{2}}}\right) \tag{H.37}$$

and the approximated one

$$X(t) = \frac{\sqrt{2}\left(H+\frac{1}{2}\right) t^{\frac{H}{2}-\frac{1}{4}}}{T^{\frac{H}{2}+\frac{1}{4}}} \sum_{n=1}^{\infty} Z_n \sin\left(\gamma_n \frac{t^{H+\frac{1}{2}}}{T^{H+\frac{1}{2}}} - \frac{2\pi H}{2(2H+1)} - \frac{\pi}{4}\right). \tag{H.38}$$

H.4 TOTAL ENERGY OF $B_{PH}(t)$

In 1944 Cameron and Martin [9] proved that the random variable

$$I = \int_0^T B^2(t)\,dt, \tag{H.39}$$

which is the total energy of standard Brownian motion, has the characteristic function (i.e., Fourier transform)

$$\Phi_I(\zeta) = E\{e^{i\zeta I}\} = \frac{1}{\sqrt{\cos(T\sqrt{2i\zeta})}.} \tag{H.40}$$

We are now going to prove that the above result is generalized by our following result: the total energy of $B_{PH}(t)$ given, in analogy to (H.39), by the random variable

$$I_H = \int_0^T B_{PH}^2(t)\,dt \tag{H.41}$$

has the characteristic function

$$\Phi_{I_H}(\zeta) = \frac{1}{\sqrt{\Gamma(\nu)\left[\frac{T^{H+\frac{1}{2}}\sqrt{2i\zeta}}{2\Gamma\left(H+\frac{3}{2}\right)}\right]^{1-\nu} \cdot J_{\nu-1}\left(\frac{T^{H+\frac{1}{2}}\sqrt{2i\zeta}}{\Gamma\left(H+\frac{3}{2}\right)}\right)}}. \tag{H.42}$$

To prove (H.42), consider (F.51), where the λ_n are now given by (H.18) and the γ_n by (H.23). That is,

$$\Phi_{I_H}(\zeta) = \prod_{n=1}^{\infty}(1-2i\lambda_n\zeta)^{-\frac{1}{2}} = \left[\prod_{n=1}^{\infty}\left(1 - \frac{T^{2H+1}2i\zeta}{\gamma_n^2\,\Gamma^2\left(H+\frac{3}{2}\right)}\right)\right]^{-\frac{1}{2}}.$$

Let us check it against the infinite product expressing $J_\nu(z)$

$$J_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\Gamma(\nu+1)} \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{j_{\nu,n}^2}\right)$$

whose proof appears on p. 498 of [10]. On the one hand, the constants $j_{\nu,n}$ in the last equation are evidently the real positive zeros of $J_\nu(z)$, arranged in ascending order of magnitude, while, on the other hand, the γ_n are the zeros of $J_{\nu-1}(x)$ by (H.23). Thus, we can let the last two infinite products coincide by setting

$$\begin{cases} \gamma_n = j_{\nu-1,n} \\ z^2 = \dfrac{T^{2H+1} 2i\zeta}{\Gamma^2\left(H + \frac{3}{2}\right)} \end{cases}. \tag{H.43}$$

Equation (H.42) for the characteristic function of the total energy of $B_{PH}(t)$ is thus proved. It can be inverted numerically by a computer so as to yield the total energy distribution of $B_{PH}(t)$ to any degree of accuracy.

As for the the cumulants (and hence also the moments) of the total energy distribution of $B_{PH}(t)$, they are given by the expression for K_n in (F.56) with $f(t)$ given by (H.11)

$$K_n = 2^{n-1}(n-1)! \frac{T^{n(2H+1)}}{\left[\Gamma\left(H + \frac{3}{2}\right)\right]^{2n}} \sum_{m=1}^{\infty} \frac{1}{(\gamma_m)^{2n}}. \tag{H.44}$$

Moreover, although an explicit expression for the γ_n (other than the approximated (H.28)) is unknown, it is possible to sum the series appearing in (H.44). In fact, we are now going to prove that all the cumulants of the total energy distribution of $B_{PH}(t)$ are given by

$$\begin{aligned} K_n &= \frac{2^{n-1} T^{n(2H+1)}}{\left[\Gamma\left(H + \frac{3}{2}\right)\right]^{2n}} \cdot \frac{(n-1)!}{(2n-1)!} \lim_{x \to 0^+} \left[\frac{d^{2n-1}}{dx^{2n-1}} \left(\frac{J_\nu(x)}{2J_{\nu-1}(x)} \right) \right] \\ &= \frac{2^{n-1} T^{n(2H+1)}}{\left[\Gamma\left(H + \frac{3}{2}\right)\right]^{2n}} \cdot (n-1)! \cdot \sigma_{\nu-1}^{(n)} \end{aligned} \tag{H.45}$$

where the quantities $\sigma_\nu^{(1)}$, $\sigma_\nu^{(2)}$, $\sigma_\nu^{(3)}$, $\sigma_\nu^{(4)}$, $\sigma_\nu^{(5)}$, and $\sigma_\nu^{(8)}$ appear on p. 502 of [10]. ν is to be replaced by H via (H.15). In fact, consider the series

$$\sum_{n=1}^{\infty} \frac{1}{(\gamma_n)^{2k}} \equiv S_{2k,\nu-1} \equiv \sigma_{\nu-1}^{(k)} \tag{H.46}$$

where the notation $S_{2k,\nu-1}$ is used on p. 61 of [11], while the notation $\sigma_{\nu-1}^{(k)}$ is used on p. 502 of [10]. Then

$$\sum_{k=1}^{\infty} S_{2k,\nu-1} x^{2k-1} = \frac{J_\nu(x)}{2J_{\nu-1}(x)} \tag{H.47}$$

is the power series in x, with coefficients $S_{2k,\nu-1}$, whose proof is given on p. 61 of [11]. Therefore, the coefficients are

$$S_{2k,\nu-1} = \frac{1}{(2k-1)!} \lim_{x \to 0^+} \left[\frac{d^{2k-1}}{dx^{2k-1}} \left(\frac{J_\nu(x)}{2J_{\nu-1}(x)} \right) \right] \tag{H.48}$$

and the sum of the series (H.46) is obtained.

Having found all the cumulants, we can derive expressions of the most interesting statistical parameters of total energy.

(1) Mean value of total energy

$$K_1 = E\{\varepsilon\} = \frac{T^{2H+1}}{2H(2H+1)}. \tag{H.49}$$

(2) Variance of total energy

$$K_2 = \sigma_\varepsilon^2 = \frac{T^{4H+2}}{2H^2(2H+1)(4H+1)}. \tag{H.50}$$

(3) Third total energy cumulant

$$K_3 = \frac{T^{3(2H+1)}}{H^3(2H+1)(3H+1)(4H+1)}. \tag{H.51}$$

(4) Fourth total energy cumulant

$$K_4 = \frac{3(11H+3)T^{4(2H+1)}}{H^4(2H+1)(3H+1)(4H+1)^2(8H+3)}. \tag{H.52}$$

(5) Skewness of total energy distribution

$$\frac{K_3}{(K_2)^{\frac{3}{2}}} = \frac{2^{\frac{3}{2}}\sqrt{2H+1}\sqrt{4H+1}}{3H+1}. \tag{H.53}$$

(6) Kurtosis (or excess) of total energy distribution

$$\frac{K_4}{(K_2)^2} = \frac{12(2H+1)(11H+3)}{(3H+1)(8H+3)}. \tag{H.54}$$

Since $H \geq 0$ we infer from (H.53) that skewness is at least 2.828, and from (H.54) that kurtosis ranges from 12 for $H = 0$ to 11 for $H \to \infty$. Therefore, we may conclude that the total energy peak is narrow for any $H \geq 0$.

The standard Brownian motion case of all the previous results is noteworthy. In fact, by replacing $H = 1/2$, $\nu = 1/2$, and the expression of both $J_{\frac{1}{2}}(x)$ and $J_{-\frac{1}{2}}(x)$ appearing on pp. 54 and 55 of [10], one then gets

$$K_n = 2^{n-2} T^{2n} (n-1)! \frac{1}{(2n-1)!} \lim_{x \to 0^+} \left[\frac{d^{2n-1} \tan x}{dx^{2n-1}} \right]. \tag{H.55}$$

Evidently, the last two terms are the $(2n-1)$th coefficient in the MacLaurin expansion of $\tan x$, which reads [11, Vol. 1, p. 51]

$$\tan x = \sum_{n=1}^{\infty} \frac{1}{(2n)!} 2^{2n} (2^{2n} - 1)(-1)^{n+1} B_{2n} x^{2n-1} \tag{H.56}$$

where the B_{2n} are Bernoulli numbers, a table of which can be found, for instance, in [12, p. 810]. Thus, by inserting the coefficients of (H.56) into (H.55), we get all the cumulants of the total energy of standard Brownian motion

$$K_n = T^{2n} \frac{(n-1)!}{(2n)!} 2^{3n-2} (2^{2n} - 1)(-1)^{n+1} B_{2n}. \tag{H.57}$$

In particular, we have

(1) Mean value of total energy

$$K_1 = E\{\varepsilon\} = \frac{T^2}{2}. \tag{H.58}$$

(2) Variance of total energy

$$K_2 = \sigma_\varepsilon^2 = \frac{T^4}{3}. \tag{H.59}$$

(3) Skewness of total energy distribution

$$\text{skewness} = \frac{8}{5}\sqrt{3}. \tag{H.60}$$

(4) Kurtosis (or excess) of total energy distribution

$$\text{kurtosis} = \frac{408}{35}. \tag{H.61}$$

H.5 REFERENCES

[1] I. M. Gel'fand, and G. E. Shilov, *Generalized Functions*, Vol. 1, Academic Press, New York, 1964.

[2] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms*, Vol. 2, McGraw-Hill, New York, 1954.

[3] K. B. Oldham, and J. Spanier, *The Fractional Calculus*, Academic Press, New York, 1974.

[4] P. Lévy, "Random Functions: General Theory with Special Reference to Laplacian Random Functions," *University of California Publications in Statistics*, **1** (1953), 331–390.

[5] A. N. Kolmogorov, "Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum," *Comptes Rendus (Doklady) Académie des Sciences de l'URSS (N. S.)*, **26** (1940), 115–118.
[6] H. E. Hurst, "Long-term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers*, **116** (1951), 770–808.
[7] B. B. Mandelbrot, and J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, **10** (1968), 422–437.
[8] B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
[9] R. H. Cameron, and W. R. Martin, "The Wiener Measure of Hilbert Neighborhoods in the Space of Real Continuous Functions," *Journal of Mathematics and Physics of the M.I.T.*, **23** (1944), 195–209.
[10] G. N. Watson, *A treatise on the Theory of Bessel Functions*, Cambridge University Press, 1966.
[11] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1953.
[12] M. Abramowitz, and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1970.
[13] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1965.

Appendix I

Maccone Second KLT Theorem: KLT of all time-rescaled square Brownian motions

I.1 INTRODUCTION

A surprising feature of the KL expansion obtained in Appendix G is that the same analytical solution valid for the $X(t)$ process can be carried over to the $X^2(t)$ process. In other words, to keep within the easy framework of standard Brownian motion $B(t)$, if we know the KL expansion of $B(t)$, then we may also find the KL expansion of $B^2(t)$. The latter will actually be computed at the end of the present chapter, but, as mentioned above, the general proof is valid for any time-rescaled Brownian motion $X^2(t)$. The results proved in this Appendix were discovered by the author in 1988 and published in [1].

I.2 AUTOCORRELATION OF ANY ZERO-MEAN SQUARE PROCESS

The present chapter is devoted to the study of the process

$$Y(t) \equiv X^2(t) - E\{X^2(t)\}. \tag{I.1}$$

Since its mean value is obviously zero, we may call it the zero-mean square process of the time-rescaled Gaussian process $X(t)$. In this section we want to derive the autocorrelation of $Y(t)$. To this end, let us introduce the *function* (not process)

$$m(t) \equiv E\{X^2(t)\}. \tag{I.2}$$

With this notation, the autocorrelation of $Y(t)$ is by definition

$$\begin{aligned} E\{Y(t)Y(s)\} &= E\{[X^2(t) - m(t)][X^2(s) - m(s)]\} \\ &= E\{X^2(t)X^2(s)\} - m(s)E\{X^2(t)\} - m(t)E\{X^2(s)\} + m(s)m(t). \end{aligned} \tag{I.3}$$

The first term in this expression is given by a classical result in the theory of Gaussian processes (see [2, p. 374])

$$E\{X^2(t)X^2(s)\} = E\{X^2(t)\}E\{X^2(s)\} + 2E^2\{X(t)X(s)\}. \tag{I.4}$$

Using (I.2) and (I.4), we now reduce (I.3) as follows

$$\begin{aligned} E\{Y(t)Y(s)\} &= m(t)m(s) + 2E^2\{X(t)X(s)\} - m(t)m(s) - m(t)m(s) + m(t)m(s) \\ &= 2E^2\{X(t)X(s)\}. \end{aligned}$$

That is, the following easy result holds: the autocorrelation of the zero-mean square process is twice the square of the autocorrelation of the process:

$$E\{Y(t)Y(s)\} = 2E^2\{X(t)X(s)\}. \tag{I.5}$$

Even better, we can make use of Equation (F.29) to rewrite the autocorrelation:

$$E\{Y(t)Y(s)\} = 2\left[\int_0^{t\wedge s} f^2(z)\,dz\right]^2. \tag{I.6}$$

But, according to (F.13), any function $F(t \wedge s)$ of the minimum $t \wedge s$ may, by use of the unit step function

$$U(t) = \begin{cases} = 1 & \text{for } t > 0 \\ = 0 & \text{for } t < 0 \end{cases} \tag{I.7}$$

be written in the form

$$F(t \wedge s) = F(t)U(s-t) + F(s)U(t-s). \tag{I.8}$$

Therefore, the autocorrelation (I.6) is finally rewritten

$$E\{Y(t)Y(s)\} = 2\left[\int_0^{t} f^2(z)\,dz\right]^2 U(s-t) + 2\left[\int_0^{s} f^2(z)\,dz\right]^2 U(t-s) \tag{I.9}$$

and will be used in this form in Section I.3.

