Third IAA Symposium
on
Searching for Life Signature

St-Petersburg, Russia, June 27-30, 2011

S.Dumas & A.Panov, eds.

July 31, 2015

Contents

Preface

The Search for Extraterrestrial Intelligence (SETI) was pioneered by Philip Morrison, Giuseppe Cocconi and Frank Drake in the early 60's. Not long after the first experiment (eg. Project OZMA), the first US conference on the subject was held in Green Bank in 1961.

Ten years later, several scientists met in Byurakan (Armenia) for what is now known as the first US-USSR conference on the subject of SETI and the search for life signature. Two more similar conferences were held in 1981 (Tallinn, Estonia) and 1991 (Santa Cruz, USA). In the mean time, SETI sessions were now held each year in other conferences related to astronomy and later astrobiology.

In 2008, the International Academy of Astronautics (IAA), decided to held a SETI only conference in Paris (at the UNESCO). This was the first of a series of conferences on the search for life signature. This document contains the papers from the third conference of the series. The third conference was held at the office of the Institute of Applied Astronomy of Russia, at St-Petersburg, during the summer of 2011. The second conference took place in Chicheley Hall, United Kingdom in 2010, hosted by the Royal Society.

Most of the original papers of the third conference were written in Russian and translated into English. The bibliography for those papers remained in there original format and language to facilitate their search. Some papers were written in English in there original version. All the papers presented in this document are not peer-reviewed and are presented as the last version of the manuscript provided by each author.

The Call of Papers is included as an introduction to the meeting. It described the event, the venue and the member of the organisation and scientific committee.

The Editors.

Call of Papers

St. Petersburg, Russia, June 27-30, 2011
http://www.ipa.nw.ru

The International Academy of Astronautics (IAA) created a SETI committee in the seventies now established in Permanent Study Group. Its role is to develop studies, establish protocols to be followed by SETI scientists in the detection, analysis, verification, announcement, and response to signals from extraterrestrial civilizations, call for symposia and cooperate with the scientific community worldwide.

Rationale

SETI (the "Search for ExtraTerrestrial Intelligence") refers to the experiments intended to find either radio or optical signals from extraterrestrial societies situated on planets around other stars. The largest radio telescopes world-wide have occasionally pursued SETI searches since 1960, in most cases hunting for signals near in frequency to 1420 MHz (the emission line of neutral hydrogen). Optical SETI searches have been pursued since the 1990s. These attempts to detect extraterrestrial signals are called "Passive SETI".

In recent years, a new kind of SETI, called "Active SETI" or "METI" (the "Messaging to ExtraTerrestrial Intelligence"), has been discussed by radio astronomers and has actually been attempted in a few cases: it consists of deliberately transmitting signals to enhance the probability of making contact with other hypothetical galactic technological civilizations. In addition, the discovery since 1995 of more than 500 extrasolar planets makes it clear that both Passive and Active SETI may now be "targeted" towards exoplanets that are situated within habitable zones, further increasing the probability of a SETI success.

In conclusion, it is now timely to gather a large conference of experts in SETI, biosignatures, the search for terrestrial exoplanets and related disciplines from all over the globe to openly discuss the strategies of both Passive and Active SETI. These are the goals of the "Searching for Life Signatures" Conference that will take place in St. Petersburg, June 27th thru 30th, 2011, in the week just prior to the "ORIGINS" IAU Commission 55 Conference that will take place at Montpellier, France, July 3rd thru 8th, 2011.

Program committee:

Jean-Michel Contant (Chairman)	International Academy of Astronautics, France
Andrey Finkelstein (co-chairman)	Institute of Applied Astronomy of Russian Academy of Sciences, Russia
Claudio Maccone (Scientific secretary)	International Academy of Astronautics, Italy
Nikolay Kardashev	Astro Space Center of Lebedev Physical Institute of Russian Academy of Sciences, Russia
Lev Gindilis	Sternberg State Astronomical Institute of the Moscow State University, Russia
Oleg Ventskovsky	State Enterprise "Yuzhnoye Design Office" (European Representation), Ukraine
Alexander Zaitsev	Kotel'nikov Institute of Radioengineering and Electronics of Russian Academy of Sciences, Russia
Stephane Dumas	SETI League Coordinator for Eastern Canada, Quebec, Canada

Local committee:

Andrey Finkelstein (Chairman)
Nadia Shuygina (Secretary)
Eugene Lysenkov
Alexandra Talvik
Valery Valyaev
Alexander Salnikov

Conference Venue

The conference will take place at the Institute of Applied Astronomy of Russian Academy of Sciences (RAS), 10 Kutuzova quay, St. Petersburg, Russia.

Format of the Conference

The language of the conference is English. Contributed paper can be submitted for oral presentation or as poster. Depending upon the number of papers offered, it is hoped to provide 30 minutes for invited papers and 15 minutes for contributed ones, including discussion. For oral presentation a multimedia projector will be in your disposition. Area up to 1.2 square meters will be provided for poster. Submitted papers are reviewed by the Program Committee and selected on the basis of their suitability for inclusion in the program. The Committee reserves the right to reject paper or to change form of its presentation. It is supposed that Proceedings of the Conference will be published soon after the end of the Conference. Abstracts (400 words max) must be submitted to seti2011@ipa.nw.ru. Deadline for Abstract Submission is 31.03.2011.

Agenda

MONDAY, 27 June

- Openning remarks from A. Finkelstein (in Russian, with English translation)
- Openning remarks and presentation of the IAA from C. Macconne (in English)
- **Finkelstein, A.**
- **Marov, M. Ya**

  1) On some problems of cosmogony

  2) On some problems of the origin of life
- **Sokulina, N.V.**

  Red dwarves' planetary systems and their civilisations.
- **Panov, A.**

  Prebiological panspermia and the hypothesis of the self-consistent Galaxy origin of life.
- LUNCH
- **Maccone, C.**

  Statistics for exoplanets and ET civilizations.
- **Panov, A.**

  Dynamical generalizations of the Drake equation: the linear and non-linear theories.
- **Efremov, Y.N.**

  The greatest mystery of the Universe.

- **Dumas, S. and Dutil, Y.**

  The possibility of an interstellar empire.

- **Yazev, S.A.**

  The end of socium and the SETI challenge.

- **Gontcharov, G.**

  Probable natural sources of the "Wow!" radio signal.

TUESDAY, 28 June

- **Maccone, C.**

  The KLT (Karhunen–Loeve Transform) to extend SETI searches to broadband and extremely feeble signals.

- **Zaitzev, A.**

  METI: Messaging to Extra-Terrestrial Intelligence.

- **Dumas, S.**

  Principal components analysis and its application.

- **Dumas, S.**

  Writing a letter to ET.

- **Gindilis, L.M.**

  Is it dangerous or not to transmit signals?

- **Maccone, C.**

  Realistic targets at 1000 AU for interstellar precursor missions.

- **Maccone, C.**

  Protected antipode circle on the Farside of the Moon.

WEDNESDAY, 29 June, Visit to the Svetloe observatory

Chapter 1

Red dwarves' planetary systems and their civilisations

by **N. V. Sokulina**
Geostroicom Ltd
sirius1nvs@yandex.ru
and **A.S. Fionov**
Joint Venture Satellite system Gonets
asf79@inbox.ru

Abstract

For 50 years the researchers' numerous attempts to establish contacts with Extraterrestrial Civilizations (EC) by "tapping" Galaxy's signals have not shown any positive results. Voluminous literature has reflected problems of the EC search. The majority of works are based on the assumption shared by the author, and this is the recognition of EC existence; of the high level of their technological progress; of their willingness and capacity to contact other civilizations. Informational exchange may become an additional knowledge source to optimize their way of development.

SETI successful advancement urges the scientists to find solutions for the two top priority tasks: search for a contact partner and a communication channel. Observations carried out by the SETI now are focused on contacts with planets of solar type stellar systems, with the use of the terrestrial civilization experience and conceptions. Our planet is relatively young and radio has been used for a little over a hundred years. There may exist in the Galaxy mature civilizations aged several billion years [1] which may have stored knowledge about the Universe, the laws of

space and physics. These may include Red dwarves stellar systems civilizations, which, according to the author, may become further preferable research objects of the SETI. Following points are indicative of this fact:

- Red dwarves belong to the majority of stars in the vicinity of the Sun,
- Red dwarves can attain the age of 7-11 gigayears, which allows to assume the presence of mature civilizations,
- Exoplanets have been discovered in Red dwarves' systems, and part of them can be located in the inhabited zone (Gl 581).

The author suggests that research objects lists should include stars of spectral types from M3 to M5.5 located up to 10 parsecs away from the Sun and forming binary systems with stars of a later spectral type. The latter fact can be explained in the following way: if there is a younger star near stars, aged 8-10 gigayears, its energy can support life of a civilization of the ageing Red Dwarf. That's why not solitary stars, but stars of binary systems, stars of late spectral types in the first place, are to be considered the most promising ones among the Red Dwarves. It's reasonable to search for Red Dwarves systems close to giant stars. Following Red dwarves star systems may be suggested as top priority research objects:

Proxima Centauri C, spectral type M5.5V, α Centauri A and B belong to spectral types G2V and K0V respectively; Sirius C – spectral type M5.5? Sirius A is attributed to A1V spectral type, Sirius B – the White dwarf – to type D2A. Alternative options include the Gl 783 stellar system, star 783B spectral type 4.0V, and star Gl 783A, spectral type K3V. The signal is supposed to be transmitted from an object identical with the civilization on a Red dwarf star system planet.

Since the terrestrial civilization is willing to contact the EC and has reached a certain level of technological and spiritual development (the latter may be just an allegation), developed civilizations may meet us halfway and transmit a signal to the earthlings.

There are several examples of the supposed EC signals received. Alain Labeque reported of one of them [2]. He relied upon the observations made by the USAF RB-47 Boeing crew made in July of 1957. An unidentified object kept following the aircraft for over an hour covering a distance of over 1,000 miles. The Boeing's electronic surveillance registered this unidentified object emitting signals within 3 GHz frequency range. The object also manoeuvred quite sensibly. The object was registered by the surface radar as well. It was suggested that the flying object gad been sent from an extraterrestrial station based in the vicinity of our Solar system. Basing upon this observation, Alain Labeque suggested that the SETI scientists transmit a 3 GHz frequency range signal unidirectionally and see if any extraterrestrial civilization reacts to it. We shall see below that this experiment setting is suggested to be complemented and specified.

The NASA has received interesting data with the help of the ARCADE radiometer while researching the Universe relict radiation. A relict radiation increase was registered. During this research the equipment was located in the stratosphere around 37 km over the Earth surface. Measurements were taken for 4 hours within the frequency range of 3, 5, 8, 10, 30 and 90 GHz. The signals intensity registered was 5-10 times higher than the expected radiation [5]. It is expected that it might be signals from extraterrestrial civilizations. In order to prove this it is proposed to receive signals at the same altitude but with wider bandwidth frequency range of instruments, including submillimetric.

In respect to mature civilizations of Red Dwarves' planetary systems observation strategy correction is advisable. Particularly, it's reasonable to use several (2 or 3) receivers of submillimetric range carried outside the terrestrial atmosphere for simultaneous signals registration. It will help to distinguish artificial signals from noisy ones. It is advisable to direct the receiver at one of the Red dwarves designated for research and wait for a signal to be transmitted.

Certain research studies are dedicated to information transmission at supraluminal speeds. In classical physics the question of information transmission at supraluminal speeds is a closed one. Theoretical research and experiments in quantum-mechanical systems affirm that information can be transmitted at a speed exceeding the electromagnetic constant. B.B.Kadomtsev, based upon the experiments and research by Y.L. Sokolov, demonstrates signal propagation at a supraluminal speed [4].

B.A.Veklenko in his study by example of dispersing an atom excited quantized electromagnetic field demonstrates a theoretical probability of information transmitting supraluminal signals [5]. So far there is neither recognized theory not experiments in information transmission at supraluminal speeds, though hopefully mature civilizations possess the means of supraluminal communication.

Nowadays the science of the Earth does not accept supraluminal communication in outer space, same as it does not accept hidden matter and energy. Research in the spheres of elementary particle physics and vacuum may provide important breakthroughs.

Brochure by M. Snegiryov "On the materiality of thought and the sixths sense" is to be published (A physicist's noted on the unknown). A quotation by Bertrand Russell is used as an epigraph to it: "The science teaches us that we can recognize but what we recognize is limited and if we forget how much there is beyond these limits, we would lose receptivity to many important factors".

There are certain suggestions and hypotheses in the brochure which may be referred to this conference subject matter. For example, is assumed that in outer space electrons disintegrate in an immense amount of particles, numbered at least 1018. If there is plasma, made up of these particles, longitudinal plasma waves may propagate in it, and there may be memory of an influence apparent as the known plasma echo effect. By example of neutrons it is evident that elementary particles lifetime depends on these particles conditions. Neutrons associated in the nuclei are stable, same as protons, the estimated lifetime of which is no less than 1030 years. However, a free neutron lifetime is about 103 seconds. Thus, the hypothesis about electrons disintegrating in outer space is quite viable and entitled to consideration.

According to the other supposition the Earth with its ionosphere, its magnetic field and plasma shell may retransmit for the earthlings the information transmitted from the outer space.

In my opinion, the SETI should make use of the ideas and hypotheses of which V.I. Vernadsky wrote:

"Seeking answers to questions frequently arising on the basis of religious meditation, philosophical thought, artistic inspiration or social life far from the science itself, may sometimes turn into a fountain of living waters nourishing research for whole generations of scientists".

The author would like to thank Mr. Andrei Fionov for the valuable help.

Literature

1. N.S. Kardashov, The Earth and the Universe, 2002, #4

2. Adam Korbitz. Paris SETI Conference at UNESCO. "Something is Here" September 25, 2008, http://estimateofthesituation.blogspot.com/2008/09/something-is-here.html

3. B.B. Kadomtsev, Physical sciences progress, 1994, vol.164, #5, p. 449-530.

4. B.A. Veklenko, Applied physics, 2010, #3, p. 10-17.

5. J. Singal, D. J. Fixsen, A. Kogut, S. Levin, M. Limon, P. Lubin, P. Mirel, M. Seiffert, T. Villela, E. Wollack and C. A. Wuensche, The ARCADE 2

Chapter 2

Prebiological panspermia and the hypothesis of the self-consistent Galaxy origin of life

by **A.D. Panov**
Moscow State University, Russia

Abstract

We argue that the panspermia may mean not only other place of the origin of life but the prebiological panspermia may mean other mechanism of the origin of life that increases the probability of the origin of life to many orders compare to a single-planet prebiological evolution. The prebiological evolution may be an all-Galaxy coherent process due to the prebiological panspermia and the origin of life is similar to Galaxy-scale second-order phase transition. This mechanism predicts life to have the same chemical base and the same chirality everywhere in the Galaxy.

2.1 Introduction

Life was to appear in the process of a natural chemical pre-biological evolution. Nobody can estimate now a “natural” duration of the pre-biological evolution on a single planet like Earth proceeding from the “first principles” or from experiment. Show how an independent phenomenological estimation of its time scale can be obtained from the timescale of the Earth biosphere evolution.

Let consider a number of the first great steps of the biological evolution (hereinafter "phase transitions"). Phase transition 0. The origin of life – about 3.9×10^9 years ago [1]. After the biosphere appeared, it was presented by anucleate anaerobic unicellular organisms – prokaryotes (and, possibly, viruses). Evidently, it existed in such a form without considerable shocks during the first 2-2.5 billion years.

Phase transition 1. The Neoproterozoic revolution [2] Anaerobic cyanobacteria enriched atmosphere in oxygen which was a strong poison for the anaerobic prokaryotes. This caused an ecological crisis, apparently, the first one in the history of the Earth. Extinction of the anaerobic prokaryotes started, and the anaerobic prokaryote fauna gave place to the eukaryote and primitive multicellular.

Phase transition 2. The Cambrian explosion (the beginning of the Paleozoic era) – 570×10^6 years ago [3, V.1]. During tens of millions of years practically all modern phylogenetic branches of multicellular organisms (including vertebrates) appeared. In the Paleozoic era the land was gradually inhabited by living creatures. When it was totally inhabited and all corresponding ecological niches were filled, the next evolution crisis occurred.

Phase transition 3. The revolution of reptiles (the beginning of the Mesozoic era) -235×10^6 years ago [3, V.1,V.2]. Practically all species of Paleozoic amphibia die out. Reptiles become leaders of evolution on land.

Phase transition 4. The revolution of mammals (the beginning of the Cenozoic era) -66×10^6 years ago [3, V.2,V.3]. Dinosaurs died out. Mammals and birds became leaders of evolution on land.

It is not difficult to see that the duration of the phases of biosphere evolution steadily decreases from the past to the present. Furthermore, the sequence of durations of the phase transitions forms a geometric series T_0/α^n with α =2.7 in good approximation (and the limit point t^* of the series fits very good the present moment of time – we are living near the point of the singularity of evolution [4]).

We see that the higher is the organization level of the biosphere, the higher is the evolution rate. Since any pre-biological system has a lower organization level than the biological one, then it seems that the pre-biological evolution rate must be even lower than the rate of the subsequent evolution of the biosphere. Furthermore, one can speculate that the duration of the pre-biological evolution belongs to the same geometric series of the phases of the biosphere evolution and estimate expected duration of the pre-biological evolution by its extrapolation back in time. It is clear that this is only an incomplete induction; our speculations are not a proof; this estimation should be considered as a conjecture.

Using the duration of the first step of the biological evolution 3.9×10^9 - 1.5×10^9 = 2.4×10^9 years, we get an estimation of the duration of the last phase of the pre-

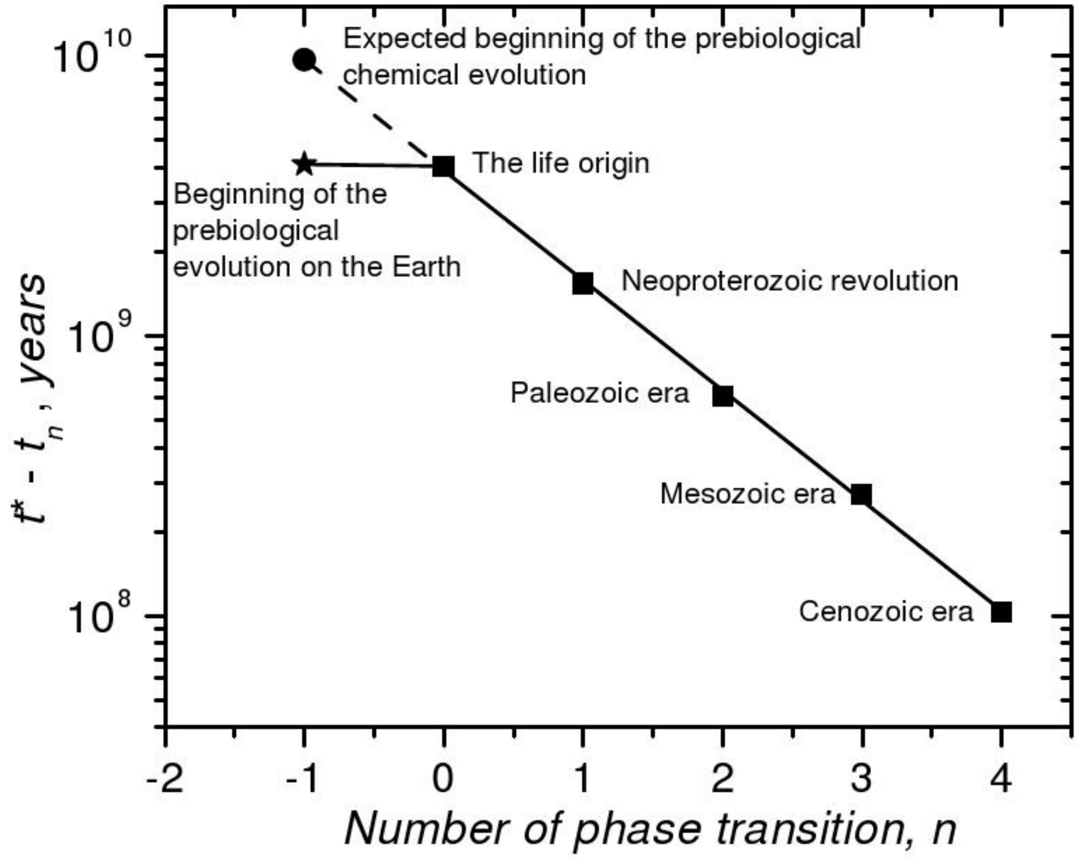


Figure 2.1: Extremely short time of the prebiological chemical evolution on the Earth produces an anomaly ‘hockey’ stick in the exponential scale of time of the evolution.

biological chemical evolution to be $\tau_{chem} = 2.4{\times}10^9{\times}2.7 = 6.4{\times}10^9$ years. This is the lower limit of the total duration of the pre-biological evolution since the latter can consist of many phases.

The value $\tau_{chem} \approx 6 \times 10^9$ years is very large. At the same time, there is evidence that the duration of the pre-biological chemical evolution on the Earth did not exceed $0.2{\times}10^9$ years [1]. An obvious contradiction is present and this contradiction is clearly seen in Fig. 2.1. It can be solved in the following way. The duration of the “natural” pre-biological chemical evolution actually is of order of 6 billion years (or even more), but it occurred not on the Earth, but on other planets near stars that are much older than the Sun. Life could get to the Earth by the process of interstellar panspermia from these old planets.

However, if the biological panspermia took place, then the pre-biological panspermia could be quite possible as well. Products of the pre-biological chemical evolution must be less sensitive to difficulties of cosmic missions (hard radiation, cold and vacuum) than any biological systems. What is a typical time scale of expansion of a pre-biological or biological "infection" over the Galaxy?

Let us refine some details of the panspermia mechanism. Suppose the question is on expansion of a biological or pre-biological product characterized by a high elasticity and competitiveness. Upon getting to a planet suitable for adaptation, such a product must expand over the planet surface in some thousands of years or even faster, replacing local weaker systems. As a result, the planet itself becomes a source of panspermia of this advanced product of evolution. If its host star flies near another star, then the latter can be infected and become an object of panspermia too. Then the spread of the product of evolution would have not the diffusion character, but the character of an autowave propagating at a constant velocity, approximately as it occurs in epidemics. The typical velocity of peculiar chaotic motion of stars is decisive. Its value – about 30 km per second – is the typical velocity of the panspermia wave in the Galaxy. To model it, the pure Huygens principle can be used. Of course, the model contains a lot of simplifications. So, for instance, the typical peculiar velocities can differ at different distances from the Galaxy center, etc. But the model is suitable to make a rough estimate of the time scale of the process.

Fig. 2.2 shows results of digital simulation of panspermia wave propagation in the Galaxy fulfilled with the above assumptions with allowance for differential rotation of the galactic disk. It can be seen from Fig. 2.1 that due to this rotation the process is practically finished in two galactic years (one galactic year – the period of rotation of the Sun around the Galaxy center – is equal to 216 million years), and 70% of the Galaxy volume is inhabited for about 300 million years.

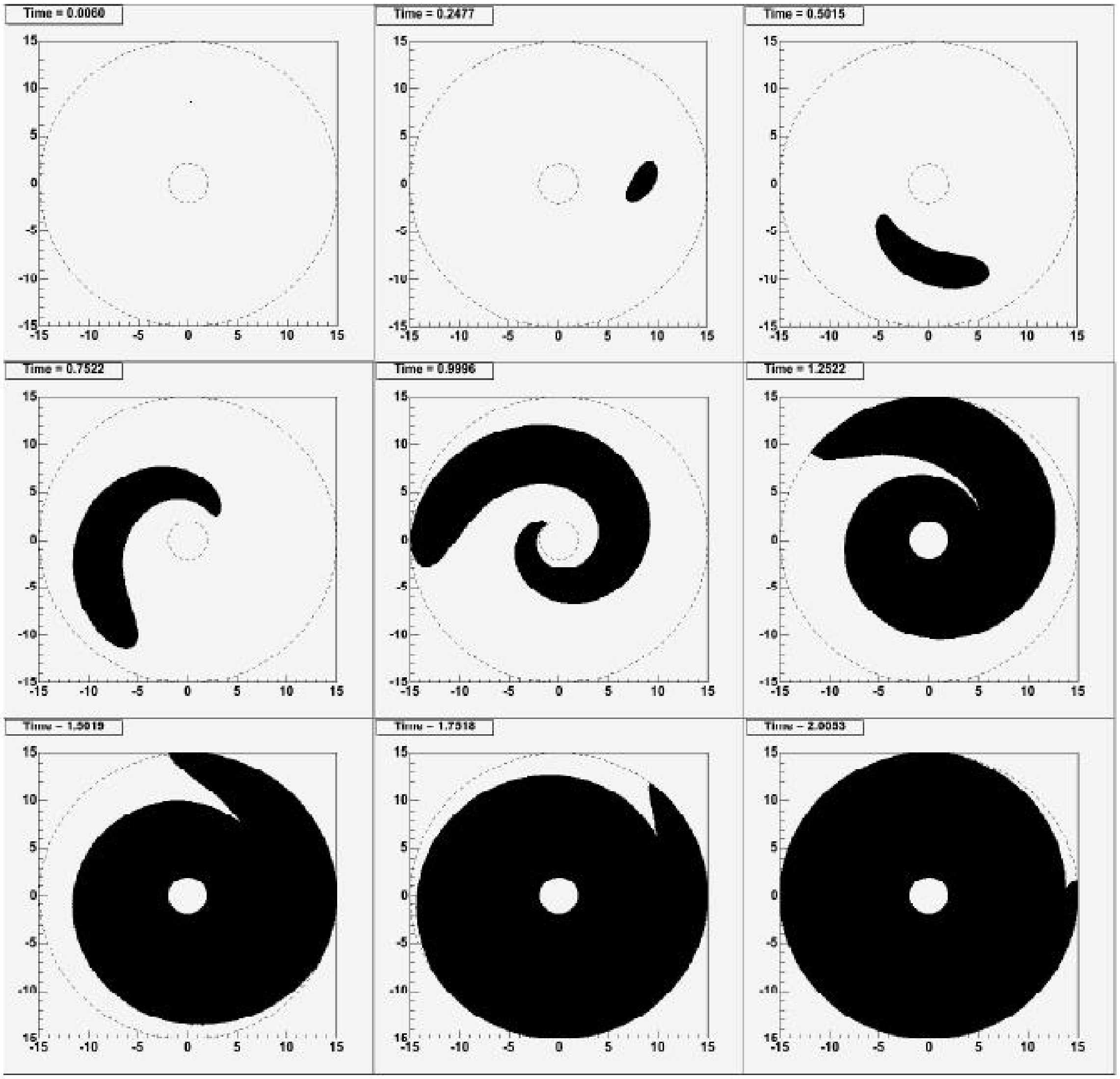


Figure 2.2: Digital model of propagation of panspermia wave in the Galaxy disc. Time in figures is shown in galactic years. The subsequent times of the phases are (from upper-left to top-down): 0.0060, 0.2477, 0.5015, 0.7522, 0.9996, 1.2522, 1.5019, 1.7518, 2.0053. The Galaxy rotates clockwise.

So, we get two time scales: one long scale of $\tau_{chem} \approx 6 \times 10^9$ years (or more), this is a scale of natural duration of the pre-biological chemical evolution on an isolated planet; the other short scale of $\tau_{pansp} \approx 0.3 \times 10^9$ years is a scale of duration of the process of galactic panspermia. From the two very different time scales it follows that the pre-biological chemical evolution on separate planets could not occur independently of processes on other planets.

Suppose a stable and competitive pre-biological system appears on a planet at the pre-biological evolution stage of the Galaxy (i.e. before life appeared for the first time). This is quite a random event. Then, during a short time, of order τ_{pansp}, this pre-biological system spreads over the whole volume of the Galaxy displacing less effective local pre-biological systems because of the ordinary natural selection. This is a mechanism of natural selection at the pre-biological level on a scale of the whole Galaxy. Thanks to the condition $\tau_{pansp} \ll \tau_{chem}$, this process must synchronize (with an accuracy of τ_{pansp} the pre-biological evolution in the whole volume of the Galaxy. As a result, life originates almost simultaneously on all planets having suitable conditions for its existence, with one molecular basis (in terms of the basis of genetic code, etc.) and with one chirality. This event resembles the non-equilibrium phase transition of second order. Thus, the pre-biological chemical evolution and the origin of life can be a self-consistent collective process, but not a process located on separate planets as is usually supposed.

If the mechanism of the self-consistent galactic origin of life operated, then a gigantic burst of inhabitance of planets with life must have taken place in the Galaxy soon after life appeared somewhere for the first time. After that life could not arise anywhere in the process of the natural pre-biological evolution since the natural pre-biological process cannot compete with much faster processes of panspermia.

It is widely believed that the probability of self-generating life on any separate planet is vanishingly small. For instance, the origin of life on an isolated planet of the Earth type with suitable conditions can take a billion billion years or some so absurdly long time. If the pre-biological evolution proceeded independently on different planets, then, at present, life would not exist at all or would be a quite unique phenomenon. However, if an effective process of the pre-biological panspermia is possible, then any random success of the pre-biological evolution on one of about 109 planets of the Galaxy becomes the property of other planets practically immediately. In other words, the probability of such an event on any separate planet increases 109 times! And the rate of the pre-biological evolution increases as well. Thus, even if the self-generating origin of life is practically improbable under the conditions of an isolated planet, it can be quite probable because of pre-biological panspermia (the last idea is from G.A. Skorobogatov, 2004, private communication). Panspermia provides other place of the origin of life of course, but it also does provide other mechanism and other way of the origin.

References

1. Orgel L. E. // Origins Life Evol. Biosph., 1998, V. 28, P. 91.

2. Rozanov A. Yu. // Paleontology magazine, 2003, No. 6, P. 41.

3. Carrol, R.L. Vertebrate Paleontology and Evolution, V.1-V.3 W.H. Freeman and Company, New York, 1988.

4. Panov A.D. // Advances in Space Research, 2005, V. 36 P. 220-225.

Chapter 3

Statistics for exoplanets and ET civilizations

by **Claudio Maccone**
International Academy of Astronautics
Via Martorelli, 43, Torino (Turin) 10155, Italy

Abstract

In this paper we provide the statistical generalization of the Fermi paradox.

We start by noting that the statistics of habitable planets may be based on a set of ten (and possibly more) astrobiological requirements first pointed out by Stephen H. Dole in his book "Habitable planets for man" (15964). We thus first provide the statistical generalization of the original and by now too simplistic Dole equation by replacing a product of ten positive numbers by the product of ten positive random variables. This we call the SEH, an acronym standing for "Statistical Equation for Habitables". Our proof is based on the Central Limit Theorem (CLT) of Statistics, stating that the sum of any number of independent random variables, each of which may be ARBITRARILY distributed, approaches a Gaussian (i.e. normal) random variable (Lyapunov form of the CLT). We then show that:

1) The new random variable N_{Hab}, yielding the number of habitables (i.e. habitable planets) in the Galaxy, follows the LOGNORMAL distribution. By construction, the mean value of this lognormal distribution is the total number of habitable planets as given by the statistical Dole equation. But now we also derive the standard deviation, the mode, the median and all the moments of this new lognormal N_{Hab} random variable.

2) The ten (or more) astrobiological factors are now positive random variables. The probability distribution of each random variable may be ARBITRARY. The CLT in the so-called Lyapunov or Lindeberg forms (that both do not assume the factors to be identically distributed) allows for that. In other words, the CLT "translates" into our SEH by allowing an arbitrary probability distribution for each factor. This is both astrobiologically realistic and useful for any further investigations.

3) An application of our SEH then follows. The (average) DISTANCE between any two nearby habitable planets in the Galaxy may be shown to be inversely proportional to the cubic root of N_{Hab}. Then, in our approach, this distance becomes a new random variable, denoted by D. We derive the relevant probability density function, apparently previously unknown and dubbed "Maccone distribution" by Paul Davies in 2008.

4) A practical example is then given of how our SEH works numerically. We work out in detail the case where each of the ten random variables is uniformly distributed around its own mean value as given by Dole back in 1964 and has an assumed standard deviation of 10%. The conclusion is that the average number of habitable planets in the Galaxy should be around 100 millions ± 200 millions, and the average distance in between any couple of nearby habitable planets should be about 88 light years ± 40 light years.

5) We match our SEH results against the results of the Statistical Drake Equation that we introduced in our 2008 IAC presentation. As expected, the number of currently communicating ET civilizations in the Galaxy turns out to be much smaller than the number of habitable planets (about 10,000 against 100 millions, i.e. one ET civilization out of 10,000 habitable planets). And the average distance between any two nearby habitable planets turns out to be much smaller that the average distance between any two neighbouring ET civilizations: 88 light years vs. 2000 light years, respectively. This means an ET average distance about 20 times higher than the average distance between any couple of adjacent habitable planets.

6) Finally, we derive our statistical model of the Fermi Paradox by applying all the above results to the coral expansion model of Galactic colonization. These equations are quite difficult and could be handled only by aid of the symbolic manipulator "Macsyma". Basically, a new random variable T_{col}, representing the time needed to colonize a new planet is introduced, and it follows the lognormal distribution, Then the new quotient random variable T_{col}/D is studied and its (difficult) probability density function is derived by Macsyma. Finally a linear transformation of random variables yields the overall time T_{Galaxy} needed to colonize the whole Galaxy. We believe that our mathematical work in deriving this STATISTICAL Fermi Paradox is highly innovative and fruitful for the future.

3.1 Introduction

The Drake equation is a now famous result (see ref. [1] for the Wikipedia summary) in the fields of SETI (the Search for ExtraTerrestial Intelligence, see ref. [2]) and Astrobiology (see ref. [3]). Devised in 1960, the Drake equation was the first scientific attempt to estimate the number N of ExtraTerrestrial civilizations in the Galaxy with which we might come in contact. Frank D. Drake (see ref. [4]) proposed it as the product of seven factors:

$$N = N_S f_p n_e f_l f_i f_c f_L \tag{3.1}$$

Where:

1. N_S is the estimated number of stars in our Galaxy.

2. f_p is the fraction (= percentage) of such stars that have planets.

3. n_e is the number "Earth-type" such planets around the given star; in other words, n_e is number of planets, in a given stellar system, on which the chemical conditions exist for life to begin its course: they are "ready for life".

4. f_l is fraction (= percentage) of such "ready for life" planets on which life actually starts and grows up (but not yet to the "intelligence" level).

5. f_i is the fraction (= percentage) of such "planets with life forms" that actually evolve until some form of "intelligent civilization" emerges (like the first, historic human civilizations on Earth).

6. f_c is the fraction (= percentage) of such "planets with civilizations" where the civilizations evolve to the point of being able to communicate across the interstellar distances with other (at least) similarly evolved civilizations. As far as we know in 2008, this means that they must be aware of the Maxwell equations governing radio waves, as well as of computers and radioastronomy (at least).

7. f_L is the fraction of galactic civilizations alive at the time when we, poor humans, attempt to pick up their radio signals (that they throw out into space just as we have done since 1900, when Marconi started the transatlantic transmissions). In other words, f_L is the number of civilizations now transmitting and receiving, and this implies an estimate of "how long will a technological

> civilization live?" that nobody can make at the moment. Also, are they going to destroy themselves in a nuclear war, and thus live only a few decades of technological civilization? Or are they slowly becoming wiser, reject war, speak a single language (like English today), and merge into a single "nation", thus living in peace for ages? Or will robots take over one day making "flesh animals" disappear forever (the so-called "post-biological universe")? No one knows...

But let us go back to the Drake equation (1). In the fifty years of its existence, a number of suggestions have been put forward about the different numeric values of its seven factors. Of course, every different set of these seven input numbers yields a different value for N, and we can endlessly play that way. But we claim that these are like... children plays!

We claim the classical Drake equation (1), as we shall call it from now on to distinguish it from our statistical Drake equation to be introduced in the coming sections, well, the classical Drake equation is scientifically inadequate in one regard at least: it just handles sheer numbers and does not associate an error bar to each of its seven factors. At the very least, we want to associate an error bar to each D_i.

Well, we have thus reached STEP ONE in our improvement of the classical Drake equation: replace each sheer number by a probability distribution!

The reader is now asked to look at the flow chart in the next page as a guide to this paper, please.

3.2 STEP 1: Letting each factor become a random variable

In this paper we adopt the notations of the great book "Probability, Random Variables and Stochastic Processes" by Athanasios Papoulis (1921-2002), now re-published as Papoulis-Pillai, ref. [5]. The advantage of this notation is that it makes a neat distinction between probabilistic (or statistical: it's the same thing here) variables, always denoted by capitals, from non-probabilistic (or "deterministic") variables, always denoted by lower-case letters. Adopting the Papoulis notation also is a tribute to him by this author, who was a Fulbright Grantee in the United States with him at the Polytechnic Institute (now Polytechnic University) of New York in the years 1977-78-79. We thus introduce seven new (positive) random variables ("D" from "Drake") defined as

$$\begin{aligned} D_1 &= N_S \\ D_2 &= f_p \\ D_3 &= n_e \\ D_4 &= f_l \\ D_5 &= f_i \\ D_6 &= f_c \\ D_7 &= f_L \end{aligned} \tag{3.2}$$

so that our **STATISTICAL Drake equation** may be simply rewritten as

$$N = \prod_{i=1}^{7} D_i \tag{3.3}$$

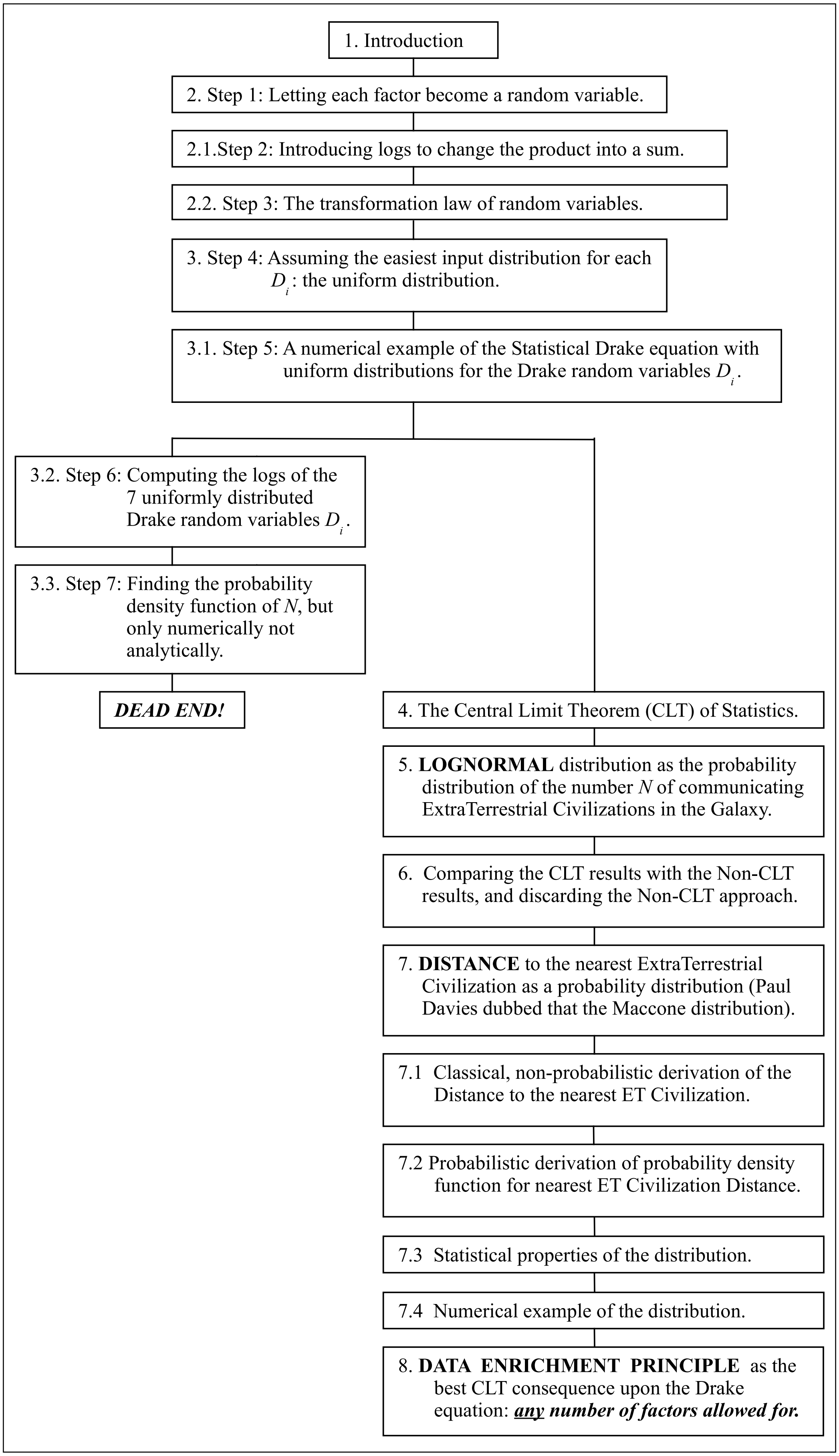
Of course, N now becomes a (positive) random variable too, having its own (positive) mean value and standard deviation. Just as each of the D_i has its own (positive) mean value and standard deviation the natural question then arises: how are the seven mean values on the right related to the mean value on the left? And how are the seven standard deviations on the right related to the standard deviation on the left? Just take the next step.

3.2.1 STEP 2: Introducing logs to change the product into a sum

Products of random variables are not easy to handle in probability theory. It is actually much easier to handle sums of random variables, rather than products, because:

1) The probability density of the sum of two or more independent random variables is the convolution of the relevant probability densities (worry not about the equations, right now).

2) The Fourier transform of the convolution simply is the product of the Fourier transforms (again, worry not about the equations, at this point)


1. Introduction
2. Step 1: Letting each factor become a random variable.
2.1.Step 2: Introducing logs to change the product into a sum.
2.2. Step 3: The transformation law of random variables.
3. Step 4: Assuming the easiest input distribution for each D_i: the uniform distribution.
3.1. Step 5: A numerical example of the Statistical Drake equation with uniform distributions for the Drake random variables D_i.
3.2. Step 6: Computing the logs of the 7 uniformly distributed Drake random variables D_i.
3.3. Step 7: Finding the probability density function of *N*, but only numerically not analytically.
DEAD END!
4. The Central Limit Theorem (CLT) of Statistics.
5. **LOGNORMAL** distribution as the probability distribution of the number *N* of communicating ExtraTerrestrial Civilizations in the Galaxy.
6. Comparing the CLT results with the Non-CLT results, and discarding the Non-CLT approach.
7. **DISTANCE** to the nearest ExtraTerrestrial Civilization as a probability distribution (Paul Davies dubbed that the Maccone distribution).
7.1 Classical, non-probabilistic derivation of the Distance to the nearest ET Civilization.
7.2 Probabilistic derivation of probability density function for nearest ET Civilization Distance.
7.3 Statistical properties of the distribution.
7.4 Numerical example of the distribution.
8. **DATA ENRICHMENT PRINCIPLE** as the best CLT consequence upon the Drake equation: ***any number of factors allowed for.***

So, let us take the natural logs of both sides of the Statistical Drake equation (3.3) and change it into a sum:

$$ln(N) = ln\left(\prod_{i=1}^{7} D_i\right) = \sum_{i=1}^{7} ln(D_i) \tag{3.4}$$

It is now convenient to introduce eight new (positive) random variables defined as follows:

$$\begin{gathered} Y = ln(N) \\ Y_i = ln(D_i), \text{ for i=1..7} \end{gathered} \tag{3.5}$$

Upon inversion, the first equation of (3.5) yields the important equation, that will be used in the sequel

$$N = e^{Y} \tag{3.6}$$

We are now ready to take STEP THREE.