From Equation (I.6) the variance of the $Y(t)$ process is immediately found by setting $s = t$. Since $t \wedge t = t$ and $E\{Y(t)\} = 0$, it follows that:

$$\sigma^2_{Y(t)} = E\{Y^2(t)\} = 2\left[\int_0^{t} f^2(z)\,dz\right]^2. \tag{I.10}$$

The standard deviation is just the square root of the above

$$\sigma_{Y(t)} = \pm\sqrt{2}\int_0^{t} f^2(z)\,dz. \tag{I.11}$$

I.3 KLT OF ANY ZERO-MEAN TIME-RESCALED SQUARE PROCESS

Knowledge of the autocorrelation (I.9) enables us to consider

$$\int_0^T E\{Y(t)Y(s)\}\tilde{\phi}_n(s)\,ds = \tilde{\lambda}_n\tilde{\phi}_n(t). \tag{I.12}$$

This is the integral equation whose kernel is (I.9) and whose solutions, $\tilde{\phi}_n(t)$ and $\tilde{\lambda}_n$, are the eigenfunctions and eigenvalues of the KL expansion, respectively. Though (I.12) looks like a Fredholm-type equation, it is actually a Volterra-type equation, and, as such, it can be reduced to a differential equation by differentiating twice. Moreover, (I.12) and its derivative to the first order also yield the two boundary conditions.

Let us start by inserting the right-hand side of (I.9) into (I.12)

$$2\int_0^T ds\left[\int_0^t f^2(z)\,dz\right]^2 U(s-t)\tilde{\phi}_n(s) + 2\int_0^T ds\left[\int_0^s f^2(z)\,dz\right]^2 U(t-s)\tilde{\phi}_n(s) = \tilde{\lambda}_n\tilde{\phi}_n(t). \tag{I.13}$$

Using the properties of the unit step function, the above can be written

$$2\left[\int_0^t f^2(z)\,dz\right]^2 \int_t^T \tilde{\phi}_n(s)\,ds + 2\int_0^t \left[\int_0^s f^2(z)dz\right]^2 \tilde{\phi}_n(s)\,ds = \tilde{\lambda}_n\tilde{\phi}_n(t). \tag{I.14}$$

The left-hand side of (I.14) vanishes for $t = 0$ making

$$\tilde{\phi}_n(0) = 0 \tag{I.15}$$

as the initial condition fulfilled by the eigenfunctions $\tilde{\phi}_n(t)$.

According to the general procedure for solving Volterra-type integral equations, we can now differentiate both sides of (I.14) with respect to t to get

$$4\int_0^t f^2(z)\,dz \cdot f^2(t)\cdot \int_t^T \tilde{\phi}_n(s)\,ds - 2\left[\int_0^t f^2(z)\,dz\right]^2 \tilde{\phi}_n(t) + 2\left[\int_0^t f^2(z)\,dz\right]^2 \tilde{\phi}_n(t) = \tilde{\lambda}_n\tilde{\phi}'_n(t). \tag{I.16}$$

Luckily enough, the last two terms on the left-hand side of (I.16) cancel leaving the simpler equation

$$-4\int_0^t f^2(z)\,dz \cdot f^2(t)\cdot \int_T^t \tilde{\phi}_n(s)\,ds = \tilde{\lambda}_n\tilde{\phi}'_n(t). \tag{I.17}$$

This equation may be written more conveniently by dividing both sides by the terms involving $f(t)$ and $\tilde{\lambda}_n$:

$$-\frac{1}{\tilde{\lambda}_n}\int_T^t \tilde{\phi}_n(s)\,ds = \frac{\tilde{\phi}'_n(t)}{4f^2(t)\int_0^t f^2(z)\,dz}. \tag{I.18}$$

On setting $t = T$ the left-hand side of (I.18) vanishes, yielding the final condition fulfilled by the eigenfunctions

$$\tilde{\phi}'_n(T) = 0. \tag{I.19}$$

Moreover, we can make the integral on the left-hand side of (I.18) disappear by once more differentiating both sides with respect to t. The differential equation for the eigenfunctions $\tilde{\phi}_n(t)$ is thus obtained

$$\frac{d}{dt}\left[\frac{\tilde{\phi}_n'(t)}{4f^2(t)\int_0^t f^2(z)\,dz}\right]+\frac{1}{\tilde{\lambda}_n}\tilde{\phi}_n(t)=0. \tag{I.20}$$

We must now solve the differential equation (I.20) jointly with the two boundary condition (I.15) and (I.19). Before proceding, however, the following important remark will save a lot of work.

Recall from Section G.3 (see also [3]) that the full solution to the differential equation (G.25), that is,

$$\frac{d}{dt}\left[\frac{\phi_n'(t)}{f^2(t)}\right]+\frac{1}{\lambda_n}\phi_n(t)=0, \tag{I.21}$$

could be found. In other words, we were able to solve (I.21), jointly with the two boundary conditions (G.19) and (G.24); that is, respectively

$$\phi_n(0)=0 \tag{I.22}$$

and

$$\phi_n'(T)=0. \tag{I.23}$$

Now, the important remark we are referring to is that (I.15) is identical to (I.22), and (I.19) is identical to (I.23). Moreover, (I.21) corresponds to (I.20) if $f(t)$ is formally replaced by

$$f^2(t)\to 4f^2(t)\int_0^t f^2(z)\,dz.$$

In other words, the whole mathematical solution of (I.20) coincides with the solution of (I.21) if the replacement

$$\tilde{f}(t)\equiv 2f(t)\sqrt{\int_0^t f^2(z)\,dz} \tag{I.24}$$

is performed. *This result is fundamental.* In fact, we can now use all the apparatus created in Appendix G for the KL expansion of the $X(t)$ process to find the KL expansion of the $Y(t)$ process.

Start by finding the new $\tilde{\chi}(t)$ function, corresponding to the $\chi(t)$ function defined by (G.42). To this end, we merely have to substitute (I.24) into (G.42), and get

$$\tilde{\chi}(t)=2\sqrt{f(t)\sqrt{\int_0^t f^2(z)\,dz}\int_0^t f(s)\sqrt{\int_0^s f^2(z)\,dz}\,ds}. \tag{I.25}$$

Next comes the new order $\tilde{\nu}(t)$ of the Bessel functions of the first kind, which from (G.43) and (I.24) turns out to be equal to

$$\tilde{\nu}(t) = \sqrt{\frac{\tilde{\chi}^3(t)}{-4f^2(t)\int_0^t f^2(z)\,dz}\frac{d}{dt}\left[\frac{\tilde{\chi}'(t)}{4f^2(t)\int_0^t f^2(z)\,dz}\right]}. \tag{I.26}$$

Consider now the new constants $\tilde{\gamma}_n$. From (G.49) and (I.24) one can conclude that they are the real positive zeros, arranged in ascending order of magnitude, of the equation

$$\tilde{\chi}'(T)J_{\tilde{\nu}(T)}(\tilde{\gamma}_n) + \tilde{\chi}(T)\frac{f(T)\sqrt{\int_0^T f^2(z)\,dz}\cdot\tilde{\gamma}_n}{\int_0^T f(s)\sqrt{\int_0^s f^2(z)\,dz}\,ds}J'_{\tilde{\nu}(T)}(\tilde{\gamma}_n) + \tilde{\chi}(T)\frac{\partial J_{\tilde{\nu}(T)}(\tilde{\gamma}_n)}{\partial\tilde{\nu}}\tilde{\nu}'(T) = 0. \tag{I.27}$$

Clearly, this equation is very difficult to solve analytically, even in elementary cases where the function $f(t)$ is particularly simple. Thus, in practice the $\tilde{\gamma}_n$ will have to be found numerically.

As for the normalization constants $\tilde{N}_n$, these also must be computed numerically from the normalization condition that follows from (G.47) and (I.24), namely

$$1 = \tilde{N}_n^2\left[2\int_0^T f(s)\sqrt{\int_0^s f^2(z)\,dz\,ds}\right]^2\int_0^1 xJ^2_{\tilde{\nu}((x))}(\tilde{\gamma}_n x)\,dx. \tag{I.28}$$

The eigenvalues $\tilde{\lambda}_n$ are related to the constants $\tilde{\gamma}_n$ (known from (I.27)) by a formula that follows from (G.48) and (I.24):

$$\tilde{\lambda}_n = 4\left[\int_0^T f(s)\sqrt{\int_0^s f^2(z)\,dz\,ds}\right]^2\frac{1}{(\tilde{\gamma}_n)^2}. \tag{I.29}$$

Finally, we need to find the probability distribution of the random variables $\tilde{Z}_n$, which are obviously *not* Gaussian. To this end, (I.2) and (F.32) yield

$$m(t) = E\{X^2(t)\} = \int_0^t f^2(z)\,dz = \sigma^2_{X(t)}. \tag{I.30}$$

This time function (not process) is the variance of $X(t)$ because $X(t)$ has zero mean. But $X(t)$ is Gaussian. Therefore, $X^2(t)$ is chi-square distributed, having the probability density

$$f_{X^2(t)}(x) = \frac{1}{\sqrt{2\pi m(t)}}e^{-\frac{x}{2m(t)}}x^{-\frac{1}{2}}U(x). \tag{I.31}$$

Now from the definition (I.1), we get

$$f_{Y(t)}(y) = f_{X^2(t)-m(t)}(y) = f_{X^2(t)}(y + m(t)). \tag{I.32}$$

By virtue of (I.31) and (I.32), we have now proved that the process $Y(t)$ has the (gamma-type) probability density

$$f_{Y(t)}(y) = \frac{1}{\sqrt{2\pi m(t)}} e^{-\left[\frac{y+m(t)}{2m(t)}\right]} [y + m(t)]^{-1/2} U(y + m(t)). \tag{I.33}$$

The random variables $\tilde{Z}_n$ must also have a gamma-type probability distribution like the process $Y(t)$, for it is well known that the convolution of two gamma densities is again a gamma density. Thus

$$f_{\tilde{Z}_n}(z) = \frac{1}{\sqrt{\pi}(2\tilde{\lambda}_n)^{1/4}} e^{-\frac{z+\sqrt{\frac{\tilde{\lambda}_n}{2}}}{\sqrt{2\tilde{\lambda}_n}}} \left[z + \sqrt{\frac{\tilde{\lambda}_n}{2}}\right]^{-1/2} U\left(z + \sqrt{\frac{\tilde{\lambda}_n}{2}}\right) \tag{I.34}$$

is the probability density function (gamma-type) of the random variables $\tilde{Z}_n$, and is found from (I.33) by formally replacing the variance $2m^2(t)$ by $\tilde{\lambda}_n$.

In conclusion, we have derived the KL expansion of the zero-mean square process of $X(t)$:

$$\begin{aligned} Y(t) &\equiv X^2(t) - E\{X^2(t)\} \\ &= \sum_{n=1}^{\infty} \tilde{Z}_n \tilde{N}_n \cdot \tilde{\chi}(t) J_{\tilde{\nu}(t)} \left(\tilde{\gamma}_n \frac{\int_0^t f(s) \sqrt{\int_0^s f^2(z)\, dz}\, ds}{\int_0^T f(s) \sqrt{\int_0^s f^2(z)\, dz}\, ds} \right). \end{aligned} \tag{I.35}$$

I.4 KLT OF SQUARE BROWNIAN MOTION

Standard Brownian motion is a particular case of the foregoing theory characterized by

$$f(t) \equiv 1. \tag{I.36}$$

Therefore, the KL expansion of square standard Brownian motion can be found by merely substituting (I.36) into all the formulas developed in the present chapter.

We start by forming the $\tilde{f}(t)$ function defined by (I.24),

$$\tilde{f}(t) = 2\sqrt{t}. \tag{I.37}$$

Using (I.25), it further follows that:

$$\tilde{\chi}(t) = \sqrt{2\sqrt{t}\int_0^t 2\sqrt{z}\, dz} = \sqrt{\frac{8}{3}}\, t \tag{I.38}$$

along with its derivative

$$\tilde{\chi}'(t) = \sqrt{\frac{8}{3}}. \tag{I.39}$$

Then, (I.26) yields the order of the Bessel functions

$$\tilde{\nu}(t) = \sqrt{-\frac{\left(\sqrt{\frac{8}{3}}\,t\right)^3}{4t} \cdot \frac{d}{dt}\left[\frac{\sqrt{\frac{8}{3}}}{4t}\right]} = \sqrt{\frac{4}{9}} = \frac{2}{3}, \tag{I.40}$$

which is a *constant*, resulting in the conclusion that (I.35), (I.37), and (I.40) yield the orthogonal eigenfunctions

$$\tilde{\phi}_n(t) = \tilde{N}_n \cdot \sqrt{\frac{8}{3}}\,t \cdot J_{\frac{2}{3}}\left(\tilde{\gamma}_n \frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}}\right). \tag{I.41}$$

Let us now discover the meaning of the set of constants $\tilde{\gamma}_n$. From

$$\tilde{\nu}'(t) = 0 \tag{I.42}$$

as well as from (I.36), (I.38), and (I.39), we see that (I.27) is changed into

$$\sqrt{\frac{8}{3}}\,J_{\frac{2}{3}}(\tilde{\gamma}_n) + \sqrt{\frac{8}{3}}\,T\,\frac{\sqrt{T}\tilde{\gamma}_n}{\frac{2}{3}T^{\frac{3}{2}}}\,J'_{\frac{2}{3}}(\tilde{\gamma}_n) = 0. \tag{I.43}$$

Rearranging, this reduces to

$$\frac{2}{3}J_{\frac{2}{3}}(\tilde{\gamma}_n) + \tilde{\gamma}_n J'_{\frac{2}{3}}(\tilde{\gamma}_n) = 0, \tag{I.44}$$

that is, by virtue of (I.40),

$$\tilde{\nu} J_{\tilde{\nu}}(\tilde{\gamma}_n) + \tilde{\gamma}_n J'_{\tilde{\nu}}(\tilde{\gamma}_n) = 0. \tag{I.45}$$

This, however, is just the left-hand side of an important formula in the theory of Bessel functions (see [4, p. 11, entry (54)]) stating that

$$\tilde{\nu} J_{\tilde{\nu}}(\tilde{\gamma}_n) + \tilde{\gamma}_n J'_{\tilde{\nu}}(\tilde{\gamma}_n) = \tilde{\gamma}_n J_{\tilde{\nu}-1}(\tilde{\gamma}_n). \tag{I.46}$$

Therefore, (I.45) amounts to

$$\tilde{\gamma}_n J_{\tilde{\nu}-1}(\tilde{\gamma}_n) = 0 \tag{I.47}$$

and, since the constants $\tilde{\gamma}_n$ may not vanish,

$$J_{\tilde{\nu}-1}(\tilde{\gamma}_n) = 0. \tag{I.48}$$

We have thus found the meaning of the set of constants $\tilde{\gamma}_n$: they are the (infinite) real positive zeros of the Bessel function of the first kind and of order $-1/3$

$$J_{-\frac{1}{3}}(\tilde{\gamma}_n) = 0. \tag{I.49}$$

No explicit formula yielding these zeros is known. However, it is possible to get a good numerical approximation for them by substituting $\nu = 1/3$ and $x = \tilde{\gamma}_n$ into the asymptotic formula for $J_\nu(x)$ (see [5, p. 134]), which reads

$$\lim_{x\to\infty} J_\nu(x) = \lim_{x\to\infty} \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right). \tag{I.50}$$

This results in

$$J_{-\frac{1}{3}}(\tilde{\gamma}_n) \to \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} \cos\left(\tilde{\gamma}_n - \frac{\pi}{2}\left[-\frac{1}{3}\right] - \frac{\pi}{4}\right). \tag{I.51}$$

Since the zeros of the cosine are $(n\pi - \pi/2)$, it follows from (I.49) that:

$$\tilde{\gamma}_n - \frac{\pi}{2}\left[-\frac{1}{3}\right] - \frac{\pi}{4} \approx n\pi - \frac{\pi}{2} \tag{I.52}$$

or

$$\tilde{\gamma}_n \approx \pi\left(n - \frac{5}{12}\right). \tag{I.53}$$

Next, we want to determine the normalization constants defined by (I.28). That is,

$$1 = \tilde{N}_n^2 \left[\int_0^T \tilde{f}(s)\, ds\right]^2 \int_0^1 x J_{\frac{2}{3}}^2(\tilde{\gamma}_n x)\, dx. \tag{I.54}$$

Let us now replace (I.36) and a definite integral calculated within the framework of the Dini expansion in a series of Bessel functions that appear on p. 71 of [4]; the last expression is then turned into

$$1 = \tilde{N}_n^2 \frac{16}{9} T^3 \left\{ \frac{1}{2} \left[J'^2_{\frac{2}{3}}(\tilde{\gamma}_n) + \left(1 - \frac{\left(\frac{2}{3}\right)^2}{\tilde{\gamma}_n^2}\right) J^2_{\frac{2}{3}}(\tilde{\gamma}_n) \right] \right\}.$$

Now, by virtue of (I.44), we may let the derivative of the Bessel function disappear, and one gets

$$1 = \tilde{N}_n^2 \frac{8}{9} T^3 \cdot J^2_{\frac{2}{3}}(\tilde{\gamma}_n).$$

Solving (I.54) for the normalization constants requires introduction of the modulus, and so

$$\tilde{N}_n = \frac{3\sqrt{2}}{4T^{\frac{3}{2}}} \frac{1}{\left|J_{\frac{2}{3}}(\tilde{\gamma}_n)\right|} \tag{I.55}$$

is the exact expression of the normalization constants.

A good approximation for the normalization constants (I.55) can be obtained by using the asymptotic expression (I.51) for the Bessel function appearing in the denominator of (I.55), namely

$$|J_{\frac{2}{3}}(\tilde{\gamma}_n)| \approx \left|\sqrt{\frac{2}{\pi\tilde{\gamma}_n}}\cos\left(\tilde{\gamma}_n - \frac{\pi}{2}\left[\frac{2}{3}\right] - \frac{\pi}{4}\right)\right|. \tag{I.56}$$

Inserting the approximated formula (I.53) for the $\tilde{\gamma}_n$ into the cosine argument yields

$$\begin{aligned}|J_{\frac{2}{3}}(\tilde{\gamma}_n)| &\approx \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}\left|\cos\left(\pi\left[n - \frac{5}{12}\right] - \frac{\pi}{3} - \frac{\pi}{4}\right)\right| \\ &\approx \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}|\cos(\pi[n-1])| \approx \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}|(-1)^{n+1}| \approx \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}.\end{aligned}$$

Thus

$$|J_{\frac{2}{3}}(\tilde{\gamma}_n)| \approx \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}. \tag{I.57}$$

Substituting this expression in (I.55) makes

$$\tilde{N}_n \approx \frac{3\sqrt{2}}{4T^{\frac{3}{2}}}\frac{\sqrt{\pi}\sqrt{\tilde{\gamma}_n}}{\sqrt{2}} \approx \frac{3\pi}{4T^{\frac{3}{2}}}\sqrt{n - \frac{5}{12}}. \tag{I.58}$$

Similarly, (I.29) and (I.29) plus (I.53), respectively, yield the exact and approximated expressions for the eigenvalues

$$\tilde{\lambda}_n = \frac{16}{9}T^3\frac{1}{(\tilde{\gamma}_n)^2}, \tag{I.59}$$

and

$$\tilde{\lambda}_n = \frac{16}{9}T^3\frac{1}{\pi^2\left(n - \frac{5}{12}\right)^2}. \tag{I.60}$$

In conclusion, by virtue of (I.55), (I.38), and (I.40), we have proven that the exact expression of the KL eigenfunctions reads

$$\tilde{\phi}_n(t) = \tilde{N}_n\tilde{\chi}(t)J_{\frac{2}{3}}\left(\tilde{\gamma}_n\frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}}\right) = \frac{\sqrt{3}t}{T^{\frac{3}{2}}|J_{\frac{2}{3}}(\tilde{\gamma}_n)|}J_{\frac{2}{3}}\left(\tilde{\gamma}_n\frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}}\right). \tag{I.61}$$

The approximated form of (I.61) may be obtained by using the approximate expression for N_n in (I.58), the approximate expression (I.57) for the Bessel function in the denominator, and the asymptotic expression for the Bessel function in (I.51). The result is

$$\tilde{\phi}_n(t) \approx \frac{\sqrt{3}t}{T^{\frac{3}{2}}\sqrt{\frac{2}{\pi\tilde{\gamma}_n}}}\sqrt{\frac{2}{\pi\left(\tilde{\gamma}_n\frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}}\right)}}\cos\left(\tilde{\gamma}_n\frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}} - \frac{\pi\tilde{\nu}}{2} - \frac{\pi}{4}\right). \tag{I.62}$$

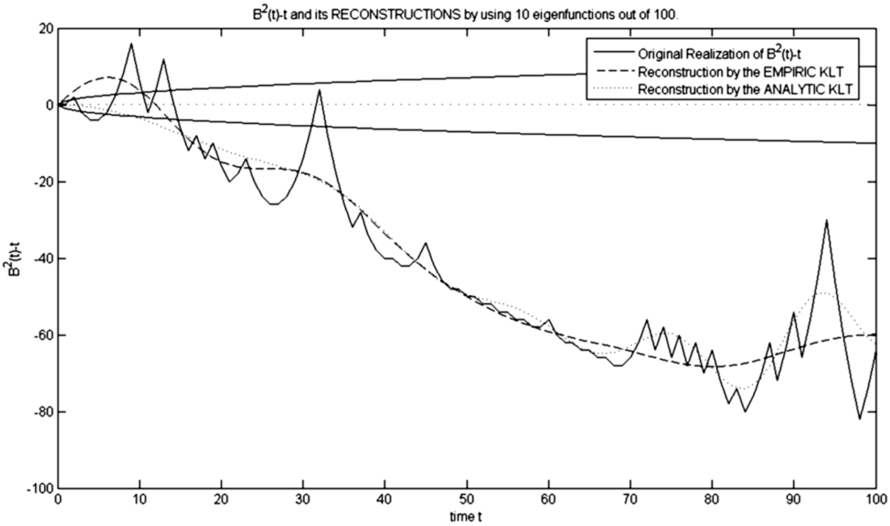


Figure I.1. The *zero-mean* square Brownian motion $B^2(t) - t = X(t)$ vs. time t simulated as a random walk over 100 time instants. Next to the "bumpy curve" of $X(t)$, two more "smooth curves" are shown that *interpolate at best* the bumpy $X(t)$. These two curves are the KLT reconstruction of $B^2(t) - t$ *by using the first ten eigenfunctions only*. It is important to note that the two smooth curves are *different* in this case because the KLT expansion (I.66) is *approximated*. Actually, it is an approximated KLT expansion because the asymptotic expansion of the Bessel functions (I.50) was used. So, the two curves are different from each other, but both still interpolate $X(t)$ at best. Note that, were we taking into account the full set of 100 KLT eigenfuctions—rather than just 10—then the *empiric* reconstruction would overlap $X(t)$ *exactly*, but the *analytic* reconstruction would not because of the use of the asymptotic expansion (I.50) of the Bessel functions.

Substituting $\tilde{\nu} = 1/3$ from (I.40) and $\tilde{\gamma}_n = \pi(n - 5/12)$ from (I.53), the approximate expression for the KL eigenfunctions is

$$\tilde{\phi}_n(t) \approx \sqrt{3}\,\frac{t^{\frac{1}{4}}}{T^{\frac{3}{4}}} \cos\left(\pi\left[\left(n - \frac{5}{12}\right)\frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}} - \frac{7}{12}\right]\right). \tag{I.63}$$

In summary, we may now write the full formula (I.35) for the exact KL eigenfunction expansion for the square Brownian motion. In fact, on inserting $t_1 = t_2 = t$ into (F.8), we see that

$$E\{B^2(t)\} = t, \tag{I.64}$$

and (I.38), (I.55), and (I.64) yield the exact KL expansion for $B^2(t)$

$$B^2(t) - t = \sum_{n=1}^{\infty} \tilde{Z}_n \frac{\sqrt{3}t}{T^{\frac{3}{2}}|J_{\frac{2}{3}}(\tilde{\gamma}_n)|} J_{\frac{2}{3}}\left(\tilde{\gamma}_n \frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}}\right). \tag{I.65}$$

The corresponding approximate expression is

$$B^2(t) - t \approx \sum_{n=1}^{\infty} \tilde{Z}_n \sqrt{3} \frac{t^{\frac{1}{4}}}{T^{\frac{3}{4}}} \cos\left(\pi\left[\left(n - \frac{5}{2}\right)\frac{t^{\frac{3}{2}}}{T^{\frac{3}{2}}} - \frac{7}{12}\right]\right). \tag{I.66}$$

Formulas (I.65) and (I.66) were published by the author in the last section of [1], but unfortunately they contained a slight computational error. The formulas appearing above are correct.

I.5 CHECKING THE KLT OF THE SQUARE BROWNIAN MOTION BY MATLAB SIMULATIONS

Just look at Figure I.1.

I.6 REFERENCES

[1] C. Maccone, "The Karhunen–Loève Expansion of the Square of a Time-Rescaled Gaussian Process," *Bollettino dell'Unione Matematica Italiana*, Series 7, **2-A** (1988), 221–229.

[2] A. Papoulis, *Signal Analysis*, McGraw-Hill, New York, 1977.

[3] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 213–219.

[4] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1953.

[5] N. N. Lebedev, *Special Functions and Their Applications*, Dover, New York, 1972.

Appendix J

KLT of the $B^2(t^{2H})$ time-rescaled square Brownian motion

J.1 INTRODUCTION

Just as Appendix H showed the application of the general results obtained in Appendix G to the particular time-rescaled stochastic process $B(t^{2H})$, the present appendix investigates the application of the general results obtained in Appendix I about the square time-rescaled process $B^2(\ldots)$ to the particular process $B^2(t^{2H})$.

Before doing so, however, we regard it useful to review briefly the main results so far obtained. Consider then the general time-rescaled Brownian motion given by (F.40); that is,

$$X(t) = B\left(\int_0^t f^2(s)\, ds\right). \tag{J.1}$$

In Appendix G, as well as in [1], it was proved that the KL eigenfunction expansion of (J.1) reads

$$X(t) = \sum_{n=1}^{\infty} Z_n N_n \sqrt{f(t)\int_0^t f(s)\, ds} \cdot J_{\nu(t)}\left(\gamma_n \frac{\int_0^t f(s)\, ds}{\int_0^T f(s)\, ds}\right) \tag{J.2}$$

and converges in mean square, and uniformly in t, for $0 \le t \le T$.

Here

(1) The time-dependent order $\nu(t)$ of the Bessel function of the first kind $J_\nu(\ldots)$ is given by

$$\nu(t) = \sqrt{-\frac{\chi^3(t)}{f^2(t)}\frac{d}{dt}\left[\frac{\chi'(t)}{f^2(t)}\right]}, \tag{J.3}$$

where the auxiliary function $\chi(t)$ is defined by

$$\chi(t) = \sqrt{f(t)\int_0^t f(s)\,ds}. \tag{J.4}$$

Moreover, in Section G.4 as well as in [2], the following straightforward expression for $\nu(t)$, in terms of the time-rescaling function $f(t)$, was proved:

$$\nu(t) = \sqrt{\frac{1}{4} + \left[\frac{\int_0^t f(s)\,ds}{f(t)}\right]^2 \left\{\frac{3}{4}\left[\frac{d\ln f(t)}{dt}\right]^2 - \frac{1}{2}\frac{d^2\ln f(t)}{dt^2}\right\}}. \tag{J.5}$$

(2) The constants γ_n appearing in the argument of the Bessel function in (J.2) are the real positive zeros, arranged in ascending order of magnitude, of the equation

$$\chi'(T)\cdot J_{\nu(T)}(\gamma_n) + \chi(T)\left[\frac{f(T)\cdot\gamma_n}{\int_0^T f(s)\,ds}J'_{\nu(T)}(\gamma_n) + \frac{\partial J_{\nu(T)}(\gamma_n)}{\partial\nu}\nu'(T)\right] = 0. \tag{J.6}$$

In general, (J.6) cannot be solved for the γ_n analytically, and one has to do so numerically. However, some particular case of the time-rescaling function $f(t)$ may exist for which (J.6) can be solved analytically. The present chapter is devoted to one such important case.

(3) The normalization constants N_n follow from the normalization condition

$$N_n^2\left[\int_0^T f(s)\,ds\right]^2 \int_0^1 x[J_{\nu((x))}(\gamma_n x)]^2\,dx = 1, \tag{J.7}$$

where the new transformed order $\nu((x))$ is obtained from the order $\nu(t)$ of either (J.3) or (J.5) via the transformation

$$\int_0^t f(s)\,ds = x\int_0^T f(s)\,ds. \tag{J.8}$$

(4) The eigenvalues λ_n are determined by the γ_n (known from (J.6)) according to

$$\lambda_n = \left[\int_0^T f(s)\,ds\right]^2 \frac{1}{(\gamma_n)^2}. \tag{J.9}$$

(5) Finally, the Z_n are independent and orthogonal Gaussian random variables having mean zero and variance equal to the eigenvalues λ_n; that is,

$$E\{Z_m Z_n\} = \lambda_n\delta_{mn}. \tag{J.10}$$

Let us next turn to the further process $Y(t)$ defined by (I.1); that is,

$$Y(t) = X^2(t) - E\{X^2(t)\}. \tag{J.11}$$

It is natural to call this the zero-mean square process of the time rescaled Brownian motion given by (J.1), for (J.11) is just the square of (J.1) centered around the latter's mean value.