3.2.2 STEP 3: The transformation law of random variables

So far we did not mention at all the problem: "which probability distribution shall we attach to each of the seven (positive) random variables D_i ?"

It is not easy to answer this question because we do not have the least scientific clue to what probability distributions fit at best to each of the seven points listed in Section 3.1.

Yet, at least one trivial error must be avoided: claiming that each of those seven random variables must have a Gaussian (i.e. normal) distribution. In fact, the Gaussian distribution, having the well-known bell-shaped probability density function

$$f_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (\sigma \geq 0) \tag{3.7}$$

has its independent variable y ranging between $-\infty$ and ∞ and so it can apply to a **real** random variable Y only, and never to **positive** random variables like those in the statistical Drake equation (3.3). Period.

Searching again for probability density functions that represent positive random variables, an obvious choice would be the gamma distributions (see, for instance, ref. [6]). However, we discarded this choice too because of a different reason: please keep in mind that, according to (3.5), once we selected a particular type of probability density function (pdf) for the last seven of equations (3.5), then we must compute the (new and different) pdf of the logs of such random variables. And the pdf of these logs certainly is not gamma-type any more.

It is high time now to remind the reader of a certain theorem that is proved in probability courses, but, unfortunately, does not seem to have a specific name. It is the transformation law (so we shall call it, see, for instance, ref. [5]) allowing us to compute the pdf of a certain new random variable Y that is a known function of another random variable X having a known pdf. In other words, if the pdf of a certain random variable X is known, then the pdf of the new random variable Y, related to X by the functional relationship

$$Y = g(X) \tag{3.8}$$

can be calculated according to this rule:

1) First invert the corresponding non-probabilistic equation $y = g(x)$ and denote by $x_i(y)$ the various real roots resulting from the this inversion.

2) Second, take notice whether these real roots may be either finitely- or infinitely-many, according to the nature of the function $y = g(x)$.

3) Third, the probability density function of Y is then given by the (finite or infinite) sum

$$f_Y(y) = \sum_i \frac{f_X(x_i(y)}{|g'(x_i(y))|} \tag{3.9}$$

where the summation extends to all roots $x_i(y)$ and $|g'(x_i(y))|$ is the absolute value of the first derivative of $g(x)$ where the i-th root $x_i(y)$ has been replaced instead of x.

Since we must use this transformation law to transfer from the D_i to the $Y_i = ln(D_i)$, it is clear that we need to start from a D_i pdf that is as simple as possible. The gamma pdf is not responding to this need because the analytic expression of the transformed pdf is very complicated (or, at least, it looked so to this author in the first instance). Also, the gamma distribution has two free parameters in it, and this "complicates" its application to the various meanings of the Drake equation. In conclusion, we discarded the gamma distributions and confined ourselves to the simpler uniform distribution instead, as shown in the nest section.

3.3 STEP 4: Assuming the easiest input distribution for each D_i : The uniform distribution

Let us now suppose that each of the seven Di is distributed UNIFORMLY in the interval ranging from the lower limit $a_i \geq 0$ to the upper limit $b_i \geq a_i$. This is the same as saying that the probability density function of each of the seven Drake random variables D_i has the equation

$$f_{uniform\_Di}(x) = \frac{1}{b_i - a_i} \quad \texttt{with} \quad 0 \leq a_i \leq x \leq b_i \tag{3.10}$$

as it follows at once from the normalization condition

$$\int_{a_i}^{b_i} f_{uniform\_Di}(x)dx = 1 \tag{3.11}$$

Let us now consider the mean value of such uniform D_i defined by

$$\begin{aligned} \langle uniformD_i \rangle &= \int_{a_i}^{b_i} x f_{uniformD_i}(x)dx = \frac{1}{b_i - a_i} \int_{a_i}^{b_i} x dx \\ &= \frac{1}{b_i - a_i} \left[\frac{x^2}{2}\right]_{a_i}^{b_i} = \frac{b_i^2 - a_i^2}{2(b_i - a_i)} = \frac{a_i + b_i}{2} \end{aligned} \tag{3.12}$$

By words (as it is intuitively obvious): the **mean value of the uniform distribution** simply is the mean of the lower plus upper limit of the variable range

$$\langle uniformD_i \rangle = \frac{a_i + b_i}{2} \tag{3.13}$$

In order to find the variance of the uniform distribution, we first need finding the second moment

$$\begin{aligned} \langle uniformD_i^2 \rangle &= \int_{a_i}^{b_i} x^2 f_{uniformD_i}(x)dx \\ &= \frac{1}{b_i - a_i} \int_{a_i}^{b_i} x^2 dx = \frac{1}{b_i - a_i} \left[\frac{x^3}{3}\right]_{a_i}^{b_i} = \frac{b_i^3 - a_i^3}{3(b_i - a_i)} \\ &= \frac{(b_i - a_i)(a_i^2 + a_i b_i + b_i^2)}{3(b_i - a_i)} = \frac{a_i^2 + a_i b_i + b_i^2}{3} \end{aligned} \tag{3.14}$$

The second moment of the uniform distribution is thus

$$\langle uniformD_i^2 \rangle = \frac{a_i^2 + a_i b_i + b_i^2}{3} \tag{3.15}$$

From (3.13) and (3.15) we may now derive the variance of the uniform distribution

$$\begin{gathered}\sigma^2_{uniformD_i} = \langle uniformD_i^2 \rangle - \langle uniformD_i \rangle^2 \\ \tfrac{a_i^2+a_ib_i+b_i^2}{3} - \tfrac{(a_i+b_i)^2}{4} = \tfrac{(b_i-a_i)^2}{12}\end{gathered} \tag{3.16}$$

Upon taking the square root of both sides of (14), we finally obtain the standard deviation of the uniform distribution:

$$\sigma_{uniformD_i} = \frac{b_i - a_i}{2\sqrt{3}} \tag{3.17}$$

We now wish to perform a calculation that is mathematically trivial, but rather unexpected from the intuitive point of view, and very important for our applications to the statistical Drake equation. Just consider the two simultaneous equations (3.13) and (3.13)

$$\begin{cases} \langle uniformD_i \rangle = \frac{a_i+b_i}{2} \\ \sigma_{uniformD_i} = \frac{b_i-a_i}{2\sqrt{3}} \end{cases} \tag{3.18}$$

Upon inverting this trivial linear system, one finds

$$\begin{cases} a_i = \langle uniformD_i \rangle - \sqrt{3}\sigma_{uniformD_i} \\ b_i = \langle uniformD_i \rangle + \sqrt{3}\sigma_{uniformD_i} \end{cases} \tag{3.19}$$

This is of paramount importance for our application the Statistical Drake equation inasmuch as it shows that: *if one (scientifically) assigns the mean value and standard deviation of a certain Drake random variable D_i, then the lower and upper limits of the relevant uniform distribution are given by the two equations (3.19), respectively.*

In other words, there is a factor of included in the two equations (3.19) that is not obvious at all to human intuition, and must indeed be taken into account.

The application of this result to the Statistical Drake equation is discussed in the next section.

3.3.1 STEP 5: A numerical example of the statistical Drake equation with uniform distributions for the Drake random variables D_i

The first variable N_s in the classical Drake equation (3.1) is the number of stars in our Galaxy. Nobody knows how many they are exactly (!). Only statistical estimates can be made by astronomers, and they oscillate (say) around a mean value of 350 billions (if this value is indeed correct!). This being the situation, we assume that our uniformly distributed random variable N_s has a mean value of 350 billions minus or plus a standard deviation of (say) one billion (we don't care whether this number is scientifically the best estimate as of August 2008: we just want to set up a numerical example of our Statistical Drake equation). In other words, we now assume that one has:

$$\begin{cases} \langle uniformD_1 \rangle = 350 \cdot 10^9 \\ \sigma_{uniformD_1} = 10^9 \end{cases} \tag{3.20}$$

Therefore, according to equations (3.19) the lower and upper limit of our uniform distribution for the random variable $N_s = D_1$ are, respectively

$$\begin{cases} a_{N_s} = \langle uniformD_1 \rangle - \sqrt{3}\sigma_{uniformD_1} = 348.3 \cdot 10^9 \\ b_{N_s} = \langle uniformD_1 \rangle + \sqrt{3}\sigma_{uniformD_1} = 351.7 \cdot 10^9 \end{cases} \tag{3.21}$$

Similarly we proceed for all the other six random variables in the Statistical Drake equation (3.3).

For instance, we assume that the fraction of stars that have planets is 50%, i.e. 50/100, and this will be the mean value of the random variable $f_p = D_2$. We also assume that the relevant standard deviation will be 10%, i. e. that $\sigma_{fp} = 10/100$. Therefore, the relevant lower and upper limits for the uniform distribution of $f_p = D_2$ turn out to be

$$\begin{cases} a_{fp} = \langle uniformD_2 \rangle - \sqrt{3}\sigma_{uniformD_2} = 0.327 \\ b_{fp} = \langle uniformD_2 \rangle + \sqrt{3}\sigma_{uniformD_2} = 0.673 \end{cases} \tag{3.22}$$

The next Drake random variable is the number ne of "Earth-type" planets in a given star system. Taking example from the Solar System, since only the Earth is truly "Earth-type", the mean value of ne is clearly 1, but the standard deviation is not zero if we assume that Mars also may be regarded as Earth-type. Since there are thus two Earth-type planets in the Solar System, we must assume a standard deviation of $1\sqrt{3} = 0.577$ to compensate the $\sqrt{3}$ appearing in (3.19) in order to finally yield two "Earth-type" planets (Earth and Mars) for the upper limit of the random variable ne. In other words, we assume that

$$\begin{cases} a_{ne} = \langle uniformD_3 \rangle - \sqrt{3}\sigma_{uniformD_3} = 0 \\ b_{ne} = \langle uniformD_3 \rangle + \sqrt{3}\sigma_{uniformD_3} = 2 \end{cases} \tag{3.23}$$

The next four Drake random variables have even more "arbitrarily" assumed values that we simply assume for the sake of making up a numerical example of our

Statistical Drake equation with uniform entry distributions. So, ***we really make no assumption about the astronomy, or the biology, or the sociology of the Drake equation: we just care about its mathematics.***

All our assumed entries are given in Table 3.1.

Please notice that, had we assumed all the standard deviations to equal zero in Table 3.1, then our Statistical Drake equation (3.3) would have obviously reduced to the classical Drake equation (3.1), and the resulting number of civilizations in the Galaxy would have turned out to be 3500:

$$N = 3500 \tag{3.24}$$

This is the important deterministic number that we will use in the sequel of this paper for comparison with our statistical results on the mean value of N, i.e.. This will be explained in Sections 3.3.3 and 3.5.

Table 3.1: Input values (i.e. mean values and standard deviations) for the seven Drake uniform random variables D_i. The first column on the left lists the seven input sheer numbers that also become the mean values (middle column). Finally the last column on the right lists the seven input standard deviations. The bottom line is the classical Drake equation (3.1).

$\mathrm{Ns} := 360 \cdot 10^9$	$\mu\mathrm{Ns}{:=}\mathrm{Ns}$	$\sigma\mathrm{Ns} := 1 \cdot 10^9$
$\mathrm{fp} := \frac{20}{100}$	$\mu\mathrm{fp}{:=}\mathrm{fp}$	$\sigma\mathrm{fp} := \frac{10}{100}$
$\mathrm{ne} := 1$	$\mu\mathrm{ne}{:=}\mathrm{ne}$	$\sigma\mathrm{ne} := \frac{10}{\sqrt{3}}$
$\mathrm{fl} := \frac{50}{100}$	$\mu\mathrm{fl}{:=}\mathrm{fl}$	$\sigma\mathrm{fl} := \frac{10}{100}$
$\mathrm{fi} := \frac{20}{100}$	$\mu\mathrm{fi}{:=}\mathrm{fi}$	$\sigma\mathrm{fi} := \frac{10}{100}$
$\mathrm{fc} := \frac{20}{100}$	$\mu\mathrm{fc}{:=}\mathrm{fc}$	$\sigma\mathrm{fc} := \frac{10}{100}$
$\mathrm{fL} := \frac{10000}{10^{10}}$	$\mu\mathrm{fL}{:=}\mathrm{fL}$	$\sigma\mathrm{fL} := \frac{1000}{10^{10}}$

$\mathrm{N} := Ns \cdot fp \cdot ne \cdot fl \cdot fi \cdot fc \cdot fL \quad \mathrm{N{=}3500}$

3.3.2 STEP 6: Computing the logs of the 7 uniformly distributed Drake random variables D_i

Intuitively speaking, the natural log of a uniformly distributed random variable may not be another uniformly distributed random variable! This is obvious from the trivial diagram of $y = ln(x)$ shown below:

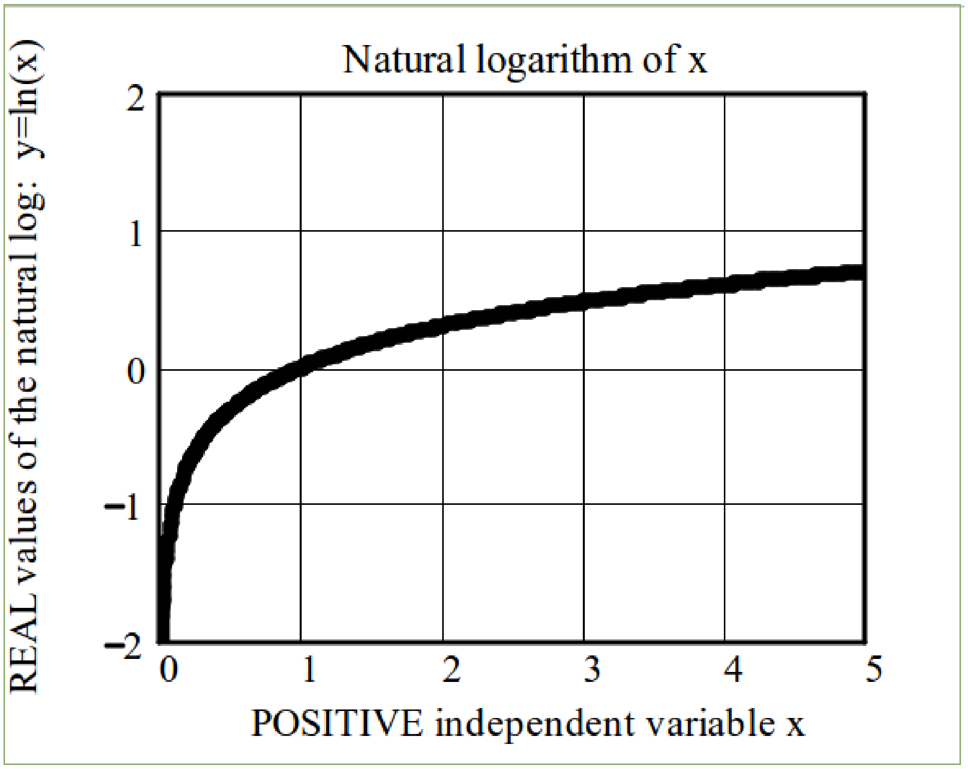


Figure 3.1: The simple function $y = ln(x)$.

So, if we have a uniformly distributed random variable D_i with lower limit a_i and upper limit b_i , the random variable

$$Y_i = ln(D_i) \quad i = 1, ..., 7 \tag{3.25}$$

must have its range limited in between the lower limit $ln(a_i)$ and the upper limit $ln(b_i)$. In other words, this are the lower and upper limits of the relevant probability density function $f_Y(y)$. But what is the actual analytic expression of such a pdf? To find it, we must resort to the general transformation law for random variables, defined by equation (3.9). Here we obviously have

$$y = g(x) = ln(x) \tag{3.26}$$

That, upon inversion, yields the **single** root

$$x_1(y) = x(y) = e^y \tag{3.27}$$

On the other hand, differentiating (3.26) one gets

$$g'(x) = \frac{1}{x} \; and \; g'(x_1(y)) = \frac{1}{x_1(y)} = \frac{1}{e^y} \tag{3.28}$$

where (3.27) was already used in the last step. By virtue of the uniform probability density function (3.10) and of (3.28), the general transformation law (3.9) finally yields

$$f_Y(y) = \sum_i \frac{f_X(x_i(y))}{|g'(x_i(y))|} = \frac{1}{b_i - a_i} \cdot \frac{1}{|\frac{1}{e^y}|} = \frac{e^y}{b_i - a - i} \tag{3.29}$$

In other words, the requested pdf of Y_i is

$$f_{Y_i}(y) = \frac{e^y}{b_i - a_i} \quad i = 1, .., 7 \quad ln(a_i) \le y \le ln(b_i) \tag{3.30}$$

Probability density functions of the natural logs of all the uniformly distributed Drake random variables D_i.

This is indeed a positive function of y over the interval $ln(a_i) \le y \le ln(b_i)$, as for every pdf, and it is easy to see that its normalization condition is fulfilled:

$$\int_{ln(a_i)}^{ln(b_i)} f_{Y_i}(y)dy = \int_{ln(a_i)}^{ln(b_i)} \frac{e^y}{b_i - a_i} = \frac{e^{ln(b_i)} - e^{ln(a_i)}}{b_i - a_i} = 1 \tag{3.31}$$

Next we want to find the mean value and standard deviation of Y_i , since these play a crucial role for future developments. The mean value $\langle Y_i \rangle$ is given by

$$\begin{aligned} \langle Y_i \rangle &= \int_{ln(a_i)}^{ln(b_i)} y f_{Y_i}(y)dy = \int_{ln(a_i)}^{ln(b_i)} \frac{ye^y}{b_i - a_i} \\ &= \frac{b_i[ln(b_i)-1] - a_i[ln(a_i)-1]}{b_i - a_i} = 1 \end{aligned} \tag{3.32}$$

This is thus the **mean value of the natural log of all the uniformly distributed Drake random variables** D_i

$$\langle Y_i \rangle = \langle ln(D_i) \rangle = \frac{b_i[ln(b_i) - 1] - a_i[ln(a_i) - 1]}{b_i - a_i} \tag{3.33}$$

In order to find the variance also, we must first compute the mean value of the square of Y_i, that is

$$\begin{aligned} \langle Y_i^2 \rangle &= \textstyle\int_{ln(a_i)}^{ln(b_i)} y^2 f_{Y_i}(y)dy = \int_{ln(a_i)}^{ln(b_i)} \frac{ye^y}{b_i - a_i} \\ &= \textstyle\frac{b_i[ln^2(b_i) - 2ln(b_i) + 2] - a_i[ln^2(a_i) - 2ln(a_i) + 2]}{b_i - a_i} \end{aligned} \tag{3.34}$$

The variance of $Y_i = ln(D_i)$ is now given by (3.34) minus the square of (3.33), that, after a few reductions, yield:

$$\sigma_{Y_i}^2 = \sigma_{ln(D_i)}^2 = 1 - \frac{a_i b_i [ln(b_i) - ln(a_i)]^2}{(b_i - a_i)^2} \tag{3.35}$$

Whence the corresponding standard deviation

$$\sigma_{Y_i} = \sigma_{ln(D_i)} = \sqrt{1 - \frac{a_i b_i [ln(b_i) - ln(a_i)]^2}{(b_i - a_i)^2}} \tag{3.36}$$

Let us now turn to another topic: the use of Fourier transforms, that, in probability theory, are called "characteristic functions". Following again the notations of Papoulis (ref. [5]) we call "characteristic function", $\Phi_{Y_i}(\zeta)$, of an assigned probability distribution Y_i , the Fourier transform of the relevant probability density function, that is (with $j = \sqrt{-1}$)

$$\Phi_{Y_i} = \int_{-\infty}^{\infty} e^{j\zeta y} f_{Y_i}(y)dy \tag{3.37}$$

The use of characteristic functions simplifies things greatly. For instance, the calculation of all moments of a known pdf becomes trivial if the relevant characteristic function is known, and greatly simplified also are the proofs of important theorems of statistics, like the Central Limit Theorem that we will use in Section 3.4. Another important result is that the characteristic function of the sum of a finite number of independent random variables is simply given by the product of the corresponding characteristic functions. This is just the case we are facing in the Statistical Drake equation (3.3) and so we are now led to find the characteristic function of the random variable Y_i , i.e.

$$\begin{aligned}\Phi_Y i(\zeta) &= \textstyle\int_{-\infty}^{\infty} e^{j\zeta y} f_{Y_i}(y)dy = \int_{ln(a_i)}^{ln(b_i)} e^{j\zeta y} \frac{e^y}{b_i-a_i} dy \\ &= \textstyle\frac{1}{b_i-a_i} \int_{ln(a_i)}^{ln(b_i)} e^{(1+j\zeta)y} dy = \frac{1}{b_i-a_i} \cdot \frac{1}{1+i\zeta} [e^{(1+j\zeta)y}]_{ln(a_i)}^{ln(b_i)} \\ &= \textstyle\frac{e^{(1+j\zeta)ln(b_i)} - e^{(1+j\zeta)ln(b_i)}}{(b_i-a_i)(1+j\zeta)} = \frac{b_i^{1+j\zeta} - a_i^{1+j\zeta}}{(b_i-a_i)(1+j\zeta)}\end{aligned} \tag{3.38}$$

Thus, **the characteristic function of the natural log of the Drake uniform random variable D_i is given by**

$$\Phi_{Y_i}(\zeta) = \frac{b_i^{1+j\zeta} - a_i^{1+j\zeta}}{(b_i - a_i)(1 + j\zeta)} \tag{3.39}$$

3.3.3 STEP 7: Finding the probability density function of n, but only numerically not analytically

Having found the characteristic functions $\Phi_Y i(\zeta)$ of the logs of the seven input random variables D_i , we can now immediately find the characteristic function of the random variable $Y = ln(N)$ defined by (3.5). In fact, by virtue of (3.4), of the well-known Fourier transform property stating that "the Fourier transform of a convolution is the product of the Fourier transforms", and of (3.39), it immediately follows that equals the product of the seven :

$$\Phi_Y(\zeta) = \prod_{i=1}^{7} \Phi_{Y_i}(\zeta) = \prod_{i=1}^{7} \frac{b^{1+j\zeta} - a^{1+j\zeta}}{(b_i - a_i)(1 + j\zeta)} \tag{3.40}$$

The next step is to invert this Fourier transform in order to get the probability density function of the random variable $Y = ln(N)$. In other words, we must compute the following inverse Fourier transform

$$\begin{aligned}f_Y(y) &= \textstyle\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\zeta y} \Phi_Y(\zeta) d\zeta \\ &= \textstyle\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\zeta y} \left[\prod_{i=1}^{7} \Phi_{Y_i}(\zeta)\right] d\zeta \\ &= \textstyle\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\zeta y} \left[\prod_{i=1}^{7} \frac{b_i^{1+j\zeta} - a_i^{1+j\zeta}}{(b_i-a_i)(1+j\zeta)}\right] d\zeta\end{aligned} \tag{3.41}$$

This author regrets that he was unable to compute the last integral **analytically**. He had to compute it **numerically** for the particular values of the 14 a_i and b_i that follow from Table 3.1 and equations 3.19. The result was the probability density function for $Y = ln(N)$ plotted in the following Figure 3.2.

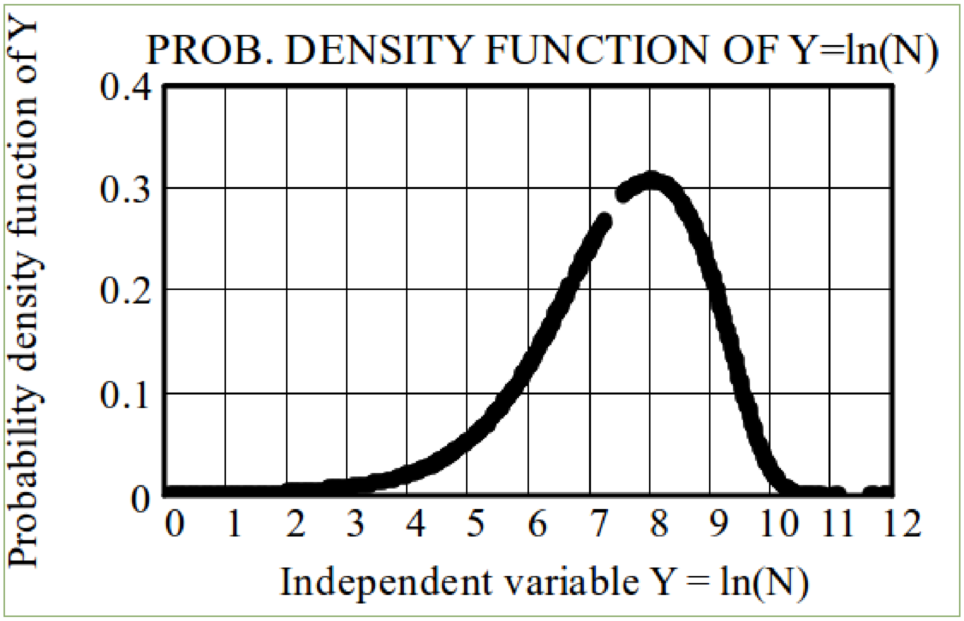


Figure 3.2: Probability density function of $Y = ln(N)$ computed numerically by virtue of the integral (3.41). The two "funny gaps" in the curve are due to the numeric limitations in the MathCad numeric solver that the author used for this numeric computation.

We are now just one more step from finding the probability density of N, the number of ExtraTerrestrial Civilizations in the Galaxy predicted by our Statistical Drake equation (3.3). The point here is to transfer from the probability density function of Y to that of N, knowing that $Y = ln(N)$, or alternatively, that $N = exp(Y)$, as stated by (3.6). We must thus resort to the transformation law of random variables (3.9) by setting

$$y = g(x) = e^x \tag{3.42}$$

This, upon inversion, yields the **single** root

$$x_1(y) = x(y) = ln(y) \tag{3.43}$$

On the other hand, differentiating (3.42) one gets

$$g'(x) = e^x \quad and \quad g'(x_1(y)) = e^{ln(y)} = y \tag{3.44}$$

where (3.43) was already used in the last step. The general transformation law (3.9) finally yields

$$f_N(y) = \sum_i \frac{f_X(x_i(y))}{g'(x_1(y))} = \frac{1}{|y|} f_Y(ln(y)) \tag{3.45}$$

This probability density function $f_N(y)$ was computed numerically by using (3.45) and the numeric curve given by (3.41), and the result is shown in Figure 3.3.

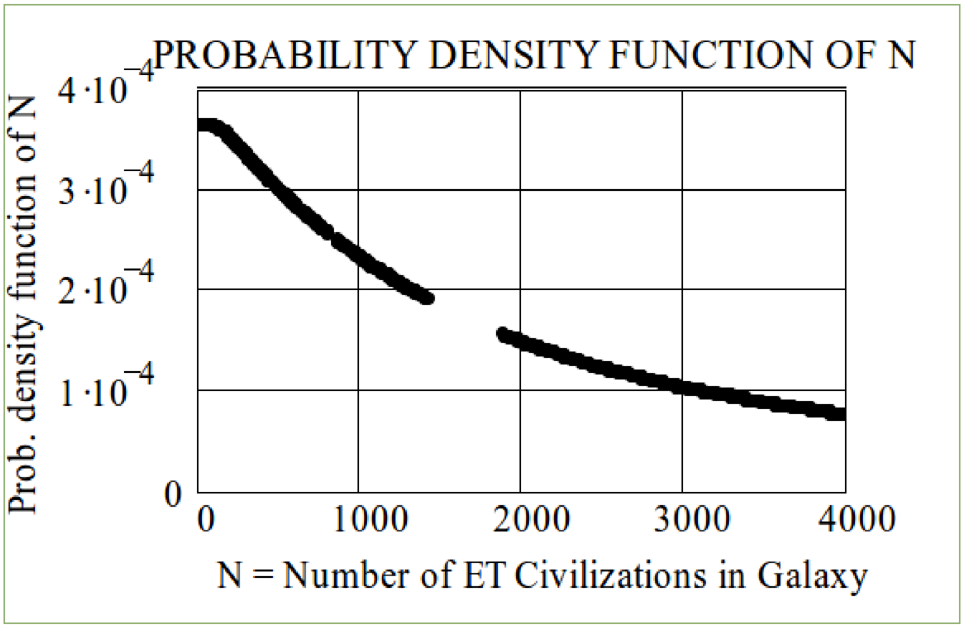


Figure 3.3: The **numeric** (and not analytic) probability density function curve $f_N(y)$ of the number N of ExtraTerrestrial Civilizations in the Galaxy according to the Statistical Drake equation (3.3). We see that the curve peak (i.e. the mode) is very close to low values of N, but the tail on the right is high, meaning that the resulting mean value $\langle N \rangle$ is of the order of thousands.

We now want to compute the mean value $\langle N \rangle$ of the probability density (3.45). Clearly, it is given by

$$\langle N \rangle = \int_0^\infty y f_N(y) dy \tag{3.46}$$

This integral too was computed numerically, and the result was a **perfect match** with N=3500 of (3.24), that is

$$\langle N \rangle = 3499.99880177509 + 0.00000124914686i \tag{3.47}$$

Note that this result was computed numerically in the complex domain because of the Fourier transforms, and that the real part is virtually 3500 (as expected) while the imaginary part is virtually zero because of the rounding errors. So, this result is excellent, and proves that the theory presented so far is mathematically correct.

Finally we want to consider the standard deviation. This also had to be computed numerically, resulting in

$$\langle \sigma_N \rangle = 3953.42910143389 + 0.000000032800058i \tag{3.48}$$

This standard deviation, higher than the mean value, implies that N might range in between 0 and 7453.

This completes our study of the probability density function of N if the seven uniform Drake input random variable D_i have the mean values and standard deviations listed in Table 3.1.

We conclude that, unfortunately, **even under the simplifying assumptions that the D_i be uniformly distributed, it is impossible to solve the full problem analytically, since all calculations beyond equation (3.40) had to be performed numerically.**

This is no good.

Shall we thus loose faith, and declare "impossible" the task of finding an analytic expression for the probability density function $f_N(y)$?

Rather surprisingly, the answer is "no", and there is indeed a way out of this dead-end, as we shall see in the next section.

3.4 The central limit theorem (CLT) of statistics

Indeed there is a good, approximating analytical expression for $f_N(y)$, and this is the following **lognormal probability density function**

$$f_N(u, \mu, \sigma) = \frac{1}{y} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(ln(y)-\mu)^2}{2\sigma^2}} \quad (y \geq 0) \tag{3.49}$$

To understand why, we must resort to what is perhaps the most beautiful theorem of Statistics: the Central Limit Theorem (abbreviated CLT). Historically, the

CLT was in fact proven first in 1901 by the Russian mathematician Alexandr Lyapunov (1857-1918), and later (1920) by the Finnish mathematician Jarl Waldemar Lindeberg (1876-1932) under weaker conditions. These conditions are certainly fulfilled in the context of the Drake equation because of the "reality" of the astronomy, biology and sociology involved with it, and we are not going to discuss this point any further here. A good, synthetic description of the Central Limit Theorem (CLT) of Statistics is found at the Wikipedia site (ref. [7]) to which the reader is referred for more details, such as the equations for the Lyapunov and the Lindeberg conditions, making the theorem "rigorously" valid.

Put in loose terms, the CLT states that, if one has a sum of random variables even NOT identically distributed, this sum tends to a normal distribution when the number of terms making up the sum tends to infinity. Also, the normal distribution mean value is the sum of the mean values of the addend random variables, and the normal distribution variance is the sum of the variances of the addend random variables.

Let us now write down the equations of the CLT in the form needed to apply it to our Statistical Drake equation (3.3). The idea is to apply the CLT to the sum of random variables given by (3.4) and (3.5) whatever their probability distributions can possibly be. In other words, the CLT applied to the Statistical Drake equation (3.3) leads immediately to the following three equations:

1) The sum of the (arbitrarily distributed) independent random variables Y_i makes up the new random variable Y.

2) The sum of their mean values makes up the new mean value of Y.

3) The sum of their variances makes up the new variance of Y.

In equations:

$$\begin{cases} Y & = \sum_{i=1}^{7} Y_i \\ \langle Y \rangle & = \sum_{i=1}^{7} \langle Y_i \rangle \\ \sigma_Y^2 & = \sum_{i=1}^{7} \sigma_Y^2 \end{cases} \tag{3.50}$$

This completes our synthetic description of the CLT for sums of random variables.

3.5 The lognormal distribution is the distribution of the number n of extraterrestrial civilizations in the galaxy

The CLT may of course be extended to products of random variables upon taking the logs of both sides, just as we did in equation (3.3). It then follows that the exponent random variable, like Y in (3.6), tends to a normal random variable, and, as a consequence, it follows that the base random variable, like N in (3.6), tends to a lognormal random variable.

To understand this fact better in mathematical terms consider again of the transformation law (3.9) of random variables. The question is: what is the probability density function of the random variable N in equation (3.6), that is, what is the probability density function of the lognormal distribution? To find it, set

$$y = g(x) = e^{x} \tag{3.51}$$

This, upon inversion, yields the **single** root

$$x_1(y) = x(y) = ln(y) \tag{3.52}$$

On the other hand, differentiating (3.51) one gets

$$g'(x) = e^{x} \quad and \quad g'(x_1(y)) = e^{ln(y)} = y \tag{3.53}$$

where (3.52) was already used in the last step. The general transformation law (3.9) finally yields

$$f_N(y) = \sum_i \frac{f_X(x_i(y))}{|g'(x_i(y))|} = \frac{1}{|y|} f_Y(ln(y)) \tag{3.54}$$

Therefore, replacing the probability density on the right by virtue of the well-known normal (or Gaussian) distribution given by equation (3.7), the lognormal distribution of equation (3.49) is found, and the derivation of the lognormal distribution from the normal distribution is proved.

In view of future calculations, it is also useful to point out the so-called "Gaussian integral", that is:

$$\int_{-\infty}^{\infty} e^{-Ax^2} B^{Bx} dx = \sqrt{\frac{\pi}{A}} e^{\frac{B^2}{4A}} \ , \ A > 0 \ B = real \tag{3.55}$$

This follows immediately from the normalization condition of the Gaussian (3.7), that is

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \tag{3.56}$$

just upon expanding the square at the exponent and making the two replacements (we skip all steps)

$$\begin{aligned} A &= \tfrac{1}{2\sigma^2} > 0 \\ B &= \tfrac{\mu}{\sigma^2} = real \end{aligned} \tag{3.57}$$

In the sequel of this paper we shall denote the independent variable of the lognormal distribution (3.49) by a lower case letter n to remind the reader that corresponding random variable N is the positive integer number of ExtraTerrestrial Civilizations in the Galaxy. In other words, n will be treated as a positive real number in all calculations to follow because it is a "large" number (i.e. a continuous variable) compared to the only civilization that we know of, i.e. ourselves. In conclusion, from now on the lognormal probability density function of N will be written as

$$f_N(n) = \frac{1}{n}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(ln(n)-\mu)^2}{2\sigma^2}} \quad (n \geq 0) \tag{3.58}$$

Having so said, we now turn to the statistical properties of the lognormal distribution (3.57), i.e. to the statistical properties that describe the number N of ExtraTerrestrial Civilizations in the Galaxy.

Our first goal is to prove an equation yielding all the moments of the lognormal distribution (3.58), that is, for every non-negative integer $k = 0, 1, 2, ...$ one has

$$\langle N^k \rangle = e^{k\mu} e^{k^2 \frac{\sigma^2}{2}} \tag{3.59}$$

The relevant proof starts with the definition of the k-th moment

$$\begin{aligned} \langle N^k \rangle &= \textstyle\int_0^\infty n^k f_N(n) dn \\ &= \textstyle\int_0^\infty n^k \frac{1}{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(ln(n)-\mu)^2}{2\sigma^2}} dn \end{aligned} \tag{3.60}$$

One then transforms the above integral by virtue of the substitution

$$ln(n) = z \tag{3.61}$$

The new integral in z is then seen to reduce to the Gaussian integral (3.55) (we skip all steps here) and (3.59) follows

$$= e^{k\mu} e^{k^2 \frac{\sigma^2}{2}} \tag{3.62}$$

Upon setting $k = 0$ into (3.58), the normalization condition for $f_N(n)$ follows

$$\int_0^\infty f_N(n) dn = 1 \tag{3.63}$$

Upon setting $k = 1$ into (3.58), the important mean value of the random variable N is found

$$\langle N \rangle = e^{\mu} e^{\frac{\sigma^2}{2}} \tag{3.64}$$

Upon setting $k = 2$ into (3.58), the mean value of the square of the random variable N is found

$$\langle N^2 \rangle = e^{2\mu} e^{2\sigma^2} \tag{3.65}$$

The variance of N now follows from the last two formulae:

$$\sigma_N^2 = e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1) \tag{3.66}$$

The square root of this is the important **standard deviation formula for the N random variable**

$$\sigma_N = e^{\mu} e^{\frac{\sigma^2}{2}} \sqrt{e^{\sigma^2} - 1} \tag{3.67}$$

The third moment is obtained upon setting $k = 3$ into (3.58)

$$\langle N^3 \rangle = e^{3\mu} e^{9/2\sigma^2} \tag{3.68}$$

Upon setting $k = 4$ into (3.58), the mean value of the square of the random variable N is found

$$\langle N^4 \rangle = e^{4\mu} e^{8\sigma^2} \tag{3.69}$$

Our next goal is to find the cumulants of N. In principle, we could compute all the cumulants K_i from the generic i-th moment by virtue of the recursion formula (see ref. [8])

$$K_i = \mu_i - \sum_{k=1}^{i-1} \left(\begin{array}{c} i-1 \\ k-1 \end{array} \right) K_k \mu_{n-k} \tag{3.70}$$

In practice, however, here we shall confine ourselves to the computation of the first four cumulants only because they only are required to find the skewness and kurtosis of the distribution. Then, the first four cumulants in terms of the first four moments read:

$$\begin{aligned} K_1 &= \mu_1 \\ K_2 &= \mu_2 - K_1^2 \\ K_3 &= \mu_3 - 3K_1K_2 - K_1^3 \\ K_4 &= \mu_4 - 4K_1K_3 - 3K_2^2 - 6K_2K_1^2 - K_1^4 \end{aligned} \tag{3.71}$$

These equations yield, respectively:

$$K_1 = e^{\mu} e^{\frac{\sigma^2}{2}} \tag{3.72}$$

$$K_2 = e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1) \tag{3.73}$$

$$K_3 = e^{3\mu} e^{\frac{3}{2}\sigma^2} (e^{\sigma^2} - 1)^2 (e^{\sigma^2} + 2) \tag{3.74}$$

$$K_4 = e^{4\mu + 2\sigma^2} (e^{\sigma^2} - 1)^3 (e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6) \tag{3.75}$$

From these we derive the skewness

$$\frac{K_3}{(K_2)^{3/2}} = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1} \tag{3.76}$$

and the kurtosis

$$\frac{K_4}{(K_2)^2} = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6 \tag{3.77}$$

Finally, we want to find the mode of the lognormal probability density function, i.e. the abscissa of its peak. To do so, we must first compute the derivative of the probability density function $f_N(n)$ of equation (3.58), and then set it equal to zero.

This derivative is actually the derivative of the ratio of two functions of n, as it plainly appears from (3.59). Thus, let us set for a moment

$$E(n) = \frac{(ln(n) - \mu)^2}{2\sigma^2} \tag{3.78}$$

where "E" stands for "exponent". Upon differentiating this, one gets

$$E'(n) = \frac{1}{2\sigma^2} 2(ln(n) - \mu)\frac{1}{n} \tag{3.79}$$

But the lognormal probability density function (3.58), by virtue of (3.78), now reads

$$f_N(n) = \frac{1}{\sqrt{2\pi}\sigma} \frac{e^{-E(n)}}{n} \tag{3.80}$$

So that its derivative is

$$\begin{aligned} \frac{df_{ET\ \ distance}(r)}{dr} &= \frac{1}{\sqrt{2\pi}\sigma} \frac{-e^{-E(n)}E'(n)n - e^{-E(n)}}{n^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \frac{-e^{E(n)}(E'(n)n+1)}{n^2} \end{aligned} \tag{3.81}$$

Setting this derivative equal to zero means setting

$$E'(n)n + 1 = 0 \tag{3.82}$$

That is, upon replacing (3.79),

$$\frac{1}{\sigma^2}(ln(n) - \mu) + 1 = 0 \tag{3.83}$$

Rearranging, this becomes

$$ln(n) - \mu + \sigma^2 = 0 \tag{3.84}$$

and finally

$$n_{mode} \equiv n_{peak} = e^{\mu} e^{-\sigma^2} \tag{3.85}$$

How likely? To find the value of the probability density function corresponding to this value of the mode, we must obviously replace (3.85) into (3.58). After a few rearrangements, one then gets

$$f_N(n_{node}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\mu} e^{\frac{\sigma^2}{2}} \tag{3.86}$$

This is "how likely" the most likely number of ExtraTerrestrial Civilizations in the Galaxy is, i.e. it is the peak height in the lognormal probability density function $f_N(n)$.

Next to the mode, the median (ref. [9]) is one more statistical number used to characterize any probability distribution. It is defined as the independent variable abscissa such that a realization of the random variable will take up a value lower than with 50% probability or a value higher than with 50% probability again. In other words, the median splits up our probability density in exactly two equally probable parts. Since the probability of occurrence of the random event equals the area under its density curve (i.e. the definite integral under its density curve) then the median (of the lognormal distribution, in this case) is defined as the integral upper limit m:

$$\int_0^m f_N(n)dn = \int_0^m \frac{1}{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(ln(n)-\mu)^2}{2\sigma^2}} = \frac{1}{2} \tag{3.87}$$

In order to find m, we may **not** differentiate (3.87) with respect to m, since the "precise" factor 1/2 on the right would then disappear into a zero. On the contrary, we may try to perform the obvious substitution

$$z^2 = \frac{(ln(n) - \mu)^2}{2\sigma^2} \quad z \geq 0 \tag{3.88}$$

into the integral (3.87) to reduce it to the following integral defining the error function $erf(z)$

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz \tag{3.89}$$

Then, after a few reductions that we skip for the sake of brevity, the full equation (3.87) is turned into

$$\frac{1}{2} + efc\left(\frac{ln(m) - \mu}{\sqrt{2}\sigma}\right) = \frac{1}{2} \tag{3.90}$$

that is

$$efc\left(\frac{ln(m) - \mu}{\sqrt{2}\sigma}\right) = 0 \tag{3.91}$$

Since from the definition (3.89) one obviously has $erf(0) = 0$, (3.91) becomes

$$\frac{ln(n) - \mu}{\sqrt{2}\sigma} = 0 \tag{3.92}$$

whence finally

$$median = m = e^{\mu} \tag{3.93}$$

This is the median of the lognormal distribution of N. In other words, this is the number of ExtraTerrestrial civilizations in the Galaxy such that, with 50% probability the actual value of N will be lower than this median, and with 50% probability it will be higher.

In conclusion, we feel useful to summarize all the equations that we derived about the random variable N in the following Table 3.2.

We want to complete this section about the lognormal probability density function (3.58) by finding out its numeric values for the inputs to the Statistical Drake equation (3.3) listed in Table 3.1.

According to the CLT, the mean value μ to be inserted into the lognormal density (3.58) is given (according to the second equation (3.50) by the sum of all the mean values $\langle Y_i \rangle$, that is, by virtue of (3.33), by:

$$\mu = \sum_{i=1}^{7} \langle Y_i \rangle = \sum_{i=1}^{7} \frac{b_i(ln(b_i) - 1) - a_i(ln(a_i) - 1)}{b_i - a_i} \tag{3.94}$$

Upon replacing the 14 a_i and b_i listed in Table 3.1 into (3.94), the following numeric mean value μ is found

$$\mu \approx 7.462176 \tag{3.95}$$

Similarly, to get the numeric variance one must resort to the last of equations (3.50) and to (3.35):

$$\sigma^2 = \sum_{i=1}^{7} \sigma_{Y_i} = \sum_{i=1}^{7} 1 - \frac{a_i b_i (ln(b_i) - ln(a_i))^2}{(b_i - a_i)^2} \tag{3.96}$$

yielding the following numeric variance σ^2 to be inserted into the lognormal pdf (3.58)

$$\sigma^2 = 1.938725 \tag{3.97}$$

whence the numeric standard deviation σ

Table 3.2: Summary of the properties of the lognormal distribution that applies to the random variable N = number of ET communicating civilizations in the Galaxy.