In Appendix I, as well as in [3], the author proved that the KL eigenfunction expansion of the process (J.11) reads:

$$Y(t) = \sum_{n=1}^{\infty} \tilde{Z}_n \tilde{N}_n \sqrt{\tilde{f}(t) \int_0^t \tilde{f}(s)\, ds} \cdot J_{\tilde{\nu}(t)}\left(\tilde{\gamma}_n \frac{\int_0^t \tilde{f}(s)\, ds}{\int_0^T \tilde{f}(s)\, ds} \right). \tag{J.12}$$

By checking (J.12) against (J.2), one sees that the KL expansion of the zero-mean square process $Y(t)$ is formally identical to the KL expansion of the original $X(t)$ process, with only two exceptions.

(1) The time-rescaling function $f(t)$ of the $X(t)$ is now replaced by the new function (no longer called "time rescaling") (for the proof, see (I.20) through (I.24))

$$\tilde{f}(t) = 2f(t)\sqrt{\int_0^t f^2(s)\, ds}. \tag{J.13}$$

(2) The gamma-type probability density

$$f_{\tilde{Z}_n}(z) = \frac{1}{\sqrt{\pi}(2\tilde{\lambda}_n)^{\frac{1}{4}}} e^{-\frac{z+\sqrt{\frac{\tilde{\lambda}_n}{2}}}{\sqrt{2\tilde{\lambda}_n}}} \left[z + \sqrt{\frac{\tilde{\lambda}_n}{2}} \right]^{-\frac{1}{2}} U\left(z + \sqrt{\frac{\tilde{\lambda}_n}{2}} \right), \tag{J.14}$$

where $U(x)$ is the unit step function, is followed by the random variables Z_n (for the proof, see (I.34)).

Consequently, we infer that

(1) the Bessel function of the first kind appearing in (J.12) has the time-dependent order

$$\tilde{\nu}(t) = \sqrt{-\frac{\tilde{\chi}^3(t)}{\tilde{f}^2(t)} \frac{d}{dt}\left[\frac{\tilde{\chi}'(t)}{\tilde{f}^2(t)} \right]}, \tag{J.15}$$

where the auxiliary function $\tilde{\chi}(t)$ has been defined by

$$\tilde{\chi}(t) = \sqrt{\tilde{f}(t) \int_0^t \tilde{f}(s)\, ds} \tag{J.16}$$

in analogy to (J.3) and (J.4), respectively. Alternatively, it is possible to express the order by virtue of a single formula, corresponding to (J.5)

$$\tilde{\nu}(t) = \sqrt{\frac{1}{4} + \left[\frac{\int_0^t \tilde{f}(s)\, ds}{\tilde{f}(t)} \right]^2 \left\{ \frac{3}{4}\left[\frac{d \ln \tilde{f}(t)}{dt} \right]^2 - \frac{1}{2} \frac{d^2 \ln \tilde{f}(t)}{dt^2} \right\}}. \tag{J.17}$$

(2) The real positive zeros, $\tilde{\gamma}_{n'}$, arranged in ascending order of magnitude, of the equation corresponding to (J.6), are defined by

$$\tilde{\chi}'(T) \cdot J_{\tilde{\nu}(T)}(\tilde{\gamma}_n) + \tilde{\chi}(T) \left[\frac{\tilde{f}(T) \cdot \tilde{\gamma}_n}{\int_0^T \tilde{f}(s)\, ds} J'_{\tilde{\nu}(T)}(\tilde{\gamma}_n) + \frac{\partial J_{\tilde{\nu}(T)}(\tilde{\gamma}_n)}{\partial \tilde{\nu}_t} \tilde{\nu}'(T) \right] = 0. \quad \text{(J.18)}$$

(3) The normalization condition, analogous to (J.7), reads

$$1 = \tilde{N}_n^2 \left[\int_0^T \tilde{f}(s)\, ds \right]^2 \int_0^1 x [J_{\tilde{\nu}((x))}(\tilde{\gamma}_n x)]^2\, dx \quad \text{(J.19)}$$

and defines the normalization constants $\tilde{N}_n$. In (J.19) the new transformed order $\tilde{\nu}((x))$ is obtained from the order $\tilde{\nu}(t)$ of either (J.15) or (J.17) via the transformation

$$\int_0^t \tilde{f}(s)\, ds = x \int_0^T \tilde{f}(s)\, ds. \quad \text{(J.20)}$$

(4) Finally, the eigenvalues are given by an expression analogous to (J.9); that is,

$$\tilde{\lambda}_n = \left[\int_0^T \tilde{f}(s)\, ds \right]^2 \frac{1}{(\tilde{\gamma}_n)^2}. \quad \text{(J.21)}$$

J.2 PREPARATORY CALCULATIONS ABOUT $B^2(t^{2\alpha+1})$

In the present section we pave the way to mathematically finding the KL eigenfunction expansion of the square process $B^2(t^{2\alpha+1})$.

Let the time-rescaling function $f(t)$ be a real-exponent power of time, multiplied by a generic real constant C

$$f(t) = Ct^{\alpha}. \quad \text{(J.22)}$$

The range of the exponent α is determined by the condition that the following pair of definite integrals, appearing in (J.2) and (J.1), respectively, must converge:

$$\int_0^t f(s)\, ds = \frac{C}{\alpha+1} t^{\alpha+1}, \quad \text{(J.23)}$$

$$\int_0^t f^2(s)\, ds = \frac{C^2}{2\alpha+1} t^{2\alpha+1}. \quad \text{(J.24)}$$

Evidently, the stricter condition on α is due to the convergence of (J.24); that is,

$$\alpha > -\frac{1}{2}. \quad \text{(J.25)}$$

Let us now go back to the problem of finding the KL eigenfunction expansion of the square process $B^2(t^{2\alpha+1})$. To this end, we must first form the new function defined by (J.13), which, by virtue of (J.1) and (J.3), turns out to be

$$\tilde{f}(t) = Kt^{2\alpha+\frac{1}{2}}, \tag{J.26}$$

where the new constant K is introduced to simplify things a little

$$K = \frac{2C^2}{\sqrt{2\alpha+1}}. \tag{J.27}$$

Our first task is to find the corresponding order of the Bessel function, alternatively defined by (J.15) plus (J.16), or by (J.17). Choosing the latter, we are led to compute the logarithm of (J.26)

$$\ln \tilde{f}(t) = \ln K + \left(2\alpha + \frac{1}{2}\right) \ln t \tag{J.28}$$

the logarithmic derivatives

$$\frac{d \ln \tilde{f}(t)}{dt} = \frac{\left(2\alpha + \frac{1}{2}\right)}{t}, \tag{J.29}$$

and

$$\frac{d^2 \ln \tilde{f}(t)}{dt^2} = -\frac{\left(2\alpha + \frac{1}{2}\right)}{t^2}, \tag{J.30}$$

and the integral

$$\int_0^t \tilde{f}(s)\, ds = \frac{K}{2\alpha + \frac{3}{2}} t^{2\alpha+\frac{3}{2}}. \tag{J.31}$$

Note that the condition (J.25) on α still holds, since the converge of (J.31) only requires that

$$\alpha > -\frac{3}{4}. \tag{J.32}$$

On substituting these results into the square of (J.17), the latter is turned into

$$\begin{aligned}\tilde{\nu}^2(t) &= \frac{1}{4} + \left[\frac{\frac{K}{2\alpha+\frac{3}{2}} t^{2\alpha+\frac{3}{2}}}{Kt^{2\alpha+\frac{1}{2}}}\right]^2 \cdot \left\{\frac{3}{4}\frac{\left(2\alpha+\frac{1}{2}\right)^2}{t^2} + \frac{1}{2}\frac{\left(2\alpha+\frac{1}{2}\right)}{t^2}\right\} \\ &= \frac{1}{4} + \frac{t^2}{\left(2\alpha+\frac{3}{2}\right)^2} \cdot \frac{\left(2\alpha+\frac{1}{2}\right)}{t^2} \left\{\frac{3}{4}\left(2\alpha+\frac{1}{2}\right) + \frac{1}{2}\right\}.\end{aligned} \tag{J.33}$$

The time t now cancels out in the expression, leaving

$$\tilde{\nu}^2(t) = \frac{1}{4} + \frac{\left(2\alpha + \frac{1}{2}\right)}{\left(2\alpha + \frac{3}{2}\right)^2} \frac{1}{4} \left\{ 3\left(2\alpha + \frac{1}{2}\right) + 2 \right\}$$

$$= \frac{\left(2\alpha + \frac{3}{2}\right)^2 + \left(2\alpha + \frac{1}{2}\right)\left\{6\alpha + \frac{3}{2} + 2\right\}}{4\left(2\alpha + \frac{3}{2}\right)^2} = \frac{(2\alpha + 1)^2}{\left(2\alpha + \frac{3}{2}\right)^2}.$$

That is,

$$\tilde{\nu} = \frac{2\alpha + 1}{2\alpha + \frac{3}{2}}, \tag{J.34}$$

with the immediate consequence

$$\tilde{\nu}'(t) = 0. \tag{J.35}$$

This circumstance helps simplify upcoming calculations considerably.

Let us now turn to finding the roots of (J.18) that define the constants $\tilde{\gamma}_n$. By virtue of (J.35), (J.18) takes the simpler form

$$\tilde{\chi}'(T) \cdot J_{\tilde{\nu}(T)}(\tilde{\gamma}_n) + \tilde{\chi}(T) \frac{\tilde{f}(T) \cdot \tilde{\gamma}_n}{\int_0^T \tilde{f}(s)\, ds} J'_{\tilde{\nu}(T)}(\tilde{\gamma}_n) = 0. \tag{J.36}$$

We next compute the function defined by (J.16)

$$\tilde{\chi}(t) = \sqrt{\tilde{f}(t) \int_0^t \tilde{f}(s)\, ds} = \frac{K}{\sqrt{2\alpha + \frac{3}{2}}} t^{2\alpha+1} \tag{J.37}$$

and its derivative

$$\tilde{\chi}'(t) = \frac{K(2\alpha + 1)}{\sqrt{2\alpha + \frac{3}{2}}} t^{2\alpha}. \tag{J.38}$$

Then, (J.38) and (J.37), after rearranging, change (J.15) into

$$\frac{2\alpha + 1}{2\alpha + \frac{3}{2}} J_{\tilde{\nu}}(\tilde{\gamma}_n) + \tilde{\gamma}_n J'_{\tilde{\nu}}(\tilde{\gamma}_n) = 0. \tag{J.39}$$

By virtue of (J.34), the above may be rewritten as

$$\tilde{\nu} J_{\tilde{\nu}}(\tilde{\gamma}_n) + \tilde{\gamma}_n J'_{\tilde{\nu}}(\tilde{\gamma}_n) = 0. \tag{J.40}$$

This form of the equation corresponds to the Bessel functions property appearing in [4, p. 12, entry (55)]. Since zero may not be a root of (J.18), this amounts to

$$J_{\tilde{\nu}-1}(\tilde{\gamma}_n) = 0. \tag{J.41}$$

The meaning of the set of constants $\tilde{\gamma}_n$ has now been found: these are the (infinite) real positive zeros, arranged in ascending order of magnitude, of the Bessel function of the first kind and order given by

$$\tilde{\nu} - 1 = -\frac{1}{4\alpha + 3}. \tag{J.42}$$

No analytic formula explicitly yielding these zeros is known. However, a good approximated expression for them may be found by resorting to the asymptotic (for $x \to \infty$) expansion for the Bessel function of the first kind (see [4, p. 134]), which reads

$$J_\nu(x) \to \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\pi\nu}{2} + \frac{\pi}{4}\right). \tag{J.43}$$

In fact, by replacing (J.42), (J.43) takes the form

$$J_{-\frac{1}{4\alpha+3}}(\tilde{\gamma}_n) \to \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} \cos\left(\tilde{\gamma}_n - \frac{\pi}{2}\left[-\frac{1}{4\alpha+3}\right] - \frac{\pi}{4}\right). \tag{J.44}$$

It then follows from (J.41) that:

$$\tilde{\gamma}_n - \frac{\pi}{2}\left[-\frac{1}{4\alpha+3}\right] - \frac{\pi}{4} \approx n\pi - \frac{\pi}{2} \tag{J.45}$$

and finally

$$\tilde{\gamma}_n \approx \pi\left[n - \frac{1}{4} - \frac{1}{2(4\alpha+3)}\right] \quad \text{valid for } n = 1, 2, \ldots. \tag{J.46}$$

Next, we want to determine the normalization constants defined by (J.19); that is,

$$1 = \tilde{N}_n^2 \left[\int_0^T \tilde{f}(s)\, ds\right]^2 \int_0^1 x[J_{\tilde{\nu}((x))}(\tilde{\gamma}_n x)]^2\, dx. \tag{J.47}$$

Let us now replace (J.26) and the integral that is calculated within the framework of the Dini expansion in a series of Bessel functions (see [4, p. 71]); (J.47) thus becomes

$$1 = \tilde{N}_n^2 \left[\frac{KT^{2\alpha+\frac{3}{2}}}{2\alpha + \dfrac{3}{2}}\right]^2 \left\{\frac{1}{2}\left[J_{\tilde{\nu}}'^2(\tilde{\gamma}_n) + \left(1 - \frac{\tilde{\nu}^2}{\tilde{\gamma}^2}\right) J_{\tilde{\nu}}^2(\tilde{\gamma}_n)\right]\right\}.$$

Now, by virtue of (J.40), the derivative of the Bessel function disappears, and one gets

$$1 = \tilde{N}_n^2 \left[\frac{KT^{2\alpha+\frac{3}{2}}}{2\alpha + \frac{3}{2}}\right]^2 \left\{\frac{1}{2} J_{\tilde{\nu}}^2(\tilde{\gamma}_n)\right\}$$

or

$$\tilde{N}_n = \frac{\sqrt{2}\left(2\alpha + \frac{3}{2}\right)}{KT^{2\alpha+\frac{3}{2}}} \frac{1}{|J_{\tilde{\nu}}(\tilde{\gamma}_n)|} \tag{J.48}$$

as the exact expression of the normalization constants.