Random variable	N = number of communicating ET civilizations in Galaxy
Probability distribution	Lognormal
Probability density function	$f_N(n) = \frac{1}{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(ln(n)-\mu)^2}{2\sigma^2}}$ $(n \geq 0)$
Mean value	$\langle N \rangle = e^{\mu}e^{\sigma^2/2}$
Variance	$\sigma_N^2 = e^{2\mu}e^{\sigma^2}(e^{\sigma^2}-1)$
Standard deviation	$\sigma_N = e^{-\mu}e^{\sigma^2/2}\sqrt{e^{\sigma^2}-1}$
All the moments, i.e. k-th moment	$\langle N^k \rangle = e^{k\mu}e^{k^2\sigma^2/2}$
Mode (= abscissa of the lognormal peak)	$n_{mode} \equiv n_{peak} = e^{-\mu}e^{-\sigma^2}$
Value of the Mode Peak	$f_N(n_{mode}) = \frac{1}{\sqrt{s\pi}\sigma}e^{-\mu}e^{\sigma^2/2}$
Median (= fifty-fifty probability value for N)	median = m = e^{μ}
Skewness	$\frac{K_3}{(K_2)^{2/3}} = (e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$
Kurtosis	$\frac{K_4}{(K_2)^2} = e^{4\sigma^2}+2e^{2\sigma^2}+3e^{2\sigma^2}-6$
Expression of μ in terms of the lower (a_i) and upper (b_i) limits of the Drake uniform input random variables D_i	$\mu = \sum_{i=1}^{7}\langle Y_i \rangle = \sum_{i=1}^{7}\frac{b_i(ln(b_i)-1)-a_i(ln(a_i)-1)}{b_i-a_i}$
Expression of σ^2 in terms of the lower (a_i) and upper (b_i) limits of the Drake uniform input random variables D_i	$\sigma^2 = \sum_{i=1}^{7}\sigma_{Y_i}^2 = \sum_{i=1}^{7}\left(1-\frac{a_i b_i(ln(b_i)-ln(a_i))^2}{(b_i-a_i)^2}\right)$

$$\sigma = 1.392381 \tag{3.98}$$

Upon replacing these two numeric values (3.88) and (3.90) into the lognormal pdf (3.58), the latter is perfectly determined. It is plotted in Figure 3.4 hereafter as the thin curve.

In other words, Figure 3.4 shows the lognormal distribution for the number N of ExtraTerrestrial Civilizations in the Galaxy derived from the Central Limit Theorem as applied to the Drake equation (with the input data listed in Table 3.1).

We now like to point out the most important statistical properties of this lognormal pdf:

1) Mean Value of N. This is given by equation (3.64) with μ and σ given by (3.95) and (3.98), respectively:

$$\langle N \rangle = e^{\mu} e^{\frac{\sigma^2}{2}} \approx 4589.559 \tag{3.99}$$

In other words, there are 4590 ET Civilizations in the Galaxy according the Central Limit Theorem of Statistics with the inputs of Table 3.1. This number 4590 is HIGHER than the 3500 foreseen by the classical Drake equation working with sheer numbers only, rather than with probability distributions. Thus equation (3.99) IS GOOD FOR NEWS FOR SETI, since it shows that the expected number of ETs is HIGHER with an adequate statistical treatment than just with the too simple Drake sheer numbers of (3.1).

2) Variance of N. The variance of the lognormal distribution is given by (3.66) and turns out to be a huge number:

$$\sigma_N^2 = e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1) \approx 125328623 \tag{3.100}$$

3) Standard deviation of N. The standard deviation of the lognormal distribution is given by (3.67) and turns out to be:

$$\sigma_N = e^{\mu} e^{\frac{\sigma^2}{2}} \sqrt{e^{\sigma^2} - 1} = 11195 \tag{3.101}$$

Again, this is GOOD NEWS FOR SETI. In fact, such a high standard deviation means that N may range from very low values (zero, theoretically, and one since Humanity exists) up to tens of thousands (4590+11195=15785 is (3.99)+(3.101)).

4) Mode of N. The mode (= peak abscissa) of the lognormal distribution of N is given by (3.85), and has a surprisingly low numeric value:

$$n_{mode} \equiv n_{peak} = e^{\mu} e^{-\sigma^2} \approx 250 \tag{3.102}$$

This is well shown in Figure 3.4 the mode peak is very pronounced and close to the origin, but the right tail is high, and this means that the mean value of the distribution is much higher than the mode: 4590⋙250.

5) Median of N. The median (= fifty-fifty abscissa, splitting the pdf in two exactly equi-probable parts) of the lognormal distribution of N is given by (3.93), and has the numeric value:

$$n_{median} \equiv = e^{\mu} \approx 1740 \tag{3.103}$$

In words, assuming the input values listed in Table 3.1, we have exactly a 50% probability that the actual value of N is lower than 1740, and 50% that it is higher than 1740.

3.6 Comparing the CLT results with the non-CLT results

The time is now ripe to compare the CLT-based results about the lognormal distribution of N, just described in Section 3.5, against the Non-CLT-based results obtained numerically in Section 3.3.3

To do so in a simple, visual way, let us plot on the same diagram two curves:

1) The numeric curves appearing in Figure 3.2 and obtained after laborious Fourier transform calculations in the complex domain, and

2) The lognormal distribution (3.58) with numeric μ and σ given by (3.95) and (3.98) respectively.

We see that the two curves are virtually coincident for values of N larger than 1500. This is a consequence of the law of large numbers, of which the CLT is just one of the many facets.

Similarly it happens for natural log of N, i.e. the random variable Y of (3.5), that is plotted in Figure 3.5 both in its normal curve version (thin curve) and in its numeric version, obtained via Fourier transforms and already shown in Figure 3.2.

The conclusion is simple: from now on we shall discard forever the numeric calculations and we'll stick only to the equations derived by virtue of the CLT, i.e. to the lognormal (3.58) and its consequences.

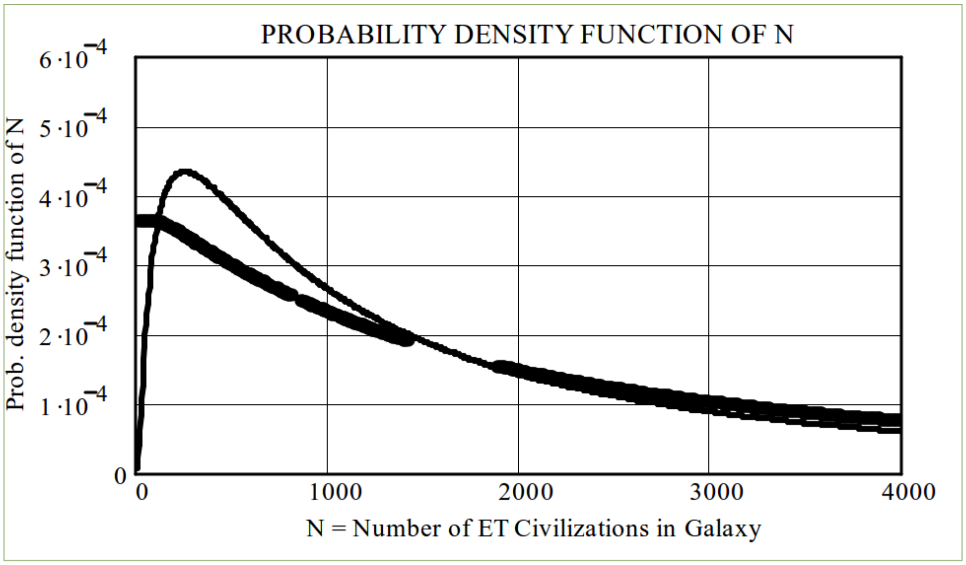


Figure 3.4: Comparing the two probability density functions of the random variable N found:
1) At the end of Section 3.3.3. in a purely numeric way and without resorting to the CLT at all (thick curve) and
2) Analytically by using the CLT and the relevant lognormal approximation (thin curve).

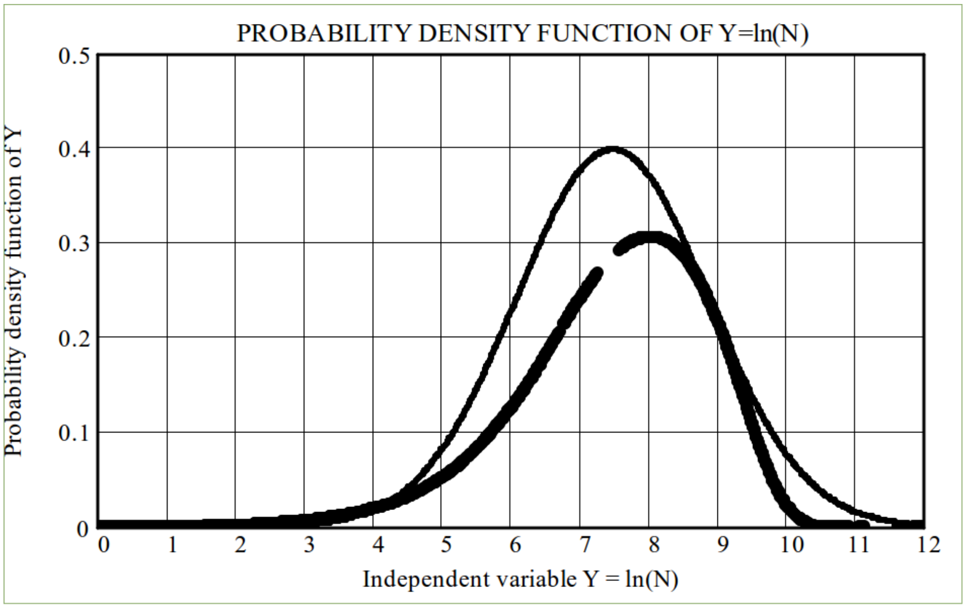


Figure 3.5: Comparing the two probability density functions of the random variable Y=ln(N) found:
1) At the end of Section 3.3.3. in a purely numeric way and without resorting to the CLT at all (thick curve) and
2) Analytically by using the CLT and the relevant normal (Gaussian) approximation (thin Gaussian curve).

3.7 Distance of the nearest extraterretrial civilization as a probability distribution

As an application of the Statistical Drake Equation developed in the previous sections of this paper, we now want to consider the problem of estimating the distance of the ExtraTerrestrial Civilization nearest to us in the Galaxy. In all Astrobiology textbooks (see, for instance, ref. [10]) and in several web sites, the solution to this problem is reported with only slight differences in the mathematical proofs among the various authors. In the first of the coming two sections (section 3.7.1) we derive the expression for this "ET Distance" (as we like to denote it) in the classical, non-probabilistic way: in other words, this is the classical, deterministic derivation. In the second section (3.7.2) we provide the probabilistic derivation, arising from our Statistical Drake Equation, of the corresponding probability density function $f_{ETdistance}(r)$: here r is the distance between us and the nearest ET civilization assumed as the independent variable of its own probability density function. The ensuing sections provide more mathematical details about this $f_{ETdistance}(r)$ such as its mean value, variance, standard deviation, all central moments, mode, median, cumulants, skewness and kurtosis.

3.7.1 Classical, non-probabilistic derivation of the distance of the nearest ET civilization

Consider the Galactic Disk and assume that:

1) The diameter of the Galaxy is (about) 100,000 light years, (abbreviated ly) i.e. its radius, R_{Galaxy} , is about 50,000 ly.

2) The thickness of the Galactic Disk at half-way from its center, h_{Galaxy}, is about 16,000 ly. Then

3) The volume of the Galaxy may be approximated as the volume of the corresponding cylinder, i.e.

$$V_{Galaxy} = \pi R^2_{Galaxy} h \tag{3.104}$$

4) Now consider the sphere around us having a radius r. The volume of such as sphere is

$$V_{oursphere} = \frac{4}{3}\pi \left(\frac{ETdistance}{2}\right)^2 \tag{3.105}$$

In the last equation, we had to divide the distance "ET Distance" between ourselves and the nearest ET Civilization by 2 because we are now going to make the unwarranted assumption that **all ET Civilizations are equally space from each other in the Galaxy!** This is a crazy assumption, clearly, and should be replaced by more scientifically-grounded assumptions as soon as we know more about our Galactic Neighbourhood. At the moment, however, this is the best guess that we can make, and so we shall take it for granted, although we are aware that this is weak point in the reasoning.

Having thus assumed that ET Civilizations are UNIFORMLY SPACED IN THE GALAXY, we can write down this proportion:

$$\frac{V_{Galaxy}}{N} = \frac{V_{oursphere}}{1} \tag{3.106}$$

That is, upon replacing both (3.104) and (3.105) into (3.106):

$$\frac{\pi R_{Galaxy}^2 h}{N} = \frac{\frac{4}{3}\pi\left(\frac{ETdistance}{2}\right)^2}{1} \tag{3.107}$$

The only unknown in the last equation is "ET Distance", and so we may solve for it, thus getting the: (AVERAGE) DISTANCE BETWEEN ANY PAIR OF NEIGHBOURING CIVILIZATIONS IN THE GALAXY

$$ETDistance = \frac{\sqrt[3]{6R_{Galaxy}^2 h}}{\sqrt[3]{N}} = \frac{C}{\sqrt[3]{N}} \tag{3.108}$$

where the positive constant C is defined by

$$C = \sqrt[3]{6R_{Galaxy}^2 h_{Galaxy}} \approx 28845 \; light \; years \tag{3.109}$$

Equations (3.108) and (3.109) are the starting point for our first application of the Statistical Drake equation, that we discuss in detail in the coming sections of this paper.

3.7.2 Probabilistic derivation of the probability density function for ET distance

The probability density function (pdf) yielding the distance of the ET Civilization nearest to us in the Galaxy and presented in this section, was discovered by this author on September 5th, 2007. He did not disclose it to other scientists until the SETI meeting run by the famous mathematical physicist and popular science author, Paul Davies, at the "Beyond" Center of the University of Arizona at Phoenix, on February 5-6-7-8, 2008. This meeting was also attended by SETI Institute experts Jill Tarter, Seth Shostak, Doug Vakoch, Tom Pierson and others. During this author's talk, Paul Davies suggested to call "the Maccone distribution" the new probability density function that yields the ET_Distance and is derived in this section.

Let us go back to equation (3.108). Since N is now a random variable (obeying the lognormal distribution), it follows that the ET_Distance must be a random variable as well. Hence it must have some unknown probability density function that we denote by

$$f_{ETdistance}(r) \tag{3.110}$$

where r is the new independent variable of such a probability distribution (it is denoted by r to remind the reader that it expresses the three-dimensional radial distance separating us from the nearest ET civilization in a full spherical symmetry of the space around us).

The question then is: what is the unknown probability distribution (3.110) of the ET_Distance?

We can answer this question upon making the two formal substitutions

$$\begin{gathered} N \to \infty \\ ETdistance \to y \end{gathered} \tag{3.111}$$

into the transformation law (3.8) for random variables. As a consequence, (3.108) takes form

$$y = g(x) = \frac{C}{\sqrt[3]{x}} = Cx^{-1/3} \tag{3.112}$$

In order to find the unknown probability density $f_{ETdistance}(r)$, we now to apply the rule (3.9) to (3.112).

First, notice that (3.112), when inverted to yield the various roots, yields a **single** real root only

$$x_1(y) = \frac{C^3}{y^3} \tag{3.113}$$

Then, the summation in (3.9) reduces to one term only. Second, differentiating (3.112) one finds

$$g'(x) = -\frac{C}{3}x^{-4/3} \tag{3.114}$$

Thus, the relevant absolute value reads

$$|g'(x)| = \left| -\frac{C}{3}x^{-4/3} \right| = \frac{C}{3}x^{-4/3} \tag{3.115}$$

Upon replacing (3.115) into (3.9), we then find

$$|g'(x)| = \frac{C}{3}x^{-4/3} = \frac{C}{3}\left(\frac{C^3}{y^3}\right)^{-4/3} = \frac{C}{3}\left(\frac{C}{y}\right)^{-4} = \frac{y^4}{3C^3} \tag{3.116}$$

This is the denominator of (3.9). The numerator simply is the lognormal probability density function (3.58) where the old independent variable x must now be re-written in terms of the new independent variable y by virtue of (3.113). By doing so, we finally arrive at the new probability density function $f_Y(y)$

$$f_Y(y) = \frac{3C^3}{y^4}\frac{1}{C^3/y^3}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(ln(C^3/y^3)-\mu)^2}{2\sigma^2}} \tag{3.117}$$

Rearranging and replacing y by r, the final form is:

$$f_{ETDistance}(r) = \frac{3}{r}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(ln(C^3/y^3)-\mu)^2}{2\sigma^2}} \tag{3.118}$$

Now, just replace C in (3.118) by virtue of (3.109). Then:

We have discovered the probability density function yielding the probability of finding the nearest ExtraTerrestrial Civilization in the Galaxy in the spherical shell between the distances r and $r + dr$ from Earth:

$$f_{ETDistance}(r) = \frac{3}{r}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(ln\left(\frac{6R^2_{Galaxy}h_{Galaxy}}{r^3}\right)-\mu\right)^2}{2\sigma^2}} \tag{3.119}$$

holding for $r \geq 0$.

3.7.3 Statistical properties of the distribution

We now want to study this probability distribution in detail. Our next questions are:

1. What is its mean value?
2. What are its variance and standard deviation?
3. What are its moments to any higher order?
4. What are its cumulants?
5. What are its skewness and kurtosis?
6. What are the coordinates of its peak, i.e. the mode (peak abscissa) and its ordinate?
7. What is its median?

The first three points in the list are all covered by the following theorem: all the moments of (3.118) are given by (here k is the generic and non-negative integer exponent, i.e. $k = 0, 1, 2, 3, ... \geq 0$)

$$\langle ET\_distance^k\rangle = \int_0^\infty r^k f_{ETDistance}(r)dr$$

$$\int_0^\infty r^k \frac{3}{r}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(ln(C^3/y^3)-\mu)^2}{2\sigma^2}} \tag{3.120}$$

$$= C^k e^{-k\frac{\mu}{3}} e^{k^2\frac{\sigma^2}{18}}$$

To prove this result, one first transforms the above integral by virtue of the substitution

$$ln\left(\frac{C^3}{r^3}\right) = z \tag{3.121}$$

Then the new integral in z is then seen to reduce to the known Gaussian integral (3.55) and, after several reductions that we skip for the sake of brevity, (3.120) follows from (3.55). In other words, we have proven that

$$\langle ET\_distance^k\rangle = C^k e^{-k\frac{\mu}{3}} e^{k^2\frac{\sigma^2}{18}} \tag{3.122}$$

Upon setting $k = 0$ into (3.122), the normalization condition for $f_{ETdistance}(r)$ follows

$$\int_0^\infty f_{ETdistance}(r)dr = 1 \tag{3.123}$$

Upon setting $k = 1$ into (3.122), the important mean value of the random variable ET_Distance is found

$$\langle ET\_distance\rangle = Ce^{-\frac{\mu}{3}} e^{\frac{\sigma^2}{18}} \tag{3.124}$$

Upon setting $k = 2$ into (3.122), the mean value of the square of the random variable ET_Distance is found

$$\langle ET\_distance^2\rangle = C^2 e^{-\frac{2\mu}{3}} e^{\frac{2}{9}\sigma^2} \tag{3.125}$$

The variance of ET_Distance now follows from the last two formulae with a few reductions:

$$\begin{aligned}\sigma^2_{ETDistance} &= \langle ET\_distance^2\rangle - \langle ET\_distance\rangle^2 \\ &= C^2 e^{-2\mu/3} e^{\sigma^2/9}\left(e^{\sigma^2/9} - 1\right)\end{aligned} \tag{3.126}$$

So, **the variance of ET_Distance is**

$$\sigma^2_{ETdistance} = C^2 e^{-2\mu/3} e^{\sigma^2/9} \left(e^{\sigma^2/9} - 1 \right) \tag{3.127}$$

The square root of this is the important **standard deviation of the ET_Distance random variable**

$$\sigma_{ETdistance} = C e^{-\mu/3} e^{\sigma^2/18} \sqrt{e^{\sigma^2/9} - 1} \tag{3.128}$$

The third moment $k = 3$ is obtained upon setting $k = 4$ into (3.122)

$$\langle ET\_distance^3 \rangle = C^3 e^{-\mu} e^{\sigma^2/2} \tag{3.129}$$

Finally, upon setting $k = 4$ into (3.122), the fourth moment of ET_Distance is found

$$\langle ET\_distance^4 \rangle = C^4 e^{-4\mu/3} e^{8\sigma^2/9} \tag{3.130}$$

Our next goal is to find the cumulants of the ET_Distance. In principle, we could compute all the cumulants K_i from the generic i-th moment μ'_i by virtue of the recursion formula (see ref. [8])

$$K_i = \mu'_i - \sum_{k=1}^{i-1} \left(\begin{array}{c} i-1 \\ k-1 \end{array} \right) K_k \mu'_{n-k} \tag{3.131}$$

In practice, however, here we shall confine ourselves to the computation of the first four cumulants because they only are required to find the skewness and kurtosis of the distribution (3.118). Then, the first four cumulants in terms of the first four moments read:

$$\begin{gathered} K_1 = \mu'_1 \\ K_2 = \mu'_2 - K_1^2 \\ K_3 = \mu'_3 - 3K_1K_2 - K_1^3 \\ K_4 = \mu'_4 - 4K_1K_3 - 3K_2^2 - 6K_2K_1^2 - K_1^4 \end{gathered} \tag{3.132}$$

These equations yield, respectively:

$$K_1 = C e^{\mu/3} e^{\sigma^2/18} \tag{3.133}$$

$$K_2 = C^2 e^{-2\mu/3} e^{\sigma^2/9} \left(e^{\sigma^2/9} - 1\right) \tag{3.134}$$

$$K_3 = C^3 e^{-\mu} \left(e^{\sigma^2/2} - 3e^{5\sigma^2/18} + 2e^{\sigma^2/6}\right) \tag{3.135}$$

$$K_4 = C^4 e^{-4\mu/3} \left(e^{8\sigma^2/9} - 4e^{5\sigma^2/9} - 3e^{4\sigma^2/9} + 12e^{\sigma^2/3} - 6e^{2\sigma^2/9}\right) \tag{3.136}$$

From these we derive the skewness

$$\frac{K_3}{(K_2)^{3/2}} = \frac{e^{2\sigma^2/9} + e^{\sigma^2/2} - 2}{\sqrt{e^{\sigma^2/9}} - 1} \tag{3.137}$$

and the kurtosis

$$\frac{K_4}{(K_2)^2} = e^{4\sigma^2/9} + e^{\sigma^2/3} + e^{2\sigma^2/9} - 6 \tag{3.138}$$

Next we want to find the mode of this distribution, i.e. the abscissa of its peak. To do so, we must first compute the derivative of the probability density function $f_{ETdistance}(r)$ of (3.118), and then set it equal to zero. This derivative is actually the derivative of the ratio of two functions of r, as its plainly appears from (3.118). Thus, let us set for a moment

$$E(r) = \frac{(ln(\frac{C^3}{r^3}) - \mu)^2}{2\sigma^2} \tag{3.139}$$

where "E" stands for "exponent". Upon differentiating, one gets

$$\begin{aligned} E'(r) &= \tfrac{1}{2\sigma^2} 2 \left(ln\left(\tfrac{C^3}{r^3}\right) - \mu\right) \tfrac{1}{C^3/r^3} C^3 (-3) r^{-4} \\ &= \tfrac{1}{\sigma^2} \left(ln\left(\tfrac{C^3}{r^3}\right) - \mu\right) (-3) \tfrac{1}{r} \end{aligned} \tag{3.140}$$

But the probability density function (3.118) now reads

$$f_{ETdistance}(r) = \frac{3}{\sqrt{2\pi}\sigma} \frac{e^{-E(r)}}{r} \tag{3.141}$$

So that its derivative is

$$\begin{aligned} \frac{df_{ETdistance}(r)}{dr} &= \tfrac{3}{\sqrt{2\pi}\sigma}\tfrac{-e^{-E(r)}E'(r)r-e^{-E(r)}}{r^2} \\ &= \tfrac{3}{\sqrt{2\pi}\sigma}\tfrac{-e^{-E(r)}(E'(r)r+1)}{r^2} \end{aligned} \tag{3.142}$$

Setting this derivative equal to zero means setting

$$E'(r)r + 1 = 0 \tag{3.143}$$

That is, upon replacing (3.140) into (3.143), we get

$$\frac{1}{\sigma^2}\left(ln\left(\frac{C^3}{r^3}\right) - \mu\right)(-3)\frac{1}{r}r + 1 = 0 \tag{3.144}$$

Rearranging, this becomes

$$-3\left(ln\left(\frac{C^3}{r^3}\right) - \mu\right) + \sigma^2 = 0 \tag{3.145}$$

that is

$$-3ln\left(\frac{C^3}{r^3}\right) + 3\mu + \sigma^2 = 0 \tag{3.146}$$

whence

$$ln\left(\frac{C}{r}\right) = \frac{\mu}{3} + \frac{\sigma^2}{9} \tag{3.147}$$

and finally

$$r_{mode} \equiv r_{peak} = Ce^{-\mu/3}e^{-\sigma^2/9} \tag{3.148}$$

This is the most likely ET_Distance from Earth.
How likely ?

To find the value of the probability density function $f_{ETdistance}(r)$ corresponding to this value of the mode, we must obviously replace (3.148) into (3.58). After a few rearrangements, which we skip for the sake of brevity, one gets

$$\begin{aligned} &Peak\ Value\ of\ f_{ETDistance}(r) \equiv f_{ETdistance}(r_{mode}) \\ &= \tfrac{3}{C\sqrt{2\pi}\sigma}e^{\mu/3}e^{\sigma^2/18} \end{aligned} \tag{3.149}$$

This is the peak height in the pdf $f_{ETdistance}(r)$.

Next to the mode, the median m (ref. [9]) is one more statistical number used to characterize any probability distribution. It is defined as the independent variable abscissa m such that a realization of the random variable will take up a value lower than m with 50% probability or a value higher than m with 50% probability again. In other words, the median m splits up our probability density in exactly two equally probable parts. Since the probability of occurrence of the random event equals the area under its density curve (i.e. the definite integral under its density curve) then the median (of the lognormal distribution, in this case) is defined as the integral upper limit m:

$$\int_0^m f_{Etdistance}(r)dr = \frac{1}{2} \tag{3.150}$$

Upon replacing (3.118), this becomes

$$\int_0^m \frac{3}{r}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(ln(C^3/r^3)-\mu\right)^2}{2\sigma^2}} = \frac{1}{2} \tag{3.151}$$

In order to find m, we may **not** differentiate (3.151) with respect to m, since the "precise" factor 1/2 on the right would then disappear into a zero. On the contrary, we may try to perform the obvious substitution

$$z^2 = \frac{\left(ln(C^3/r^3)-\mu\right)^2}{2\sigma^2} \quad z \geq 0 \tag{3.152}$$

into the integral (3.151) to reduce it to the following integral (3.89) defining the error function erf(z). Then, after a few reductions that we leave to the reader as an exercise, the full equation (3.150), defining the median, is turned into the corresponding equation involving the error function erf(x) as defined by (3.89):

Table 3.3: Summary of the properties of the probability distribution that applies to the random variable ET_Distance yielding the (average) distance between any two neighboring communicating civilizations in the Galaxy.

Random variable	ET_Distance between any two neighboring ET Civilizations in Galaxy assuming they are UNIFORMLY distributed throughout the whole Galaxy volume.
Probability distribution	Unnamed (Paul Davies suggested "Maccone distribution")
Probability density function	$f_{ET\_distance}(n) = \frac{3}{r}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(ln\left(\frac{6R_{Galaxy}h_{Galaxy}}{r^3}\right)-\mu\right)^2}{2\sigma^2}}$
(Defining the positive numeric constant C)	$C = \sqrt[3]{6R_{Galaxy}^2 h_{Galaxy}} \approx 28845\ light\ years$
Mean value	$\langle ET\_distance\rangle = Ce^{\mu/3}e^{\sigma^2/18}$
Variance	$\sigma_{ET\_distance}^2 = C^2e^{2\mu/3}e^{\sigma^2/9}(e^{\sigma^2/9}-1)$
Standard deviation	$\sigma_{ET\_distance} = Ce^{-\mu/3}e^{\sigma^2/18}\sqrt{e^{\sigma^2/9}-1}$
All the moments, i.e. k-th moment	$\langle ET\_distance^k\rangle = Ce^{-k\mu/3}e^{k^2\sigma^2/18}$
Mode (= abscissa of the lognormal peak)	$r_{mode} \equiv r_{peak} = Ce^{-\mu/3}e^{-\sigma^2/9}$
Value of the Mode Peak	Peak Value of $f_{ET\_distance}(r) =$ $f_{ET\_distance}(r_{mode}) = \frac{3}{C\sqrt{2\pi}\sigma}e^{\mu/3}e^{\sigma^2/18}$
Median (= fifty-fifty probability value for ET_distance)	median = m = $Ce^{-\mu/3}$
Skewness	$\frac{K_3}{(K_2)^{2/3}} = \frac{e^{2\sigma^2/9}+e^{\sigma^2/9}-2}{\sqrt{e^{\sigma^2/9}-1}}$
Kurtosis	$\frac{K_4}{(K_2)^2} = e^{4\sigma^2/9} + 2e^{\sigma^2/3} + 3e^{2\sigma^2/9} - 6$
Expression of μ in terms of the lower (a_i) and upper (b_i) limits of the Drake uniform input random variables D_i	$\mu = \sum_{i=1}^{7}\langle Y_i\rangle = \sum_{i=1}^{7}\frac{b_i(ln(b_i)-1)-a_i(ln(a_i)-1)}{b_i-a_i}$
Expression of σ^2 in terms of the lower (a_i) and upper (b_i) limits of the Drake uniform input random variables D_i	$\sigma^2 = \sum_{i=1}^{7}\sigma_{Y_i}^2 = \sum_{i=1}^{7}\left(1-\frac{a_ib_i(ln(b_i)-ln(a_i))^2}{(b_i-a_i)^2}\right)$

$$\frac{1}{2} + erf\left(\frac{ln(C^3/r^3) - \mu}{\sqrt{2}\sigma}\right) = \frac{1}{2} \tag{3.153}$$

$$erf\left(\frac{ln(C^3/r^3) - \mu}{\sqrt{2}\sigma}\right) = 0 \tag{3.154}$$

Since from the definition (147) one obviously has erf(0)=0, (149) yields

$$\frac{ln(C^3/r^3) - \mu}{\sqrt{2}\sigma} = 0 \tag{3.155}$$

whence finally

$$median = m = Ce^{-\mu/3} \tag{3.156}$$

This is the median of the lognormal distribution of N. In other words, this is the number of ExtraTerrestrial civilizations in the Galaxy such that, with 50% probability the actual value of N will be lower than this median, and with 50% probability it will be higher.

In conclusion, we feel useful to summarize all the equations that we derived about the random variable N in the following Table 3.2.

3.7.4 Numerical example of the distance distribution

In this section we provide a numerical example of the analytic calculations carried on so far.

Consider the Drake Equation values reported in Table 3.1. Then, the graph of the corresponding probability density function of the nearest ET_Distance, $f_{ET\_distance}$, is shown in Figure 3.6.

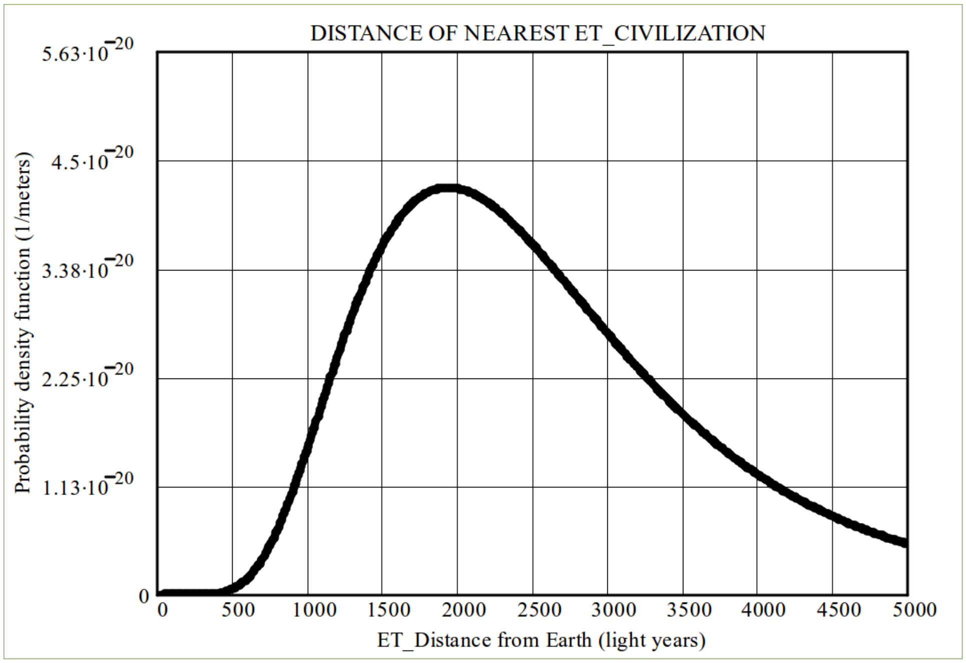


Figure 3.6: his is the probability of finding the nearest ExtraTerrestrial Civilization at the distance r from Earth (in light years) if the values assumed in the Drake Equation are those shown in Table 3.1. The relevant probability density function is given by equation (3.138). Its mode (peak abscissa) equals 1933 light years, but its mean value is higher since the curve has a high tail on the right: the mean value equals in fact 2670 light years. Finally, the standard deviation equals 1309 light years: **THIS IS GOOD NEWS FOR SETI, inasmuch as the nearest ET Civilization might lie at just 1 sigma = 2670-1309 = 1361 light years from us.**

From Figure 3.6, we see that the probability of finding ExtraTerrestrials is practically zero up to a distance of about 500 light years from Earth. Then it starts increasing with the increasing distance from Earth, and reaches its maximum at

$$r_{mode} \equiv r_{peak} = Ce^{-\mu/3}e^{-\sigma^2/9} \approx 1933 \; light \; years \tag{3.157}$$

This is the MOST LIKELY VALUE of the distance at which we can expect to find the nearest ExtraTerrestrial civilization.

It is not, however, the mean value of the probability distribution (3.138) for $f_{ETdistance}(r)$. In fact, the probability density (3.138) has an infinite tail on the right, as clearly shown in Figure 3.6, and hence its mean value must be higher than its peak value. As given by (3.124), its mean value is

$$r_{mean\_value} = Ce^{-\mu/3}e^{-\sigma^2/18} \approx 2670 \ light \ years \tag{3.158}$$

This is the MEAN (value of the) DISTANCE at which we can expect to find ExtraTerrestrials.

After having found the above two distances (1933 and 2670 light years, respectively), the next natural question that arises is: "what is the range, forth and back around the mean value of the distance, within which we can expect to find ExtraTerrestrials with "the highest hopes ?". The answer to this question is given by the notion of standard deviation, that we already found to be given by (123)

$$\sigma_{ETdistance} = Ce^{-\mu/3}e^{-\sigma^2/18}\sqrt{e^{\sigma^2/9}-1} \approx 1309 \ light \ years \tag{3.159}$$

More precisely, this is the so called 1-sigma (distance) level. Probability theory then shows that the nearest ExtraTerrestrial civilization is expected to be located within this range, i.e. within the two distances of (2670-1309) = 1361 light years and (2670+1309) = 3979 light years, with probability given by the integral of $f_{ETdistance}(r)$ taken in between these two lower and upper limits, that is:

$$\int_{1361 \ light \ years}^{3979 \ light \ years} f_{ETdistance}(r)dr \approx 0.75 = 75\% \tag{3.160}$$

In plain words: with 75% probability, the nearest ExtraTerrestrial civilization is located in between the distances of 1361 and 3979 light years from us, having assumed the input values to the Drake Equation given by Table 3.1. If we change those input values, then all the numbers change again.

3.8 The "data enrichment principle" as the best CLT consequence upn the statistical Drake equation (any number of factors allowed)

As a fitting climax to all the statistical equations developed so far, let us now state our "DATA ENRICHMENT PRINCIPLE". It simply states that "The Higher the Number of Factors in the Statistical Drake equation, The Better".

Put in this simple way, it simply looks like a new way of saying that the CLT lets the random variable Y approach the normal distribution when the number of terms in the sum (3.4) approaches infinity. And this is the case, indeed. However, our

"Data Enrichment Principle" has more profound methodological consequences that we cannot explain now, but hope to describe more precisely in one or more coming papers.

3.9 Conclusions about our statistical Drake equation

We have sought to extend the classical Drake equation to let it encompass Statistics and Probability.

This approach appears to pave the way to future, more profound investigations intended not only to associate "error bars" to each factor in the Drake equation, but especially to increase the number of factors themselves. In fact, this seems to be the only way to incorporate into the Drake equation more and more new scientific information as soon as it becomes available. In the long run, the Statistical Drake equation might just become a huge computer code, growing up in size and especially in the depth of the scientific information it contained. It would thus be Humanity's first "Encyclopaedia Galactica".

Unfortunately, to extend the Drake equation to Statistics, it was necessary to use a mathematical apparatus that is more sophisticated than just the simple product of seven numbers.

When this author had the honour and privilege to present his results at the SETI Institute on April 11th, 2008, in front of an audience also including Professor Frank Drake, he felt he had to add these words: "My apologies, Frank, for disrupting the beautiful simplicity of your equation".

3.10 Ackowledments about our statistical Drake equation

The author is grateful to Drs. Jill Tarter, Paul Davies, Seth Shostak, Doug Vakoch, Tom Pierson, Carol Oliver, Paul Shuch and Kathryn Denning for attending his first presentation ever about these topics at the "Beyond" Center of the University of Arizona at Phoenix on February 8th, 2008. He also would like to thank Dan Werthimer and his School of SETI young experts for keeping alive the interplay between experimental and theoretical SETI. But the greatest "thanks" goes of course to the Teacher to all of us: Professor Frank D. Drake, whose equation opened a new way of thinking about the past and the future of Humans in the Galaxy.

3.11 Habitable planets for man

Let us now change topics completely!

Rather than seeking for ETs in the Galaxy, we now seek for Habitable Planets for Man in the Galaxy. How many are there ? And how far from us is the nearest such a Habitable Planet ?

These topics seem to have been faced "seriously" for the first time in 1964 by Stephen H. Dole, then with the Rand Corporation (unfortunately, this author was unable to find when and where Dole was born and died, although he knows that Dole passed away from his friend Dr. Laurance Doyle of the SETI Institute).

Back in 1964, only three years had elapsed since Frank Drake had made known his now famous Drake equation. Dole learned the lesson of the Drake equation perfectly, and in his now famous book entitled "Habitable Planets for Man" (ref. [12]) he used the same mathematical structure as the Drake equation (1) in order to find the NUMBER OF HABITABLE PLANETS FOR MAN IN THE GALAXY. In other words, on page 82 of his book, he wrote the same mathematical thing as the Drake equation, but he applied it to Habitable Planets.

Figure 3.7 reproduces this crucial page of Dole's 1964 book, that nowadays can be freely downloaded from the Rand Corporation web site. This equation we shall now call "the classical Dole equation".

As we can see from Figure 3.7, the classical Dole equation is made up by TEN factors (instead of SEVEN factors as in the Drake equation):

$$N_{Hab} = N_s \cdot P_p \cdot P_i \cdot PM \cdot P_e \cdot PB \cdot PR \cdot PA \cdot PL \tag{3.161}$$

Here is the total number of Habitable Planets for Man in the Galaxy, and it is given by the product of the following TEN input numbers:

8. Ns is the number of stars in the suitable mass range 0.35 to 1.43 solar masses (this is Dole's assumption about to the mass of "habitable stars").

9. Pp is the probability that a given star has planets in orbit around it.

10. Pi is the probability that the inclination of the planet's equator is correct for its orbital distance.

11. PD is the probability that at least one planet orbits within an ecosphere.

12. PM is the probability that the planet has a suitable mass, 0.4 to 2.35 Earth masses (again, this is Dole's assumption in this regard).

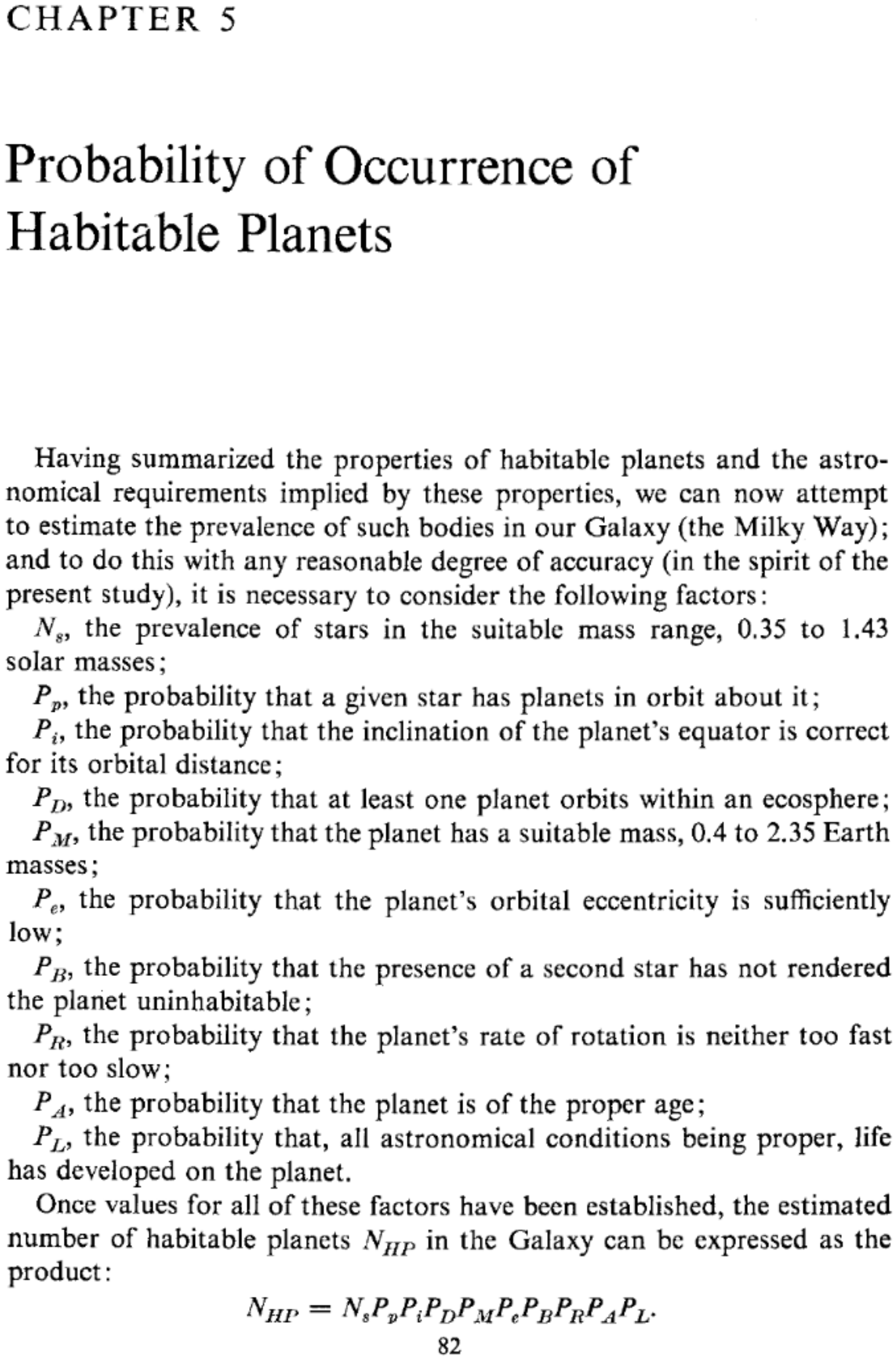
13. Pe is the probability that the planet's orbital eccentricity is sufficiently low.

14. PB is the probability that the presence of a second star has not rendered the planet uninhabitable.

15. PR is the probability that the planet's rate of rotation is neither too fast nor too slow.

16. PA is the probability that the planet is of the proper age.

17. PL is the probability that, all astronomical conditions being proper, life has developed on the planet.



CHAPTER 5

Probability of Occurrence of Habitable Planets

Having summarized the properties of habitable planets and the astronomical requirements implied by these properties, we can now attempt to estimate the prevalence of such bodies in our Galaxy (the Milky Way); and to do this with any reasonable degree of accuracy (in the spirit of the present study), it is necessary to consider the following factors:

N_s, the prevalence of stars in the suitable mass range, 0.35 to 1.43 solar masses;

P_p, the probability that a given star has planets in orbit about it;

P_i, the probability that the inclination of the planet's equator is correct for its orbital distance;

P_D, the probability that at least one planet orbits within an ecosphere;

P_M, the probability that the planet has a suitable mass, 0.4 to 2.35 Earth masses;

P_e, the probability that the planet's orbital eccentricity is sufficiently low;

P_B, the probability that the presence of a second star has not rendered the planet uninhabitable;

P_R, the probability that the planet's rate of rotation is neither too fast nor too slow;

P_A, the probability that the planet is of the proper age;

P_L, the probability that, all astronomical conditions being proper, life has developed on the planet.

Once values for all of these factors have been established, the estimated number of habitable planets N_{HP} in the Galaxy can be expressed as the product:

$$N_{HP} = N_s P_p P_i P_D P_M P_e P_B P_R P_A P_L.$$

82

Figure 3.7: Reproduction of page 82 of Stephen H. Dole's book "Habitable Planets for Man", first edition published in 1964, as it can be freely downloaded today from the web site of the Rand Corporation.

3.12 The statistical Dole equation

It is now natural to rename the above ten input variables of the classical Dole equation (3.161) as follows:

$$\begin{gathered}
D_1 = Ns \\
D_2 = Pp \\
D_3 = Pi \\
D_4 = PD \\
D_5 = PM \\
D_6 = Pe \\
D_7 = PB \\
D_8 = PR \\
D_9 = PA \\
D_{10} = PL
\end{gathered} \tag{3.162}$$

so that our classical Dole equation may be simply rewritten as

$$N_{Hab} = \prod_{i=1}^{10} D_i \tag{3.163}$$

We now let (3.163) undergo exactly the same changes that we applied to the classical Drake equation (3.1). In other words:

1) All the input variables on the right-hand side of (3.161) now become POSITIVE RANDOM VARIABLES.

2) All these random variables are supposed to be UNIFORMLY DISTRIBUTED with assigned mean values μ_{D_i} and standard deviations σ_{D_i}. It can then be shown that assigning them actually amounts to assigning the lower and upper limits (a_i and b_i, respectively) of each uniform random variable D_i.

3) As a consequence of these assumptions, the total number of Habitable Planets in the Galaxy, N_{Hab}, also becomes a RANDOM VARIABLE, that we already know to be LOGNORMALLY DISTRIBUTED from our previous similar work about the Drake equation.

Thus, we may now call (3.163) the STATISTICAL DOLE EQUATION. The notation obviously comes from "Dole", but the lucky coincidence that both Frank Drake's and Stephen Dole's family names both start with a "D" will save us from introducing new notations other than these D_i!

It is true that the classical Drake equation (3.1) and the classical Dole equation (3.161) have a different number of factors (7 and 10, respectively), but... frankly speaking, who cares ? This perfectly in line with what we did already for the Drake equation, and so THE NUMBER OF FACTORS IN BOTH (3.1) AND (3.163) IS TOTALLY IRRELEVANT, THANKS TO THE CENTRAL LIMIT THEOREM !

3.13 The number of habitable planets for man in the galaxy follows the lognormal distribution

We now just repeat the same arguments developed for the Drake equation to immediately conclude that: THE TOTAL NUMBER OF HABITABLE PLANETS IN THE GALAXY FOLLOWS THE LOGNORMAL DISTRIBUTION given in Table 3.1.

3.14 The distance between any two nearby habitable planets follows the maccone distribution

Again we now just repeat the same arguments developed for the Drake equation to immediately conclude that THE DISTANCE BETWEEN ANY TWO NEARBY HABITABLE PLANETS FOLLOWS THE MACCONE DISTRIBUTION given in Table 3.2.

3.15 a numerical example: some a hundred million habitable planets exist in the galaxy !

We just need to complete this paper by giving a numerical example of how our Statistical Dole equation (3.163) works, and this we will do in the present section.

Consider the following Input Table 3.4. This is in principle comparable to Input Table 3.1 for the Statistical Drake equation. In fact, the arguments developed by Dole in Chapter 5 of ref. [12] do provide the mean values of each , but only such mean values, and not the relevant standard deviations, of course.

To set up a working example of the Statistical Dole Equation, however, we must assign the ten standard deviations also, that were not given by Dole and are unknown to this author from the current scientific literature about these matters.

No problem. In order to cut short, this author thus simply assigned the value of 1/10 (i.e. 10%) to each of the ten standard deviations listed in Input Table 3.4, and the Input Table 3.2 is now complete.

Having assumed all the values listed in Input Table 3.4 as the input values, a new (unpublished) MathCad code was created by this author for the Statistical Dole Equation. For the input values of Input Table 3.4, this code yielded the results described hereafter.

First of all, the lognormal probability density for the random variable is shown in Figure 3.8. We see that the peak (i.e. the mode) corresponds to about ten million planets, but the tail is rather long.

To quantify these remarks, let us first point out that the author's MathCad code yields the following numerical values for the two parameters μ and σ given by the last two rows in both Tables 3.1 and 3.2:

$$\begin{aligned} \mu_{Hab} &= 1.76268289631314 \cdot 10^{1} \\ \sigma_{Hab} &= 1.27010132908265 \cdot 10^{0} \end{aligned} \tag{3.164}$$

Then, the mean value of the random variable, given by the fourth row in Table 1, is given by

$$\langle N_{Hab} \rangle = e^{\mu_{Hab}} e^{\sigma_{Hab}^2/2} = 1.012 \cdot 10^{8} \approx 100 \; millions \tag{3.165}$$

In other words, our statistical (and thus more serious, scientifically speaking) treatment of the Dole equation yields 100 million expected Habitable Planets in the Galaxy.

This figure is higher than the 35 million given by the classical Dole equation, and much higher than the value of the mode (10 million) shown by the lognormal curve in Figure 3.8.

Table 3.4: Input values (i.e. mean values and standard deviations) for the ten Dole uniform random variables D_i. The first column on the left lists the ten input sheer numbers that also are the mean values (middle column). The last column on the right lists the ten input standard deviations. The bottom line is the classical Dole equation (3.161). So, the number of Habitable Planets in the Galaxy, given by the classical Dole equation just as a sheer number, is 35 millions 171 hundred thousand and 930.

$\mathrm{Ns} := 6.448 \cdot 10^{8}$	$\mu\mathrm{Ns} := \mathrm{Ns}$	$\sigma\mathrm{Ns} := 1 \cdot 10^{7}$
$\mathrm{Pp} := 1.0$	$\mu\mathrm{Pp} := \mathrm{Pp}$	$\sigma\mathrm{Pp} := \frac{10}{100}$
$\mathrm{Pi} := 0.81$	$\mu\mathrm{Pi} := \mathrm{Pi}$	$\sigma\mathrm{Pi} := \frac{10}{100}$
$\mathrm{PD} := 0.63$	$\mu\mathrm{PD} := \mathrm{PD}$	$\sigma\mathrm{PD} := \frac{10}{100}$
$\mathrm{PM} := 0.19$	$\mu\mathrm{PM} := \mathrm{PM}$	$\sigma\mathrm{PM} := \frac{10}{100}$
$\mathrm{Pe} := 0.94$	$\mu\mathrm{Pe} := \mathrm{Pe}$	$\sigma\mathrm{Pe} := \frac{10}{100}$
$\mathrm{PB} := 0.95$	$\mu\mathrm{PB} := \mathrm{PB}$	$\sigma\mathrm{PB} := \frac{10}{100}$
$\mathrm{PR} := 0.9$	$\mu\mathrm{PR} := \mathrm{PR}$	$\sigma\mathrm{PR} := \frac{10}{100}$
$\mathrm{PA} := 0.7$	$\mu\mathrm{PA} := \mathrm{PA}$	$\sigma\mathrm{PA} := \frac{10}{100}$
$\mathrm{PL} := 1$	$\mu\mathrm{PL} := \mathrm{PL}$	$\sigma\mathrm{PL} := \frac{10}{100}$

$$N_{Hab} := Ns \cdot Pp \cdot Pi \cdot PD \cdot PM \cdot Pe \cdot PB \cdot PR \cdot PA \cdot PL$$

$$N_{Hab} := 3.5171930508624 \times 10^{7}$$

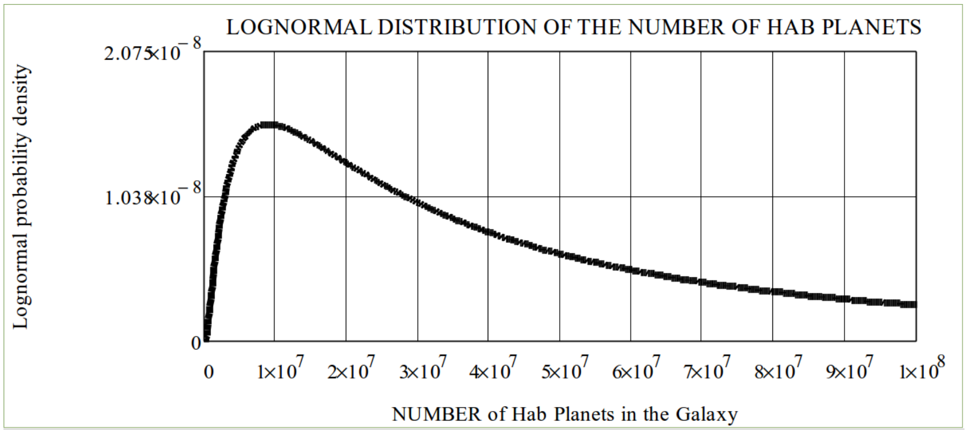


Figure 3.8: The lognormal probability density of the overall NUMBER of Habitable Planets in the Galaxy as described in Stephen H. Dole's book "Habitable Planets for Man", first edition published in 1964, and implemented by assigning a 10% standard deviation to all the ten input random variables listed in Table 3.4.

The last result, stating that there are about 100 million Habitable Planets in the Galaxy, is of course good news for the future "human conquest of the Galaxy" (if there will ever be one!), since it raises to 100 million the expected number of "Earths" to land on!

But what about the standard deviation around the mean value given by (160) ? Table 3.1, row 6, shows that such a standard deviation of the random variable is given by

$$\sigma_{N_{Hab}} = e^{\mu_{Hab}} e^{\sigma^2_{Hab}/2} \sqrt{e^{\sigma^2_{Hab}} - 1} = 2.0 \cdot 10^8 \approx 200 \; millions \qquad (3.166)$$

In other words, the standard deviation of the number of Habitable Planets is 200 millions. And so, with probability 1-sigma, we might expect the actual number of Habitable Planets to rise up 100 million plus 200 million = 300 million.

Finally, the median (fifty-fifty probability of the lognormal distribution shown in Figure 3.7) yields a value of

$$median = m = e^{\mu_{Hab}} = 4.521 \cdot 10^7 \approx 45 \; millions \qquad (3.167)$$

3.16 Distance (Maccone) distribution of the nearest habitable planet to us according to the previous

Next comes the DISTANCE distribution of the nearest Habitable Planet to us (of course under the easy hypothesis that the distribution of Habitable Planets in the Galaxy is UNIFORM). Well, from the third row of Table 3.2 it follows that the relevant probability density is given by the Maccone distribution, and this is plotted in Figure 3.9.

The mean value of the Maccone distribution is given by the fifth row in Table 3.2, that is, for the data given by the Input Table 3.4

$$\langle Hab \; Distance \rangle = Ce^{-\mu/3} e^{\sigma^2_{Hab}} = 8.8 \cdot 10^1 ly \approx 88 ly \qquad (3.168)$$

The relevant standard deviation is given by the seventh row in Table 3.2, and reads

$$\sigma_{Hab \; Distance} = Ce^{-\mu_{Hab}/3} e^{\sigma^2/18} \sqrt{e^{\sigma^2/9} - 1} = 3.9 \cdot 10^1 ly \approx 40 ly \qquad (3.169)$$

Thus, with probability 1 sigma, it should not be hopeless to expect a detection of a Habitable Planet even at, say, just 88-40 $= 48 \sim 50$ light years from us.

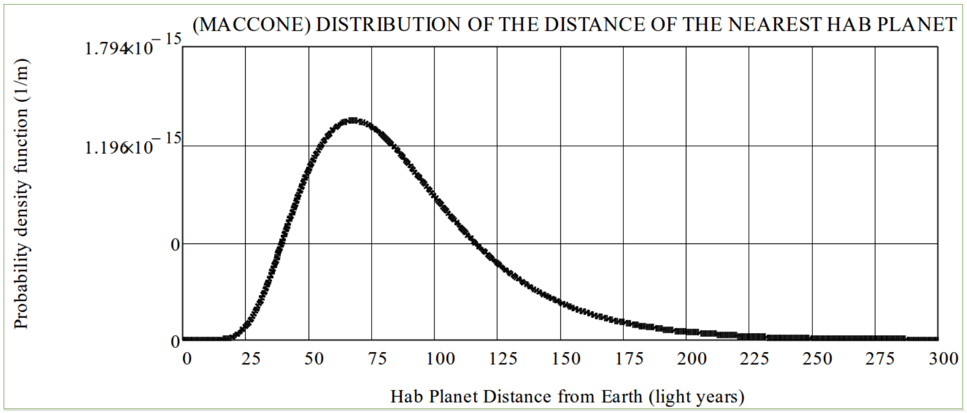


Figure 3.9: The Maccone PROBABILITY DISTRIBUTION OF THE DISTANCE OF THE NEAREST HABITABLE PLANET TO US IN THE GALAXY for the data of the Input Table 3.4 assumed as inputs to the Statistical Dole Equation (3.163). A glance to this plot immediately reveals that it is "hopeless" to expect to detect a Habitable Planet at distances smaller than 25 light years from us, since the value of the Maccone distribution is practically zero at such distances. Thus, future Interstellar Spacecraft designers should keep this lower bound in mind wished they land on Habitable Planets, rather than just on "any Planet". Also, the curve reaches its peak (mode) at about 67 light years from us, its mode (fifty-fifty probability) at about 80 light years and, above all, its mean value at 88 light years from us. The relevant standard deviation turns out to be about 40 light years, since the distribution tail is rather "short".

3.17 Comparing the statistical dole and drake equations: number of habitable planets vs. number of et civilizations in this galaxy

It is now appropriate to make a comparison between the number of Habitable Planets and the number of expected ET Civilizations in the Galaxy.

In other words, we want the "get the feeling" of the numbers that we have worked out in this paper just to see if the comparison among them "makes sense".

This we can do by putting on a same table

Table 3.5: Comparing the results of the Statistical Dole and Drake equation found by inputting to them the Input Table 3.4 and 3.1, respectively.

	Statistical Dole Equation	Statistical Drake Equation
Mean Value of the TOTAL NUMBER of	Habitable Planets in the Galaxy ~ 100 million	ET Civilizations in the Galaxy ~ 4590
Standard Deviation of the TOTAL NUMBER of	Habitable Planets in the Galaxy ~ 200 million	ET Civilizations in the Galaxy ~ 11195
Mean Value of the DISTANCE of	Nearest Habitable Planet ~ 88 light years	Nearest ET Civilization ~ 2670 light years
Standard Deviation of the DISTANCE of	Nearest Habitable Planet ~ 40 light years	Nearest ET Civilization ~ 1309 light years

1) the mean value and standard deviation of the total number of both Habitable Planets and ET Civilizations, and

2) the mean value and standard deviation of their respective distances from us (of course, under the hypothesis that both of them are uniformly scattered throughout the Galaxy).

The result is the following Table 3.3, clearly showing that how much "more rare" the ET Civilizations are with respect to the Habitable Planets. Roughly, one has:

$$\frac{\langle N_{Hab} \rangle}{\langle N_{ET} \rangle} = \frac{100 \ millions}{4950} \approx 20,202 \tag{3.170}$$

so that the Habitable Planets seem to 20,000 more frequent than ET Civilizations, or, if you wish, only one ET Civilization emerges out of 20,000 Habitable Planets.

As for the distances, the ratio is the other way round:

$$\frac{\langle Hab \ Distance \rangle}{\langle N_{ET \ Distance} \rangle} = \frac{2670 ly}{88 ly} \approx 30.340 \tag{3.171}$$

meaning that ETs are, on the average, 30 times further out that Habitable Planets.

And all these results, however, are just statistical, of course!

3.18 SEH, the "statistical equation for the habitables" is just the statistical dole equation

So far we have referred to (3.163) as to the Statistical Dole equation. In view of further improvements in the mathematical analysis of this equation, however, it appear to be suitable to rename it "SEH", an acronym standing for "STATISTICAL EQUATION FOR THE HABITABLES". This will be clear in the future papers by the author, where a number possibly higher than ten will be the new number of independent, uniform random variables describing the equation inputs.

These topics have to be deferred to a further paper, though.

3.19 The classical coral model of galactic colonization

The following description of the Coral Model of Galactic Colonization is taken from the book "Life in the Universe" – second edition – 2007, by Jeffrey O. Bennett and G. Seth Shostak, ref. [10], see especially pages 459 and 476 there.

Let's start by assuming that another civilization decided to start sending out spacecraft to colonize other habitable planets. How long would it take for this civilization to colonize the entire Galaxy?

The answer clearly depends on the civilization's technological capabilities. For example, if it has the technology to build spacecraft that can travel at speeds close to the speed of light, then it could add colonies throughout the Galaxy fairly quickly, since trips between nearby stars would take only a few years. Perhaps surprisingly, the conclusion is not that much different if we assume much lower speeds.

In fact, consider a civilization that has nuclear rockets such as the Project Orion (1958-63, see the site[1]) or Project Daedalus (1973-78, see the site[2]) rockets: such rockets do not seem that much beyond our technological grasp and they might attain speeds of about 10% of the speed of light (0.1 c). Given that a typical distance between star system in our region of the Galaxy is about 5 light-years, a nuclear spacecraft travelling at 10% of the speed of light could journey from one star system to the next in about 50 years. This trip would be possible in a human lifetime and might be practical is the colonizers have found ways to hibernate during the voyage or if they have somewhat longer life spans than we do (either naturally or through

[1] http://en.wikipedia.org/wiki/Project_Orion_(nuclear_propulsion)

[2] http://en.wikipedia.org/wiki/Project_Daedalus

medical intervention).

After arriving at a new star system, the colonists establish themselves and begin to increase the population. Once the population has grown sufficiently, these colonists send their own pilgrims into space, adding yet more star systems to the growing civilization. Thus, the process starts at the home star system and the first few colonies are located within just a few light-years. These colonies then lead to other colonies at greater distances, as well as at unexplored locations in between. The growth tends to expand the empire around the edges of the existing empire, much like to growth of coral in the sea. For this reason, this type of colonization model is often called a coral model of Galactic colonization.

The overall result is a gradually expanding region in which all habitable planets are colonized. The colonization rate depends on the speed of spacecraft and the time it takes each colony to start sending their own spacecraft to other stars. For travel 10% the speed of light and assuming that it takes 150 years before each colony's population grows enough to send out more colonists, the calculations that we shall make in the next section show that the inhabited region of the Galaxy expands outwards from the home world at about 1% of the speed of light. Thus, if the home star is near one edge of the Galactic disk, so that colonizing the entire Galaxy means inhabiting star systems 100,000 light-years away, the civilization would expand through the entire Galaxy in about 10 million years. The required time would be a few million years less if the home star is in a more central part of the Galaxy.

For an even more conservative estimate, suppose the colonists have rockets that travel at only 1% of the speed of light and that it takes each new colony 5,000 years until it is ready to send out additional colonists. Even in this case, the region occupied by this civilization would grow at a rate of roughly 1/1000 (0.1 %) the speed of light and the entire Galaxy would be colonized in 100 million years. This is still a very short time compared to the time that has been available for civilizations to arise (4.5 billion years for Humanity), further deepening the mystery of why we see no evidence that anyone else has done it by now.

This is, of course, the now well-known Fermi Paradox, first stated by Enrico Fermi (1901-1954) to his colleagues during a lunch discussion at Los Alamos back in 1950 (for a good summary, see the Wikipedia site:[3]).

But let us go back to the coral model of Galactic colonization: we now want to cast it into a sound mathematical theory. First of all, let us write down the fact that the overall expansion speed of the empire, v_{exp} , is the ratio of the average distance among any two nearby stars, D, to the sum of two times:

1) the time of actual spaceflight from one star to the next one, t_{flight}, plus

[3] http://en.wikipedia.org/wiki/Fermi_paradox

2) the time t_{col} requested to colonize a planet, i.e. to develop there a civilization until the time is ripe for one more spaceflight jump to the next star. That is, we assume that the equation holds

$$v_{exp} = k\frac{D}{t_{flight} + t_{col}} \tag{3.172}$$

We wrote a factor k in front of (3.172) just to take into account the "zigzag" motion of expansion from one star to the next in three-dimensional space. This k is explained by Bennett and Shostak in their book [10] on page 476, in a way that we rephrase as follows. The purely numerical factor k would be equal to 1 only if the colonization was always directed straight outward from the home star. In reality, the colonists will sometimes go to uncolonized star systems in other directions, so we will introduce a constant k that accounts for this zigzag motion. For zigzag in three-dimensional space, we assume that

$$k = \frac{1}{2} \tag{3.173}$$

Next to (3.172) and (3.173) one more equation is needed to cast our coral expansion model in mathematical form: this is the obvious relationship between the flight time from one star to the next one, and the corresponding spaceship (average) speed, v_{ss}, that is

$$t_{flight} = \frac{D}{v_{ss}} \tag{3.174}$$

Inserting then (3.173) and (3.174 into (3.172), a little rearranging yields

$$v_{exp}(D, t_{col}, v_{ss}) = k\frac{v_{ss}D}{D + v_{ss}t_{col}} \tag{3.175}$$

This is the expression of the expansion speed of the empire throughout the Galaxy that we want to concentrate on in the coming sections.

An immediate consequence of (3.175) is the Galaxy colonization time, denoted T_{Galaxy}, i.e. the overall time that our expanding empire will need to colonize the whole Galaxy. If a civilization starts conquering the Galaxy from the outskirts (more or less like ours!), the largest possible amount of time is clearly given by

$$T_{Galaxy} = \frac{2R_{Galaxy}}{v_{exp}} = \frac{2R_{Galaxy}}{k} \cdot \frac{D + v_{ss}t_{col}}{v_{ss}D} \tag{3.176}$$

As we said, it could take a little less if the conquerors lived nearby the center of the Galaxy, so, to be conservative, let us take (3.176) for granted and just rewrite it as

$$T_{Galaxy} = \frac{2R_{Galaxy}}{k} \cdot \left(\frac{1}{v_{ss}} \frac{t_{col}}{D} \right) \tag{3.177}$$

or

$$T_{Galaxy} = \frac{2R_{Galaxy}}{k} \cdot \frac{t_{col}}{D} + \frac{2R_{Galaxy}}{kv_{ss}} \tag{3.178}$$

Let us now introduce two positive constants a and b:

$$\begin{gathered} a = \tfrac{2R_{Galaxy}}{k} \approx 200000 \; light \; years \\ b = \tfrac{2R_{Galaxy}}{kv_{ss}} \approx 20 \; millions \; years \; for \; v_{ss} \approx 0.01c \end{gathered} \tag{3.179}$$

Then, the whole Galaxy colonization time T_{Galaxy} takes the final form that we will use in the coming sections

$$T_{Galaxy} = a\frac{t_{col}}{D} + b \tag{3.180}$$

3.20 The classical Fermi paradox (1950)

Let us now consider three different numerical cases of (174) and check them against each other.

1) First, suppose that one has

$$\begin{gathered} k = 1/2 \\ v_{ss} = 0.1c \\ D = 5ly \\ t_{col} = 150yr. \end{gathered} \tag{3.181}$$

Then, (3.175) and (3.180) yield, respectively:

$$\begin{gathered} v_{exp} = 3747 \; km/s = 0.0125c \\ T_{Galaxy} \approx 8 \; millions \; years \end{gathered} \tag{3.182}$$

2) Second, suppose that one has

$$\begin{gathered} k = 1/2 \\ v_{ss} = 0.01c \\ D = 5ly \\ t_{col} = 1000yr. \end{gathered} \tag{3.183}$$

Then, (3.175) and (3.180) yield, respectively:

$$\begin{gathered} v_{exp} = 500\ km/s = 0.001c \\ T_{Galaxy} \approx 60\ millions\ years \end{gathered} \tag{3.184}$$

3) Third, assume the HUMAN CASE. By this we mean that the HABITABLE PLANETS are just those planets HABITABLE BY HUMANS, and not planets of any other kind! Thus, we must apply the classical Dole equation (3.161) of the "Habitable Planets for Man" book, reach the important conclusion that the average distance between planets habitable by Humans is 84 light years. In other words, let us assume the inputs:

$$\begin{gathered} k = 1/2 \\ v_{ss} = 0.01c \\ D = 84ly \\ t_{col} = 1000yr. \end{gathered} \tag{3.185}$$

Then, (3.175) and (3.180) yield, respectively:

$$\begin{gathered} v_{exp} = 1339\ km/s = 0.004c \\ T_{Galaxy} \approx 22\ millions\ years \end{gathered} \tag{3.186}$$

So, about 22 million years would be the overall time necessary for Humankind to colonize the whole Milky Way (was no alien civilization trying to stop us, of course!) had Humans spaceships capable of traveling at 1% of the speed of light and was the average colonization time for every new planet about 1000 years.

The basic difference between the Humanity expansion model and the two previous models was of course the difference in the average distance among habitable extrasolar planets, Actually, it is interesting to take the limit of (3.180) for the distance D increasing more and more, i.e. $D \to \infty$. This yields

$$\lim_{D\to\infty} T_{Galaxy} = \frac{2R_{Galaxy}}{k} \cdot \frac{1}{v_{ss}} \tag{3.187}$$

meaning that the spaceship speed v_{ss} plays an increasing role in the Galaxy colonization when the average distance D increases more an more. In other words still, if ETs of a certain "race" can live on fewer planets only, then they must have much faster spaceships to colonize the Galaxy than ETs than can live on a variety of planets! Not a small result at all...

In conclusion, from the above three examples we see that the time of colonization of the whole Galaxy seems to be of the order of some tens million year: just a blink compared the Galaxy age of about 10 billion years, and that is the Fermi paradox, of course. Many papers and books have been written about the Fermi paradox, especially in recent years, but none has been able to solve it so far. In the coming sections, however, we are going to present the STATISTICAL Fermi paradox.

3.21 The statistical coral model of galactic colonization

The goal of this paper is to present for the first time the statistical generalization of the classical Fermi paradox described in the previous section.

This is a difficult mathematical job. In fact, consider first the statistical expansion speed (3.175) of the empire, that we rewrite here in the form

$$V_{exp}(D, T_{col}) = k \frac{v_{ss} D}{D + v_{ss} + T_{col}} \tag{3.188}$$

We follow here the convention of denoting all random variables by capitals, contrary to the ordinary (deterministic) variables that we keep denoting by lower-case letters. Thus, in (3.188) we have three random variables in the game, V_{exp}, D and T_{col}, while v_{vss} is just a real positive, known parameter. Why do we proceed this way? Because:

1) The (average) spaceship speed, allowing us to jump from a star to the next one, is entirely under human (or ET's) control, and so can be regarded just as a sheer number rather than a random variables. This assumption simplifies things greatly from the mathematical point of view making the problem still mathematically solvable.

2) On the contrary, the colonization time T_{col} is indeed a random variable, inasmuch as we don't know in advance what difficulties we will have to face in order to colonize a new planet, and so we don't know how long it will actually take to develop facilities on this planet that finally make us ready for the next jump. Because of the random character of T_{col} we denoted it in capitals in (3.188).

3) Finally, D, the (average) distance in between any two nearby Habitable Planets follows the Maccone distribution, as we already know. We thus have for its probability density function (pdf) the Maccone distribution

$$f_D(d) = \frac{3}{d}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(ln(C^3/d^3)-\mu\right)^2}{2\sigma^2}} \tag{3.189}$$

But then, what about the probability density function of the new (positive) random variable T_{col} yielding the amount of time needed to colonize a new planet? Well, we are of course free to choose any density function we wish, but the best choice seems to be a LOGNORMAL pdf because (as we already know from the Drake and Dole equations) this can be thought of as the multiplicative product many random variables each of which is positive. Thus, we assume that

$$f_{T_{col}}(t) = \frac{1}{yr}\frac{yr}{t}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(ln(t/yr)-\mu)^2}{2\sigma^2}} \tag{3.190}$$

A question now raises: is any lognormal pdf fully specified if we know in advance both its mean value and its standard deviation? The answer to this question is "yes". In fact, consider the relevant mean value and standard deviation, given by the 4th and 6th line in Table 3.1, respectively, and set up the two simultaneous equations

$$\begin{aligned} \langle N\rangle &= e^{\mu}e^{\sigma^2/2} \\ \sigma_N &= e^{\mu}e^{\sigma^2/2}\sqrt{e^{\sigma^2}-1} \end{aligned} \tag{3.191}$$

This system may indeed be inverted by dividing the first equation by the second one, solving for σ^2, replacing this σ^2 into the first equation and finally solving that one for μ. One thus gets:

$$\begin{aligned} \mu &= ln\left(\frac{\langle N\rangle^2}{\sqrt{\langle N\rangle^2+\sigma_N^2}}\right) \\ \sigma &= \sqrt{ln\left(\frac{\sigma_N^2}{\langle N\rangle^2}+1\right)} \end{aligned} \tag{3.192}$$

But that was good for the lognormally-distributed random variable N, whereas here we must apply (3.192) to the lognormally-distributed new random variable T_{col}. Thus, $\langle N\rangle$ must be replaced by the mean value of the colonization time, μ_{t_col}, and σ_N must be replaced by the standard deviation of the colonization time, σ_{t_col}. Thus, for the colonization time, (3.192) becomes

$$\begin{aligned} \mu &= ln\left(\frac{\mu^2_{t\_col}}{\sqrt{\mu^2_{t\_col}+\sigma^2_{t\_col}}}\right) \\ \sigma &= \sqrt{ln\left(\frac{\sigma^2_{t\_col}}{\mu^2_{t\_col}}+1\right)} \end{aligned} \tag{3.193}$$

For instance, suppose that the mean colonization time equals 1000 years, with a standard deviation of ±500 yr. Then replacing these two values (previously divided by yr to make things dimensionally correct) into (3.193), yields

$$\begin{aligned} \mu &= ln\left(\frac{\mu^2_{t\_col}}{\sqrt{\mu^2_{t\_col}+\sigma^2_{t\_col}}}\right) = 6.7961835 \\ \sigma &= \sqrt{ln\left(\frac{\sigma^2_{t\_col}}{\mu^2_{t\_col}}+1\right)} = 0.4723807 \end{aligned} \tag{3.194}$$

and the relevant lognormal distribution (3.190) is thus perfectly determined. This lognormal pdf of the time (in years) needed to colonize a new planet if one assumes a mean value of 1000 years plue or minus a standard deviation of 500 years is plotted in the following Figure 3.9. In our opinion, much of History on Earth, such as the colonization of America, is similar.

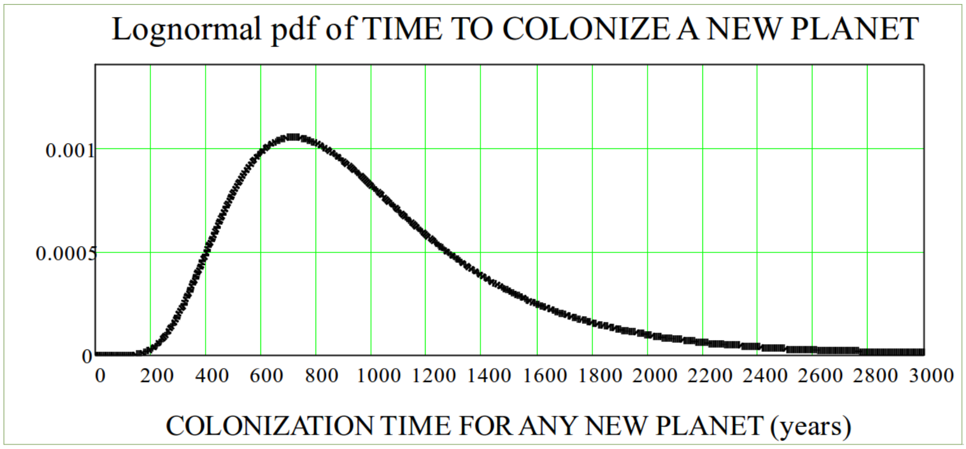


Figure 3.10: The lognormal distribution of the time needed to colonize each new planet in our statistical extension of the coral Galactic expansion model, assuming that, for instance, it takes 1000 years plus or minus 500 years to colonize that planet. Our assumption that this probability distribution must be lognormally distributed is not "arbitrary", but is ensured by the fact that the lognormal distribution can be thought of as the multiplicative product many random variables each of which is positive and unknown (central limit theorem of statistics in its multiplicative version, rather than in its additive version, where the lognormal would be replaced by a Gaussian). This author thinks that just the same model could be applied to Human History on Earth also (like the colonization of the Americas by the Europeans), with the result of creating a new scientific discipline called "Mathematical History".

3.22 Finding the probability distribution of the overall time needed to colonize the whole galaxy

In this section we face the (difficult) mathematical problem of finding the probability distribution (i.e. the probability density function, or pdf) of the overall time needed to colonize the Galaxy. This is given by the positive random variable T_{Galaxy} defined by (3.180) where the variable T_{col} is now a positive random variable lognormally distributed as described in the last section, while D is the random variable yielding the average distance between any two nearby "Habitable Planets for Man" (as Steve Dole would have said) and given by the Maccone distribution (3.119). In other words, in this section we are going to find the pdf of the positive random variable

$$T_{Galaxy} = a\frac{T_{col}}{D} + b \tag{3.195}$$

where all capitals denote random variables while a and b are just the two positive constants defined by (3.179). In other words still, apart from the constants a and b, we really must find the pdf of the quotient of two random variables defined, in loose terms, by the fraction

$$\frac{T_{col}}{D} = \frac{lognormal}{Maccone} \tag{3.196}$$

Standard textbooks about Probability Theory (for instance, see the book by Papoulis and Pillai, "Probability Random Variables and Stochastic Processes", ref. [6], in particular pages 186-187, equation (6-59)) tell us that the random variable Z, quotient of the two random variables X and Y

$$Z = \frac{X}{Y} \tag{3.197}$$

has its pdf given by the integral

$$f_Z(z) = \int_{-\infty}^{\infty} |y| f_{XY}(yz, y)dy \tag{3.198}$$

where the function $f_{XY}(..., ...)$ is the joint pdf of the two random variables X and Y. Now, the two random variables

$$\begin{array}{c} X = T_{col} \\ Y = D \end{array} \tag{3.199}$$

are for sure statistically independent of each other, inasmuch as the average time to colonize a planet T_{col} is a "Human Thing" (or an "Alien Thing", if referred to an Alien Civilization), while the average distance D among any two nearby Habitable Planets is an "Astrophysical Thing", depending on how the Galaxy formed billion of years ago. Thus, their joint pdf $f_{XY}(..., ...)$ simply is the product of the two pdfs, i.e. the lognormal one (3.190) and the Maccone one (3.189), i.e.

$$\begin{aligned} & f_{T_{col}D}(t,d) = f_{T_col}(t) f_D(d) = \\ & = \frac{1}{t}\frac{1}{\sqrt{2\pi}\sigma_{t\_col}} e^{-\frac{\left(ln(t/yr) - \mu_{t\_col}\right)^2}{2\sigma_{t\_col}}} \frac{3}{d}\frac{1}{\sqrt{2\pi}\sigma_D} e^{-\frac{\left(ln(C^3/d^3) - \mu_D\right)^2}{2\sigma_D}} \end{aligned} \tag{3.200}$$

Rearranging, this becomes

$$f_{T_col}(t)f_D(d) = \frac{3}{2\pi t d\sigma_{t\_col}\sigma_D} e^{-\frac{\left(ln(t/yr)-\mu_{t\_col}\right)^2}{2\sigma_{t\_col}}} e^{-\frac{\left(ln(C^3/d^3)-\mu_D\right)^2}{2\sigma_D}} \tag{3.201}$$

This is the joint pdf that must be introduced into the integral (3.198). Notice, however, that the integral actually ranges from 0 to infinity only, since both t and d do so. The modulus affecting y in (3.198) thus disappears also, and we are just left with the computation of the definite integral

$$f_{\frac{T_{col}}{D}}(z) = \int_0^\infty y f_{T_{col}D}(zy, y)dy \tag{3.202}$$

That is

$$f_{\frac{T_{col}}{D}}(z) = \frac{3}{2\pi\sigma_{T_{col}}\sigma_D} \int_0^\infty y\frac{1}{zy}\frac{1}{y} e^{-\frac{\left(ln(zy/yr)-\mu_{t\_col}\right)^2}{2\sigma_{t\_col}^2}} e^{-\frac{\left(ln(C^3/d^3)-\mu_D\right)^2}{2\sigma_D^2}} \tag{3.203}$$

This is a tough integral to compute. Basically, it can be reduced to the Gauss integral, i.e. to the normalization condition of the ordinary Gaussian or normal curve, but many, many steps are required to perform the integration with respect to y. This author, when faced with its computation, turned to Macsyma, the wonderful Computer Algebra code that was created at the MIT Artificial Intelligence Laboratory back in the 1960s to let NASA re-compute analytically the orbits requested for the Apollo astronauts to safely reach the Moon and come back. So, Macsyma was able to perform the integration in (197) in a matter of seconds. The outcome was the function of z shown in Figure 10. As one can see, this function of z is a complicated mix of exponentials in z through the natural log of z squared, times a power of z at the denominator, times many other constants, like the dimensional yr = year.

Though the pdf (3.204) is difficult to handle by hand, it can be easily handled by Macsyma. Thus, one can prove that it fulfills indeed the normalization condition

$$\int_0^\infty f_{\frac{T_{col}}{D}}(z)dz = 1 \tag{3.205}$$

A similar calculation then shows that the mean value of the quotient of random variables T_{col}/D reads

$$\mu_{\frac{T_{col}}{D}} = \int_0^\infty z f_{\frac{t_{col}}{D}}(z)dz = e^{\frac{\sigma_D^2+6\mu_D+9\sigma_{t\_col}^2+18\mu_{t\_col}}{18}} \frac{yr}{C} \tag{3.206}$$

$$\boxed{f_{T_{Galaxy}}(t)=\frac{1}{a}\cdot\frac{3\sqrt{2}\,C^{\frac{3\mu_D+9\mu_{t\_col}}{o_D^2+9\sigma_{t\_col}^2}}\cdot e^{-\frac{\left(\mu_D+3\mu_{t\_col}\right)^2+\left(3\cdot\log\left(C\,yr\,\frac{t-b}{a}\right)\right)^2}{2\left(o_D^2+9\sigma_{t\_col}^2\right)}}}{2\sqrt{\pi}\,yr^{\frac{3\mu_D+9\mu_{t\_col}-9\log C-9\log\left(\frac{t-b}{a}\right)}{o_D^2+9\sigma_{t\_col}^2}}\cdot\left(\frac{t-b}{a}\right)^{\frac{o_D^2+9\sigma_{t\_col}^2-3\mu_D-9\mu_{t\_col}+9\log C}{o_D^2+9\sigma_{t\_col}^2}}\sqrt{o_D^2+9\sigma_{t\_col}^2}}} \tag{3.204}$$

Figure 3.11: The probability density function (pdf), i.e. a function of z, of the quotient of two positive random variables: T_{col}, the time requested to colonize a new planet (in years = yr) over D, the average distance between nearby planets belonging to different stellar system (in units of the constant C=28845 light years, typical of the Milky Way Galaxy).

The corresponding variance was again found by Macsyma through a similar calculation, and reads

$$\sigma^2_{\frac{T_{co}}{D}} = e^{\sigma^2/9+2\mu_D/3+\sigma^2_{t\_col}+2\mu_{t\_col}}\left(e^{\sigma_D^2/9+\sigma^2_{t\_col}}-1\right)\frac{yr^2}{C^2} \tag{3.207}$$

Its square root is thus the relevant standard deviation:

$$\sigma_{\frac{T_{col}}{D}} = e^{\frac{\sigma^2/9+2\mu_D/3+\sigma^2_{t\_col}+2\mu_{t\_col}}{2}}\sqrt{e^{\sigma_D^2/9+\sigma^2_{t\_col}}-1}\,\frac{yr}{C} \tag{3.208}$$

In order to find the mode of the pdf (3.204), i.e. the abscissa of its peak, we must find out the first derivative of (3.204) with respect to z and then set the resulting equation equal to zero. Good old Macsyma did a good job again, and the two results are the two abscissas of the minimum of (3.61), obviously at $z = 0$, and of the maximum (i.e. the peak, or mode) at

$$z_{mode} = e^{-\sigma_D^2/9+\mu_D/3-\sigma^2_{t\_col}+\mu_{t\_col}}\cdot\frac{yr}{C} \tag{3.209}$$

Finally, we had Macsyma prove that the two inflexion points of (3.204), that is the one before and the one after the peak (or mode), are found as the two roots of a quadratic algebraic equation in $log(z)$ that is found after equalling to zero the second derivative of (3.204) with respect to z. Thus, it is found that the abscissas of such two inflexion points of (3.204) read, respectively

$$z_{inflexion\_1} = e^{\frac{\sqrt{\sigma_D^2+9\sigma_{t\_col}^2}\sqrt{\sigma_D^2+9\sigma_{t\_col}^2+36}+3\sigma_D^2-6\mu_D+27\sigma_{t\_col}^2-18\mu_{t\_col}}{18}} \cdot \frac{yr}{C} \tag{3.210}$$

$$z_{inflexion\_2} = e^{\frac{\sqrt{\sigma_D^2+9\sigma_{t\_col}^2}\sqrt{\sigma_D^2+9\sigma_{t\_col}^2+36}-3\sigma_D^2-6\mu_D-27\sigma_{t\_col}^2-18\mu_{t\_col}}{18}} \cdot \frac{yr}{C} \tag{3.211}$$

We shall not derive here further statistical properties of (3.61), though a series of lengthy calculations performed by Macsyma would probably enable us to do so.