A well-approximated expression of the normalization constants (J.48) may be obtained by first working out the asymptotic expression—via (J.43)—for the Bessel function appearing in the denominator of (J.48). To this end, (J.43) and (J.34) yield

$$|J_{\tilde{\nu}}(\tilde{\gamma}_n)| \approx \left| \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} \cos\left(\tilde{\gamma}_n - \frac{\pi}{2}\left[\frac{2\alpha+1}{2\alpha+\frac{3}{2}}\right] - \frac{\pi}{4} \right) \right| \tag{J.49}$$

and the approximated formula (J.46) produces

$$\begin{aligned} |J_{\tilde{\nu}}(\tilde{\gamma}_n)| &\approx \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} \left| \cos\left(\pi\left[n - \frac{1}{4} - \frac{1}{2(4\alpha+3)}\right] - \frac{\pi}{2}\left[\frac{2\alpha+1}{2\alpha+\frac{3}{2}}\right] - \frac{\pi}{4} \right) \right| \\ &= \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} \left| \cos\left(\pi\left[n - \frac{1}{2} - \frac{1+2(2\alpha+1)}{2(4\alpha+3)}\right] \right) \right| \\ &= \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} |\cos(\pi[n-1])| \\ &= \sqrt{\frac{2}{\pi\tilde{\gamma}_n}} |(-1)^{n+1}| = \sqrt{\frac{2}{\pi\tilde{\gamma}_n}}. \end{aligned} \tag{J.50}$$

By substituting (J.50) and (J.46) into (J.48), the desired approximated expression for the normalization constants becomes

$$\begin{aligned} \tilde{N}_n &\approx \frac{\sqrt{2}\left(2\alpha + \frac{3}{2}\right)}{KT^{2\alpha+\frac{3}{2}}} \cdot \frac{\sqrt{\pi}\sqrt{\tilde{\gamma}_n}}{\sqrt{2}} \\ &= \frac{\pi\left(2\alpha + \frac{3}{2}\right)}{KT^{2\alpha+\frac{3}{2}}} \sqrt{n - \frac{1}{4} - \frac{1}{2(4\alpha+3)}}. \end{aligned} \tag{J.51}$$

Similarly, (J.21) yields the exact eigenvalues

$$\tilde{\lambda}_n = \frac{K^2 T^{4\alpha+3}}{\left(2\alpha + \frac{3}{2}\right)^2} \cdot \frac{1}{(\tilde{\gamma}_n)^2}, \tag{J.52}$$

while (J.46) gives the approximate expression

$$\tilde{\lambda}_n \approx \frac{K^2 T^{4\alpha+3}}{\left(2\alpha + \frac{3}{2}\right)^2} \cdot \frac{1}{\pi^2 \left(n - \frac{1}{4} - \frac{1}{2(4\alpha+3)}\right)^2}. \tag{J.53}$$

In conclusion, using (J.48), (J.37) and (J.34), we have proven that the exact expression of the KL eigenfunctions reads

$$\tilde{\phi}_n(t) = \tilde{N}_n \cdot \tilde{\chi}(t) \cdot J_{\tilde{\nu}}\left(\tilde{\gamma}_n \frac{t^{2\alpha+\frac{3}{2}}}{T^{2\alpha+\frac{3}{2}}}\right) = \frac{\sqrt{2}\sqrt{2\alpha + \frac{3}{2}}\, t^{2\alpha+1}}{T^{2\alpha+\frac{3}{2}} |J_{\tilde{\nu}}(\tilde{\gamma}_n)|} J_{\tilde{\nu}}\left(\tilde{\gamma}_n \frac{t^{2\alpha+\frac{3}{2}}}{T^{2\alpha+\frac{3}{2}}}\right). \tag{J.54}$$

The approximated counterparts to (J.54) may be obtained by resorting to the exact expression (J.37), to the approximated expression (J.51) for the normalization constants, to the approximated (J.50) for the Bessel function at the denominator, and finally to (J.43) for the asymptotic expansion of the Bessel function of the first kind. The result is

$$\begin{aligned} \tilde{\phi}_n(t) &\approx \frac{\sqrt{4\alpha+3}\, t^{2\alpha+1}}{T^{2\alpha+\frac{3}{2}} \sqrt{\frac{2}{\pi \tilde{\gamma}_n}}} \cdot \sqrt{\frac{2}{\pi\left(\tilde{\gamma}_n \frac{t^{2\alpha+\frac{3}{2}}}{T^{2\alpha+\frac{3}{2}}}\right)}} \cdot \cos\left(\tilde{\gamma}_n \frac{t^{2\alpha+\frac{3}{2}}}{T^{2\alpha+\frac{3}{2}}} - \frac{\pi\tilde{\nu}}{2} - \frac{\pi}{4}\right) \\ &\approx \sqrt{4\alpha+3}\, \frac{t^{\alpha+\frac{1}{4}}}{T^{\alpha+\frac{3}{4}}} \cos\left(\tilde{\gamma}_n \frac{t^{2\alpha+\frac{3}{2}}}{T^{2\alpha+\frac{3}{2}}} - \frac{\pi(2\alpha+1)}{4\alpha+3} - \frac{\pi}{4}\right), \end{aligned} \tag{J.55}$$

where (J.34) was used to replace the order in the last expression. Finally, using (J.46) we get

$$\tilde{\phi}_n(t) \approx \sqrt{4\alpha+3}\, \frac{t^{\alpha+\frac{1}{4}}}{T^{\alpha+\frac{3}{4}}} \cdot \cos\left(\pi\left[\left(n - \frac{1}{4} - \frac{1}{2(4\alpha+3)}\right) \frac{t^{2\alpha+\frac{3}{2}}}{T^{2\alpha+\frac{3}{2}}} - \frac{12\alpha+7}{4(4\alpha+3)}\right]\right) \tag{J.56}$$

as the approximate expression for the KL orthonormalized eigenfunctions.

J.3 KL EXPANSION OF THE SQUARE PROCESS $B^2(t^{2H})$

In the present section, we derive the full expression for the KL expansion of the $B^2(t^{2H})$ process by resorting to the results obtained in Section J.2. The notation in this section is thus consistent with that adopted in Appendix H for the $B(t^{2H})$ process, and this will allow useful comparisons to be made, particularly with regard to the self-similarity of both processes.

We start by setting

$$2\alpha + 1 = 2H \tag{J.57}$$

from which

$$\alpha = H - \frac{1}{2}. \tag{J.58}$$

The range (J.25) is then replaced by the new range

$$H > 0. \tag{J.59}$$

The advantage of this notation lies in that the Gaussian process

$$X(t) = B\left(\int_0^t f^2(s)\,ds\right) = B\left(\frac{C^2}{2H}t^{2H}\right) \tag{J.60}$$

now plainly reveals its H-self-similarity. By this, we mean the $\frac{1}{2}$-self-similarity of the Brownian motion expressed by the formula (F.6); that is,

$$B(c \cdot t) = \sqrt{c} \cdot B(t) \tag{J.61}$$

(valid for any real positive constant c) is carried over to the process $X(t)$ by way of the generalization

$$X(c \cdot t) = c^H \cdot X(t). \tag{J.62}$$

The proof of this fact immediately follows from (J.60) and (J.61)

$$X(c \cdot t) = B\left(\frac{C^2}{2H} \cdot c^{2H} \cdot t^{2H}\right) = c^H \cdot B\left(\frac{C^2}{2H}t^{2H}\right) = c^H \cdot X(t). \tag{J.63}$$

Since the publication of Benoit Mandelbrot's book about fractals [5], the importance of H-self-similarity (J.63) is apparent. An investigation about a possible relationship between the processes studied in the present chapter and Mandelbrot's fractional Brownian motions deserves much deeper investigation. We confine ourselves to pointing out the further $2H$-self-similarity fulfilled by the $Y(t)$ process of (J.11)

$$Y(c \cdot t) = c^{2H} \cdot Y(t). \tag{J.64}$$

The relevant proof follows at once from (J.11) and (J.62):

$$\begin{aligned} Y(c \cdot t) &= X^2(c \cdot t) - E\{X^2(c \cdot t)\} = c^{2H}X^2(t) - E\{c^{2H}X^2(t)\} \\ &= c^{2H}[X^2(t) - E\{X^2(t)\}] = c^{2H}Y(t). \end{aligned} \tag{J.65}$$

Let us now turn to the KL expansion of the $Y(t)$ process. The autocorrelation of the $X(t)$ process is given by (F.30), that is

$$E\{X(t_1)X(t_2)\} = \int_0^{t_1 \wedge t_2} f^2(s)\,ds. \tag{J.66}$$

Upon setting $t_1 = t_2 = t$ in (J.66), the obvious formula $t \wedge t = t$, (J.22) and (J.24) yield

$$E\{X^2(t)\} = \int_0^t f^2(s)\,ds = \frac{C^2}{2H}t^{2H}. \tag{J.67}$$

Therefore, the definition (J.11) of $Y(t)$, by virtue of (J.60) and (J.67), becomes

$$Y(t) = X^2(t) - E\{X^2(t)\} = \frac{C^2}{2H}[B^2(t^{2H}) - t^{2H}]. \tag{J.68}$$

Since the KL expansion (J.12) has the form

$$Y(t) = \sum_{n=1}^{\infty} \tilde{Z}_n \tilde{\phi}_n(t) \tag{J.69}$$

from (J.68) and (J.69) the following KL expansion for the process $B^2(t^{2H})$ is inferred:

$$B^2(t^{2H}) - t^{2H} = \frac{2H}{C^2}\sum_{n=1}^{\infty} \tilde{Z}_n \tilde{\phi}_n(t). \tag{J.70}$$

Inserting now the exact orthonormalized eigenfunctions (J.54), rewritten by aid of the substitution (J.22), we get the explicit KL expansion

$$B^2(t^{2H}) - t^{2H} = \frac{2H}{C^2}\sum_{n=1}^{\infty} \tilde{Z}_n \frac{\sqrt{4H+1}\,t^{2H}}{T^{2H+\frac{1}{2}}|J_{\tilde{\nu}}(\tilde{\gamma}_n)|} J_{\tilde{\nu}}\left(\tilde{\gamma}_n \frac{t^{2H+\frac{1}{2}}}{T^{2H+\frac{1}{2}}}\right). \tag{J.71}$$

Here

(1) The order of the Bessel function of the first kind is constant in time, and reads

$$\tilde{\nu} = \frac{2H}{2H + \dfrac{1}{2}}. \tag{J.72}$$

(2) The set of constants $\tilde{\gamma}_n$ are defined as the real positive zeros of the Bessel function of the first kind and of order given by (J.34) minus 1; that is,

$$J_{\tilde{\nu}-1}(\tilde{\gamma}_n) = 0. \tag{J.73}$$

The approximate expression is derived by substituting the approximated orthonormalized eigenfunctions (J.56) into (J.70):

$$B^2(t^{2H}) - t^{2H} \approx \frac{2H}{C^2}\sum_{n=1}^{\infty} \tilde{Z}_n \sqrt{4H+1}\,\frac{t^{H-\frac{1}{4}}}{T^{H+\frac{1}{4}}} \cdot \cos\left(\pi\left[\left(n - \frac{1}{4} - \frac{1}{2(4H+1)}\right)\frac{t^{2H+\frac{1}{2}}}{T^{2H+\frac{1}{2}}} - \frac{12H+1}{4(4H+1)}\right]\right). \tag{J.74}$$

J.4 CHECKING THE KLT OF $B^2(t^{2H})$ BY MATLAB SIMULATIONS

Just look at Figure J.1.

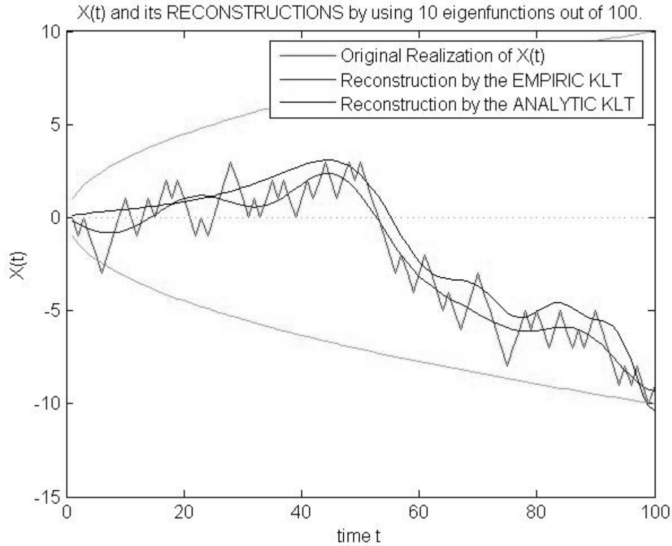


Figure J.1. The *zero-mean* time-rescaled square Brownian motion $B^2(t^{2H}) - t^{2H} = X(t)$ vs. time t simulated as a random walk over 100 time instants. Next to the "bumpy curve" of $X(t)$, two more "smooth curves" are shown that *interpolate at best* the bumpy $X(t)$. These two curves are the KLT reconstruction of $X(t)$) *by using the first ten eigenfunctions only*. It is important to note that the two smooth curves are *different* in this case because the KLT expansion (J.74) is *approximated*. Actually, it is an approximated KLT expansion because the asymptotic expansion of the Bessel functions (J.43) was used. So, the two curves are different from each other, but both still interpolate $X(t)$ at best. Note that were we taking into account the full set of 100 KLT eigenfuctions, rather than just 10, then the *empiric* reconstruction would overlap $X(t)$ *exactly*, but the *analytic* reconstruction would not because of the use of the asymptotic expansion (J.43) of the Bessel functions.

J.5 REFERENCES

[1] C. Maccone, "Eigenfunctions and Energy for Time-Rescaled Gaussian Processes," *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A** (1984), 213–219.

[2] C. Maccone, "Special Relativity and the Karhunen–Loève Expansion of Brownian Motion," *Il Nuovo Cimento*, **100-B** (1987), 329–341.

[3] C. Maccone, "The Karhunen–Loève Expansion of the Square of a Time-Rescaled Gaussian Process," *Bollettino dell'Unione Matematica Italiana*, Series 7, **2-A** (1988), 221–229.

[4] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. 2, McGraw-Hill, New York, 1953.

[5] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.

Appendix K

A Matlab code for KLT simulations

K.1 INTRODUCTION

After so much mathematics, it is natural to think of some computer code capable of simulating the KLTs derived analytically in this book.

The well-known Matlab environment is well-suited for such simulations inasmuch as it can handle both the eigenvalues and the eigenvectors of symmetric matrices that are at the heart of the KLT.

But this author is hardly an expert in Matlab programming! So he turned to one of his pupils, Dr. Nicolò Antonietti, and together they wrote the set of Matlab routines described in this appendix. This does not mean that such routines are "optimized", nor even that they are error-free! Thus, readers of this book might wish to improve on our work, and we would be most grateful if they could let us have their new Matlab codes by sending them to the author's e-mail address: *clmaccon@libero.it*. Thanks!

K.2 THE MAIN FILE "STANDARD_BROWNIAN_MOTION_MAIN.M"

The main file of our Matlab 7.1 set of routines is called "Standard_Brownian_Motion_MAIN.m" and is listed hereafter. It has plenty of comments, so our reader should not have difficulties in following what is going on.

```
% Matlab 7.1 Code for the SIMULATION OF THE KLT.
% Authors: Drs. Nicolò Antonietti & Claudio Maccone.
% This version was completed in January 2008.

% Clear the memory & worksheet.
clear all, close all, clc
```

```
% Define the initial value of the loop variable "runtime" as the
% initial cpu time. At the end of this simulation, the runtime will
% be equal to the final cpu time minus the initial cpu time.
run_time = cputime;

% Do you prefer to run a new simulation (creating a new and different
% realization of the stochastic process X(t)) or do you wish to load an
% existing data file (produced by a previous simulation) ?
[t, n, Input_Process_data, flag] = input_data_toggle;

% Decide how many eigenfuctions in the KL expansion you wish to take into
% account for the reconstruction of the process X(t) in the time interval
% between 0 and T. Clearly, the number of eigenfunctions taken into account
% is at most equal to the number of instants considered in the simulation.
% In practice, however, you may wish to use FEWER eigenfunctions, or even
% just VERY FEW eigenfuctions. The reconstruction of X(t) will thus be
% rougher and rougher, but the computation burden will still be affordable
% by your machine. THIS IS THE TRADE-OFF that the KLT offers to you as
% A LOSSY COMPRESSION ALGORITHM.
How_many_eigenfunctions = input('How many eigenfunctions ? \n');

% Computation of the ANALYTIC autocorrelation matrix of the Brownian
% motion, defined as min(t1, t2). This autocorrelation matrix is fed into
% the code only if you previously selected to run an entirely new
% simulation. If you previously selected to load a pre-existing data file
% (as it happens in all EXPERIMENTAL applications of the KLT), then the
% data file of the values of X(t), for t ranging between 0 and T, is fed
% into the code.
% The autocorrelation of the Brownian motion of size n is defined as
% min(t1, t2) by the function (i.e. by the subroutine)
% Brownian_Autocorrelation(n), hereby called by the Main code.
Autocorrelation_matrix = Brownian_Autocorrelation(t);

% The next step is the most important step in this Main code.
% By virtue of the "eigs" subroutine of Matlab, we avoid getting entangled
% in the computation of the eigenvalues Lambda and of the eigenfunctions Phi
% of the KLT. Quite simply, we feed in the Autocorrelation matrix (whether
% it was ANALYTIC or NUMERIC = EXPERIMENTAL) and "eigs" returns both Lambda
% and Phi! Clearly, in non-Matlab simulations, this "eigs" routine must be
% very carefully written!
[Phi,Lambda] = eigs(Autocorrelation_matrix,How_many_eigenfunctions);

% We now compute the EMPIRIC KLT (as opposed to the ANALYTIC KLT derived in
% the book analytically) for the simulation of X(t) under consideration.
%
% This EMPIRIC KLT we obtain in the following LOOP by:
%
% 1) PROJECTING the vector of the Input_Process_data (i.e. the vector
% representing the stochastic process X(t) to be KL-expanded) ONTO THE
% RELEVANT i-th EIGENVECTOR Phi(i). THIS PROJECTION IS THE RANDOM VARIABLE
% Z(i) of the KL expansion (as it follows by INVERTING the KL expansion,
% just as one does for the Fourier series).
%
% 2) DEFINING the ith term of the KL expansion as the product of Z(i) times
% Phi(i).
%

for i = 1:How_many_eigenfunctions,
  Z(i) = Input_Process_data.' * Phi(:,i);
  KL_EXPANSIONs_ith_term(:,i) = Z(i) * Phi(:,i);
end
```

```
% We now create the DATA VECTOR of the EMPIRIC RECONSTRUCTION of X(t)
% achieved by the KLT numerically. This is simply the sum of all the
% KL_EXPANSIONs_ith_term obtained in the previous step of this Main code.
% EMPIRIC_Data_vector = sum(KL_EXPANSIONs_ith_term,2);
EMPIRIC_Data_vector = sum(KL_EXPANSIONs_ith_term,2);

% Next we create the DATA VECTOR of the ANALYTIC RECONSTRUCTION of X(t) as
% given by the formulae mathematically demonstrated in the book.
% This requires a separate routine (named hereafter ANALYTIC_KLT) to be
% called up by this Main code. The text of this routine clearly changed
% according to which formula in the book we refer to.
if flag == 1
  [ANALYTIC_Data_vector] = ANALYTIC_KLT(Input_Process_data, n,
How_many_eigenfunctions, t, Lambda);
elseif flag == 2
  [ANALYTIC_Data_vector] = ANALYTIC_KLT_decelerated(Input_Process_data, n,
How_many_eigenfunctions, t, Lambda);
elseif flag == 3
  [ANALYTIC_Data_vector] = ANALYTIC_KLT_square_brow_motion(Input_Process_data, n,
How_many_eigenfunctions, t, Lambda);
elseif flag == 4
  [ANALYTIC_Data_vector] =
ANALYTIC_KLT_square_brow_dec_motion(Input_Process_data, n,
How_many_eigenfunctions, t, Lambda);
elseif flag == 5
  [ANALYTIC_Data_vector, t] = ANALYTIC_KLT_uniform_rel(Input_Process_data, n,
How_many_eigenfunctions, t, Lambda);
end

% Plot the EIGENVALUES of the EMPIRIC RECONSTRUCTION of X(t).
h2 = figure;

% Plots of:
% 1) The ORIGINAL REALIZATION of X(t).
% 2) The EMPIRIC RECONSTRUCTION OF X(t) by the KLT.
% 3) The ANALYTIC RECONSTRUCTION OF X(t) by the KLT.
graphic(Input_Process_data, EMPIRIC_Data_vector, ANALYTIC_Data_vector, flag,
How_many_eigenfunctions, t, n)

% Save the Input_Process_data as the Matlab file i01mat.
save i01 Input_Process_data

% How long it took to do all these calculations.
run_time = cputime - run_time
```

K.3 THE FILE "INPUT_DATA_TOGGLE.M"

The file "input_data_toggle.m" allows the user to choose which stochastic process to select for the KLT computation. Here is this file's listing.

```
% This subroutine allows you either to:
% 1) Create a brand-new REALIZATION of the input stochastic process X(t) or
% 3) Load an existing matlab file (.mat) where the input variables are
% saved
% 2) Load an existing file with all the numeric data of the input
% stochastic process. Clearly, this arises when you do EXPERIMENTAL work, %
such as getting the input of a radiotelescope, etc.
```

```
function [time_range,final_instant_T,process_values_vector, flag] = input_data_toggle

while true,
  flag = input('What process is going to be analized? \n 1. A standard brownian motion
from a still source \n 2. A standard brownian motion from a source in a decelerated
motion \n 3. A square standard brownian motion from a still source \n 4. A square
standard brownian motion from a source in decelerated motion \n 5. A uniform
relativistic motion \n');
  if flag==1 | flag==2 | flag==3 | flag==4 | flag==5 |flag==6
    break
  end
end

while true,
  case_number = input(' Enter 1 to create a NEW REALIZATION of the Brownian
motion X(t).\n Enter 2 to load an existing Brownian motion matlab file (.mat). \n Enter
3 to load an existing Brownian motion data file. \n');
  if case_number==1 | case_number==2 | case_number==3
    break
  end
end
  if case_number == 1
    % Creating the NEW REALIZATION of the stochastic process X(t)
    final_instant_T = input('Please, type the final time unit. (Suggested: no more than
1000) \n');
    time_range = (1:1:final_instant_T)';
    process_values_vector = process_path(final_instant_T);
    % Plot the Original Realization of X(t) to be later expanded and
    % reconstructed by virtue of the KLT.
    t = [0; time_range];
    random_walk = [0, process_values_vector];
    h0 = figure;
    parabola = sqrt(t);
    plot(t, random_walk,'-k', t,parabola,'-k', t,-parabola,'-k', t,0,'-k'), title(['REALIZATION of
B(t) over ', num2str(final_instant_T), ' time instants.']), xlabel('time t'), ylabel('X(t)')
    process_values_vector = process_values_vector';

    % Loading an existing matlab file (.mat) with its input stochastic
    % process.
  elseif case_number == 2
    nome = input('Please TYPE the full path and file name.\n','s');
    load(nome);
    process_values_vector = Input_Process_data;
    final_instant_T = length(Input_Process_data);
    time_range = (1:1:final_instant_T)';

    % Plot the Original Realization of X(t) to be later expanded and
    % reconstructed by virtue of the KLT.
    t = [0; time_range];
    random_walk = [0, process_values_vector'];
    h0 = figure;
    parabola = sqrt(t);
    plot(t, random_walk,'-k', t,parabola,'-k', t,-parabola,'-k', t,0,'-k'), title(['REALIZATION of
B(t) over ', num2str(final_instant_T), ' time instants.']), xlabel('time t'), ylabel('X(t)')

    % Loading an existing EXPERIMENTALLY OBTAINED input stochastic
    % process.
  elseif case_number == 3
    nome = input('Please TYPE the full path and file name.\n','s');
    fid = fopen(nome);
    A = fscanf(fid, '%4d %4d', [2, inf]);
```

```
    A = A';
    time_range = A(:,1);
    final_instant_T = length(time_range);
    process_values_vector = A(:,2);
    fclose(fid);

    % Plot the Original Realization of X(t) to be later expanded and
    % reconstructed by virtue of the KLT.
    t = [0; time_range];
    random_walk = [0, process_values_vector'];
    h0 = figure;
    parabola = sqrt(t);
    plot(t, random_walk,'-k', t,parabola,'-k', t,-parabola,'-k', t,0,'-k'), title(['REALIZATION of
B(t) over ', num2str(final_instant_T), ' time instants.']), xlabel('time t'), ylabel('X(t)')
  end

  if flag == 1
    process_values_vector = process_values_vector;

  elseif flag == 2
    process_values_vector = process_values_vector;

  elseif flag == 3
    process_values_vector = process_values_vector.^2 - time_range;

  elseif flag == 4
    process_values_vector = process_values_vector.^2;

  elseif flag == 5
    process_values_vector = process_values_vector;

  end
```

K.4 THE FILE "BROWNIAN_AUTOCORRELATION.M"

The file "Brownian_Autocorrelation.m" simply translates the Brownian motion autocorrelation formula (i.e., the minimum—smallest—of t_1 and t_2) into a Matlab file ready for further applications. Please notice the "lucky circumstance" that the Brownian motion autocorrelation is known in its analytical form, rather than in some purely numerical form. It is by virtue of this analytical form, coded in the routine below, that all the KLT simulations described in this appendix can be performed.

```
% This subroutine computes the AUTOCORRELATION of the Brownian motion B(t)
% by translating its analytical definition min(t1, t2) into a numeric
% matrix. The entries of such a matrix are each the MINIMUM between the
% relevant row and column numbers.

function C = corr_brow(t)

n = length(t)
C = zeros(n);

for row = 1:n,
  for column = 1:n,
    C(row, column) = min(t(row),t(column));
  end
end
```

K.5 THE FILE "PROCESS_PATH.M"

The subroutine "process_path.m" creates the Brownian motion RANDOM WALK. In other words, this subroutine adds $+1$ to or subtracts -1 from the value of the $B(t)$ stochastic process at every new instant t. Its Matlab listing is the following.

```
% Subroutine creating the RANDOM WALK path of the Brownian motion B(t).
function X = process_path(T)

% Set to zero all the initial values of the T-element vector B that will
% contain the (random) values of the X(t) process when the new realization
% of X(t) will have been computed.
X = zeros(1,T);

% Create a vector with random entries and as many elements as are the time
% instants between 1 and the final instant T.
random_vector = rand(1,T);

for i = 1:T
  while random_vector(i) == 0.5
    random_vector(i) = rand(1);
  end

  if random_vector(i) < 0.5
    Increment(i) = - 1;
  elseif random_vector(i) > 0.5
    Increment(i) = + 1;
  end

  if i == 1
    X(i) = 0 + Increment(i);
  else
    X(i) = X(i-1) + Increment(i);
  end
end
```

K.6 THE FILE "GRAPHIC.M"

The subroutine "graphic.m" provides all the graphic instructions enabling Matlab to DRAW the plot of the selected stochastic process as a function of time. There are basically THREE "curves" for each realization of the Brownian motion—or time-rescaled Brownian motion—drawn by our Matlab code:

(1) The original "peaky" Brownian motion "curve", which is the actual Brownian motion realization.
(2) Its EMPIRIC RECONSTRUCTION, performed by the KLT, BY USING ONLY A FEW (OR EVEN ALL) THE KLT EIGENFUNCTIONS.
(3) Its ANALYTIC RECONSTRUCTION, performed by the KLT, BY USING ONLY A FEW (OR EVEN ALL) THE KLT EIGENFUNCTIONS.

In this way, we may clearly see the extent to which the two KLT RECONSTRUCTIONS actually FIT the original Brownian motion process.