We only need finding the pdf and basic statistical properties of the random variable T_{Galaxy} defined by (3.180). So, recall that, if one has the pdf of a random variable X and wants to find the pdf of the new, linearly-transformed random variable $aX+b$, where a and b are just constants (i.e. non-random-variables), then the two pdfs are related to each other by

$$f_{aX+b}(x) = \frac{1}{|a|} f_X \left(\frac{x-b}{|a|} \right) \tag{3.212}$$

Now, the constant a defined by (3.179) is positive, and so no absolute value is needed. Thus, (3.180) and (3.212) yield at once

$$f_{T_{Galaxy}}(t) = \frac{1}{a} f_{\frac{T_{col}}{D}} \left(\frac{t-b}{a} \right) \tag{3.213}$$

We thus conclude that the pdf of the random variable T_{Galaxy} is obtained from (3.204) by letting (3.204) undergo the transformation given by (3.213). In other words, the pdf of the overall time needed to colonize the whole Galaxy is given by:

Again, although the pdf (208) is difficult to handle by hand, it can be easily handled by Macsyma. Notice that this probability distribution holds only for positive values of the time that also are larger than the constant b defined by the second equation in (3.179). This is of course requested to avoid imaginaries that would otherwise be brought in by the real power of $(t-b)$ at the denominator. In other words, for values of t ranging between zero and b, the above pdf is understood to be equal to zero.

Thus, one can prove (by virtue of Macsyma) that it fulfills indeed the normalization condition

$$\int_b^{\infty} f_{T_{Galaxy}}(t)dt = 1 \tag{3.215}$$

$$f_{T_{Galaxy}}(t)=\frac{1}{a}\cdot\frac{3\sqrt{2}\,C^{\frac{3\mu_D+9\mu_{t\_col}}{o_D^2+9\sigma_{t\_col}^2}}\cdot e^{-\frac{(\mu_D+3\mu_{t\_col})^2+\left(3\cdot\log\left(C\,yr\frac{t-b}{a}\right)\right)^2}{2\left(o_D^2+9\sigma_{t\_col}^2\right)}}}{2\sqrt{\pi}\,yr^{\frac{3\mu_D+9\mu_{t\_col}-9\log C-9\log\left(\frac{t-b}{a}\right)}{o_D^2+9\sigma_{t\_col}^2}}\cdot\left(\frac{t-b}{a}\right)^{\frac{o_D^2+9\sigma_{t\_col}^2-3\mu_D-9\mu_{t\_col}+9\log C}{o_D^2+9\sigma_{t\_col}^2}}\sqrt{o_D^2+9\sigma_{t\_col}^2}} \tag{3.214}$$

Figure 3.12: The probability density function, i.e. a function of t, of the positive random variables: T_{Galaxy}, the time requested to colonize the whole Milky Way Galaxy. Notice that this probability distribution holds only for positive values of the time t such that they also are larger than the constant b defined by the second equation in (3.179). This is of course requested to avoid imaginaries that would otherwise be brought in by the real power of $(t-b)$ at the denominator. In other words, for values of t ranging between zero and b, the above pdf is understood to be equal to zero.

Now we want to derive the mean value of the time T_{Galaxy} needed to colonize the whole Galaxy. Since the mean value operator is a linear operator, the requested mean value is found immediately by letting (3.206) undergo the linear transformation (3.213) and so one gets at once

$$\mu_{T_{Galaxy}} = ae^{\frac{\sigma_D^2+6\mu_D+9\sigma_{t\_col}^2+18\mu_{t\_col}}{18}}\cdot\frac{yr}{C}+b \tag{3.216}$$

Notice that this equation is dimensionally correct since $a = 2R_{Galaxy}/k = 4R_{Galaxy}$ has the dimension of a length, that cancels against the length C at the denominator, leaving just a time (in years) as requested. Of course, b, given by (3.35), has the dimension of a time and depends only on the speed v_{ss} of the pure interstellar flight to hop between planets.

The corresponding variance was again found by Macsyma through a similar calculation, and reads

$$\sigma_{T_{Galaxy}}^2 = a^2e^{\sigma_D^2/9+2\mu_D/9+\sigma_{t\_col}^2+2\mu_{t\_col}}\left(e^{\sigma_D^2+\sigma_{t\_col}^2}-1\right)\cdot\frac{yr^2}{C} \tag{3.217}$$

Its square root is thus the relevant standard deviation:

$$\sigma_{T_{Galaxy}} = a e^{\frac{\sigma_D^2 + 2\mu_D/3 + \sigma_{t\_col}^2 + 2\mu_{t\_col}}{2}} \sqrt{e^{\sigma_D^2/9 + \sigma_{t\_col}^2} - 1} \cdot \frac{yr}{C} \tag{3.218}$$

The mode of the pdf (3.214), i.e. the abscissa of its peak, is found by setting to zero the first derivative of (3.214) with respect to z. Macsyma did a good job again, and the two results are the two abscissas of the minimum of (3.214), obviously at $z = 0$, and of the maximum (i.e. the peak, or mode) at

$$z_{mode\_T_{Galaxy}} = a e^{-\sigma_D^2 + \mu_D/3 - \sigma_{t\_col}^2 + \mu_{t\_col}} \cdot \frac{yr}{C} + b \tag{3.219}$$

Finally, we had Macsyma prove that the two inflexion points of (3.214), that is the one before and the one after the peak (or mode), are found as the two roots of a quadratic algebraic equation in log(z) that is found after equalling to zero the second derivative of (3.214) with respect to z. Thus, it is found that the abscissas of such two inflexion points of (3.214) read, respectively

$$z_{inflexion\_1} = e^{\frac{\sqrt{\sigma_D^2 + 9\sigma_{t\_col}^2}\sqrt{\sigma_D^2 + 9\sigma_{t\_col}^2 + 36} + 3\sigma_D^2 - 6\mu_D + 27\sigma_{t\_col}^2 - 18\mu_{t\_col}}{18}} \cdot \frac{yr}{C} \tag{3.220}$$

$$z_{inflexion\_2} = e^{\frac{\sqrt{\sigma_D^2 + 9\sigma_{t\_col}^2}\sqrt{\sigma_D^2 + 9\sigma_{t\_col}^2 + 36} - 3\sigma_D^2 - 6\mu_D - 27\sigma_{t\_col}^2 - 18\mu_{t\_col}}{18}} \cdot \frac{yr}{C} \tag{3.221}$$

We shall not derive here further statistical properties of (58), though a series of lengthy calculations performed by Macsyma would probably enable us to do so.

3.23 Conclusions about the statistical Fermi Paradox

We extended the classical Fermi paradox to let it encompass Statistics and Probability. We gave (difficult) statistical equations that are related to both the Statistical Drake equation and the Statistical Dole equation for habitable planets for Man. This was the analytical theory, but now a good and portable numeric code should be written to let our results be applied to cases of practical interest.

This approach appears to pave the way to future, more profound investigations intended not only to associate "error bars" to each factor in the Drake and Dole equations, but especially to increase the number of factors themselves. In fact, this seems to be the only way to incorporate into these equations more and more new scientific information as soon as it becomes available. As we said, in the long run, our Statistical results might just become a huge computer code, growing in size and especially in the depth of the scientific information it contained. It would thus be Humanity's first "Encyclopaedia Galactica."

Unfortunately, to extend the Drake and Dole equation to Statistics, it was necessary to use a mathematical apparatus that is more sophisticated than just the simple product of numbers.

When this author had the honour and privilege to first present his results at the SETI Institute on April 11th, 2008, in front of an audience also including Professor Frank Drake, he felt he had to add these words: "My apologies, Frank, for disrupting the beautiful simplicity of your equation."

Acknowledgements

The author is grateful to Drs. Hal Puthoff and Eric Davis of the Institute for Advanced Study at Austin for the reading the author's paper #IAC-08-A4.1.4 and suggesting its publication made up by both the paper itself and its popularized version. Thanks are also due to Professor Frank Drake and Drs. Jill Tarter, Seth Shostak, and Doug Vakoch of the SETI Institute, as well as Dr. H. Paul Shuch, for their appreciation of the statistical work described in this paper. Finally, special thanks go to Marc Millis and Paul Gilster of the Tau Zero Foundation for realizing the importance of our statistical extension of the Drake and Dole equations and helping to have our results popularized within the international scientific community.

References

1. http://en.wikipedia.org/wiki/Drake_equation

2. http://en.wikipedia.org/wiki/SETI

3. http://en.wikipedia.org/wiki/Astrobiology

4. http://en.wikipedia.org/wiki/Frank_Drake

5. Athanasios Papoulis and S. Unnikrishna Pillai, "Probability, Random Variables and Stochastic Processes", Fourth Edition, Tata McGraw-Hill, New Delhi, 2002, ISBN 0-07-048658-1.

6. http://en.wikipedia.org/wiki/Gamma_distribution

7. http://en.wikipedia.org/wiki/Central_limit_theorem

8. http://en.wikipedia.org/wiki/Cumulants

9. http://en.wikipedia.org/wiki/Median

10. Jeffrey Bennett and Seth Shostak, "Life in the Universe", Second Edition, Pearson – Addison-Wesley, San Francisco, 2007, ISBN 0-8053-4753-4. See in particular page 404.

11. Claudio Maccone, "The Statistical Drake Equation", paper #IAC-08-A4.1.4 presented on October 1st, 2008, at the 59th International Astronautical Congress (IAC) held in Glasgow, Scotland, UK, September 29th thru October 3rd, 2008.

12. Stephen H. Dole, "Habitable planets for Man", first edition, 1964, by the RAND Corporation, Library of Congress Catalogue Card Number 64-15992. See in particular page 82, i.e. the beginning of Chapter 5, entitled "Probability of Occurrence of Habitable Planets".

13. Athanasios Papoulis and S. Unnikrishna Pillai, "Probability, Random Variables and Stochastic Processes", Fourth Edition published by Tata-McGraw-Hill, New Delhi, 2002. See in particular pages 186-187.

14. Claudio Maccone, "The Statistical Drake Equation", Acta Astronautica, Vol. 67 (2010), pages 1366-1383.

15. Claudio Maccone, "SETI and SEH (Statistical Equation for Habitables)", Acta Astronautica, Vol. 68 (2011), pages 63-75.

16. Claudio Maccone, "The Living Drake Equation of Tau Zero Foundation", Acta Astronautica, Vol. 68 (2011), pages 582-590.

17. Claudio Maccone, "The Statistical Fermi Paradox", Journal of the British Interplanetary Society, Vol. 63 (2010), pages 222-239.

18. Claudio Maccone, "The statistical Drake equation and A. M. Lyapunov theorem in problem of search for extraterrestrial intelligence, Part I. International scientific Journal "Actual problems of aviation and aerospace systems", 1(32), Vol. 16, 2011, pages 38-63 (in Russian).

Chapter 4

Dynamical generalizations of the Drake equation: the linear and non-linear theories

by **A.D. Panov**
Moscow State University, Russia

Abstract

The Drake equation pertains to the essentially equilibrium situation in a population of communicative civilizations (CCs) of the Galaxy, but it does not describe dynamical processes which can occur in it. Both linear and non-linear dynamical population analysis is build out and discussed instead of the Drake equation.

4.1 Introduction

The crucial question of the SETI problem is how far the nearest CC from us is. Its answer depends on the number of CCs existing in the Galaxy at present. Fig. 4.1 shows how the distance between the Sun and the nearest CC depends on the number of CCs in the Galaxy. The calculation was fulfilled by the Monte Carlo method with the use of a realistic model of the distribution of stars in the Galaxy [1] and for the actual location of the Sun in the Galaxy (8.5 kpc from the center).

The best known way to answer the question about the number of CCs is the formula by F. Drake

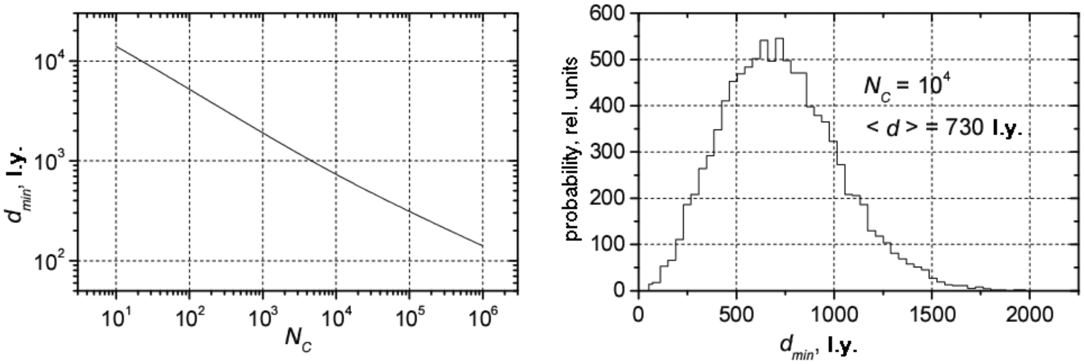


Figure 4.1: The expected distance to the nearest CC as a function of the number of CCs in the Galaxy (left panel) and the probability distribution of distances to the nearest CC for the case of NC = 10000 (right panel). The distribution function profile for other values of NC is analogous; only the most probable distance differs.

$$N_C = R_* f_p n_e f_l f_i f_c L \tag{4.1}$$

where R_* is a star-formation rate in the Galaxy averaged with respect to all time of its existence, f_p is the part of stars with planet systems, n_e is the average number of planets in systems suitable for life, f_l is the part of planet on which life did appear, f_i is the part of planets on which life developed to intelligent forms, f_c is the part of planets on which life reached the communicative phase, L is the average duration of the communicative phase. The Drake formula gives the number of CCs only in a rather rough approximation. According to the formula, N_C does not depend on time. Meanwhile, it is evident that formerly there were no CCs in the Galaxy at all. Then there was a transition period when its number was increasing somehow. In fact, the Drake formula describes only the essentially stable situation, which can be very remote from the truth.

It is necessary to modify the Drake formula to allow for the development times, the variability of star formation rate, etc. Our paper develops this approach both in linear and non-linear dynamical theory.

4.2 Linear population analysis

In the linear theory it is supposed that the CCs develop independently from each other and that CCs cannot effect the star formation rate and evolution of life on other planets in the Galaxy. The following model functions and parameters are used in the model. $R(M,T)$ is the star formation rate as a function of star mass M and galactic time T. The star lifetime is determined by the survival probability $L_S(M,\tau)$ of the star mass M on the Main Sequence at the moment τ reckoned from the moment of its birth. $B(M,\tau)$, determines the density of the probability that a CC appears in the time τ after formation of a star of the mass M. The function $B(M,\tau)$ is normalized by $\int B(M,\tau)d\tau = \alpha(M)$, where $\alpha(M)$ gives a probability that conditions suitable for origin of a CC near the star of the mass M will be implemented someday having infinitely long lifetime for the parent star. The duration of the communicative phase of CC evolution is determined by the function $L_C(M,\omega)$ that gives the probability of maintenance of the communicative phase in the time ω after its origin. The population of stars is described by distribution $n_S(M,\tau,T)$ specifying the number of stars by their mass M and age τ with the galactic time T. The population of CCs is described by distribution $n_C(M,\tau,\omega,T)$ specifying the number of communicative civilizations of the age ω at the galactic time T which appeared near the star having the mass M and the age τ. The total number of civilizations N_C is:

$$N_C(T) = \int_0^\infty dM \int_0^T d\tau \int_0^{T-\tau} d\omega n_c(M,\tau,\omega,T) \tag{4.2}$$

The complete system of equations together with margin conditions that determines the distributions $n_S(M,\tau,T)$ and $n_C(M,\tau,\omega,T)$ is

$$\begin{cases} \frac{\partial n_S}{\partial T} = -\frac{\partial n_S}{\partial \tau} - \Lambda_S(M,\tau)n_S \\ \\ -\Lambda_S(M,\tau) \equiv \frac{\partial ln L_s(M,\tau)}{\partial \tau} \end{cases} \tag{4.3}$$

$$n_s(M,\tau,0) = 0 \tag{4.4}$$

$$n_s(M,0,T) = R(M,T) \tag{4.5}$$

$$\frac{\partial n_c}{\partial T} = -\frac{\partial n_c}{\partial \omega} - [\Lambda_c(M,\omega) + \Lambda_s(M,\tau+\omega)]n_c \tag{4.6}$$

$$n_c(M,\tau,\omega,0) = 0 \tag{4.7}$$

$$n_c(M,\tau,0,T) = n_s(M,\tau,T)B(M,\tau) \tag{4.8}$$

The definition of Λ_C in (4.6) is obvious from (4.3)). The system (4.3-4.8) has exact solution for the density distribution of CCs as follows:

$$n_c(M, \tau, \omega, T) = R(M, T - \tau - \omega) L_s(M, \tau + \omega) B(M, \tau) L_c(M, \omega) \qquad (4.9)$$

The obtained solution (4.9) together with formula (4.2) allows us to investigate a huge number of various tasks. It is worth to restrict this variety by some reasonable limits. For this purpose, some simplifications will be used. We suppose additionally: the star formation rate may be factorized: $R(M, T) = R_*(T)F(M)$, where the initial mass function $F(M)$ is supposed to be independent of time; the star survival probability to be a step-like function: $L_S(M, \tau) = \Theta[\tau_0(M) - \tau]$, where $\tau_0(M)$ is the lifetime of the star with mass M on the Main Sequence; the time of development before CC formation to be independent of the star mass M: $B(M, \tau) = \alpha(M)b(\tau)$ with $\int b(\tau)d\tau = 1$; the lifetime of CC does not depend on the star mass M: $L_C(M, \omega) = L_c(\omega)$. Expression (4.2) with using of (4.9) and the introduced simplified expressions for the model functions may be rewritten as

$$N_c = \int_0^\infty dM \alpha(M) F(M) \int_0^T d\tau b(\tau) \int_0^{\omega_{max}(M)} d\omega R_*(T - \tau - \omega) L_c(\omega) \qquad (4.10)$$

The Drake equation (4.1) may be obtained from eq.(4.10) with further simplifications: R_* =const, $\tau_0(M) \equiv \infty$, time of development before CC formation is small. But we will investigate more realistic scenarios.

The initial spectrum of star masses according to [2] and the relation between star lifetimes on the Main Sequence and mass approximated in the [3, p. 58] were used in calculations. Fig. 4.2 shows corresponding functions $F(M)$ and $\tau_0(M)$. For the star formation rate function $R_*(T)$ in calculations we used averaged and interpolated data from the papers of [4] and [5] (shown by the dotted line in Fig. 4.3, left panel). The relative rate data of [3,4] were normalized to obtain correct number of stars in the Galaxy at the present time. The linear function equal to zero at $M = 0.5M_*$, equal to 1 at $M = 2M_*$ and $\alpha(M) = 1$ at $M > 2M_*$ was taken as the probability of realization of suitable conditions. The value $\alpha(M) = 1$ for $2M_*$ was chosen rather arbitrarily and it does not restrict the generality due to linearity of the theory. With such choice of $\alpha(M)$ the average probability of implementation of suitable conditions with star masses from $0.5M_*$ to $2M_*$ turns out to be about 0.02. For the distribution density of CC development times $b(\tau)$ we tested here three functions shown in Fig. 4.3, right panel. The distribution of durations of the communicative phase was taken in the form of the falling exponent $L_c(\omega) = \exp(-\omega/L_0)$ with $L_0 = 1000$ years. The choice of L_0 practically does not limit the generality of results (due to linearity).

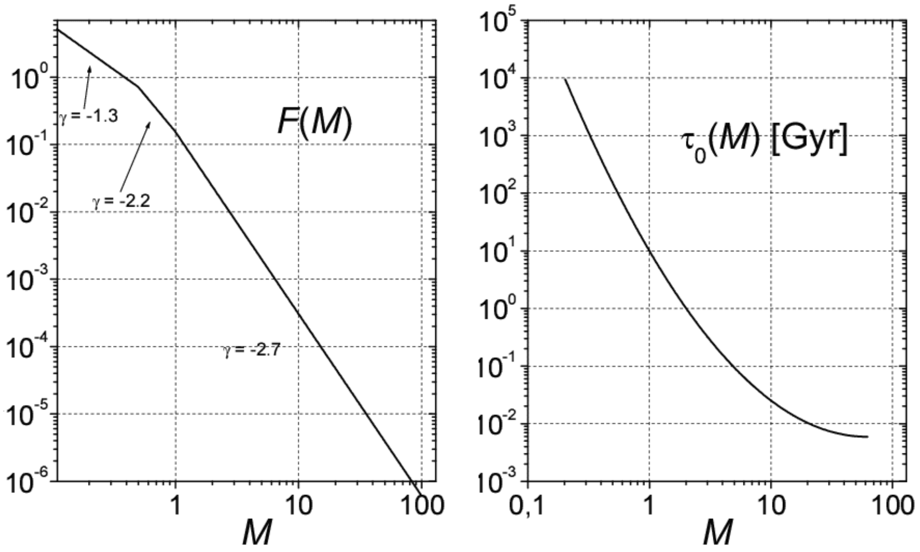


Figure 4.2: Left panel: The initial spectrum of star masses. Right panel: The star lifetimes. The star mass is in the solar mass. The quantity γ in the diagram of F(M) shows an index of the power function corresponding to different parts of the spectrum

Fig. 4.4 (left panel) shows results of calculations carried out with the above assumptions with the formulae (9,10). The results correspond to different distributions of CC development times. All curves have a strongly pronounced maximum associated with an SFR peak at T≈5 billion years (see Fig. 4.3). The peak in the number of civilizations is a linear response to it and can be called a linear demographic wave. For the basic variant of calculations (the solid line) the present time (12 billion years) falls within the region of the maximum of the linear demographic wave.

Note that though the relations in Fig. 4.4 (left panel) are constructed for a very limited set of parameters, they can be used to estimate within the context of many other scenarios. So, the curve amplitude will be proportional to the average CC lifetime (the parameter L_0) and the curve amplitude will also be proportional to the maximum probability of realization of suitable conditions (the maximal value in $\alpha(M)$ function, see above).

Up to this point the conditions leading to the origin of a CC have been supposed to be unchanged during the history of the galactic disk. Actually, variations of them are possible for a number of reasons (variable background of cosmic rays, etc). The conditions change for sure if the hypothesis about the self-consistent galactic origin of life and related phase transition [6] is true. In this case a great “Big Bang” of life

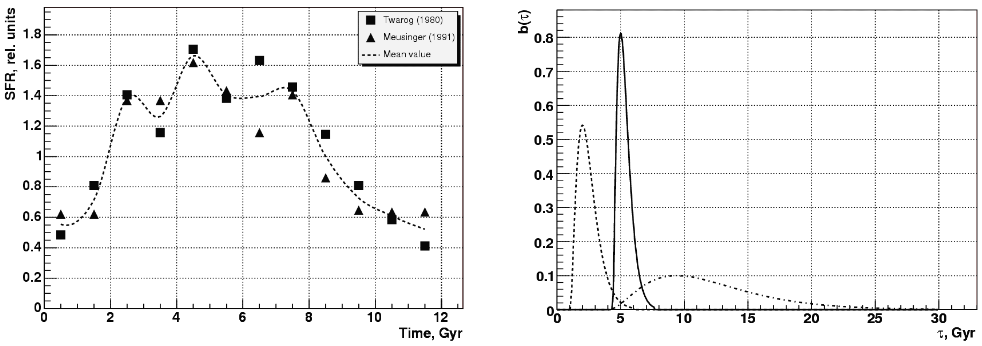


Figure 4.3: Left panel: star formation rate as a function of time. Right panel: the choice of probability of CC development times. Solid line represents the most probable case.

origin took place in the history of the Galaxy and if the development time to CC is more or less standard (like $b(\tau)$) presented by solid line in Fig. 4.3, left panel – about 5 billion years), then "Big Bang" of CCs origins should be followed as well. The theory describing this phase demographic peak may be deduced from the described above linear theory (we omit the details) and the results of calculation are presented in Fig. 4.4 (right panel). It was supposed the "Big Bang" of life origin to be 6 billion years after the start of formation of the Galaxy disk and the average time of development for CC to be 5 billion years in this calculation. The dashed line in Fig. 4.4 (right panel) shows the partial distribution for planets with the origin of life after the "Big Bang" of life origin (as Earth). One can see that we can live both before and after the phase peak.

Thus, the population analysis based on the linear theory and real astrophysical data predicts no-trivial dynamical patterns of evolution of CCs like the linear demographic wave and the phase peak (Fig. 4.3). Non-linear generalization of the formalisms leads to even more interesting picture.

4.3 Non-linear population analysis

In the linear theory given above the distributions $B(M,\tau)$ and $L_C(\omega)$ describing the origin and life of communicative civilizations were supposed to be independent on the number of available civilizations. The function $R(M,T)$ describing the "natural" star formation rate was also considered to be independent of the CC population. This is true until civilizations have no effect on one another, nor on conditions of origin

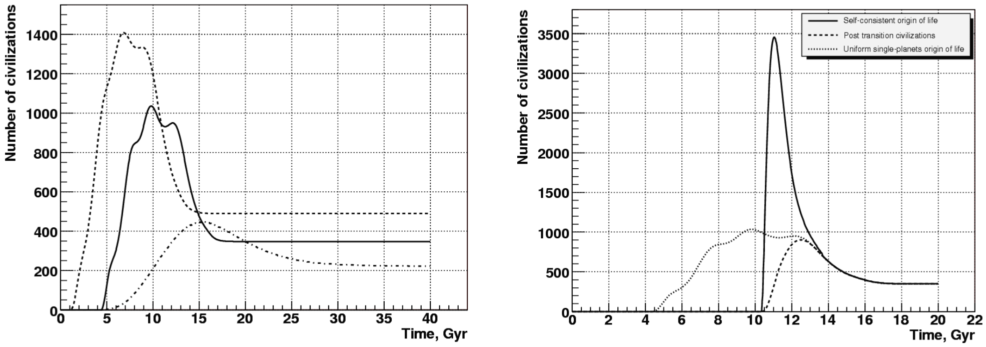


Figure 4.4: Left panel: results of calculations within the framework of the simple linear theory corresponding to different distributions $b(\tau)$ of CC development time (see Fig. 4.3). Right panel: The linear dynamics of the CC population at the origin of life in the Galaxy in the process of the self-consistent phase transition [6] in 6 billion years after the beginning of the formation of the galactic disk and its comparison with the simple linear dynamics at the constant formation of CCs with the development time 5 billion years (for references: actual age of the Galaxy disc is 12 billion years).

of other civilizations, nor on conditions of origin of stars. The theory accounting for this effect ceases to be linear.

The first possibility for a non-linear theory is related to the influence on the function $R(M,T)$ – "the artificial creation of stars". The second possibility – the influence on the distribution $B(M,\tau)$ – must imply some sort of directed panspermia of life or intelligent life. The third kind of non-linear phenomena related to the changing of the probability $L_C(\omega)$ by a mutual influence of civilizations through contacts by communication channels. We thoroughly study only the last possibility here. Other options may be studied by similar methods.

Without limiting generality, the CCs may be thought to be divided into three categories: the CCs for which the contact is "harmful", because it reduces the duration of the communicative phase, the neutral CCs, and the CCs for which the contact is "useful", because it prolongs the communicative phase. We will call the last category extrovert civilizations and will denote them as ECC. In the following we will consider the dynamics of the subpopulation of the extrovert civilizations only.

It can be supposed additionally that one of the most important properties of ECCs is an increase in efficiency of search for partners and establishment of communication under the influence of the already established contacts (we call it civilization range).

This circumstance will be substantially used below.

It is important that if ECCs do exist, then a process with a positive feedback can begin. The larger is the number of ECCs in the Galaxy, the higher is the contact probability. The contact increases the lifetime of the ECC and its civilization range, which leads to increasing the ECC population, which rises the contact probability again, and so on. The positive feedback loop can lead to an avalanche-like phase transition in the Galaxy-scale accompanied by a powerful burst of the number of ECCs. ECCs become prevailing in the Galaxy even if the situation was different before the transition. Some details of this phenomenon are described by the formalism proposed below.

In the linear theory the current state of a separate civilization was described only by age of the communicative phase ω, which, in combination with the star lifetime and the moment of the civilization origin, made it possible to statistically predict the fate of a civilization. To account for the mutual influence through communication channels they should be described in greater details. We will consider that every civilization is described by age ω and by a vector of parameters q that will be called "a quality". This is a set of characteristics of ECCs which affects, first of all, an expected duration of the communicative phase and civilization range. It is supposed that the contact increases the ECC quality in a sense, and thanks to that the communicative phase prolongs and civilization range increases. Thus, the probability of civilization survival should be considered as dependent on its quality which must be also one of the arguments of the civilization distribution function:

$$\begin{gathered} L_c(M,\omega) \to L_c(M,\mathbf{q},\omega) \\ n_C(M,\tau,\omega,T) \to n_c(M,\tau,\mathbf{q},\omega,T) \end{gathered} \tag{4.11}$$

To describe the influence of contacts of a civilization A with a number of other civilizations B1,B2,... on the quality of A we suppose the effect to be additive:

$$\frac{d\mathbf{q}_A}{dt} = \sum_i K(\mathbf{q}_A,\omega_A,\mathbf{q}_{B_i},\omega_{B_i}) \tag{4.12}$$

where $K(\mathbf{q}_A,\omega_A,\mathbf{q}_B,\omega_B)$ is some universal function representing the contact model. Obviously the additive model of contacts is a simplification that may be reasonable only in a case of low number of contacts per civilization.

Equations (4.3–4.5) for the star distribution function and equations (4.6–4.8) for the civilization distribution function remain valid in non-linear dynamics. Only a new term appears in it describing "the current" of the civilization quality in the q-space due to interaction between them. Besides, now the edge condition must describe weights of ECCs quality starting the communicative phase. The total system of equations for the distribution function $n_C(M,\tau,q,\omega,T)$ is written in the following way:

$$\frac{\partial n_c}{\partial T} = -\frac{\partial n_c}{\partial \omega} - [\Lambda_C(M,\mathbf{q},\omega) + \Lambda_S(M,\tau+\omega]n_c - \Delta_q[j(\mathbf{q},\omega,T)n_c] \tag{4.13}$$

$$n_c(M,\tau,\mathbf{q},0,T) = 0 \tag{4.14}$$

$$n_c(M,\tau,\mathbf{q},0,T) = n_s(M,\tau,T)B(M,\tau,\mathbf{q}) \tag{4.15}$$

The problem of calculation of the q-current $j(q,\omega,T)$ generally is very difficult but it may be solved for the additive model of contacts (11) and for the model of a large homogeneous galaxy (edge effects can be neglected):

$$\begin{aligned} j(\mathbf{q},\omega,T) = \frac{4\pi c^3}{V_G}\int d\omega' \int d\mathbf{q}' K(\mathbf{q},\omega,\mathbf{q}',\omega') \int_{T-r(\mathbf{q},\omega,\mathbf{q}',\omega')}^{T} dT'(T-T')^2 \times \\ \int dM \int d\tau n_c(M,\tau,\mathbf{q}',\omega',T') \end{aligned} \tag{4.16}$$

In formula (4.16) V_G is the galaxy volume, c is the velocity of light, and $r(\mathbf{q},\omega,\mathbf{q}',\omega')$ is the range of communication between two civilizations with the qualities and the ages $(\mathbf{q},\omega)$ and $(\mathbf{q}',\omega')$. Due to the term corresponding to the quality current, (4.13) turns out to be very complicated integro-differential equation. However, it may be solved numerically for simple models of contact under some additional simplifying assumptions as described below.

The civilization quality will be considered to be presented by the only scalar parameter q. It is supposed the average value of the quality for an isolated civilization (without any contacts) is q = 1. We transfer from the detailed description of a civilization by its quality and age to the average value of quality upon the whole lifetime of the civilization and averaged upon all star masses. Further, we consider the number of civilizations per unit of volume of a uniform galaxy. That is, instead of the exact distribution $n_C(M,\tau,q,\omega,T)$ we consider averaged distribution $\rho(q,T)$ such that $V_G\int\rho(q,T)dq = N_C(T)$. We consider the civilization origin rate normalized per galaxy volume unit to be a given function of the galaxy time $f(T)$, and the distribution density of the parameter q for isolated civilizations to be $\phi_0(q)$ such that $\int\phi_0(q)dq = 1$ and the mean value of $\phi_0(q)$ is equal to 1. Then equations (4.13–4.15) can be transcribed in the form of a single equation

$$\frac{\partial\rho(q,T)}{\partial T} = -\Lambda_c(q)\rho(q,T) + f(T)\Phi_0(q) - \frac{\partial}{\partial q}[j(q,T)\rho(q,T)] \tag{4.17}$$

Initial conditions for the function $\rho(q,T)$ can be specified at any time $T = T_0$, and equation (4.17) can be solved as the Cauchy initial-value problem.

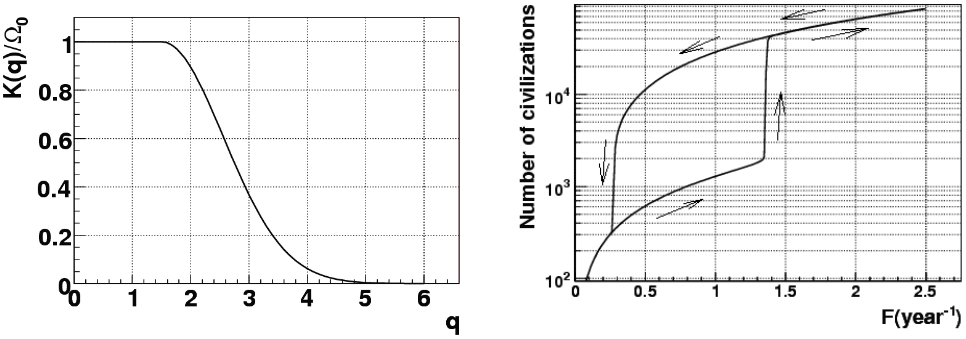


Figure 4.5: Left panel: The function $k(q)$ used in calculations. Right panel: The bistability in a ECC population obtained by numerical solution of the equation (4.17).

To calculate the divergence term in (4.18 we adopted the following simple model of contact

$$\frac{dq_A}{df} = k(q_A)q_A \sum_i q_{B_i} \tag{4.18}$$

where the function $k(q)$ is shown in Fig. 4.5 (left panel) with $\Omega_0 = 0.001$ years^{-1}. With (4.18) the equation (4.16) is simplified to

$$j(q,T) = 4\pi c^3 qk(q) \int dq'q' \int_{T-r(q,q')/c}^{T} dT'(T-T')^2\rho(q',T') \tag{4.19}$$

For the inverse lifetime function $\Lambda_c(q)$ and the range function $r(q,q')$ various assumptions may be taken but we adopted the following ones here:

$$\begin{cases} \Lambda_c(q) = \Omega_0/q^2 \\ \\ \Omega_0 = 0.001 years^{-1} \end{cases} \tag{4.20}$$

$$\begin{cases} r(q_A,q_B) = r_0(q_A,q_B)^{1/5} \\ \\ r_0 = 400 l.y. \end{cases} \tag{4.21}$$

The expression for $r(q_A, q_B)$ was obtained under the assumption that reception and transmission are fulfilled only by a beam antenna (we have to omit the details of the explanation). The distribution $\phi_0(q)$ was taken to be Gaussian one with mean value 1 and dispersion 0.2.

Some results of calculations are shown in Fig. 4.5 (right panel). Let us elucidate the computing technique and sense of the obtained results. It was supposed that at the initial time T = 0 there were no civilizations, $\rho(q, 0) = 0$. After that the civilization origin rate F begins increasing slowly, so that at any time an almost complete equilibrium is achieved in the population of ECC. Fig. 4.5 (right panel) shows the relation between the number of civilizations and F (both normalized to the volume of our Galaxy). The equilibrium number of civilizations increases as F increases. In the process, first a point in the diagram moves along the lower branch of hysteresis loop from left to right and the number of civilizations is still small (less than 2000). This is the silence epoch, the probability P to find a partner to contact for any civilization is $P \ll 1$.

However, due to the increasing number of civilizations, the situation becomes unstable, and when F achieves a value of about 1.35 civilizations per year, and $P \approx 0.05$, then the equilibrium is broken. Due to the positive feedback between the number of contacts, civilization ranges and lifetimes, the number of civilizations and the probability of their interaction start increasing as an avalanche. As this takes place, the number of civilization increases sharply by about an order, and the average number of partners per one civilization achieves 10. This phase transition ends because the possibility of "improving" is exhausted at large values of the quality q (see Fig. 4.5, left panel). The saturation of contacts epoch starts (F >1.4).

Then, in calculation, the civilization origin rate stops increasing (at F =$2.5 years^{-1}$) and begins the slow. First, a point in the diagram moves backwards, repeating the trajectory of F growth. However, when reaching a critical value of F = 1.35 per year the reverse transition does not occur. This is prevented by the positive feedback “number of contacts – lifetime and range”. The contact saturation epoch continues. Here two different stable states of the civilization population correspond to every value of F: one on the lower branch of hysteresis loop, the other on the upper branch. This is the bistability phenomenon. Only when P approaches a value of about 0.5, the positive feedback already cannot keep the contact saturation phase from destruction, the number of civilizations sharply fall, and the silence epoch returns.

We neglected by fluctuations of density of civilization, but fluctuations can create the saturation of contacts phase locally with subsequent growth.

References

1. Allen C.W. Astrophysical quantities. University of London, The Athlone Press 1973.

2. Meyer M. R., Adams F. C., Hillenbrandt L. A., Carpenter J. M., Larson R. B. // arXiv:astro-ph/9902198, 1999

3. Surdin V. G., Rozhdenie zvesd (Birth of stars), Moscow, URSS, 2001 (in Russian).

4. Twarog B. A. Astrophys. J., V. 242, P. 242, 1980.

5. Meusinger H. Astrophys. Space Sci., V. 182, P. 19, 1991

6. Panov A.D. Prebiological panspermia and the hypothesis of the self-consistent Galaxy origin of life.// This conference paper, 2011

Chapter 5

The greatest problem of science

by **Yuri N. Efremov**
Sternberg Astronomical Institute, MSU

5.1 Introduction

We reached the edge of the universe in space and in time, we understand the evolution of stars, found the planets around them – but found no trace of another Mind. Are we alone in the desert of the Universe? This issue is becoming a serious challenge to the modern scientific knowledge. "The eternal silence of these infinite spaces terrifies me" – wrote Blaise Pascal yet in 1669 – and it should scare us even more now.

We, men, have visited the Moon; our devices have been orbiting other planets of Solar system for decades, and our apparatus are travelling along the Mars surface as long too – and nothing like Life is noted there. Also, we for more than 50 years are looking for a voice from heaven; about half of a century ago, the radio telescope was first directed to the sky specifically to search for signals from extraterrestrial civilizations. The search continue, but do not give results - in spite of finding a hundred of another worlds, where the life, rather similar to terrestrial one might exist...

The problem of silence of the Universe – and general lack of observable signs of the existence of other intellectual beings, apart from us – excites us more and more. Anyway, there is in fact one absolutely true evidence for the possibility of intelligent life in our Galaxy – it is a simple fact of our own existence. This is a fact of great value, just as the greatest secret of the atomic bomb was already in what that it is possible to construct it. We exist and we are reasonable; moreover, our Earth is already "lit up" in the Cosmos until distance of about 70 light-years – due to television and radar, short-wave radio emission of which penetrates the Earth's atmosphere – and there is no response yet. (However, in recent years the radio luminosity of the Earth decreases, as communications becomes mostly through cable lines, whereas military radars began to quickly change the frequency of its impulses.

First experiments to search for ETI signals were carried out by F. Drake in April 1960 (US National Radio Astronomy Observatory) in Green Bank. Radio telescope was directed on Tau Ceti and Epsilon Eri – nearby stars similar to the Sun. Observations of these stars were lasted for three months; as is well known, signals were not detected. Since then, in various countries were carried out dozens of short-term scans of the radio sky, used different strategy and search techniques, – but there were no results. All events and objects in the sky, we seem to be able to explain in terms of modern science – ie by definition there are natural and are not produced by an activity of other civilizations....

However, after only a few million years for a civilization that is technologically developed with our current pace, will occupy the entire galaxy. In our star System there are tens of billions of stars older than the Sun and the Earth by several billion years. If other civilizations exist, even within our solar system would have been obvious signs of their presence – so where are they all? This question has long been asked Enrico Fermi. How, emphasizes N.S. Kardashev, the Fermi Paradox – this is the greatest mystery of nature...

5.2 The Fermi Paradox

In 1975, M. Hart and I.S. Shklovsky offered a radical solution to this paradox: They are silent simply because they are absent... However, the opinions of the astronomers of the reasons for this non-existence were very different. Hart [1] wrote that although "it is possible that one or two of the evolved civilization have destroyed themselves in a nuclear war, it is incredibly, that each of the 10 000 other Civilizations made the same. " They are not here; therefore they do not exist. Hart concluded that "the fact no evidence of the existence of extraterrestrial intelligence is a strong witness that we are the first civilization in the local galaxy".

On the same 1975 I.S. Shklovsky [2] at a conference on extraterrestrial civilizations in stanitsa Zelenchukskaya (near the site of 6-m telescope) concluded that the lack of "space miracles" – i.e. our ability to explain all the observed phenomena and objects – means our loneliness in the universe. The silence of the Cosmos implies that having reach the certain stage of development, the Mind always dies – there was his conclusion. One may, if he likes, call Hart's view optimistic (we – the very first!), and Shklovsky's one –pessimistic...

It's was a time of nuclear missiles confrontation and the inevitable doom of mankind seemed likely for this reason. For terrestrial civilization opportunity to give to others know about itself appeared simultaneously with the ability to self-destruct. The conclusion drawn by I.S. Shklovsky, was tragic – the Mind is something like over – hypertrophying tools like sabertoothed tiger fangs, first helping in the fight for survival, but only causing harm as conditions change. He concluded that "assuming that the mind – is just one of the many inventions of the evolutionary process, and besides, it is possible, resulting in a kind of an evolutionary dead-end, we, firstly, better understand man's place in the universe, and secondly, explain why we do not observed cosmic miracles "- wrote Joseph Samoylovich in an article published in the journal "Earth and Universe" in 1984.

Silence of the Universe can, however, be explained with many other considerations as well. In his monograph on the problem of SETI (Searches for Extraterrestrial Civilizations) L.M. Gindilis [3] gives about 20 possible causes of the Great Silence. One of the most likely explanations was suggested many years ago, by S. Lem, a Polish philosopher and writer. He noted that the characteristic scale of technological development in the world – from the emergence of a new theory to create on its basis ubiquitous devices – (for example, from Maxwell to the present day) – is only 100-150 years. Even if you start counting from the ancient Greeks, – took only about 20 centuries, whereas the age of the oldest stars in our Galaxy disk, with approximately the solar abundance of heavy elements, is larger by more than seven orders of magnitude. We can not imagine the scientific and technical potential of mankind, even after 100 years, not to mention the billions of years – of course, if the development of science will continue...

S. Lem thought that we might already observe some results of another civilization activity, but are not aware of this, since this activity can produce phenomena that we will inevitably be considered natural, if they are beyond the horizon of our present knowledge ... For example, if we saw before 1939 nuclear explosion on the Moon, we could not explain it in another way as by an asteroid impact or a volcanic eruption ... However, to conclude that there is not a natural phenomenon but a process or an object created by another civilization, can only be the case if the periods with similar levels of development of science and technology are in the same line for us and them.

Age of steam continued on Earth less than a hundred years – its successor, the age of electricity and electromagnetic field theory was no longer; the present age of quantum mechanics and nuclear energy is unlikely to be longer – and in fact the differences between the ages of civilizations on other worlds may be billions of years... "A window of contact is a cosmic moment," – wrote S. Lem in the novel "Fiasco."

Thus, the absence of any clear (to us) signals from ETI does not necessarily mean the absence of their own. There was no one to take another radio message of any civilization, if it came to the middle of the XX century. Now we take the background radiation from the first moments of life of our own Universe – in some sense, from a distance of about 13 billion light-years); we may catch neutrino radiation from the Sun (8 light-minutes) – and there are already receivers elusive until gravitational waves. It is impossible to imagine what we will have in a hundred years, and even more so in a thousand. And in a five billion ?! After all, most of the stars - and hence planets – older than our Sun for a few billion years.