This is one of the neater results provided by our simulations code, inasmuch as it "proves" in a neat, graphical way, "how good" the KLT is according to the number of eigenfunctions that we wish to take into account.

```
function graphic(Input_Process_data, EMPIRIC_Data_vector, ANALYTIC_Data_vector, flag,
How_many_eigenfunctions, t, N)

if flag == 1
t = [0; t];
Input_Process_data = [0; Input_Process_data];
EMPIRIC_Data_vector = [0; EMPIRIC_Data_vector];
ANALYTIC_Data_vector = [0; ANALYTIC_Data_vector];

plot(t, Input_Process_data,'-k', t ,EMPIRIC_Data_vector,'-k', t, ANALYTIC_Data_vector,':k'),

title(['B(t) and its RECONSTRUCTIONS by using ', num2str(How_many_eigenfunctions), '
eigenfunctions out of ', num2str(N), '.']),
xlabel('time t'), ylabel('B(t)'), legend('Original Realization of B(t)', 'Reconstruction by the
EMPIRIC KLT', 'Reconstruction by the ANALYTIC KLT')
hold on
parabola = sqrt(t);
plot(t, parabola,'-k', t, -parabola,'-k', t, 0, '-k')
hold off

elseif flag == 2
t = [0; t];
Input_Process_data = [0; Input_Process_data];
EMPIRIC_Data_vector = [0; EMPIRIC_Data_vector];
ANALYTIC_Data_vector = [0; ANALYTIC_Data_vector];

plot(t, Input_Process_data,'-k', t ,EMPIRIC_Data_vector,'-k', t, ANALYTIC_Data_vector,':k'),

title(['B(t^{2H}) and its RECONSTRUCTIONS by using ',
num2str(How_many_eigenfunctions), ' eigenfunctions out of ', num2str(N), '.']),
xlabel('time t'), ylabel('B(t^{2H})'), legend('Original Realization of B(t^{2H})',
'Reconstruction by the EMPIRIC KLT', 'Reconstruction by the ANALYTIC KLT')
hold on
parabola = sqrt(t);
plot(t, parabola, '-k', t, -parabola, '-k', t, 0, '-k')
hold off

elseif flag == 3
t = [0; t];
Input_Process_data = [0; Input_Process_data];
EMPIRIC_Data_vector = [0; EMPIRIC_Data_vector];
ANALYTIC_Data_vector = [0; ANALYTIC_Data_vector];

plot(t, Input_Process_data,'-k', t ,EMPIRIC_Data_vector,'-k', t, ANALYTIC_Data_vector,':k'),

title(['B^2(t)-t and its RECONSTRUCTIONS by using ',
num2str(How_many_eigenfunctions), ' eigenfunctions out of ', num2str(N), '.']),
xlabel('time t'), ylabel('B^2(t)-t'), legend('Original Realization of B^2(t)-t', 'Reconstruction
by the EMPIRIC KLT', 'Reconstruction by the ANALYTIC KLT')
hold on
parabola = sqrt(t);
plot(t, parabola, '-k', t, -parabola, '-k', t, 0, '-k')
hold off
```

```
elseif flag == 4
t = [0; t];
Input_Process_data = [0; Input_Process_data];
EMPIRIC_Data_vector = [0; EMPIRIC_Data_vector];
ANALYTIC_Data_vector = [0; ANALYTIC_Data_vector];

plot(t, Input_Process_data,'-k', t ,EMPIRIC_Data_vector,'-k', t, ANALYTIC_Data_vector,':k'),

title(['B^2(t^{2H}) and its RECONSTRUCTIONS by using ',
num2str(How_many_eigenfunctions), ' eigenfunctions out of ', num2str(N), '.']),
xlabel('time t'), ylabel('B^2(t^{2H})'), legend('Original Realization of B^2(t^{2H})',
'Reconstruction by the EMPIRIC KLT', 'Reconstruction by the ANALYTIC KLT')
hold on
parabola = sqrt(t);
plot(t, parabola, '-k', t, -parabola, '-k', t, 0, '-k')
hold off

elseif flag == 5
t = [0; t];
Input_Process_data = [0; Input_Process_data];
EMPIRIC_Data_vector = [0; EMPIRIC_Data_vector];
ANALYTIC_Data_vector = [0; ANALYTIC_Data_vector];

plot(t, Input_Process_data,'-k', t ,EMPIRIC_Data_vector,'-k', t, ANALYTIC_Data_vector,':k'),

title(['B(\tau) and its RECONSTRUCTIONS by using ',
num2str(How_many_eigenfunctions), ' eigenfunctions out of ', num2str(N), '.']),
xlabel('time \tau'), ylabel('B(\tau)'), legend('Original Realization of B(\tau)',
'Reconstruction by the EMPIRIC KLT', 'Reconstruction by the ANALYTIC KLT')
hold on
parabola = sqrt(t);
plot(t, parabola, '-k', t, -parabola, '-k', t, 0, '-k')
hold off

end
```

K.7 THE FILE "ANALYTIC_KLT.M"

The file "ANALYTIC_KLT.m" finds the KLT of the ordinary Brownian motion $B(t)$ as described and proven in Section F.3. We recall here that this is an EXACT ANALYTICAL solution of the KLT integral equation (10.18). Therefore, the reconstruction of $B(t)$ by the EMPIRIC KLT and the reconstruction of $B(t)$ by the ANALYTIC KLT are exactly the same thing; that is, the two reconstructed curves just OVERLAP EXACTLY in the $B(t)$ plot. This happens independently of the number of eigenfunctions that we decide to take into account for the reconstruction of $B(t)$.

This subroutine's listing is as follows.

```
% This subroutine computes the ANALYTIC RECONSTRUCTION of X(t) according to
% the analytic (either exact or approximated formulae given in the book.
function [ANALYTIC_Data_vector] = ANALYTIC_KLT(Input_Process_data, N,
How_many_eigenfunctions, t, Lambda)

ANALYTIC_Data_vector = zeros (N,1);
```

```
% Final instant T, i.e. the largest value in the "t" set.
T = N;

% This is the KEY SUBROUTINE YIELDING THE ANALYTIC RECONSTRUCTION of X(t).
for n = 1:How_many_eigenfunctions,
  arg = t.*pi * (2 * n -1) / (2*(T+1));
  NN = sqrt(2 /(T+1));
  lambda(n) = 4 * (T+1)^2 / (((pi)^2) * (2*n - 1)^2);
  phi = NN*sin(arg);
  zed = sum(Input_Process_data.*phi);
  ANALYTIC_Data_vector = ANALYTIC_Data_vector + zed * phi;
end

x = [1:1:How_many_eigenfunctions];

% Plot the EIGENVALUES of the EMPIRIC AND ANALYTIC RECONSTRUCTIONS of X(t).
figure;
plot(x, diag(Lambda(1:n,1:n)), '-k', x, lambda, ':k'), title('EIGENVALUES of the EMPIRIC
and ANALYTIC Reconstruction of B(t).'), ylabel('Eigenvalues');
legend('EMPIRIC eigenvalues','ANALYTIC eigenvalues')
```

Now we would like to show how all this performs graphically.

Figure K.1 shows a simple realization of the ordinary Brownian motion $B(t)$ over 500 time instants, starting from the origin of the axes according to the initial condition $B(0) = 0$—see Equation (F.3). The solid horizontal parabola with axis coinciding with the time axis is of course the STANDARD DEVIATION parabola,

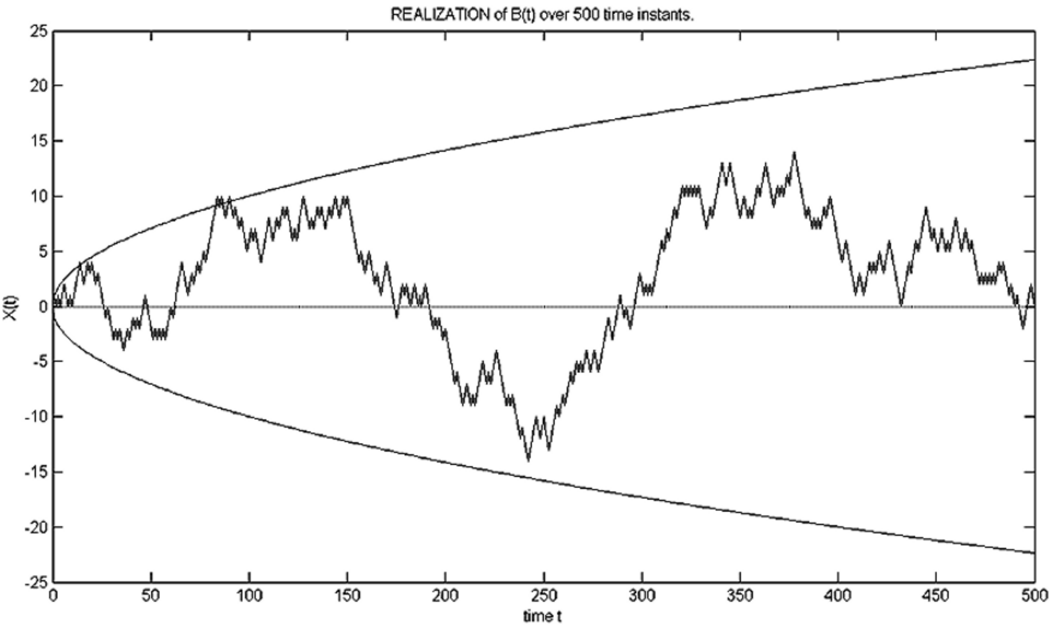


Figure K.1. A simple realization of the ordinary Brownian motion over 500 time instants.

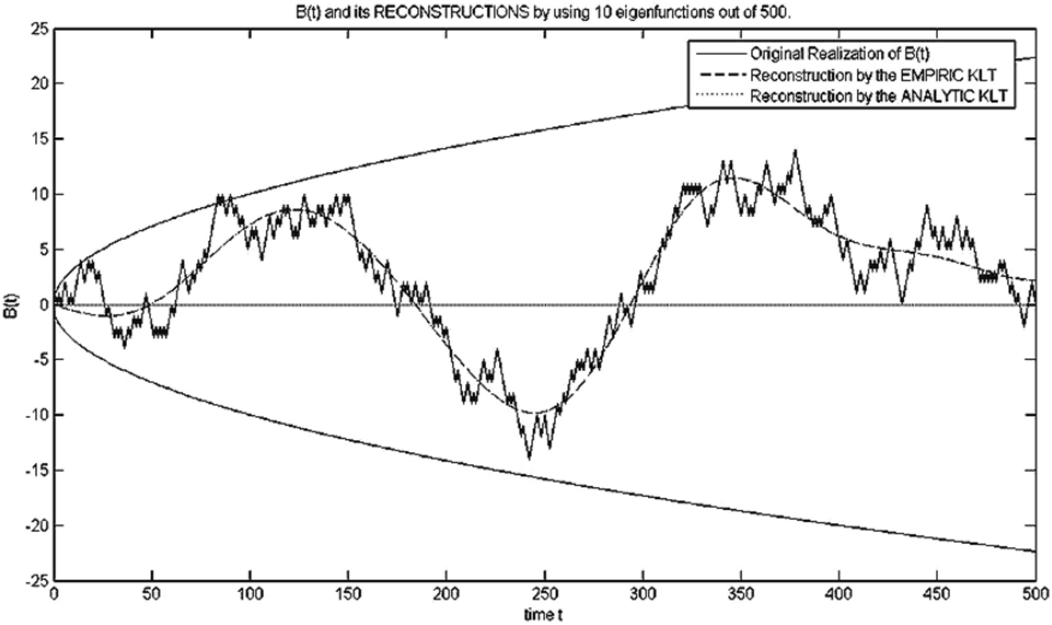


Figure K.2. The same ordinary Brownian motion realization over 500 time instants, as shown in the previous Figure K.1, plus its two KLT RECONSTRUCTIONS made by taking into account just the first 10 eigenfunctions out of 500. These two curves actually COINCIDE with the single, "smooth curve" INTERPOLATING the "peaky" Brownian motion realization. They coincide because, as shown in Section F.3, the analytical KLT for the ordinary Brownian motion is an EXACT solution, rather than a numerically approximate solution. So, the EMPIRIC and ANALYTICAL RECONSTRUCTIONS cannot fail to coincide!

having the two equations $\pm\sqrt{t}$ according to Equation (F.5). In plain words, the ordinary Brownian motion $B(t)$ "oscillates at random" above and below its mean value (i.e., the time axis, since $E\{B(t)\} \equiv 0$) and its average distance from the mean value equals approximately the standard deviation $\pm\sqrt{t}$. But this does not mean at all that $B(t)$ will always stay above or below the time axis: it actually shifts periodically, as this simulation clearly shows over just 500 time instants.

Figure K.2 shows how well the KLT reconstructs the given stochastic process according to the number of eigenfunctions taken into account for the reconstruction. It should be clear that if the user employs ALL of the KLT eigenfunctions for the reconstruction, the EMPIRICAL reconstruction will overlap exactly the original realization (100% reconstruction). In contrast, the ANALYTICAL reconstruction will overlap the original realization exactly ONLY if the solution to the KLT integral equation (10.18) is EXACT. This is precisely the case for the ordinary Brownian motion, as described in Section F.3.

K.8 THE FILE "ANALYTIC_KLT_SQUARE_BROW_MOTION.M"

The subroutine "ANALYTIC_KLT_square_brow_motion" finds the KLT of the square of the ordinary Brownian motion—that is, $B^2(t)$—as described and proven in Section I.4.

However, we must point out a "surprise" here: the mean value of the $B^2(t)$ stochastic process is NOT zero, but, rather, it is t. Thus, we may NOT compute the KLT of the $B^2(t)$ process alone, but, rather, we must compute the KLT of the new process $X(t) = B^2(t) - t$, the mean value of which is indeed zero, so that the assumption $E\{X(t)\} = 0$, upon which all the KLT theorems of Chapter 10 are based, is indeed fulfilled. Having said this, the listing is as follows.

```
% This subroutine computes the ANALYTIC RECONSTRUCTION of X(t) according to
% the analytic (either exact or approximated formulae given in the book.
function [ANALYTIC_Data_vector] = ANALYTIC_KLT(Input_Process_data, N,
How_many_eigenfunctions, t, Lambda)

ANALYTIC_Data_vector = zeros (N,1);

% Final instant T, i.e. the largest value in the "t" set.
T = N;

%Input_Process_data = Input_Process_data - mean(Input_Process_data);

% This is the KEY SUBROUTINE YIELDING THE ANALYTIC RECONSTRUCTION of
X^2(t). for n = 1:How_many_eigenfunctions,
  gamma(n) = pi * (n-5/12);
  arg = gamma(n) * t.^(3/2)/(T+1)^(3/2);
  NN = sqrt(3)*t /(T+1)^(3/2) / abs(besselj(2/3, gamma(n)));
  lambda(n) = 16/9*(T+1)^3 / ((gamma(n))^2);
  phi = NN.*besselj(2/3,arg);
  zeta = sum(Input_Process_data.*phi);
  ANALYTIC_Data_vector = ANALYTIC_Data_vector + zeta * phi ;
end

x = [1:1:How_many_eigenfunctions];

% Plot the EIGENVALUES of the EMPIRIC RECONSTRUCTION of X(t).
figure;
plot(x, diag(Lambda(1:n,1:n)), '-k', x, lambda, ':k'), title('EIGENVALUES of the EMPIRIC
and ANALYTIC Reconstruction of B^2(t)-t.'), ylabel('Eigenvalues');
legend('EMPIRIC eigenvalues','ANALYTIC eigenvalues')
```

Let us now see the graphs produced by this routine.

Figure K.3 shows a new, simple realization of the ordinary Brownian motion over 500 time instants (different from the realization shown in Figure K.1).

But the graphs shown in Figure K.4 are "unexpected"! In fact, the vertical axis now plots $X(t) = B^2(t) - t$ for the same realization of $B(t)$ shown in Figure K.3, and one sees that the numerical values on the vertical axis are of course much higher than those in Figure K.3. For instance, the parabola neatly drawn in Figure K.3 is now so much "squashed" in Figure K.4 that it looks nearly like two "parallel" straight lines above and below the time axis!

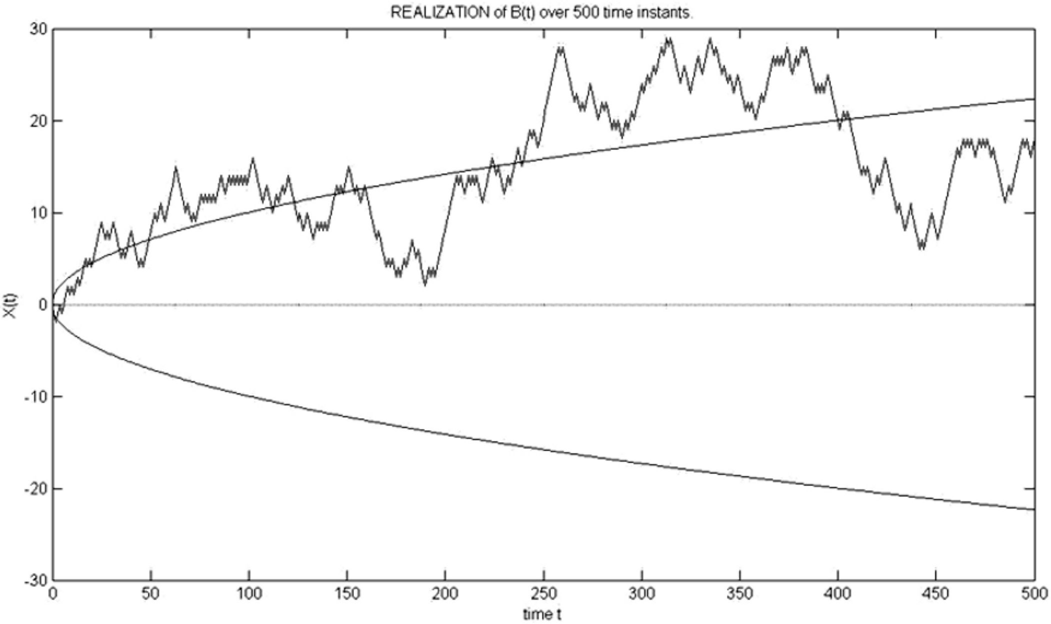


Figure K.3. A new, simple realization of the ordinary Brownian motion $B(t)$ over 500 time instants.

There is one more important feature of Figure K.4 that we must point out. The two reconstructed lines (drawn by taking into account just the first 10 eigenfunctions out of 500) DO NOT COINCIDE EXACTLY because the analytical solution given by Equation (I.53) is an APPROXIMATE analytical solution, and NOT an exact one. In fact, Equation (I.50) is the ASYMPTOTIC (for $t \to \infty$) expansion for the Bessel functions $J_\nu(t)$ and so it cannot give accurate values of $J_\nu(t)$ near the origin $t \approx 0$. This explains why, for $t \approx 0$, the two reconstructions shown in Figure K.4 are actually different from each other.

K.9 THE FILE "ANALYTIC_KLT_UNIFORM_REL.M"

Let us now turn to the RELATIVISTIC KLT, as described and proven in Chapters 11, 12, and 13. The simplest possible case of relativistic motion is of course UNIFORM motion (i.e., the motion of a spaceship at a constant speed which is also a significant fraction of the speed of light). The KLT for signals received back on Earth from a spaceship moving (away from or towards the Earth) with a UNIFORM motion was obtained in Section 11.2. The subroutine "ANALYTIC_KLT_uniform_ rel.m" translates the results of Section 11.2 into a Matlab file. Its listing is as follows.

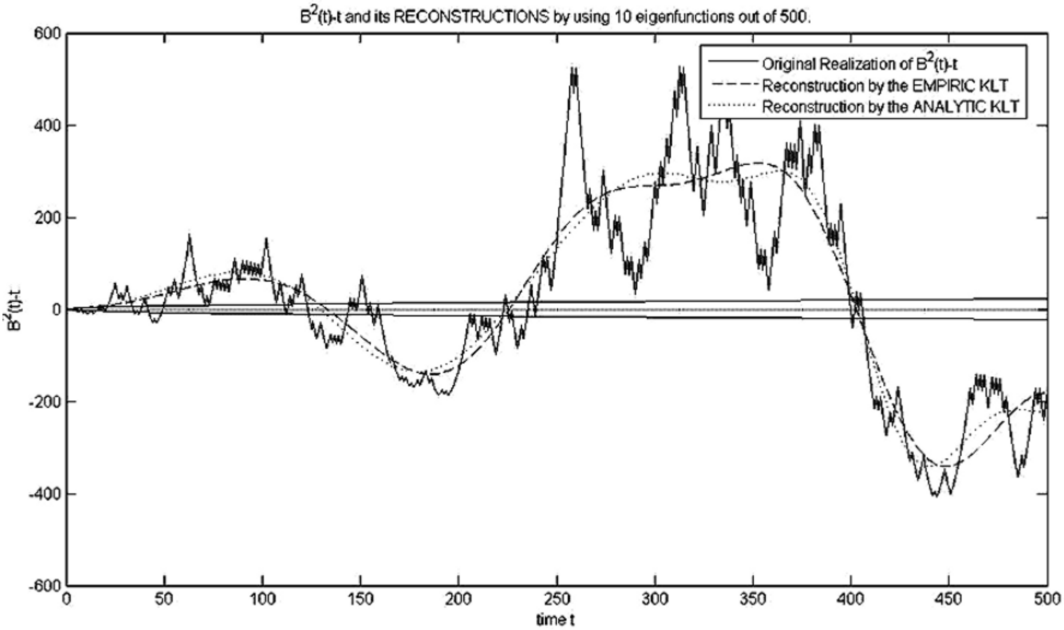


Figure K.4. Showing the stochastic process $X(t) = B^2(t) - t$ for the same realization of $B(t)$ shown in Figure K.3. Notice that the standard deviation parabola of Figure K.3 is now very much "squashed" around the time axis because of the changes in the vertical scale of the diagram. Also, the two KLT-reconstructed curves (using only the first 10 eigenfunctions out of 500), do not overlap exactly for small values of time t because the analytical reconstruction is just an approximate (i.e., inexact) formula. In fact, the asymptotic expansion formula for the Bessel functions (I.50) was used in the mathematical derivation. But for high values of time, the two reconstructions of course overlap.

```
% This subroutine computes the ANALYTIC RECONSTRUCTION of X(t) according to
% the analytic (either exact or approximated formulae given in the book.
function [ANALYTIC_Data_vector, tau] = ANALYTIC_KLT_uniform_rel(Input_Process_data,
N, How_many_eigenfunctions, t, Lambda)

ANALYTIC_Data_vector = zeros (N,1);

% Final instant T, i.e. the largest value in the "t" set.
T = N;

ratio = input('What ratio of the speed of light is the uniform velocity? (For instance
0.2) \n');
K = (1-ratio^2)^(1/4);

% This is the KEY SUBROUTINE YIELDING THE ANALYTIC RECONSTRUCTION of X(t).
for n = 1:How_many_eigenfunctions,
  gamma(n) = n*pi - pi/2;
  arg = gamma(n) * t / (T+1);
  NN = K * sqrt(2 /(T+1));
  lambda(n) = K^2 * T^2 / gamma(n)^2;
```

```
  phi = NN*sin(arg);
  zed = sum(Input_Process_data.*phi);
  ANALYTIC_Data_vector = ANALYTIC_Data_vector + zed * phi;
end

tau = sqrt(1-ratio^2)*t;
x = [1:1:How_many_eigenfunctions];

% Plot the EIGENVALUES of the EMPIRIC RECONSTRUCTION of X(t).
figure;
plot(x, diag(Lambda(1:n,1:n)), '-k', x, lambda, ':k'), title('EIGENVALUES of the EMPIRIC
and ANALYTIC Reconstruction of B(\tau).'), ylabel('Eigenvalues');
legend('EMPIRIC eigenvalues','ANALYTIC eigenvalues')
```

Again, we now want to see the graphs produced by this subroutine.

Figure K.5 shows the new realization of the Brownian motion (different of course from those in Figures K.1 and K.3). Figure K.6 shows the PROPER time τ on the horizontal axis (i.e., the time measured aboard the spaceships). The proper time τ is related to the coordinate time t by virtue of Equation (11.1). Since we have assumed in Figure K.5 that t ranges from 0 to 500, then we must compute the integral (11.1) for $t = 500$. Also, we must select at which fraction of the speed of light our spaceship is advancing in empty space. To fix ides, let us assume that it advances at 50% of the speed of light. Equation (11.1) then yields $\tau = 433.013 \approx 433$. Thus, the τ axis in Figure K.6 ranges only from 0 to 433.

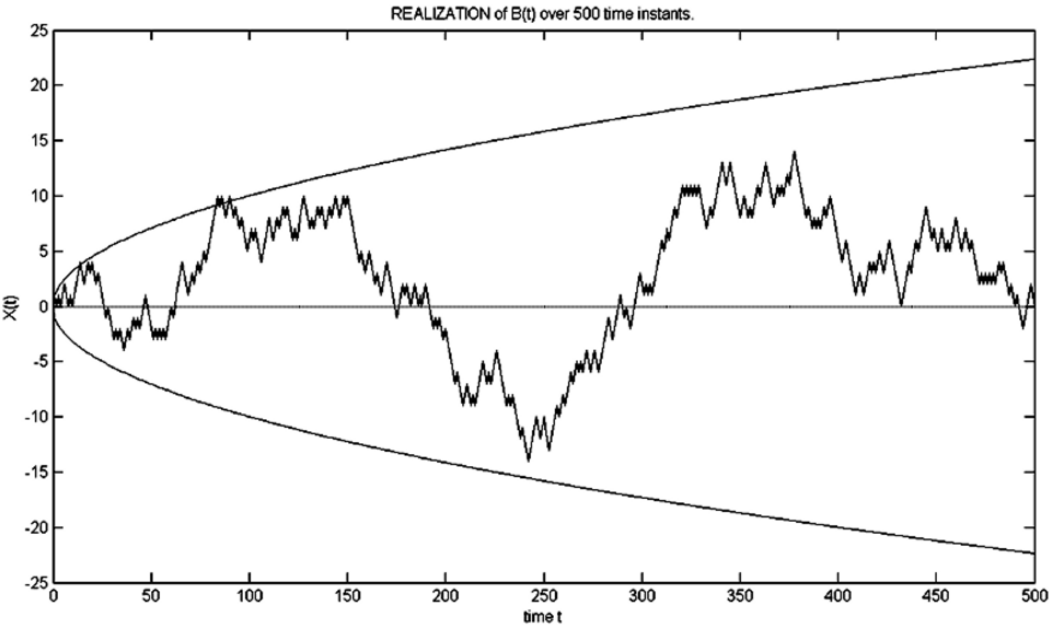


Figure K5. A new, simple realization of the ordinary Brownian motion $B(t)$ over 500 time instants.

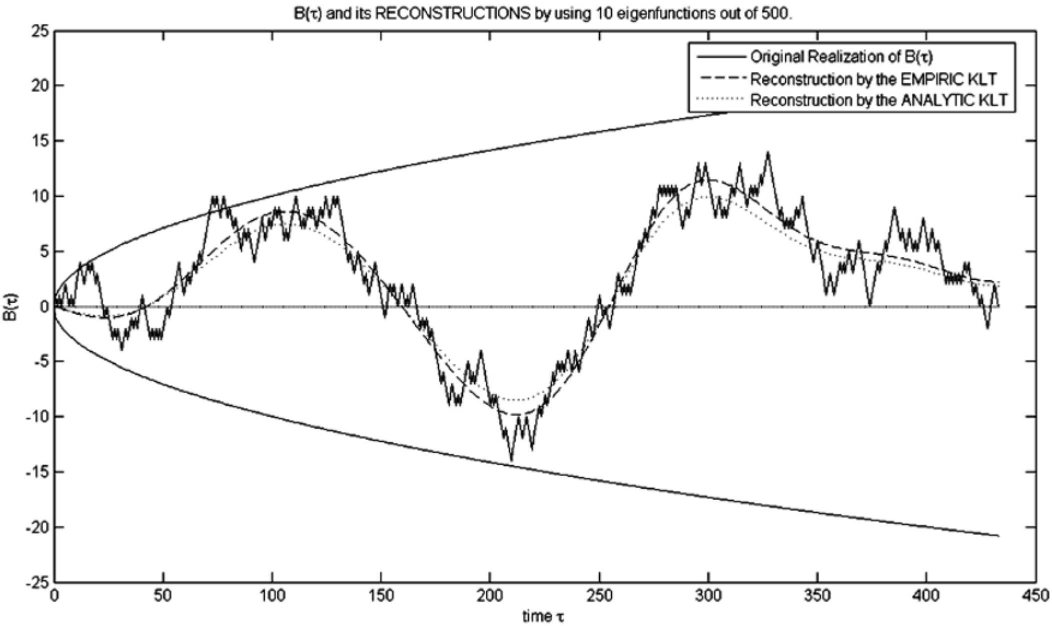


Figure K6. The time dilation effect of special relativity forces the proper time τ to range only from 0 to 433—because of Equation (11.1). Having said that, this figure shows the original input Brownian motion as in Figure K.5 plus the two reconstructed KLT curves by taking into account only the first 10 eigenfunctions.

K.10 CONCLUSIONS

We have provided the readers with a Matlab code showing the KLT of the Brownian motion in a graphical fashion. This code might be extended and improved in a number of ways, of course. Any volunteers?

Index

Numbers and dates