The knowledge and capabilities of older civilizations are impossible for us to imagine. They can control the movement of the stars (such a possibility has long been mentioned N.S. Kardashev [4], they might create new galaxies and even new universes ... Why not, if even within the framework of modern physics one can already tell what should be the energy of the collision of two elementary particles for the resulting black hole would began to expand into another space as a new universe. So for super-civilizations we are no more interesting than for us – ants; in any case, we do not try to enter into contact with them. So, some of the phenomena that we consider natural, in fact, may be the result of their activities...

Note that in the discourse of the brevity of the window of contact, period commensurably between their and our knowledge of the Universe, it is assumed inexhaustible scientific knowledge. But if there is, even in the asymptote, complete physical theory – Final Theory of Everything, it must be fair to all of our universe (in countless other universes, physics is quite different, but contacts with them are impossible), and if you continue to develop the civilization and science, you will attain this theory sooner or later. Realizing everything in our universe, we will be able to distinguish between natural phenomena from artificial ones.

Of course, we have to do everything possible to find a natural explanation of things. Of course, even the super-intelligence is subject to the physical laws of our universe. Determine the nature of man-made objects and signals is not always easy. The principle of "presumption of naturalness" (nominated by I.S. Shklovsky), rightly demands to the utmost to seek "a natural-governmental" explanation of objects and phenomena, but it should not be absolute, or turned into a ban on flights of fancy.

5.3 Strange objects

Thus, in the search for extraterrestrial intelligence is no other way but to continue the search and research all the strange objects – always keeping in mind the possibility that we may encounter the activities of sentient entities (see, e.g. [4, 5]). It must be remembered that the nature and purpose of such objects or phenomena can be completely outside of our circle of knowledge and concepts – so you have to pay as much attention to phenomena that now we can not explain.

Window opens in a while ... What can make others continuously send radio signals in all directions, to report on their existence? We do not do and can not do this in the foreseeable future – perhaps they, too ... Once I have said to J.S. Shklovsky that one can only hope for a random intercept a focused "conversation" between the two civilizations – and, therefore, it is necessary to pay attention to the unusual radio sources at diametrically opposite points of the sky, – Joseph Samoylovich only smiled sadly – "Well, how do you not understand, – it would mean that there exist three civilizations and all they are on the same direct line." Alas, by that time the founder of the study of the problem of extraterrestrial civilizations in our country was already convinced that they commit by suicide (in nuclear wars) before to discover existence of others...

At any rate, people should not rely on the suggestion that the signals of all civilizations (better to say, technical abilities) are from civilizations who are at such a low level as ours .. We may ever to consider the hypothesis that some cosmic phenomena and objects can be artificial – if not good at all attempts to explain the origin of their "natural" way. Of course, before we talk seriously about this possibility, we must exhaust in attempts to explain all of our current knowledge, this is the only way of science. If the Final Theory of Everything will be comprehended (no one knows whether it exists), each phenomenon, unexplained by this theory can be considered artificial – or theory can not recognize the final ... Anyway, we must pay particular attention to the strange objects the origin of which we can not explain.

Extreme case of such a situation may be, for example, very rare stellar regular configurations for the occurrence of which there is no plausible explanation. For example, two giant arcs of the high luminosity young stars are known in the Large Magellanic Cloud (LMC) side by side of each other – and their origin is unknown (see Fig. 5.1). Authors [5] have suggested an interpretation that goes beyond sound of modern knowledge; it was an idea by V. Lefebvre, an American psychologist (and a former young amateur of astronomy in Moscow Planetarium), who always insists to take into account a possibility of an artificial origin of strange objects. Anyway, recently the author [6] have suggested another explanation for this giant arcs, based on the idea of the ram pressure of a strong (in the past) jet from the MW core to a LMC giant gas cloud. This explanation, suggested in Efremov's paper [6], is supported by similarities of the radial velocities of stars located along the large arc. It is nothing but an attempt to save phenomenon by a natural way...

There is no a really good idea to explain why intelligent creatures might have

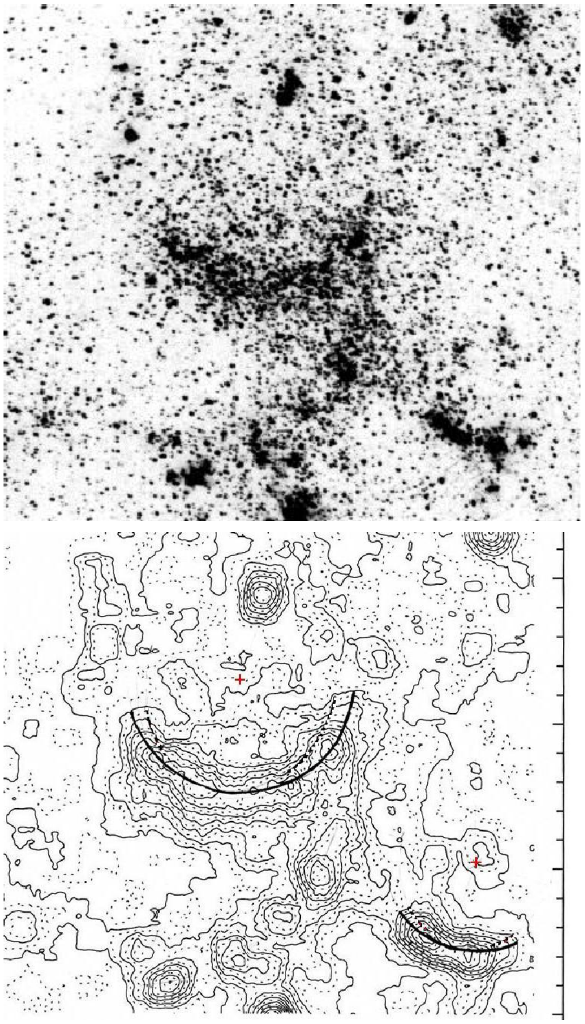


Figure 5.1: Two giant arcs of young stars in the LMC, including many thousands of stars of the same age for each arc. The major arc is somewhat older, its length – about 3,000 light-years. Left fig. Is in the UV range, right – the results of the stellar photometry. Segments of circles are drawn through the densest part of the two arcs. If these are located in the plane of the arc LMC, their external parts must be segments of ellipses (shown in phantom in the right figure), which is not observed.

need to constructs such arcs of massive stars; anyway, see the paper [5]. Either way, you must not only (and maybe not so much) to look for signals from beacons or transmitters, specifically created to alert neighbourhoods of their creature, but also pay close attention to all the unusual structures or radiation coming from space.

5.4 Brothers in mind?

But searches for the brothers in the mind, chemistry and physical organization are surely justified – and, apparently, in the foreseeable future only these have a chance of success. More distant relatives, who are much older than us – for example, those who had their minds in super-computers, or black holes, we will probably are unable to recognize, even if they exist ... To reiterate, – the search for signals, even from those who are close to us in space and on the level of development – and who, therefore, are expected to live on the planets similar to the Earthy – such searches can be successful only if THEY are active and disinterested distribute political and scientific knowledge – or, at least, send signals, the artificiality of that is obvious. The probability of such a situation is hardly great, although with strong arguments in its favour were given by one of the pioneers of the problem, F. Drake. He noted that the likelihood to survive is higher for the civilization in which the altruistic sentiment won. This is a very important consideration, based on the experience of the evolution of life (especially humans) on Earth – where the chances to survive and evolve are higher on the community where there was mutual help. The dream of the Great Ring, the humanitarian Community (and, I would like to dream, humanoid) civilizations of our Galaxy, continually exchanging information – dream, glorified in the novel *Andromeda Nebula* by known writer Ivan Efremov; the imagination from this novel the curious young people were inspired fifty years ago.

It should be noted that this writer (no relation to the author) was a specialist-palaeontologist; his arguments for extraterrestrial intelligent beings could not too much different from people (deployed, for example, in the story "Starships") must be heeded. Among the arguments for it are the exclusive property of carbon, the main carrier of life on Earth and our minds. Carbon compounds are observed in all gas clouds throughout the universe. Since the beginning of the 20th century, it is known that carbon atoms have the ability to be assembled in long chains or rings that can bind the atoms. Therefore, there are hundreds of thousands of compounds of carbon, whereas the number of compositions of all other elements taken together (including silicon), not to exceed twenty thousand (see [7]).

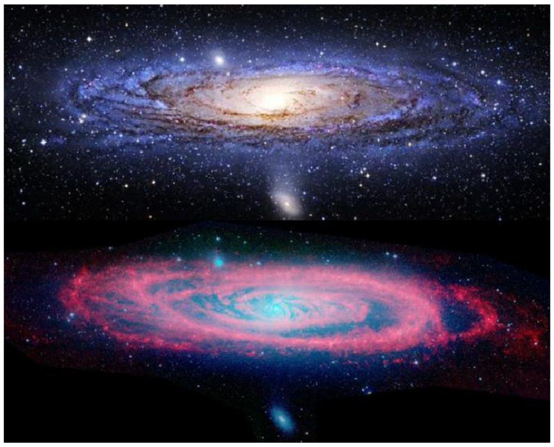


Figure 5.2

5.5 Organic molecules in space

Not so long ago, it was shown that a mysterious band of radiation found by the Spitzer Space Telescope in infrared radiation are observed in dust clouds everywhere, in our Galaxy as well as in others, including the distances in billions of light years away – these bands are due to organic molecules-polycyclic aromatic hydrocarbons (PAH, Hydrocarbonates Aromatic Policyclic Hydrocarbonates... Upper part of Fig. 5.2 represents the Andromedea galaxy in common (blue) light, -star shining – whereas the lower part gives at the same scale the purple emission of PAH, Hydrocarbonates Aromatic Policyclic Hydrocarbonates.

It was only not so long proven, that a mysterious band of radiation found by the Spitzer Space Telescope in infra-red radiation are observed in dust clouds everywhere, in our Galaxy as well as in others, including the distances in billions of light years away – and these bands are due to organic molecules-polycyclic aromatic hydrocarbons (PAH, Hydrocarbonates Aromatic Policyclic).

The similarity of the spectra has been proved in ground-based experiments, in which these molecules are placed in conditions similar to the ones in interstellar space. This discovery, due to the NASA specialists, in essence means that the prerequisites for the emergence of life exist everywhere in the universe.

But polycyclic aromatic hydrocarbons, including nitrogen are molecules of life! This type of molecule is DNA. These molecules not only are ubiquitous in the universe; they are resistant to all outer environment. When the accumulation of planets from dust disks around newborn stars, a process not only calculated theoretically, but recently became available to the direct observations, these molecules are stored. The necessary ingredients for life are everywhere! Enjoin tight curls of spiral branches, which includes the young stars in galaxies, we always see as well flanking them dark dust lanes that match the atomic especially molecular hydrogen clouds. We now know that these strips glow in the infrared (purpure) range, they are not only specks of carbon black and silica, but PAH molecules as well.

Organic molecules began to be discovered in space since the late 1960s, and now there are known about 120 species; the largest of these is HC11N. Allamandola [8] notes that appearing connection between the PAH molecules and astrobiology is an important part of ... "Total revolution in our understanding of the chemistry and biochemistry of space... In the cold molecular clouds within which stars and planets are born, interstellar molecules inside ice consisting of water, methanol, ammonia, carbon monoxide and dioxide and PAHs. In these clouds, especially near areas of formation of stars and planets, these pieces of ice exposed UV and cosmic rays, are formed on the surface thereof much more complex molecules, many of which are interesting for Biogenesis. In the formation of stars and planets, many of these components are integrated into comets and meteorites, which eventually seeded primary planet, where they take part in the budding chemistry on these young worlds" [8].

5.6 Conclusion

We ought to look first of all for our brothers (even very distant in organization and biology) in mind, the inhabitants of planets, seek out those who probably like us. Should again pay special attention to stars like the Sun, especially those which are surrounded by planets, similar to the Earth. It now becomes possible as comes into operation system of Allen radio telescopes specifically designed to search for their signals. Should be checked the hypothesis by V.G. Surdin about the possibility of detecting radio signals from planets inside star clusters – which are goaled not to us but to each others. Judging from the current rate of development of science, may be in a few thousand years we (or rather, our devices) will fly to the nearest stars. Likely may exist narrow focused radio everywhere and it has long been used to communicate with their starship. In stellar clusters the distance between the stars light up the weeks and months (not years to come and the centuries as to the nearest star to the Sun), and the age of the stars in them is almost the same. If inside a star clusters are emerging civilizations, they can develop as a result of synchronous proximity ages of stars in clusters, and because they have enough opportunity of the rapid exchange of relevant information. If the signal is too high or is designed to communicate with the spaceships of these civilizations, and we were on the continuation of their radio beam – then headed to a cluster of radio telescope, we can hope for now to catch other civilisations conversations. With the use of Allen radio telescopes, it would be possible over the years to keep under observation star clusters, close enough and also opposing them point in the sky ...

Thus, the successes of the last two decades astronomy lead to the following critical conclusions:

1. the emergence of highly complex organic molecules – an integral part of the process of the birth of stars and planets,
2. there are all conditions for these molecules gave rise to the simplest biological structures,
3. comets and meteorites are able to deliver these structures to already existing planets,
4. everywhere in the universe there must be a life built on the basis of carbon.

Certainly, it is better certainly say "primary life" – but, as already stated, we will not be able to reliably detect distant results of its evolution – a mind that is already moved to supercomputers, or controls the movement of the stars, etc. etc., – the possibility of what scientists say is quite serious ... Activity of such a kind we most probably will consider as natural processes or phenomena. And note again that the principle of "presumption of naturalness" explicitly formulated by I.S. Shklovsky – we should should consider the phenomena as natural one to the latest extreme, as it is a scientific work, but not science fiction ...

The discovery of another mind will be the greatest event in the history of mankind, that will change our destiny. Most likely, it will happen in the normal astronomical observations – when they become infinitely more extensive and time-ridden observations, and in its technical facilities. The entire spectrum of electromagnetic radiation became available to us completely only fifty years ago. The number of large (our present standards) telescopes on Earth, both radio and optical, does not reach fifty so far – and they are not used for long track one and the same object. We are still in the very beginning of the road to the stars – and to the possible habitants of their planets...

Literature

1. Hart M.N. Explanation for the Absence of Extraterrestrials on Earth. 1975, QJ RAS, v.16, p.128.

2. Shklovsky J.S. On the possible uniqueness of intelligent life in the universe Problems Philosophy, 1976, p. 80.

3. Gindilis L.M. SETI: Searches for Extrterrestrial Minds, Moscow, Fiz.-Mat., 2004 (in Russian).

4. Kardashev N.S. Вопр. Философии , 1976 no12, с. 43

5. Lefebvre V.A., Efremov Yu.N., Cosmic Intelligence and Black Holes. 2000. http://arxiv.org/pdf/astro-ph/0005546v1.pdf.

6. Efremov Yu.N. Giant stellar arcs in the Large Magellanic Cloud: a possible link with past activity of the Milky Way nucleus. 2013, MNRAS Letters, 429, p.75

7. Ичас М.И. О природе живого. Изд. “Мир”, 1994, с. 41

8. Allamandola, L.J. PAHs and the Universe. C. Joblin and A.G.G.M. Tielens (eds), EAS Publications Series, 46 (2011), 305-317.

Chapter 6

The possibility of an interstellar empire

by **Stephane Dumas**
The SETI League, inc.
jgsdumas@gmail.com
and **Yvan Dutil**
Yvan.Dutil@sympatico.ca

Abstract

For as long as people have looked in the sky, they have imagined extraterrestrial civilisations populating the planets around other stars. The idea of interplanetary empires has already entered the realm of science-fiction and many writers of the genre have even elaborated on such empires. The reality may be otherwise. Is it possible that an Alien civilisation could have really established colonises outside its own star system ? This paper will make a survey of possible motivations and technological capabilities, within the reality of Physics, for the establishment a interstellar empire.

6.1 Introduction

In 1950, Enrico Fermi postulated its famous paradox : where are they? The probability of alien civilisations being older than our own is not 0. Fermi reasoned that they should have explored the galaxy by now. This idea is also found in several science-fiction stories. It is popular concept that an alien civilisation must have established an empire.

There may be another solution to this question. It is assumed, too easily, that interstellar travels are easy for an advanced civilisation. It may be otherwise and will be discussed in this paper.

Any long lived civilisation would have needed to solve important problems related to its existence (i.e. overpopulation, food and energy supply).

Those aspects will be also discussed in this paper. The authors will make a survey of possible motivations and technological capabilities, within the reality of Physics, for the establishment a interstellar empire.

6.2 Longevity and collapse

The longevity of a civilisation depends greatly on its capacity to obtain resource and hold a stable socio-political structure. Earth History gave us some examples of long lived civilisations (i.e. Babylon, Ancient Rome and Ancient Egypt) but each of them finally collapsed for a reason or another. Even during the Egyptian civilisation (that last over 5,000 years), there were three periods where it almost collapsed (i.e. intermediate periods).

Natural evolution requires several billions year from the appearance of live to a Human type civilisation. Emergence of life may be quick. On Earth, early sign of life appeared 3.8 billions years ago. Complex life form (i.e. plants and animals) appeared on Earth between 200 and 600 millions years ago. But a real technological civilisation may take much longer. [1] gives three different scenarios for the evolution of a civilisation. He estimated that a civilisation will require between 1 and 6 billions to appear on an Earth-type planet.

With the apparition of technologies (e.g. tools, wheel, electronics), every aspects of the civilisation will grow and become interdependent. However it cannot sustain an exponential growth for long time even when neglecting numerous laws of physics. The growth, mainly characterised by resource consumption and increase in population, cannot last forever.

Technology may not always bring solution to problems. A technological solution to solve a problem requires resources and energy. It may produce pollution in the form of reject from the manufacture and from the discarded devices. A recent example would be the solar panels that is often presented as a source of renewable energy but required heavy metal to produce and does not have a long operational life.

An alien civilisation would have to retain its stability over a thousand year (or even a million years), controls its growth and develop in the same time space technology in order to establish its empire.

6.2.1 Population Growth

Population growth is the major concerned for a civilisation, even for a non technological (i.e. electronic gadgetry) civilisation. The simplest economic model for growth has been proposed by Malthus [2]. It simply states that population growth will be exponential while resources growth must be arithmetic which lead to a reduce wealth unless the population growth is stopped.

Since then various economists have studied the population behaviour and resources stocks for the case with renewable resources. These models are of the Ricardo-Malthus type [3] and are a subtype of Lotka-Voltera predator-prey models [4, 5]. These models produce three types of solutions: extinction, oscillation around a fixed point and stable steady states.

Such models have been successfully applied to population collapse of the Easter Island [6]. Generalisations of this model [7, 8] indicates than the *only escape of Malthusian trap is through institutions restricting utilization of resources.* This restriction itself is very difficult to implement effectively.

6.2.2 Sustainability

Sustainability is the capacity to endure. It is an important aspect of a long lived civilisation. Sustainability is even harder to achieve when non renewable resources are modelled. Some authors argue that technology will compensate for the natural capital loss [9] but others consider this as impossible [10, 11]. Using a completely different approach (system dynamic analysis), Meadows [12] came to the same conclusion.

Limitation of the action of the technology to insure sustainability has already being pointed out by the English economist Jeavons in 1865. Any technological amelioration leading to an improve efficiency will increase the affordability of this technology, which will then increase resources consumption. A modern formulation, known as the Khazzoom-Brookes Postulate [13, 14], argues that energy saving innovations can end up causing even more energy to be used as the money saved is spent on other goods and services which themselves require energy in their production.

Anthropologists have uncovered a rare example of strong sustainability for human civilization on a small pacific island: Tikopia. Against unfavourable odds, Tikopian have managed to survive on this isolated ecosystem for three millennia. Archaeological records show a first phase of sharp decline in forest areas, increased erosion, depletion of fish stocks and extinction of bird species, closely paralleling indigenous population growth. However, a striking finding by archaeologists is that the phase of environmental degradation was followed by a progressive historical change in society's resource-management practices.

Tikopian took effective measures between the years 1000 and 1800 to stabilize their population at approximately 1,281 to 1,323 people. They accomplished their goal by infanticide, abortion, and decreeing that only first-born sons could have children. In addition, the inhabitants shifted from *slash and burn* practices to sustainable agriculture. Doing so, they have replaced the island natural ecosystem by an artificial one that mimics the structure and interrelationship found in natural ecologies. Finally, they eliminated pigs, despite the value Polynesians placed on them, because they damaged gardens and ate food than human could consume [15, 16].

Amazingly, Tikopia sits on the cyclone belt, so every year its inhabitants deal with cyclones, every five or ten years, bringing heavy winds. Not only Tikopia's society is sustainable but is also very resilient.

6.2.3 Changes in the Ecosystem

Ecosystems are by definition sustainable. Selective pressures are hypothesised to drive evolution in one of two stereotyped directions: r- or K-selection [17, 18]. These terms, r and K, are derived from standard ecological algebra, as illustrated in the simple Verhulst equation (6.1) of population dynamics.

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \tag{6.1}$$

where r is the growth rate of the population N, and K is the carrying capacity of its local environmental setting. Typically, r-selected species produce many offspring, which are, comparatively, less likely to survive to adulthood. Whereas K-selected species invest more heavily the nurture of fewer offspring, which has a better chance of surviving to adulthood.

In unstable or unpredictable environments r-selection predominates, where the ability to reproduce quickly is crucial, and there is little advantage in adaptations that permit successful competition with other organisms, because the environment is likely to change again. In stable or predictable environments K-selection predominates, as the ability to compete successfully for limited resources is crucial. Populations of K-selected organisms typically are very constant and close to the maximum that the environment can bear. It should be pointed out than in natural ecosystem biodiversity tend to increase both the stability and the productivity of ecosystem [19, 20].

Given that the Sun is the ultimate source of heat, the equilibrium temperature T (288 K) of Earth atmosphere is

$$\frac{\pi R^2(1-A)k}{r^2} = \epsilon\sigma T^4 4\pi R^2 \tag{6.2}$$

k is the solar constant of Earth (i.e. 1,370 W/m^2), r is the distance form the Sun, A is the Earth's Albedo (0.31), R is Earth's radius, ϵ is the effective emissivity (0.61) and σ is the Stefan's constant. All activities generate heat that is absorbed in the atmosphere. Some predictions propose that in 300 years, the mean temperature of the atmosphere could be as high as 291 K (3 K of increase). This increase will have significant impacts on the climate and the ecosystem.

It is likely than any sustainable civilisations would follow a similar evolution trajectory. Therefore, alien civilizations are likely to be extremely complex, very efficient with a very low rate of growth.

6.2.4 Energy Requirement

Any civilisation requires some form of energy (i.e. coal, nuclear, oil) for its activities. In 2007, our civilisation consumed around 1.7×10^{13}W of power. We might wonder what the physical limits to our power consumption are and what rate of growth is sustainable over time.

Solar radiation reaching Earth is estimated to 5×10^{16} W. Kardashev [21] proposed an upper limit of the output of human activities to 1% of the total solar radiation to avoid climatic catastrophe. With a 2% growth rate, we will reach this limit within 170 years.

Nuclear does not seems to be a viable source of energy on very long term. To supply our near term needs (i.e. 1.7×10^{13}W) of power with nuclear, we require some 15,000 nuclear reactors (with a mean output of 3.75×10^{6}W). At that rate of consumption, there is around 80 years of uranium reserve in the Earth crust. However, there are large quantity in sea water (harder to extract) that could give us 5,700 years at an average of 1.7×10^{13}W. The problem is that other material (i.e. rare element) are needed to build a reactor. It is not possible to maintain all those nuclear reactors with the current reserve of those rare elements. Furthermore, the live of a nuclear installation is limited (e.g. around 100 years). Even switching to thorium-base nuclear reactors will not be a better solution at long term. Thermonuclear power sources are not yet operational.

Conventional sources of fuel (i.e. oil, coal, nuclear) have limited reserve. The renewable source of energy (i.e. solar, wind, hydro) have a limited power output.

These physical limitations leave few possibilities for sustainable civilisation.

They could be photosynthesis limited ($\sim 10^{13}$W), climatically limited (1.27×10^{14}W) or solar flux limited (1.74×10^{18}W).

In the Mojave Dessert in southern California, when the Ivapah project will be completed the total output power predicted is 3.7×10^{8}W (3,600 acres of solar panels). Power density of this installation is 25.4 W/m^2. 6.693x10^5 km^2 of solar panel would be required to provide a total power of 1.7×10^{13}W.

The Roscoe wind farm in Texas has a capacity of 7.82×10^{8}W (154 sq-miles). 8.677×10^{8} km^2 would be required to provide a total power of 1.7×10^{13}W.

6.2.5 Singularity

The fusion of biological and artificial entities is often referred to the singularity. Any advanced alien civilisations may have taken this path to improve there condition, by necessity (i.e. climatic changes) or choice. We may not even recognise its motivations. A civilisation in post-biologic stage of evolution may have completely different motivation than us.

6.3 The Complexity of Space Travel

Travelling between stars is a very difficult process [22, 23]. It requires huge amount of energy, is not without danger and it is a very long process. The technology used by the Alien is not known but Physics must be respected regardless of the technology use to move the ship.

For chemical furled rockets, the final velocity after burnout of all its fuel, V, the exhaust velocity generated by the propellant, S, and the so-called mass ratio, $\mathfrak{M} = M_i/M_f$, are connected by the well-known rocket formulas described in equation 6.3.

$$\frac{V}{S} = ln\mathfrak{M} \tag{6.3}$$

where M_i is the total initial mass of the space ship (including fuel), and M_f is the mass of the after burnout. With a 90% of the initial being fuel, equation 6.3 gives $V = 2.3S$. The best value of V would be around $V = 6.9S$ meaning that 99.9% of the initial mass is fuel. It is difficult to achieve high speed with conventional chemical fuel and the payload is a small fraction of the total rockets mass.

6.3.1 Relativistic Treatment

When the ship travels near the speed of light, the ratio of masses most be computed by equation 6.4.

$$\mathfrak{M} = \frac{M_i}{M_f} = \left(\frac{c + V_{max}}{c - V_{max}}\right)^{c/2V_{ex}} \tag{6.4}$$

Chemical fuel is not sufficiently powerful for this task. For fusion propellant (e.g. deuterium, tritium), $V_{max} \approx 0.99c$ and $V_{ex} \approx c/8$, $\mathfrak{M} = 1.6 \times 10^9$. While for anti-matter propellant, $V_{ex} \approx c$ and $\mathfrak{M} = 14$. For a complete trip, it is more like $\mathfrak{M}^4 = 4 \times 10^4$. To take a payload of 10 tons in a round-trip to a star, the ship needs to carry 400,000 tons of anti-matter fuel.

Since anti-matter fuel is out of reach for out technology (but maybe not for the alien's), the only realistic way to achieve high velocity is using nuclear or thermonuclear propulsion system (i.e. Orion or Daedalus type of ships). But those may only reach a fraction of the speed of light. The consequence is that the journey will take much more time.

6.3.2 Power-Mass Ratio

The acceleration of a rocket, b, is given by equation 6.5.

$$b = \frac{\text{thrust of engine}}{\text{total mass of rocket}} = \frac{2P}{S} \tag{6.5}$$

where P is the ratio of power of engine to total mass of rocket. If we are working with a high exhaust velocity S, we need a high power-mass ratio P, otherwise we would get only a small acceleration. In order to maintain $b = 1g$ with $S \approx c$ (for Human crew), we must have $P = 3 \times 10^6$W/g. A typical nuclear reactor with 15 MW and 800 tons will gives $P = 0.02$W/g.

6.3.3 Protection for the crew

The interstellar medium is not empty ([24], number density in gaseous cloud$=10^7 - 10^9 m^{-3}$, interstellar media $= 2 - 3 \times 10^5 m^{-3}$ and near solar system $= 10^6 m^{-3}$). The ship will need to be shielded against impact and radiation. Proton interaction could be neglected for $V < 0.9c$. According to [25], electron may not be a problem (e.g. Compton Effect) for $V < 0.8c$. However, particle impact (i.e. proton, electron and dust) will generate heat and possible erosion. To avoid radiation problem, the ship will need to move slower and thus increasing the length of the journey. A faster ship will need to sacrifice a fraction of the payload for shielding.

6.3.4 Project Orion

In the 50's, a group of American scientist were assembled together in think tank called General Atomics. On of their many projects was to build a space vehicle using nuclear explosion as a propulsion mean. It was called The Project Orion [26, 27, 28, 29].

The idea was to used nuclear devices (around 1kT yield) to push on the vehicle generating a huge acceleration over a small interval of time. The bigger the vehicle, the smaller the acceleration. It was then decided to build a huge (i.e. several thousands tons) ships capable to travelling between planets.

The ship was capable to handle a crew of 8 to 20 people. Several missions scenarios were planned for the Moon, Venus, Mars and Jupiter.

The first generation ships would have been designed for interplanetary journeys only but a design for interstellar mission was planned. To reach Alpha Centauri, it would have required some 25 millions nuclear devices to accelerate and the trip would have taken 150 years. Clearly, it would have been a multi-generation ship and huge. Another 25 millions nuclear devices are required to decelerated.

6.3.5 Project Daedalus

In the 70's, a groups of scientists took part of a feasibility study for an interplanetary vehicle. It is since been known as the Project Daedalus [24]. The plan was to build a vehicle capable of travelling to a neighbour star using the known technology of the time (1970). The payload would have been around 500 tons and provision for a crew was done. The launch mass was around 150,00 tons. The propulsion was based on thermonuclear reactions using tritium as the main fuel. The final velocity, after acceleration, would have been around 0.167c. The design called for a two-stage vehicle.

The mission was a fly-by of the Barnard's star at 6 light-years and would have taken 40 years to the vehicle to reach it.

6.4 Motivations for Colonisation

The alien civilisation need a powerful motivation in order to spend the huge amount of resource to send ships and colonised other world.

Overpopulation may be on the top of the list. This scenario requires more than simple cargo ships to carry the mineral. Moving large number of people to other world required huge transport ship capable of supporting the crew for a very long time.

Von Hoerner [30] has remarked that even interstellar colonisation at the speed of light cannot solve the present human population explosion on the planet Earth. With an exponential growth rate of g = 0.02/year and the colonisation sphere expanding at the speed of light, in 500 years the expansion volume will have a radius of 50 pc with all habitable planets in that volume reaching the Earth's present population density.

A single expansionist power will colonise the galaxy in ∼107 years if the starship velocities are 0.1c.

The apparent absence on Earth of representatives of advanced galactic civilisations thus argues against the presence of any such expansionist civilisations devoted to domination of the Galaxy.

6.5 Signs of Interstellar Empire

Any alien civilisation capable of establishing colonies in an interstellar level, will require a colossal source of power. One of the best solution to that problem is to collect the energy from its own star. This type of structure would have an infra-red signature that may be visible with our current state of technology.

They have been no report of such observation [31]. This is not an absolute proof that no empire exist but rather that they did not build such structure in the vicinity of Earth.

6.6 Conclusion

A very old civilisation (i.e. older than us) would have to have solved all the different problems listed in this paper (and probably more). That is to be able to maintain its own structure. Otherwise, it would be an eternal cycle between rise and fall.

Energy requirement, climate, food crisis and population growth may be incentive to colonise other worlds. But it may be more practicable to find a local solution to those problems then move to another star systems. However, colonising planets of its own solar system may be a viable alternative.

Civilisation on Earth took roughly 4 billions years from the formation of the planet. The Universe is about 14 billions years old. Given that Earth is an example of mean value, then many civilisations could have appeared and disappeared during that time.

Some of those civilisations would older than us. They would have plenty of time to travel to the stars and established colonised. The fact that no space ship has been detected in our solar system, nor any alien artefact, nor that any alien ambassador has landed on Earth may be a sign that there is no Interstellar Empire, or at least that interstellar journeys are very difficult.

But it does not mean that there are no alien civilisations out there! Contrary to the Fermi Paradox conceding that the only solution to this absence of evidence is the absence of alien.

Bibliography

[1] R.A. MacGowan and F.I. Ordway. *Intelligence in the Universe.* Prentice-Hall Inc.,Englewood Cliffs, NJ., 1966.

[2] T. Malthus. *An Essay on the Principle of Population.* Penguin Classics, 1798.

[3] D. Ricardo. *Principles of Political Economy and Taxation.* London: John Murray, 1817.

[4] A. J. Lotka. *Elements of physical biology.* Baltimore: Williams and Wilkins Co., 1925.

[5] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[6] J.A. Brander and M.S. Taylor. The simple economics of Easter Island: a Ricardo Malthus model of renewable resource use. *Am. Econ. Rev.*, 88:119–138, 1998.

[7] R. Reuveny and C.S. Decker. Easter Island: Historical Anecdote or Warning for the Future? *Ecological Economics*, 35(2):271–287, 2000.

[8] J.C.V Pezzey and J.M. Anderies. The effect of subsistence on collapse and institutional adaptation in population-resource societies. *Journal of Development Economics*, 2003.

[9] R.M. Solow. Georgescu-Roegen versus Solow/Stiglitz. *Ecological Economics*, 22:267–268, 1997.

[10] N. Georgescu-Roegen. *The Entropy Law and the Economic Process.* Cambridge MA USA: Harvard Univ. Press., 1971.

[11] H.E. Daly. Georgescu-Roegen versus Solow/Stiglitz. *Ecological Economics*, pages 261–266, 1997.

[12] D. H. Meadows, D. L. Meadows, and J. Randers. *The limits to growth: a report for The Club of Rome's project on the predicament of mankind.* New York : Universe Books, 1972.

[13] L. Brookes. Energy Efficiency and Economic Fallacies. *Energy Policy*, 18(2):199–201, 1990.

[14] J. D. Khazzoom. Economic Implications of Mandated Efficiency Standards for Household Appliance. *Energy Journal*, 1(4):21–39, 1980.

[15] R. Firth. *We, the Tikopia.* Stanford University Press, Palo Alto, CA, 1983.

[16] P. Kirsch. *On the road of the winds: an archeological history of the Pacific Islands before European contact.* University of California Press, Berkeley, CA, 2000.

[17] R. MacArthur and E. O. Wilson. *The Theory of Island Biogeography.* Princeton University Press, ISBN 0-691-08836-5M, 1967.

[18] E. R. Pianka. On r and K selection. *American Naturalist*, 104:592–597, 1970.

[19] K. H. Johnson, K. A. Vogt, H. J. Clark, O. J. Schmitz, and D. J. Vogt. Biodiversity and the productivity and stability of ecosystems. *Trends in ecology and evolution*, 11(9):372–377, 1996.

[20] J. McGrady-Steed, P.M. Harris, and P.J. Morin. Biodiversity regulates ecosystem predictability". *Nature*, 390:162–165, 1997.

[21] N.S. Kardashev. The astrophysical aspect of the search for signals from extraterrestrial civilizations. In S.A. Kaplan, editor, *Extraterrestrial civilizatons : problemes of interstellar communication*, 1971.

[22] E. Purcell. Interstellar communication. In A.G.W. Cameron, editor, *Interstellar communication - The Search for extraterrestrial life*, chapter Radioastronomy and communication through space, pages 121–143. Benjamin inc. New York, 1962.

[23] S. Hoerner. In The Search for Extraterrestrial Life. In A.G.W. Cameron, editor, *Interstellar communication - The Search for extraterrestrial life*, chapter The General Limits of Space Travel, pages 144–159. W.A. Benjamin Inc., 1962.

[24] A.R. Martin. *Project Daedalus.* Journal of the British Interplanetary Society, 1978.

[25] E.T. Benedikt. Disintegration barriers to extremely high-speed space travel. *Advances in the Astronautical Sciences*, 6:571–588, 1961.

[26] J.C. Nance. *Nuclear Pulse Space Vehicle Study, Vol.1 - Summary*, volume 1. General Atomic, 1964. Enter text here.

[27] J.C. Nance. *Nuclear Pulse Space Vehicle Study, Vol.3 - Conceptual Vehicle Designs and Operational Systems*, volume 3. General Atomics, 1964.

[28] J.C. Nance. *Nuclear Pulse Space Vehicle Study, Vol.4 - Mission Velocity Requirements and System Comparisons*, volume 4. General Atomics, 1964.

[29] G. Dyson. *Project Orion.* Owl Books, NY, 2002.

[30] S. Hoerner. The Number of Advanced Galactic Civilizations. In C.Sagan, editor, *Communication with extraterrestrial intelligence.* MIT Press, Cambridge, 1973.

[31] R.A. Carrigan. Starry Messages: Searchinf for Signatures of Interstellar Archaeology. *JBIS*, 63:90, 2010.

Chapter 7

The end of socium and the SETI challenge

by **S.A. Yazev**
Astronomical Observatory of Irkutsk State University, Irkutsk, Russia
Institute of Solar-Terrestrial Physics of SD RAS, Irkutsk, Russia

Abstract

Evolution trends of terrestrial civilization testify that information encoding methods typical of human brain will be accessible in the foreseeable future. This implies that development of a uniform network of people's mind (brain-net) will be possible. These achievements will lead to disappearance of socium (in the common sense of the word), since all this will give rise to common mind and common base of human memory united in the brain-net. On the one hand this implies creation of an immortal supraperson with unlimited memory, but on the other, this will result in a new type of mind based on genetically accelerated minds of many people united together. This development will cause disappearance of socium and traditional culture, giving rise to the common mind of the second generation. The SETI problem will most probably be solved after transition to the common mind, when we tackle an issue related to our inability to analyse and understand the surrounding reality (including existence of other civilizations) on the new level.

7.1 Introduction

The current tendencies for a rapid advancement of the technological capabilities of our terrestrial civilization testify that in the foreseeable future it will be possible to puzzle out the mysterious method of encoding information which has been achieved by human brain. What actually happens is that an understanding of the main principles of this method is now surfacing and there is reason to hope that success might be right around the corner. The problems associated with these tendencies were addressed in futurological publications by Stanislaw Lem [1], and by many other authors. The objective of this paper is to assess the possible consequences of future technological achievements along this line within the context of the range of issues relating to SETI.

7.2 The anticipated near future

It cannot be doubted that as soon as the puzzles of method of encoding information used by human brain are unravelled, this will give birth to new types of computers capable of implementing this principle. Also, whether or not the new computers will feature higher speed when compared to the existing ones is of little importance. A much more important point is that, unlike the situation with our current understanding of the workings of human brain, the operation algorithms of the new computers will be thoroughly known. As a result, it will be possible to achieve an integration of memory units and exchange of information for new and old computers. In essence, this signifies that there will emerge a possibility of interfacing the computer and human brain. The new type of computer will be able to operate as a peculiar kind of modem for an exchange of information between the computer and brain. This revolutionary breakthrough in the technosphere would cause civilization to alter radically. The integration of brain and the computer will in fact open up the way for the integration of many people's consciousnesses into unified networks.

One further rapidly evolving direction of scientific and technological progress places us on the verge of mastering the technologies of manipulating with the human genome. The present trial and error epoch will give way to the epoch witnessing the mastering of genetic operations performed on planned results. These technologies would furnish an opportunity to implement eugenic procedures in practice enabling, for example, an enhancement in the intellectual capabilities of a human individual to a genius-level intellect, and an alteration to human physiology and anatomy in a prescribed manner. The varied mutations that are realized in nature exemplify that, in principle, not only random results but also consciously planned similar and other results must be possible. It seems reasonable to say that the avenue of such an evolution runs into not fundamental but purely technical and, hence, temporal and, in principle, remediable difficulties.

Of course, there are also ethic problems: the case in point is for the first time in the entire history of terrestrial biosphere we are witnessing the start of a conscientious intervention of man into nature, and this is taking place at the genotype level rather than at the phenotype level [2]. Furthermore, one would expect an avalanche differentiation of the human genotype.

Unquestionably there must arise political and social forces which make such research difficult, much like the popular trend today to ban all forms of human cloning. Nevertheless, experience suggests that any technological achievements which became possible at a particular evolution stage of mankind were always put into practice immediately against all the odds [2]. It makes no difference whether the achievements involved will be made in top-secret military laboratories or at secretly financed medical centers on coral islands far from and beyond the legislative and public control; an important point is that sooner or later the result will be obtained, and this author does not have any doubts about that, because there are no laws of nature whatsoever which would contradict this, but on the other hand the regularities of civilization evolution fit this scenario quite well. The characteristic time needed to resolve most technical problems along these basic directions of development, starting from the current evolution level, is estimated by this author at 100 (and even 200) years, which is a negligibly short time span when compared to the mankind's lifetime. There is a good probability that it is precisely the revolutionary leap forward which is pointed out by S.Kapitsa, A.Panov and G.Beskin, who predict profound qualitative changes in the life of mankind in the immediate future [3-5].

7.3 Consequences of anticipated changes

The achievements under discussion will result in the disappearance of the socium in a general sense. The unification of the consciousnesses of individual personal entities will proceed via computer networks; after that, it is likely to lead directly to emergence of a common consciousness and a common memory of those personal entities, who would wish to unite into a brain-net. This signifies, on the one hand, the creation of an immortal suprapersonal entity, or a suprapersonality endowed with an actually unbounded and eternal memory and, on the other, the emergence of a new type of consciousness based on combining the many genetically-driven consciousnesses of individual personal entities.

The properties of such a consciousness are difficult to envision. How will the recollections of different personal entities interact in a unified consciousness? What will the personality qualities of the new type of consciousness be like and Will they exist altogether in a general sense? An attempt can be made at the construction of the model for such a state, but it is hard to suggest with certainty the particular relevance (if any) of this model to reality. It seems likely that these questions could only be properly answered based on future experience.

The heuristic possibilities of the new consciousness are also hard to imagine. Obviously they must multiply exceed the possibilities of a separate individual. What we term intuition (the unconscious solution of problems with the involvement of our brain's subcortical regions) would also be evolving further, given the combined well-concerted operation of the system of joined brains.

As a result of the evolution of this kind, the socium and traditional culture are doomed to disappear upon uniting into a common mind, as well as politics and economics which constitute essentially the methods of organizing the interaction of separate personal entities. Disappearance of individual personalities would entail disappearance of their interaction.

The switch-over to a unified consciousness brings up an exceptionally important and uncertain question as to the transformation of the gender-specific nature of an individual's personality which determine, in many respects, behavior, life strategy, cultural codes, values, and stimuli. We can reason that, in a broad sense, most human problems are due to lack of information regarding the other people's thoughts and sensations. This problem would disappear, once the unified conscience phenomenon originates, perhaps giving birth to new ones. What will happen to the sexuality phenomenon in particular as well as to the instinctive and emotional sphere in general? Will it be, in some way or other, replaced or transformed or will it disappear? These questions remain incomprehensible (to this author). In all likelihood, this sphere will also become a rationally controlled one.

It is apparent that brain will undergo restructuring in order to improve effectiveness of a unified conscience. In order for the multitude of consciences to be involved in a unified process of thinking would require some analogue of a central processor featuring advanced performance characteristics (superbrain), capable of "requesting" and "connecting" the individual consciences. An alternative must not be ruled out, namely a distributed parallel thinking by separate consciences. These are very crude estimates and reflect our current problem solving algorithms for today's electronic devices. It is likely that the future structure of unified thinking might be organized by a different, presently unknown method.

It might be anticipated that the new conscience will permit transformation of the physiology of the bearers of conscience elements (humans) in accord with the new state. If conscience is interfaced with the new (and through them, with the old) computers, brain as the only variant of material bearer of consciousness may well be not the most effective and the most convenient variant of implementation of the "thinking processor". The variants of embodiment of this physiology can be quite varied, including< for example, the abandoning of the biological body as the material bearer of conscience of structured physical fields, etc. It seems likely that it will become possible to use the various biological objects as a peculiar kind of interfaces of unified conscience where its connection to a dolphin's, bat's or bird's brain will make possible the sensual perception of ultrasound or a magnetic field. A future particular engineering problem may well involve the creation of new sensory organs, such as for perception of radiation, radio waves, and the like.

In the SETI context, what this means is that an intelligence capable of making contact (that has evolved on the basis of the terrestrial or some other civilization) will constitute with a high degree of probability not a socium but some unified second-generation intelligence on the material basis unknown to us and with unknown spatiotemporal characteristics. In any event we can be assured that they will be not the little green men flying interstellar cruisers – such an approach simply extrapolates to the future the contemporary state of socium which (the socium) will no longer be existent.

The ongoing (within SETI programs) quest for signs of extraterrestrial civilizations existing in the form of sociums similar to the terrestrial one is, in this author's opinion, strategically disadvantageous, as it is highly probable that an advanced socium would transform to a state of unified conscience. The external manifestations of the activity of such an intelligence must be dramatically different from those which are characteristic for a socium. The technological type of evolution of civilization will most likely cease at this stage of development.

7.4 Conclusion

One would expect a gigantic gap between the level of intellect of an individual personal entity the second-generation intelligence under discussion which has gone through the stage of integration of individual consciousnesses. The SETI problem will most probably be solved immediately the socium turns into a state of unified consciousness when our present inability to properly analyse and understand the surrounding reality that, perhaps, includes the manifestations of other intelligences in the Universe, the microworld, and even in our immediate environment, will be easily coped with.

I am profoundly grateful to Sergei Alexandrovich Shumsky (FIAN) for a number

of helpful discussions, and to Yuri Nikolayevich Yefremov (GAISh MGU), who has kindly agreed to present this paper at the conference.

References

1. Lem S. The Moloch. Moscow: Tranzitkniga, 2005 (in Russian).

2. Yazev S.A. On the phenomenon of hypothetical function-civilizations/ Bull. Special Astroph.Obs., 2007, V.60-61, P.195-199 (in Russian).

3. Kapitsa S.P. Global Population Blow up and After. The demographic revolution and information society. A Report to the Club of Rome. Hamburg: Global Marshall Plan Initiative; Moscow: Tolleganza, 2007.

4. Panov A.D. Scaling law of the biological evolution and the hypothesis of self-consistent galaxy origin of life / Advances in Space Research, V. 36, P. 220-225, 2005.

5. Beskin G.M. The demographic transition and great silence – does sociocosmological constant exist? // Bull. Special Astroph. Obs., 2007, V.60-61, P.187-194 (in Russian).

6. Efremov Yu.N. Where are they? // Bull. Special Astroph.Obs., 2007, V.60-61, P.158-161 (in Russian).

Chapter 8

Probable natural sources of the "Wow!" radio signal.

by **G.A. Gontcharov**
The Main (Pulkovo) astronomical observatory
of the Russian Academy of Science,
Saint-Petersburg, 196140, Russia;
e-mail: georgegontcharov@yahoo.com;
phone: +7-921-3228899

Abstract

In order to find the natural source of the "Wow!" radio signal, received on August 15, 1977 at a frequency of about 1420 MHz, the celestial objects up to the 18th magnitude are analyzed in the respective region of the sky. Six candidates of various nature are found to be able to produce rare powerful radio pulses in the narrow band of frequencies: 1) suspected radio star HIP 95865 of the 5.6 magnitude under certain circumstances, 2) weak radio source NVSS B192505-265140 amplified in the close orbital motion around an unseen massive body, 3) stars V905 Sgr and HD 182460 on an unstable stage of their evolution, 4) suspected maser about the forming star TYC 6884-2359, 5) suspected old supernova remnant, i.e. suspected rotating radio transient IRAS 19224-2707. Permanent monitoring for several years would confirm or deny the relation of these objects with the "Wow!" signal.

8.1 Introduction

August 15, 1977 at 23.16 Eastern Daylight Savings Time the Big Ear radio telescope at Ohio State, USA [1] received an unusual radio signal at a wavelength of about 21 cm. Astronomer Jerry R. Ehman, who found the signal on the computer printout several days after its receiving, immediately appreciated the unusual characteristics of the signal and labelled the signal by the "Wow!" legend on the printout. Since then the signal is known as a "Wow!" signal [2].

At that time the Big Ear telescope fulfils a sky survey in order to find quite powerful (more than a few Jy) narrowband (frequency resolution of 10 kHz) radio signals. The observations were made in the drift-scanning mode: the receiver is installed in the meridian, and the rotation of the sky provides the transits of different objects in front of them. In this mode the time of the observation corresponds uniquely to the right ascension (this value is hereinafter referred to as "time/RA"). The observations and their primary processing were largely automated. The main result of the observations, the intensity of radio emission depending on the frequency channel and time/RA was printed out by the computer and viewed by the observer every few days.

This sky survey made by the Big Ear was the largest narrowband one. Up to now the "Wow!" signal is the most powerful and longest narrowband signal with source not identified until now, though not such a unique signal.

The area of the sky with the "Wow!" signal repeatedly scanned as by the Big Ear (including the scan exactly one day after the signal) and by other radio telescopes, including the attempts to find the repeating of the "Wow!" signal: by META system in 1986-1991 [3], by VLA in 1995-1996 [4], by 43-meter Green bank radio telescope (West Virginia, USA) and 30-meter Woodbury telescope (Georgia, USA) in the Phoenix project in 1997-1998 [5], by 26-meter radio telescope of the Hobart University of Tasmania in 1998-1999 [6]. These observations found many radio sources at a frequency of about 1420 MHz in the field of "Wow!" signal but none of them is superior to the power of 0.1 Jy.

More than a thousand times during the survey in 1977-1983 the Big Ear received single narrowband (band width not exceeding 10 kHz) pulse (lasting 10-30 seconds) signals of low power (a few Jy) not related to a known radio source [7]. An example of a surprise to the observers is a signal on June 9, 1994 at a frequency of 1612.54 MHz. It appeared as a OH (hydroxyl) maser [8]. Such a maser is usually a part of an extended object (cloud) in different parts of which the masers of various power and duration turn on and off from time to time. It is perceived by observer as one source, changing frequency, power and position on the celestial sphere in a small range, both gradually and discretely. Apparently, the "Wow!" signal can also be attributed to this class of quite common, interesting, but poorly understood pulse radio sources.

Numerous speculations about this signal as a one of an extraterrestrial civilization force us to consider firstly possible natural sources of the signal. In this way the progress of radio astronomy in the detection and understanding of unusual radio sources made in recent decades is important.

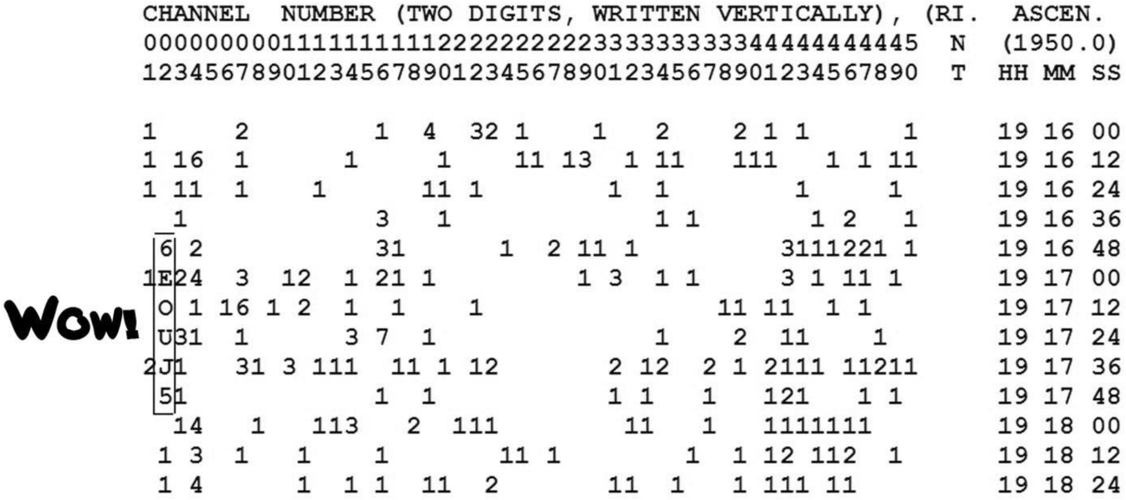
8.2 Features of the signal

Figure 8.1 shows the contents of the printout with the "Wow!" signal. The signal itself is the several unusual digits and letters boxed in the second column at left and marked by the legend "Wow!". Two vertical arranged digits at top mark the number of the frequency channel from 01 to 50. This forms the horizontal axis. Each channel has a width of 10 kHz. The channel 01 is the frequency 1420.4456 whereas the channel 50 is 1420.9356 MHz. Unfortunately, the initial hydrogen line frequency of 1420.4056 MHz (not shifted by the effect of Doppler) is out of the range of the observed frequencies. The respective radial velocities of the sources displaced due to the Doppler effect with respect to the observer are from -8 km/s for the channel 01 to -112 km/s for the channel 50. The "Wow!" signal appears in the channel 02, i.e. at the frequency of 1420.4556 MHz. Consequently, the source moved to the observer with a velocity of about 10 km/s.

At the time of the observation of the signal the Earth was located and moved relative to the Sun so that 20 km/s must be subtracted from the specified values if we want to consider the barycentric velocity of the source, i.e. the velocity with respect to the center of the masses of the Solar system. Thus, the channels from 01 to 50 correspond to the barycentric velocity from -28 to -132 km/s. And if the "Wow!" signal was radiated at a frequency of 1420.4056 MHz, the source has a barycentric radial velocity of about -30 km/s. By the way, the frequency range for observations was selected unsuccessfully: a minority of space objects in the nearest part of the Galaxy have such velocities.

The right columns at the figure are the values of time/RA about 19 hours and 17 minutes (namely, hours, minutes and seconds – the printout appears on the line every 12 seconds). As shown below, these values of the time/RA can not be used directly and must be corrected for the number of effects.

During the night the telescope was fixed in declination. But it took the next declination strip for the next night. The observations of whole band of the sky at declinations from -35° to +64° were repeated periodically. All data shown in Fig. 1 refer to one strip of declinations.



```
CHANNEL  NUMBER (TWO DIGITS, WRITTEN VERTICALLY), (RI.  ASCEN.
00000000011111111112222222222333333333344444444445  N  (1950.0)
12345678901234567890123456789012345678901234567890  T  HH MM SS

1     2        1  4  32 1    1   2    2 1 1      1     19 16 00
1 16  1      1     1    11 13  1 11   111   1 1 11     19 16 12
1 11  1    1      11 1        1  1        1     1      19 16 24
  1            3   1             1 1       1 2   1     19 16 36
 6 2           31      1  2 11 1         3111221 1     19 16 48
1E24  3  12  1 21 1         1 3  1 1     3 1 11 1      19 17 00
 O 1 16 1 2  1  1    1               11 11  1 1        19 17 12
 U31  1      3 7  1              1    2  11    1       19 17 24
2J1   31 3 111  11 1 12       2 12  2 1 2111 11211     19 17 36
 51            1  1           1 1   1   121   1 1      19 17 48
  14   1   113   2  111        11   1   1111111        19 18 00
 1 3  1   1    1       11 1        1  1 12 112  1      19 18 12
 1 4      1  1 1  11  2       11  1   1 111 11         19 18 24
```

Figure 8.1: The content of the original printout of the "Wow!" signal.

Numbers and letters in the printout reflect the signal strength, or rather the dimensionless signal/noise ratio, and the noise is obtained by averaging a few minutes. The minimum level is marked by spaces, the levels from 1 to 9 – by digits as well as the levels from 10 to 35 – by letters from A to Z. Thus, the series of signal digits and letters boxed in Fig. 8.1 correspond to the intensity profile (the combination of the signal with the antenna pattern) shown in Fig. 8.2.

The duration of the signal or rather its manifestation in the antenna pattern is just over a minute. Fairly symmetrical waveform suggests that we are dealing with a point source signal fixed on the celestial sphere. During the observational period this source had a constant power, participated in the diurnal rotation of the sky and just over a minute superimposed on the main lobe of the antenna. The maximum signal/noise ratio corresponding to the letter U on the printout was 30.5. This corresponds to the signal flux density from 54 to 212 Jy variously estimated. In any case, this is a quite powerful signal. Very few sources in the sky have such flux density in a wide range of wavelengths. However, narrowband signal of such flux density could go unnoticed for a long time. It is important that the "Wow!" signal does not appear in the adjacent frequency bands. Due to this concentration in a narrow band, the energy for signal generation, as shown hereinafter, is quite common for some classes of natural space objects. On the other hand, the frequency band of 10 kHz is extraordinarily narrow for natural radio source. It makes to consider primarily the maser mechanism of radiation generation.

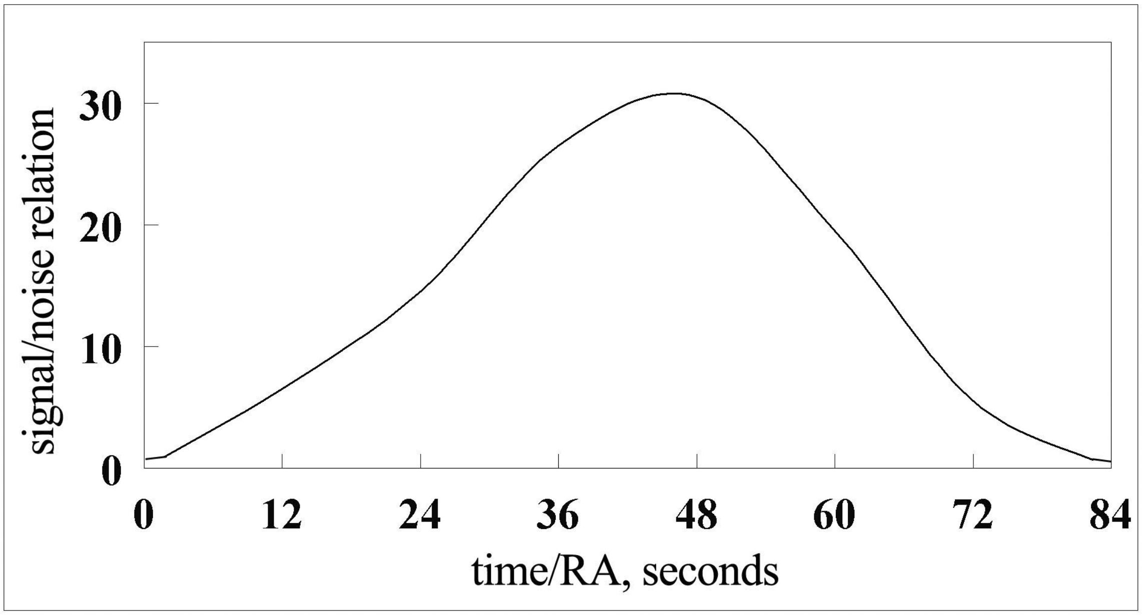


Figure 8.2: The profile of the "Wow!" signal.

The overall design of the Big Ear telescope is shown schematically in Fig. 8.3 and includes two reflectors and two metal horn receivers marked with the arrow. The horns are shown in the schematic Fig. 8.4 in more details. The horns are spaced apart. Therefore, each signal of distant source is registered twice. The "Wow!" signal was registered by only one horn, and, because of the pre-processing features, it is not clear exactly by which one. Therefore, the source of the signal may be in one of two regions of the sky. Its appearance in one horn means that it changed for a minute of watching either a position on the celestial sphere, or the radiation frequency or intensity (the appearance or disappearance during a minute). Almost perfect shape profile of the signal eliminates its significant gradual shift on the celestial sphere (at least, in right ascension) and significant smooth change of flux in a minute of observation. The lack of the signal in the adjacent frequencies eliminates a significant gradual change in frequency. Thus, we have to admit a quite rare for a natural source an abrupt change in at least one of these values (position, flux or frequency). In fact, the "Wow!" signal is more or less long narrowband radio pulse.

The value of time/RA marked on the printout of the signal is less than the true time by 5 minutes 10 seconds due to improperly introduced correction for the displacement of the horns relative to the meridian [2]. In addition, when considering the appropriate sky area one must take into account that the coordinates listed

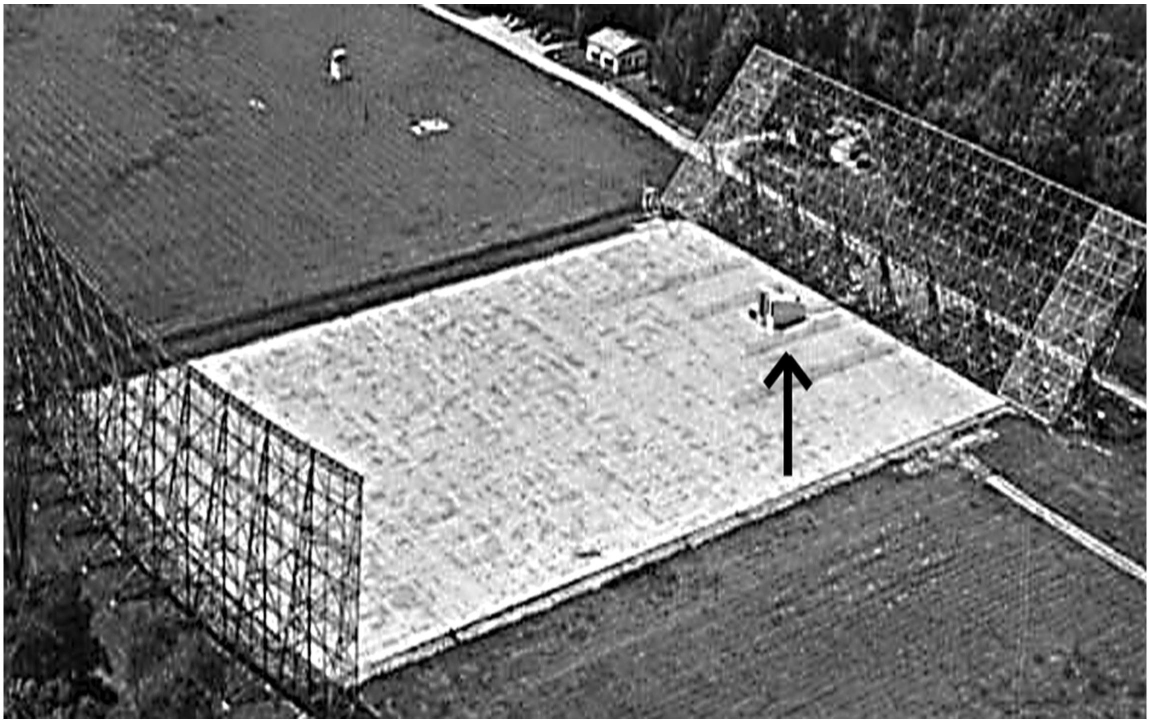


Figure 8.3: Schematic drawing of the Big Ear telescope. The arrow points to a pair of horns.

on the printout are attributed either to the equinox B1950 (mean coordinates, i.e. the coordinate system of the equator and ecliptic of B1950) or to the equator and equinox of the observational date (apparent coordinates, i.e. the coordinate system of the equator and ecliptic of 15 August 1977). Unfortunately, now it is impossible to determine which coordinate system, mean or apparent is used for the observations, reflected on the printout and used then in repeated observations of this region of the sky. There are two possibilities: 1) the telescope, as is usually done in such observations, set up by use of the apparent coordinates, and before the printout these coordinates are automatically converted into mean ones (although such conversion does not make sense during the pretreatment of observations), but it is possible that 2) the apparent coordinates are used as in set up and on printout. In the later case, the legend "RI. ASCEN. (1950.0)" on the printout is incorrect, and, moreover, wrong fields of the sky have been observed in all repeating observations and mentioned in the literature. Both cases of coordinate systems are discussed below.

In the first case (apparent and mean coordinates) after conversion to the coor-

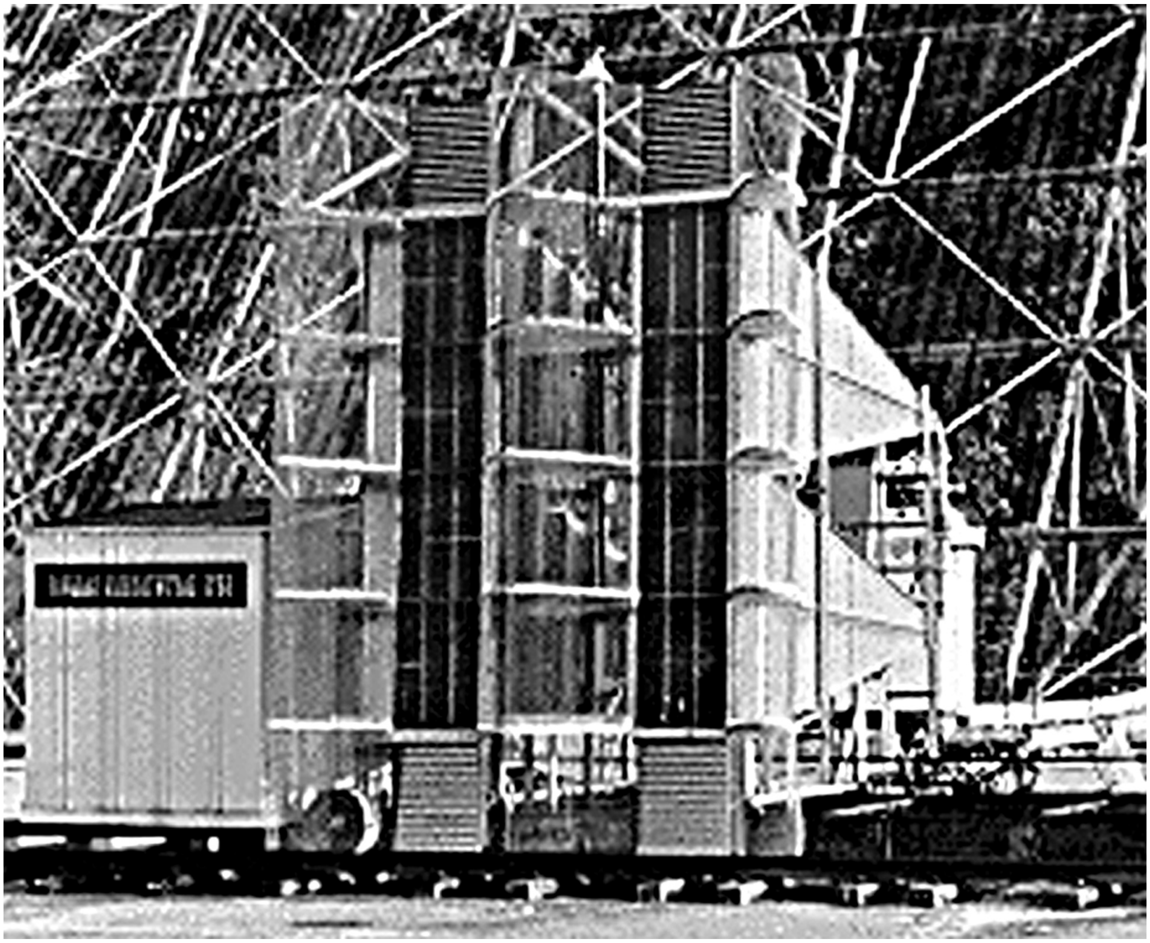


Figure 8.4: Schematic drawing of two horns of the Big Ear telescope.

dinate system J2000 one has 2 lanes in the sky with coordinates $\alpha=$ 19h 28m 22s ±10s or $\alpha=$ 19h 25m 31s ±10s, $\delta=$ -26° 57’ ±20’. These lanes are shown in Fig. 8.5 along with all the objects to V=18m. Fig. 5 shows that the considered area of the sky does not contain any extended or exotic objects (large galaxies, large clouds, etc.). In the second case (apparent and apparent coordinates) the lanes are slightly displaced, above all for right ascension, so that the bright star at the left side of Fig. 8.5 falls exactly in one lane. The second case is much less probable, but it is interesting by the presence of a bright star which could be the source of the "Wow!" signal.

The source of the "Wow!" signal is in the constellation Sagittarius, 21° from the center of the galaxy, which, consequently, can not relate to the signal. The Sun also

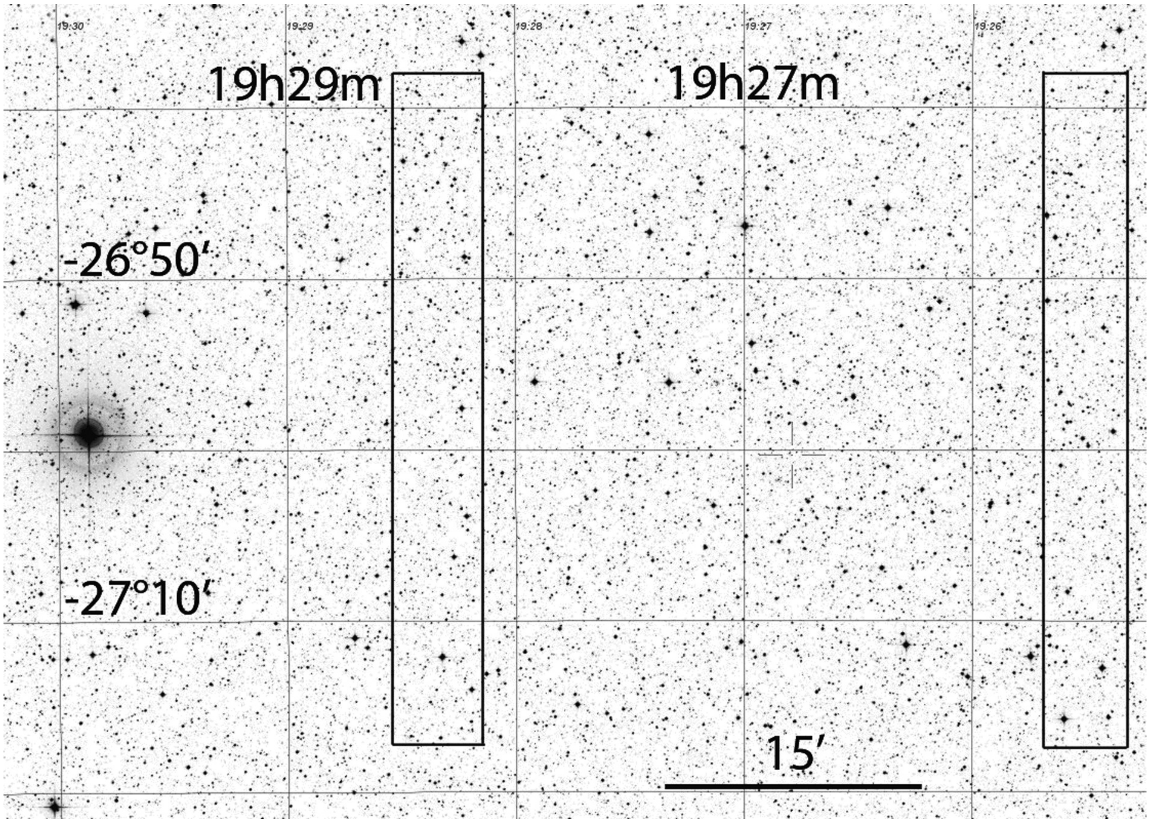


Figure 8.5: The sky area of the "Wow!" signal.

can not relate to the signal since it was under the horizon away from the meridian. Any radio source closer than Moon, with the exception of the polar satellite or other similar object slowly moving in a north-south direction (and so shifting only in declination), would have a noticeable shift with respect to the antenna pattern. As a result, the observed waveform would not be so symmetrical and consistent with the main lobe.

Of course, one can create an artificial source on or near the Earth (e.g., polar satellite), which changes the characteristics so as to look like a distant point source. However, it is difficult to create an artificial source, looking as a natural one for all types of radio receivers. Thus, it has little sense. It is important that the unintentional creation of such a source is impossible because the frequency range around 1420 MHz was forbidden to use by international treaties, and, besides, the "Wow!" signal had such a high flux density that can not be a result of a reflection or interference.

So, the main features of the "Wow!" signal in need of explanation, regardless of whether the signal is artificial or natural in origin: 1) high flux density, 2) a narrow band of frequencies, 3) point source in outer space, and 4) the appearance in only one horn of the radio telescope. Let us try to find the appropriate natural sources in this area of the sky.

8.3 Star HIP 95865

The bright star in Fig. 8.5 is double star HIP 95865 = HR 7398 = HD 183275 = PPM 269861 = SAO 188192 = ADS 12506 A and B = CCDM 19299-2659 A and B = TYC 6884-2463 and TYC 6884-2465 = CoD -27 14004 = CPD -27 6772 = TDSC 51183 A and B = WDS 19299-2659 A and B. The components are of V=5.6m and 9.0m. Their spectra are K1-2III and probably GIV or GV. According to the Hipparcos catalogue [9, 10] for the moment 1991.25 the angular distance between the components was r=7.6" whereas the positional angle was θ=143.1°. The both values changed within 20 years according to the CCDM [11] and TDSC [12] catalogues: for the moment 2011.36 r=6.6", θ=148°. Thus, probably we see an orbital motion of the fainter component with respect to the brighter one. The color indices from the Tycho catalogue [9]: (BT-VT)=1.34m and (B-V)=1.12m for the brighter component (red giant) and (BT-VT)=0.78m. for the fainter component. According to the Hipparcos catalogue [10], the parallax of the star is 0.014" ±0.001", i.e. its distance is 71 pc. Then the absolute magnitudes of the components are MVT=1.4m and 4.6m, which fits to the spectra K1III and G5IV-V. Suspecting masses of the components of 1.5 and 1 solar masses and their circular orbit the semi-major axis is 600 AU whereas the orbital period is of about 9000 years. In such case the orbital motion indeed could be detected for 20 years of observations but only in favorable circumstances, for example, near periastron.

The barycentric radial velocity of the star is -32 ±2 km/s according to the [13] and -31.6 ±0.3 km/s according to the Pulkovo Compilation of Radial Velocities [14]. These values agree well with each other (although they are not completely independent) and coincide with the previously mentioned barycentric radial velocity of the signal source (-30 km/s). This coincidence is an argument in favor of the star HIP 95865 as the source of the "Wow!" signal taking into the account that a few celestial objects have such a velocity in this sky area (21° from the center of the Galaxy).

Radio observations of this star at a frequency of 8.4 GHz in 1987-1988 [15] show the flux of less than 5.7 mJy. Systematic radio observations at frequencies around 1420 MHz have not been conducted. If the star is the source of the "Wow!" signal, for the explanation of its narrow band one can assume that the bright component of the pair, a red giant has a shell with hydrogen maser.

8.4 Radio source NVSS B192505-265140

Weak radio source may be amplified by the orbital motion close to a massive invisible body. In this region of the sky the example is the radio source NVSS B192505-265140 with the flux density of 20 mJy at a frequency of about 1420 MHz [4] and the magnitudes of B=19m and R=18.6m. Available astrometric observations suggest a non-linear movement of the source on the celestial sphere.

8.5 Stars V905 Sgr and HD 182460

Star on an unstable stage of its evolution can also be a source of radio pulses. Not all relevant stages of stellar evolution are studied. So the sources of the radio pulses can be among relatively bright stars in this area of the sky.

The RR Lyr type variable star V905 Sgr = 2MASS J19282166-2644467 of V=14.9m-15.4m is an interesting example of an unstable star in the area of the "Wow!" signal. It has the period of 0.654762 and a distance of order of 1 kpc from the Sun. According to the 2MASS [16], UCAC3 [17], PPMXL [18] and XPM [19] catalogues, the star shows non-linear motion on the celestial sphere and, therefore, can have an unseen massive component. Interaction in such a close binary can be a source of radio pulses. Such RR Lyr type variables are quite interesting in relation to the "Wow!" signal because their characteristics can change considerably within a time interval of order of several minutes. That would explain appearance and disappearance of the "Wow!" signal within a minute of observation.

Another interesting star in the field of the "Wow!" signal is HD 182460 = CD-27 13937 = CPD-27 6759 = GSC 06884-01944 = 2MASS J19254475-2712085 = PPM 735190 = TYC 6884-1944 of V=10m. Its spectral class is B9III-A0III from the Tycho Spectral Types catalogue [20] does not fit to its color indices (B-V)=1.1m and (J-Ks)=0.6m, which correspond to a red giant. According to Gontcharov [21] this star is a clump red giant (i.e. a star with the helium fusion in its core) at a distance of about 700 pc from the Sun. The disagreement between the spectral classification and other characteristics of this star can be explained if 1) it is peculiar, significantly changing the spectrum within a short time interval, or 2) indeed it is a pair of a sdB hot subdwarf and a KIII clump red giant of similar ages, with both stars at the stage of helium fusion. In both cases, a star or a pair of stars can create conditions for powerful radio pulses.

8.6 Star TYC 6884-2359

The star TYC 6884-2359 = CD-27 13981 = 2MASS J19281971-2712118 = PPM 735243 of V=9.8m is another interesting candidate for radio pulses. Judging by the very red color indices (B-V)=2m, (BT-VT)=2.4m, (J-Ks)=1.0m, the distance of 115 pc [22] and the corresponding absolute magnitude of MV=5m the star remains in the stage of its initial formation. In this case, the source of the radio pulses can be a maser arising in the gas and dust shell of the forming stars.

8.7 Radio source IRAS 19224-2707

Another class of objects that can be sources of rare radio pulses are old supernova remnants, manifesting themselves as the rotating radio transients with synchrotron radiation. In this region of the sky there is, at least one candidate for the objects of this class: infrared source IRAS 19224-2707 = USNO B1 0629-1065389 = NOMAD 0629-1139789. It is very weak in the visual range (B=21m, R=18m). It probably has a large (for an object at a distance of at least hundreds pc) proper motion of about μ=0.2" ±0.1" per year, according to the comparison of its coordinates in various catalogues. This rapid motion appears to be a consequence of the explosion of this object as a supernova.

According to the theory of stellar evolution, these objects must be quite common, being an inevitable stage in the development of the old supernova remnant (see references in [23]). Indeed, next to the considered region of the sky there is another object of this class: the infrared source IRAS 19219-2702 (B=21.6m, R=17.9m. It also has quite high proper motion of μ=0.7" ±0.2" per year as follows from the comparison of its coordinates in various catalogues. The well-known supernova remnant in the Crab Nebula (M1) with pulsars and expanding shell is at a previous stage of its development. Therefore, in future it must become a rotating radio transient and probably accept the characteristics similar to those of IRAS 19224-2707 and IRAS 19219-2702. Now the Crab pulsar has high proper motion of μ=0.013" ±0.001" per year [23].

According to modern concepts, the period of pulses of radio source in the supernova remnant becomes longer over time. The pulses become non-periodic and very rare. In addition, the nebula around this object expands, becomes less bright and dissipate. In this situation the appearance of masers is possible. Their high power is compensated by the rarity of the pulses.

A sure sign of a supernova remnant is a visible nebula. Indeed, there is such a nebula around the object IRAS 19224-2707. It has a size of about 10 arcsec and is marked in Fig. 8.6 by the big arrow. This figure represents an image area of 59" x 51" around the object IRAS 19224-2707 (not in the center of the picture and is marked by the big arrow) taken from the Digital Sky Survey ESO MAMA [24].

At 1420 MHz the synchrotron emission of the supernova remnant usually quite intense. For example, the Crab Nebula pulsar emits stronger at 1420 MHz than in the infrared range. The source IRAS 19224-2707 has no radio observations, but a comparison of its infrared radiation to the one of the Crab indirectly confirms the synchrotron nature of the radiation in both cases: exponential increase in intensity with decreasing frequency. This can be seen from Fig. 8.7, where the distribution of energy in the infrared part of the spectrum of the Crab is shown by the diamonds as well as of IRAS 19224-2707 is done by the squares whereas the solid lines show the trends. The synchrotron nature of the radiation justifies the extrapolation of the data from the infrared region to lower frequencies. This extrapolation for IRAS 19224-2707 is shown in the Fig. 8.7 by the dotted line. And the estimated so the radio flux density of the source at 1420 MHz is of about 100 Jy marked by a circle in the figure 8.7. It is consistent with the flux density of the "Wow!" signal. However, the object IRAS 19224-2707 (and any other rotating radio transient) may not be detected as powerful sources within radio sky surveys being the sources of rare short pulses.

Here the main problem is the narrow bandwidth of the "Wow!" signal. The only explanation for it seems to be a maser in the shell of a supernova remnant, though in the future one may identify a new mechanisms of such pulses. If the old supernova remnants is really capable of such powerful, although very rare pulses, they should attract the attention of researchers as a source of information about the last stages of stellar life, also as possible standard candles visible at a great distance, and as the sources of concentrated energy, which probably is reproducible and useful at the Earth.

8.8 Conclusion

The detail consideration of the circumstances of the "Wow!" signal leaves little doubt that it is a powerful radio pulse from the distant space. Therefore, the scientific community should not ignore it: it is possible for this signal is hiding a little-studied class of objects. The analysis of the area of the sky from which the signal is received has revealed the several candidates of various nature. Thus, a natural explanation of the signal is possible in addition to the popular hypothesis of a signal of an alien civilization. At the same time, the compiled list of potential candidates is short and probably complete to the 18th visual magnitude, since one has to explain the high power, narrow bandwidth, and pulse nature of the signal as well as the lack of repetitions. The classes of possible candidates: radio stars; radio source amplified by an invisible close massive object; star at unstable stage of its evolution; maser in the shell of a stars during its formation and old supernova remnant with rotating radio transient. In all cases, the most difficult to explain is the narrow bandwidth of about 10 kHz, which is now, apparently, can be explained only by the maser mechanism.

In recent decades, many rare narrow-band radio pulses of cosmic origin are found. They are common, but little studied phenomenon. The most effective approach here is a permanent long-term monitoring of the few reliable objects of this class. The current research allows us to reduce considerably the number of candidates for such monitoring in one interesting region of the sky. Such monitoring needs a relatively insensitive technique, thus making the problem much more manageable. The Table 8.1 shows the coordinates of the candidates under consideration in the coordinate system J2000.

Table 8.1: coordinates of the candidates under consideration in the coordinate system J2000.

Object	**Right ascension**	**Declination**
HIP 95865	19h 29m 52s	-26° 59' 08"
NVSS B192505-265140	19h 28m 10s	-26° 45' 31"
V905 Sgr	19h 28m 22s	-26° 44' 47"
HD 182460	19h 25m 45s	-27° 12' 08"
TYC 6884-2359	19h 28m 20s	-27° 12' 11"
IRAS 19224-2707	19h 25m 35s	-27° 01' 58"

References

1. http://www.bigear.org/

2. J.R. Ehman, 1998, http://www.bigear.org/wow20th.htm

3. R.H. Gray, A search of the 'Wow' locale for intermittent radio signals, Icarus 112 (1994) 485-489.

4. R.H. Gray, K.B. Marvel, A VLA Search for the Ohio State "Wow", Astrophys. J. 546 (2001) 1171-1177.

5. T.J.W. Lazio, J. Tarter, P.R. Backus, Megachannel Extraterrestrial Assay Candidates: No Transmissions from Intrinsically Steady Sources, Astron. J. 124 (2002) 560-564.

6. R.H. Gray, S. Ellingsen, A Search for Periodic Emissions at the Wow Locale, Astrophys. J. 578 (2002) 967-971.

7. http://www.bigear.org/oldseti.htm

8. http://www.bigear.org/lobes.htm

9. European Space Agency, Hipparcos and Tycho catalogues, 1997.

10. F. van Leeuwen, Validation of the new Hipparcos reduction, Astron. Astrophys. 474 (2007) 653-664, http://cdsweb.u-strasbg.fr/viz-bin/Cat?I/311.

11. J. Dommanget, O. Nys, Catalogue des composantes d'étoiles doubles et multiples, Observat. Royal Belgique, 2000, `http://cdsweb.u-strasbg.fr/viz-bin/Cat?I/274`.

12. C. Fabricius, E. Hog, V.V. Makarov, et al., The Tycho double star catalogue, Astron. Astrophys. 384 (2002) 180-189, http://cdsweb.u-strasbg.fr/viz-bin/Cat?I/276.

13. M. Barbier-Brossat, P. Figon, Mean radial velocities catalog of galactic stars, Astron. Astrophys. Suppl. Ser. 142 (2000) 217-223, `http://cdsweb.u-strasbg.fr/viz-bin/Cat?III/213`.

14. G.A. Gontcharov, Pulkovo Compilation of Radial Velocities for 35495 Hipparcos Stars in a Common System, Astronomy Letters 32 (2006) 759-771, `http://cdsweb.u-strasbg.fr/viz-bin/Cat?III/252`.

15. H.J. Wendker, Radio continuum emission from stars: a catalogue update, Astron. Astrophys. Suppl. Ser. 109 (1995) 177-179.

16. M.F. Skrutskie, R.M. Cutri, R. Stiening, et al., The Two Micron All Sky Survey (2MASS), Astron. J. 131 (2006) 1163-1183, `http://www.ipac.caltech.edu/2mass/releases/allsky/index.html`.

17. N. Zacharias, C. Finch, T. Girard, et al., The Third US Naval Observatory CCD Astrograph Catalog, Astron. J. 139 (2010) 2184-2199.

18. S. Roeser, M. Demleitner, E. Schilbach, The PPMXL Catalog of Positions and Proper Motions on the ICRS. Combining USNO-B1.0 and the Two Micron All Sky Survey, Astron. J. 139 (2010) 2440-2447.

19. P.N. Fedorov, A.A. Myznikov, V.S. Akhmetov, The XPM Catalogue: absolute proper motions of 280 million stars, Monthly Notices of Royal Astronomical Society 393 (2009) 133-138.

20. C.O. Wright, M.P. Egan, K.E. Kraemer, et al., The Tycho-2 Spectral Type Catalog, Astron. J. 125 (2003) 359-363.

21. G.A. Gontcharov, Red Giant Clump in the Tycho-2 Catalogue, Astronomy Letters 34 (2008) 785-796.

22. S.M. Ammons, S.E. Robinson, J. Strader, et al., The N2K Consortium. IV. New Temperatures and Metallicities for More than 100,000 FGK Dwarfs, Astrophys. J. 638 (2006) 1004-1017.

23. D.L. Kaplan, S. Chatterjee, B.M. Gaensler, et. al., A Precise Proper Motion for the Crab Pulsar, and the Difficulty of Testing Spin-Kick Alignment for Young Neutron Stars, Astrophys. J. 677 (2008) 1201-1215.

24. M. Cullum, The MAMA detector, European Southern Observatory (ESO),Garching, 1990.

Vitae: George A. Gontcharov is a senior researcher at the Main (Pulkovo) astronomical observatory of the Russian Academy of Science, Saint-Petersburg, Russia; PhD; the fields of interests: the Galaxy, structure of the Milky Way, radial velocities of stars, interstellar extinction, stellar evolution; the author of more than 30 papers including the Pulkovo Compilation of Radial Velocities at http://cdsarc.u-strasbg.fr/viz-bin/Cat?III/252; e-mail: georgegontcharov@yahoo.com

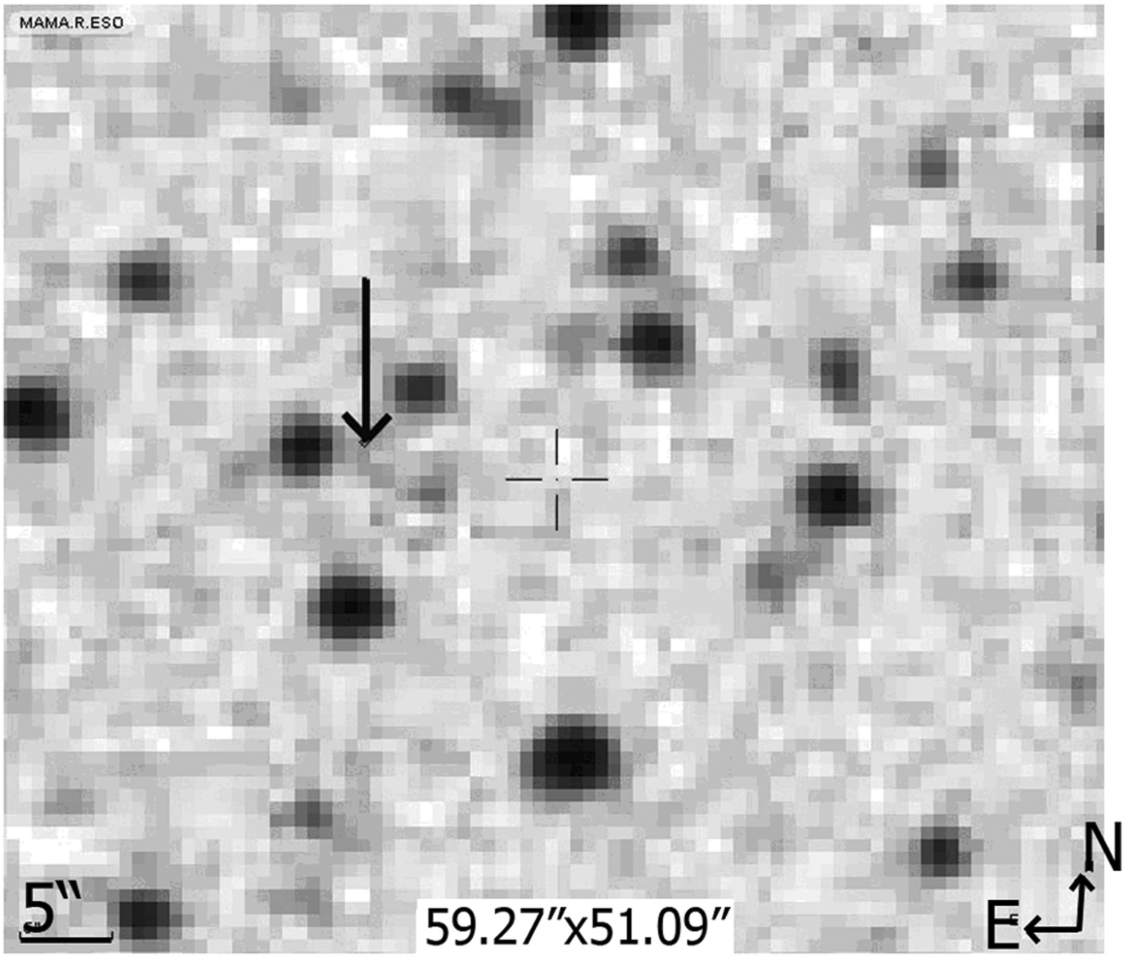


Figure 8.6: The sky area around the source IRAS 19224-2707 (marked by the big arrow)..

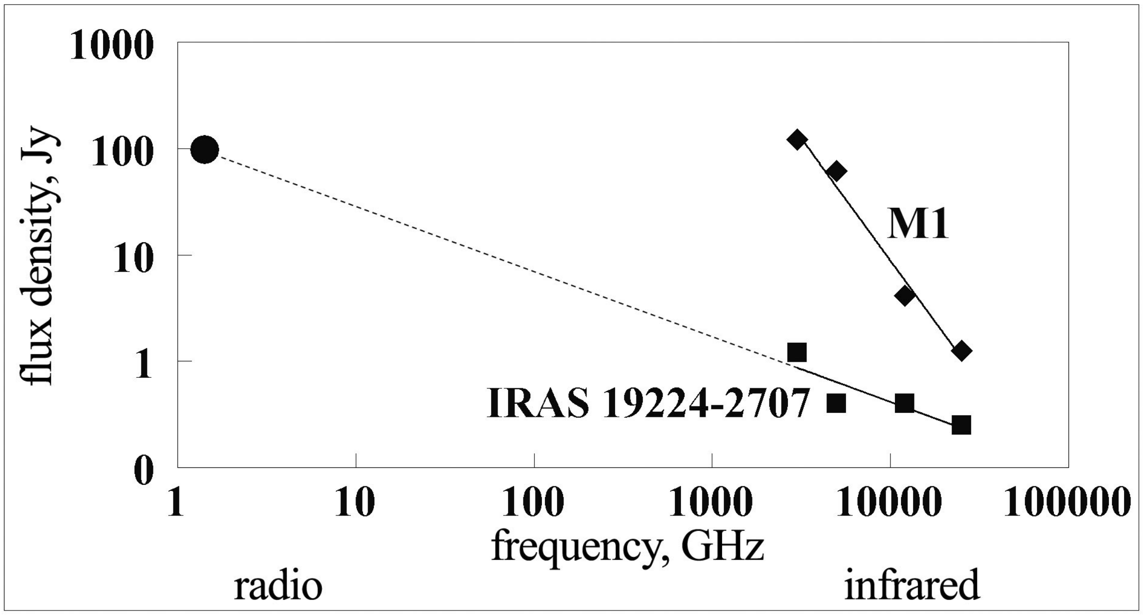


Figure 8.7: The energy distribution in the infrared range for the Crab (M1) source (diamonds) and for the IRAS 19224-2707 source (squares) as well as extrapolation for the 1.4 GHz (circle).

Chapter 9

The KLT (Karhunen–Loève Transform) to extend SETI searches to broad-band and extremely feeble signals

by **Claudio Maccone**
International Academy of Astronautics
Via Martorelli, 43, Torino (Turin) 10155, Italy

Abstract

The KLT (acronym for Karhunen–Loève Transform) is a mathematical algorithm superior to the classical FFT in many regards:

1) The KLT can filter signals out of the background noise over both wide and narrow bands. This is in sharp contrast to the FFT that rigorously applies to narrow-band signals only.

2) The KLT can be applied to random functions that are non-stationary in time, i.e. whose autocorrelation is a function of the two independent variables t_1 and t_2 separately. Again, this is a sheer advantage of the KLT over the FFT, inasmuch as the FFT rigorously applies to stationary processes only, i.e. processes whose auto-correlation is a function of the absolute value of the difference of t_1 and t_2 only.

3) The KLT can detect signals embedded in noise to unbelievably small values of the Signal-to-Noise Ratio (SNR), like 10^{-3} or so. This particular feature of the KLT is studied in detail in this paper.

An excellent filtering algorithm like the KLT, however, comes with a cost that one must be ready to pay for especially in SETI: its computational burden is much higher than for the FFT. In fact, it can be shown that no fast KLT transform can possibly exist and, for an autocorrelation matrix of size N, the calculations must be of the order of N^2, rather than N log(N). Nevertheless, for moderate values of N (in the hundreds), the KLT dominates over the FFT, as shown by the numerical simulations. Finally, an important and recent (2007–2008) development in the KLT theory, called the "Bordered Autocorrelation Method" (BAM), is presented. This BAM-KLT method gets around the difficulty of the N^2 brunt calculations and ends up in the following unexpected theorem: the KLT of a feeble sinusoidal carrier embedded into a lot of white stationary noise is given by the Fourier transform of the derivative of the largest KLT eigenvalue with respect to the bordering index. This basic result is fully proved analytically in the final sections of this paper by virtue of a new theorem discovered by this author in May 2007 and called "The Final Variance Theorem".

9.1 Introduction

This paper 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 not just white, and over wide bandwidths and 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 is 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 important practical formulae that can be applied to improve SETI, the detection of exoplanets, the asteroidal radar, and also other fields of knowledge like economics, genetics, biomedicals, etc. to which the KLT can be equally well applied with success.

We believe that these improvements in the mathematical ways of handling the KLT will increase the interest of scientists into this algorithm that may well replace the Fourier transform in the near future.

9.2 A bit of history

We argue that 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 to be fair to each contributing author. In addition, mathematicians, both pure and applied, often speak such a "clumsy" language of their own that even learned scientists find sometimes hard to understand them. This unfortunate situation hides the aesthetic 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 to improve our understanding of physical phoenomena, far superior to the classical Fourier Transform (FT). The KLT is named after two mathematicians, the Finnish actuary, Kari Karhunen (1915–1992) [1] and the French–American mathematician, Michel Loève (1907–1979) [2,3], who proved, independently and about the same time (1946), that the series (9.2) 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 had the American statistician and economist Harold Hotelling (1895–1973) used the KLT (for discrete time, rather than for continuous time) that the KLT is sometimes called the "Hotelling Transform". Even much earlier than these three authors had the Italian geometer Eugenio Beltrami (1835–1899) discovered as early as 1873 the Singular Value Decomposition (SVD) that 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 the first three sections of this paper, we will 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. The same approach is kept also in the second part of this paper (Section 7 and following) where we describe the recently discovered (2007–2008) "Bordered Autocorrelation Method" (BAM) to easily compute the KLT.

9.3 A heuristic derivation of the KLT

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 $x(t)$ unto a basis of orthonormal functions (sines and cosines, in this case), typified by the classical Fourier series (firstly put forward by the French mathematician Jean Baptiste Joseph Fourier (1768–1830) around 1807),

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n cos(\omega_n t) + b_n sin(\omega_n t)] \tag{9.1}$$

2) The need to extend to probability and statistics this too narrow and deterministic view. The much larger variety of phenomena called "noise" by physicists and engineers will thus be encompassed by the new transform. This enlarged view means to consider 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

$$X(t) = \sum_{n=1}^{\infty} Z_n \phi_n(t) \tag{9.2}$$

that is called 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 (9.2)? To find out, let us start by recalling what "orthonormality" means for the Fourier series (9.1). Leonhard Euler (1707–1783) had already laid the first stone towards the Fourier series (9.1) by proving that if $T = t_2 - t_1$ is the assumed period of $x(t)$ and one sets $\omega_n = n(2\pi/T_0)$, then the coefficients a_n and b_n in (3.1) are obtained from the known function (or "signal") $X(t)$ by virtue of the equations ("Euler formulae")

$$a_n = \frac{2}{T}\int_{t_1}^{t_2} x(t)cos(\omega_n t)dt \quad b_n = \frac{2}{T}\int_{t_1}^{t_2} x(t)sin(\omega_n t)dt \tag{9.3}$$

If the same result is going to be true for the Karhunen–Loève expansion, the functions of the time, $\phi_n(t)$ in (3.2) must be orthornormal, i.e. both orthogonal and normalized to one. That is,

$$\int_0^T \phi_m(t)\phi_n(t)dt = \delta_{mn} \tag{9.4}$$

where the δ_{mn} are the Kronecker symbols, defined by $\delta_{mn} = 0$ for $m \neq n$ and $\delta_{nn} = 1$.

But what are then the Z_n appearing in (9.2)? Well, a random function $X(t)$ can be thought of as something made by two parts: its behavior in time, represented by the functions $\phi_n(t)$, and its behavior with respect to probability and statistics that must therefore be represented by Z_n . In other words, 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 (2), one then gets (we "freely" interchange here the average operator $E\{\}$ with the infinite summation sign, bypassing the complaints of "subtle" mathematicians!) one gets

$$E\{X(t)\} = \sum_{n=1}^{\infty} E\{Z_n\}\phi_n(t) \tag{9.5}$$

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 $0 \leq t \leq T$

$$E\{X(t)\} \equiv 0 \tag{9.6}$$

In fact, was not this true, 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

$$E\{Z_n\} \equiv 0 \tag{9.7}$$

This equation has a simple consequence: since the variance $\sigma^2_{Z_n}$ of the random variables Z_n is given by

$$\sigma^2_{Z_n} = E\{Z_n^2\} - E^2\{Z_n\} \tag{9.8}$$

by inserting (9.7) into (9.8) we get

$$\sigma^2_{Z_n} = E\{Z_n^2\} \tag{9.9}$$

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

$$\sigma^2_{Z_n} = \lambda_n = E\{Z_n^2\} > 0 \tag{9.10}$$

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 (9.4) 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 (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

$$E\{Z_m Z_n\} = \lambda_n \delta_{mn} \tag{9.11}$$

In this equation, we were forced to introduce the positive λ_n in the right-hand side in order to let (9.11) reduce to (9.10) in the special case $m = n$. As for the KL equivalent of the Euler formulae (9.3) of the Fourier series, from the KL series (9.2) and the orthonormality (9.4) of the $\phi_n(t)$ one immediately finds that

$$Z_n = \int_0^T X(t)\phi_n(t)dt \tag{9.12}$$

In other words: the random variables Z_n are obtained from the given stochastic process $X(t)$ by "projecting" this $X(t)$ over the corresponding 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", that 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 $0 \le t \le T$ and it equals 1, as in (9.4).

To sum up, we have actually achieved a remarkable generalization of the Fourier series by defining the Karhunen–Loève expansion (9.2) 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

$$R_{XX}(t_1, t_2) \equiv R_X(t_1, t_2) = E\{X(t_1)X(t_2)\} \tag{9.13}$$

If we assume, according to (5) that the mean value of $X(t)$ vanishes identically in the interval $0 \le t \le T$, the autocorrelation (9.13) reduces to the variance of $X(t)$ when the two instants are the same

$$\sigma^2_{X(t)} = E\{X^2(t)\} = E\{X(t)X(t)\} = R_X(t, t) \tag{9.14}$$

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 (9.2) 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.

9.4 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 $E\{X(t_1)X(t_2)\}$ and that the $lambda_n$ 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 possesses an intrinsic set of three rectangular axes that describes its dynamics at best, 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 firstly seeks for 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 (9.2). This goal is achieved by writing down (9.2) 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

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \phi_m(t_1)\phi_n(t_2)E\{Z_m Z_n\} \tag{9.15}$$

Taking advantage of the statistical orthogonality of the Z_n, given by (9.11), (9.15) simplifies to

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty} \lambda_m \phi_m(t_1)\phi_m(t_2) \tag{9.16}$$

Finally, we now want to let the $\phi_m(t)$ "disappear" from the right-hand side of (9.16) by taking advantage of their orthonormality (9.4). To do so, we multiply both sides of (9.16) by $\phi_n(t_1)$ and then take the integral with respect to t_1 between 0 and T. One then gets

$$\begin{aligned} &\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)dt_1 \\ &= \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2) \int_0^T \phi_m(t_1)\phi_n(t_1)dt_1 \\ &= \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2)\delta_{mn} = \lambda_m \phi_m(t_2) \end{aligned} \tag{9.17}$$

that is

$$\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)dt_1 = \lambda_n \phi_n(t_2) \tag{9.18}$$

This basic result is an integral equation, called by mathematicians "of the Fredholm type". Once the correlation $E\{X(t_1)X(t_2)\}$ of X(t) is known, the integral Eq. (9.18) yields (upon its solution, that may not be easy at all to find analytically!) both the Karhunen–Loève eigenvalues λ_n and the corresponding eigenfunctions $\phi_n(t_1)$. Readers familiar with quantum mechanics will also recognize in (9.18) 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 9.3 and 9.4, and let us use the language of signal processing, that will lead us directly to SETI, the main theme of this paper.

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 KL expansion) given by (9.2), 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,...) 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 new reference frame in the Hilbert space the basis spanned by the eigenfunctions, $\phi_n(t)$, of the autocorrelation of the (signal+noise), $X(t)$.

This self-adapting capability of the KLT is probably its main advantage over the Fourier transform as well as over other transforms, like Wigner–Ville, Hilbert, etc.

9.5 Continuous 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 the time obviously must be a continuous variable. For instance, this author spent fifteen years of his life (1980–1994) to investigate 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 (that 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 it is science fiction to imagine that an alien spaceship might approach the Earth slowing down from relativistic speeds to zero speed. So, two mathematical physics books like refs. [4] and [26] 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.

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 the time is a continuous variable, then one must solve the integral Eq. (9.18), and this may require considerable mathematical skills. In fact, (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. Much easier may be the task if one is able to reduce the Fredholm integral equation to a Volterra integral equation, just as shown in the books [4] and [26] 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 (9.2). If this variable is discrete, rather than continuous, then the picture changes completely. In fact, the integral Eq. (9.2) 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). And huge also is the characteristic equation, i.e. the algebraic equation the roots of which are the KL eigenvalues. Can you imagine solving directly an algebraic equation of degree 1 million? So, the KLT is practically impossible to find numerically, unless we resort to simplifying tricks of some kind, as was done by the SETI-Italia team [22] since 2007. This was a follow-on of the first implementation ever of the KLT for SETI made at Medicina in 2003 [5].

9.6 The KLT: just a linear transformation in the Hilbert space

We explained the KL expansion (9.2), but we did not explain what the KL Transform is yet! 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 one, and is called the "dominant" eigenvalue by mathematicians.

We are now ready to compute the Direct KLT of the (signal+noise). 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 transform properly called just is a linear transformation of axes, and nothing is easier than that! (Incidentally, this accounts for the title of Karhunen's first paper "Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung" = "On the Linear Methods in the Calculus of Probabilities" [1] that obviously refers to the linear character of the transformation of axes in the Hilbert space).

9.7 A breakthrough about the KLT: the "Final Variance Theorem"

The importance of the KLT as a superior mathematical tool than the FFT was already pointed out. However, the implementation of the KLT by a numerical code running on computers has always been a difficult problem. Both Francois Biraud in France [6] and Bob Dixon in the USA [16] failed to do so in the 1980s because all computers then available got stuck by the solution of 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 Eq. (9.18). At the SETI Italia facilities at Medicina we faced the same problem, of course. But we did better than our predecessors because this author 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. We remind here that **this final instant T simply does not exist in the ordinary Fourier theory**, because this T equals infinity by definition in the Fourier theory. Therefore, the final instant T in itself is possibly the most important "novelty" introduced by the KLT with respect to the classical FFT. With respect to T, we may take derivatives (called "final derivatives" in the sequel of this paper because they are time derivatives taken with respect to the final instant T) and integrals that have no analogues 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"[1]. So, by removing this silly $T = 1$ convention we opened up new prospects in the KLT theory, as we now show by proving our "Final Variance Theorem".

Consider the eigenfunction expansion of the autocorrelation again, Eq. (9.16), with the traditional dummy index n rewritten instead of m. Upon replacing $t_1 = t_2 = t$, and invoking (9.10), this equation becomes

[1] http://en.wikipedia.org/wiki/Karhunen-Lo%C3%A8ve_theorem

$$E\{X^2(t)\} = \sum_{n=1}^{\infty} \lambda_n \phi_n^2(t) \tag{9.19}$$

Since the eigenfunctions $\phi_n(t)$ are normalized to one, we are prompted to integrate both sides of (9.19) with respect to t between 0 and T, so that the integral of the square of the $\phi_n(t)$ becomes just one

$$\int_0^T E\{X^2(t)\}dt = \sum_{n=1}^{\infty} \lambda_n \int_0^T \phi_n^2(t)dt = \sum_{n=1}^{\infty} \lambda_n \tag{9.20}$$

On the other hand, since the mean value of $X(t)$ is identically equal to zero, one may now introduce the variance $\sigma^2_{X(t)}$ of the stochastic process $X(t)$ defined by

$$\sigma^2_{X(t)} = E\{X^2(t)\} - E^2\{X(t)\} = E\{X^2(t)\} \tag{9.21}$$

Replacing (9.21) into (9.20), one gets

$$\int_0^T \sigma^2_{X(t)} dt = \sum_{n=1}^{\infty} \lambda_n \tag{9.22}$$

This formula was already given by this author in his 1994 book, and it is Eq. (1.13) on page 12 of Ref. [4]. At that time, however, (22) was regarded as interesting inasmuch as (upon interchanging the two sides) it proves that the series of all the eigenvalues l 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, this author had not yet understood that (22) has a more profound meaning, that is: 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

Back in 1994, however, this author had not yet understood that (22) has a more profound meaning, that is: 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

$$\lambda_n \equiv \lambda_n(T) \tag{9.23}$$

This new remark is vital in order to make new progress. In fact, one is now prompted to let the integral on the left-hand side of (22) disappear by differentiating both sides with respect to the final instant T. One thus gets

$$\sigma^2_{X(t)} = \sum_{n=1}^{\infty} \frac{\partial \lambda_n(T)}{\partial T} \tag{9.24}$$

This result we call the **Final Variance Theorem**. It is the key new result put forward in this paper. 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 (9.24) becomes in the practice the finite sum

$$\sigma^2_{X(T)} \approx \sum_{n=1}^{N} \frac{\partial \lambda_n(T)}{\partial T} \tag{9.25}$$

In numerical simulations, however, one always wants to cut as short as possible with the computation time! Therefore one might be led to consider the first (or dominant) eigenvalue only in (9.25), that is

$$\sigma^2_{X(T)} \approx \frac{\partial \lambda_1(T)}{\partial T} \tag{9.26}$$

This clearly is "the roughest possible" approximantion to the full $X(t)$ process since we are actually replacing the full $X(t)$ by its first KLT term $Z_1\phi_1(t)$. However, using (3.28) instead of the N-term sum (9.25) is indeed a good short-cut for the 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 (9.25) simplifies even further. In fact, by definition, the stationary processes have the same final variance at any time, i.e. for stationary processes s 2 X is a constant. Then (9.25) immediately shows that, **for stationary processes only, all the KLT eigenvalues are LINEAR functions of the final instant T**

$$\lambda_n(T) \propto T \quad for\ stationary\ processes\ only \tag{9.27}$$

As a consequence, the first-order partial derivatives of all the λ_n with respect to T for stationary processes are just **constants**. In other words still, for stationary processes only, (9.25) becomes

$$\sum_{n=1}^{N} \frac{\partial \lambda_n(T)}{\partial T} \approx \textit{ a constant with respect to } T \tag{9.28}$$

In particular, if one sticks again to the first, dominant eigenvalue only (i.e. to the roughest possible approximation), then (9.28) reduces to

$$\frac{\partial \lambda_1(T)}{\partial T} \approx \textit{ a constant with respect to } T \tag{9.29}$$

In the next section 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 (that is, so as to say, "the easiest of the non-stationary processes"), one has

$$\lambda_n(T) = \frac{4T^2}{\pi^2(2n-1)^2)} \quad (n = 1, 2, ...) \tag{9.30}$$

and so the dependence on T is quadratic. For the proof, just replace the Brownian motion variance $\sigma^2_{B(t)}$ into (9.22) and perform the integration, yielding the T^2 directly. Of course, this is in agreement with (9.30) that is proven in ref. [4], p. 17, or in ref. [26], p. 311, when finding the KLT of the standard Brownian motion.

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 it is proven in [26], p. 359, and so on for more complicated processes, like the time-rescaled squared Brownian motions whose KLT will found in [11].

9.8 BAM ("Bordered Autocorrelation Method") to find the 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 5. The BAM has its mathematical foundation in the Final Variance

Theorem already proved in the previous section. In this section we described the BAM in detail. Finally, in the next section, we will provide the results of numerical simulations showing that, by virtue of the BAM, the KLT succeeds in extracting a sinusoidal carrier embedded in lot of noise when the FFT utterly fails.

Let us start by reminding 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 $Nln(N)$ amount of calculations required by the FFT and that is just why the FFT was preferred to the KLT in the last 50 years!

Because of the Final Variance Theorem proved in the previous section, one is tempted to confine oneself to the study of the dominant eigenvalue only by virtue of use of (9.29). This means to study (9.29) 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, ..., 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[2]) that we denote by $R_{Toeplitz}$.

$$R_{Toeplitz} = \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & R_{XX}(2) & \cdots & R_{XX}(N) \\ R_{XX}(1) & R_{XX}(0) & R_{XX}(1) & \cdots & R_{XX}(N-1) \\ R_{XX}(2) & R_{XX}(1) & R_{XX}(0) & \ddots & \vdots \\ \cdots & \cdots & \ddots & R_{XX}(0) & \vdots \\ R_{XX}(N) & R_{XX}(N-1) & \cdots & \cdots & R_{XX}(0) \end{bmatrix} \tag{9.31}$$

This theorem was already 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 is, the more accurate the KLT of $X(t)$ is. 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 will do in a moment.

But let us now go back to the Toeplitz autocorrelation (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 N. This means to solve again the system of linear algebraic equations of the KLT for N+1, rather than for N. **So, for each different values 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**

[2] http://en.wikipedia.org/wiki/Toeplitz_matrix

again, for as many values as we wish (or, more correctly, for as many values of N as our computer can still handle!) is 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 to a new set of applications of the KLT to all fields of science!

In fact, numeric simulations lead to the results shown in the first 4 plots to follow. The first plot (Fig. 3.1) 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[3]. Please notice two facts: (1) this is about the lowest SNR below which the FFT starts failing to denoise a signal, a well-known fact to electrical and electronic engineers. (2) This Fourier spectrum is obviously computed by taking the Fourier transform of the stationary autocorrelation of X(t), as well-known from the Wiener–Khinchin Theorem (for a concise description of this theorem, see[4]). 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 non-stationary processes there are other "tricks" to compute the spectrum from the autocorrelation, like the Wigner–Ville Transform, but shall not consider them here.

The second plot (Fig. 2) shows the first (i.e. the dominant) KLT eigenvalue $\lambda_1(N)$ over N=1000 time samples. Clearly, this $\lambda_1(N)$ is proportional to N, as predicted by our Final Variance Theorem (9.27).

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 clearly we 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 shows such a BAM-reconstructed peak (Fig. 3.3).

Finally, this plot is of course identical to the following fourth plot (Fig. 3.4), showing the ordinary FFT of first KLT eigenfunction 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 (Fig. 9.5)!

No classical FFT spectrum can be identified at all for such a terribly low SNR!.

[3] http://en.wikipedia.org/wiki/Signal-to-noise_ratio

[4] http://en.wikipedia.org/wiki/Wiener%E2%80%93Khinchin_theorem

But for the KLTy no problem!

The next plot (Fig. 6) shows that $\lambda_1(N) \propto N$, as predicted by our Final Variance Theorem (9.27).

The third plot (Fig. 9.7) (KLT FAST way via the BAM) is the neat KLT spectrum of the 300 Hz tone obtained by computing the FFT of the constant $\partial\lambda_1(N)/\partial N$.

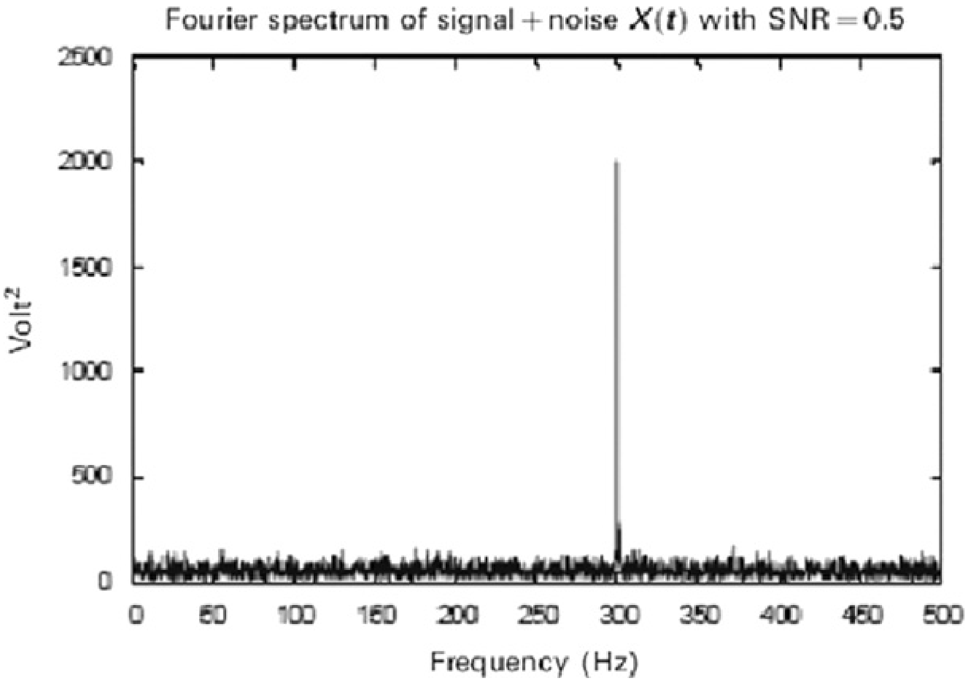


Figure 9.1: Fourier spectrum of a pure tone (i.e. just a sinusoidal carrier) with frequency at 300 Hz buried in stationary noise with a signal-to-noise ratio of 0.5.

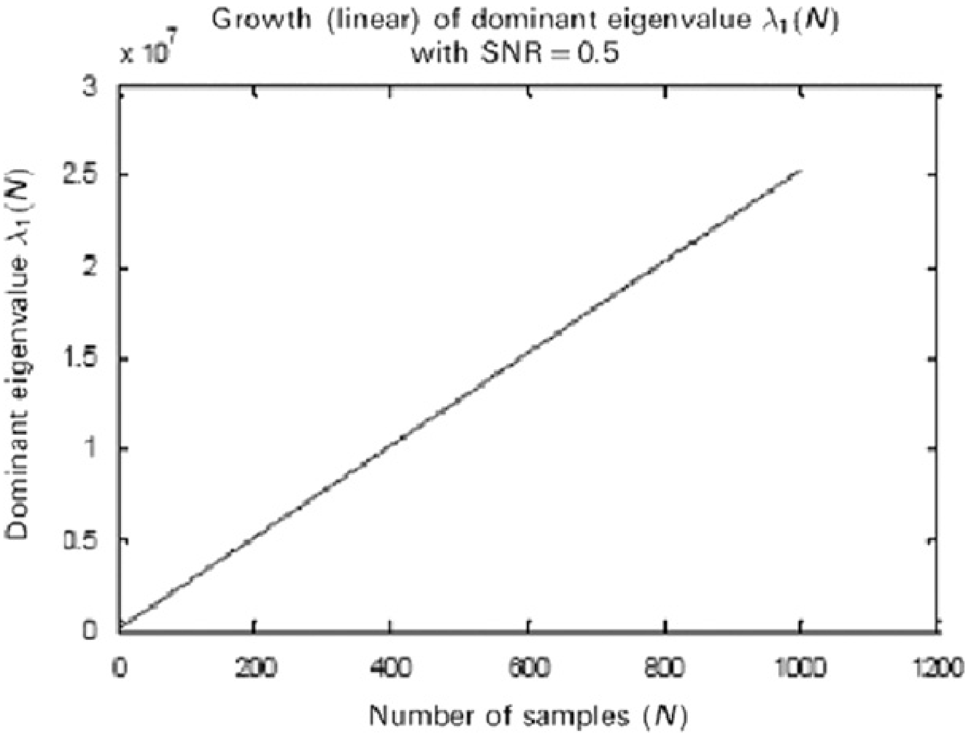


Figure 9.2: The KLT dominant eigenvalue $\lambda_1(N)$ over N = 1000 time samples, computed by virtue of the BAM, the Bordered Autocorrelation Method.

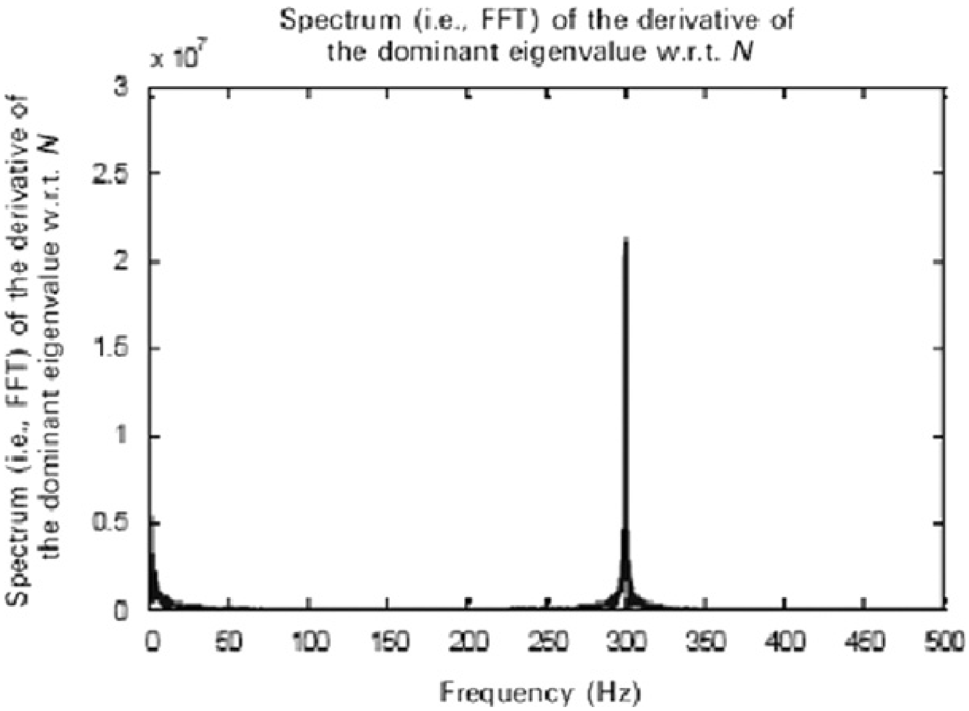


Figure 9.3: The spectrum (i.e. the Fourier Transform) of the CONSTANT derivative of the KLT dominant eigenvalue $\lambda_1(N)$ with respect to N as given by the BAM. This is clearly a Dirac delta function, i.e. a peak, at 300 Hz, as expected.

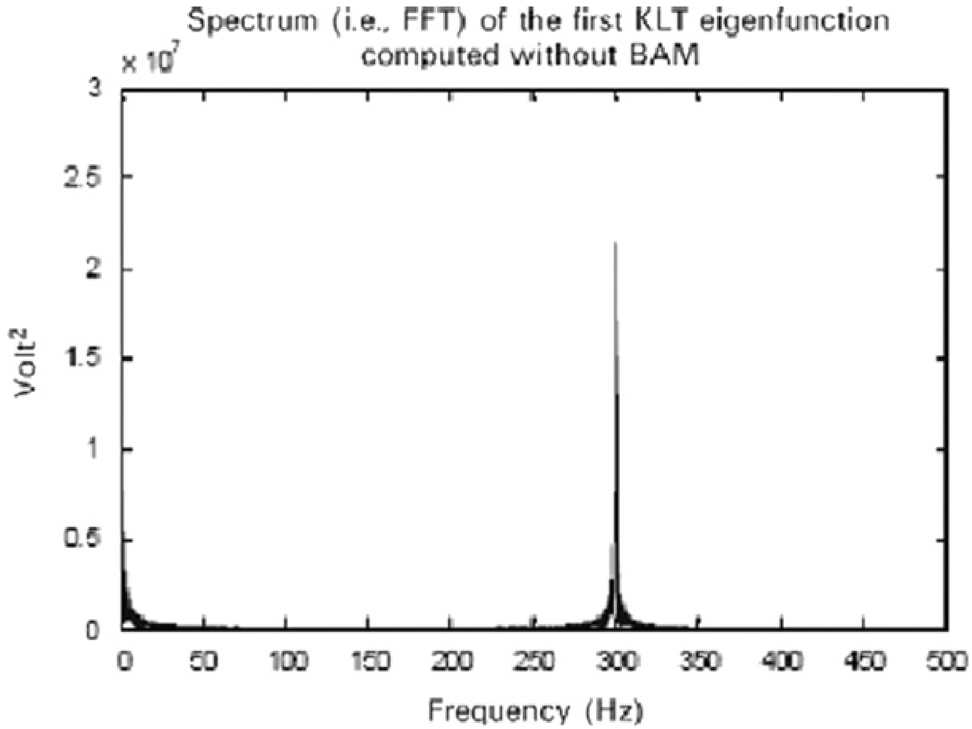


Figure 9.4: The spectrum (i.e. the Fourier Transform) of the first KLT eigenfunction NOT obtained by the BAM, but rather by the very long procedure of solving the N linear algebraic equations corresponding, in discrete time, to the integral Eq. (9.18). Clearly, the result is the same as obtained in Fig. 9.3 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 the N 2 calculations requested by the "straight" KLT theory.

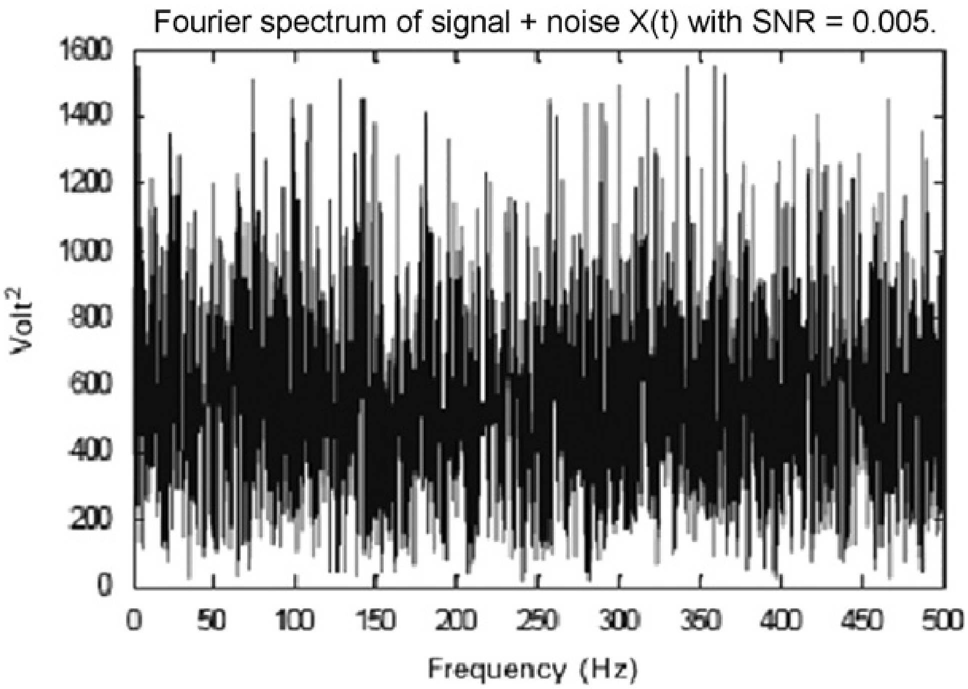


Figure 9.5: Fourier spectrum of a pure tone (i.e. just a sinusoidal carrier) with frequency at 300 Hz buried in stationary noise with the terribly low signal-to-noise ratio of 0.005. This is clearly beyond the reach of the FFT, since we know there should just be one peak only at 300 Hz. Fourier FAILS at such a low SNR.

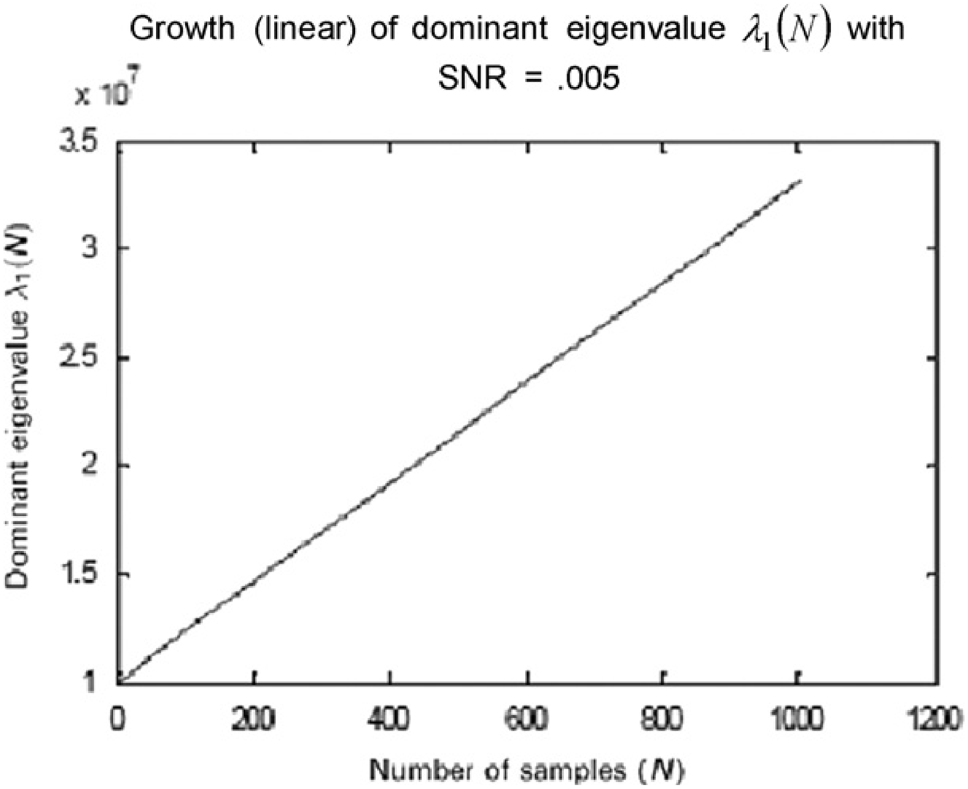


Figure 9.6: The KLT dominant eigenvalue $\lambda_1(N)$ over N=1000 time samples, computed by virtue of the BAM, for the very low SNR= 0.005.

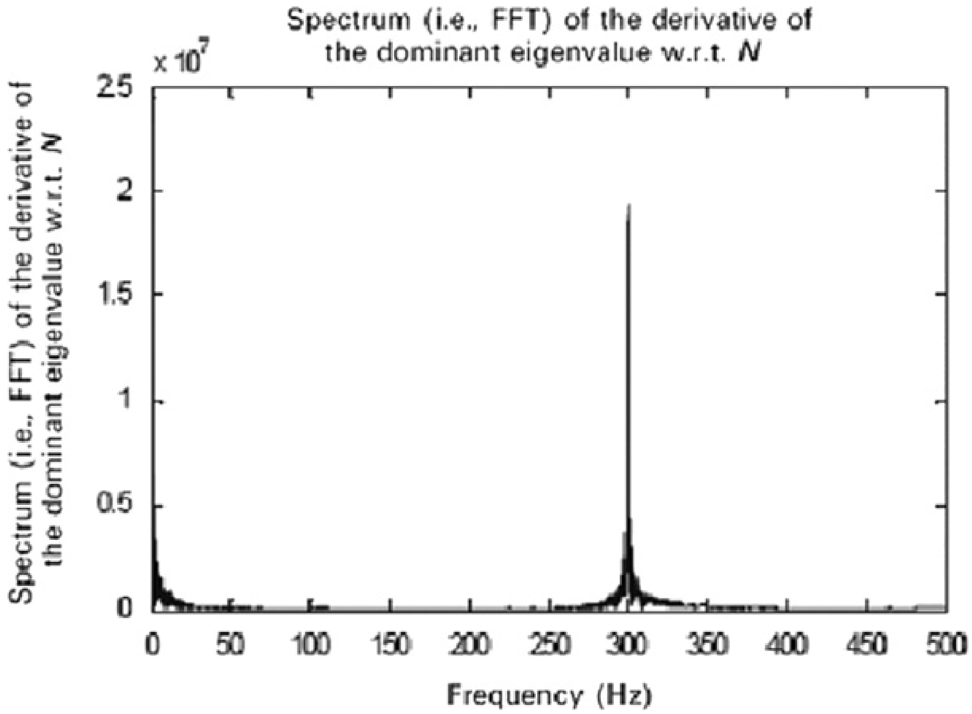


Figure 9.7: The spectrum (i.e. the Fourier Transform) of the CONSTANT derivative of the KLT dominant eigenvalue $\lambda_1(N)$ with respect to N as given by the BAM. This is a neat Dirac delta function, i.e. a peak, at 300 Hz, as expected.

And this is just the same as the last plot (Fig. 9.8) of the dominant KLT eigenfunction obtained by KLT SLOW way of doing N 2 calculations.

This proves the superior behavior of the KLT.

9.9 Developments in 2007 and 2008

The numerical simulations described in the previous section were performed at Medicina during the winter 2006–2007 by Francesco Schilliro 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, $\partial\lambda_1(T)/\partial T$, is just the frequency of the feeble sinusoidal carrier buried into the mountain of noise. In SETI language, if we are looking for a simple sinusoidal carrier sent by ET and buried into a lot of cosmic noise, then the frequency we are looking for is given by the FFT of $\partial\lambda_1(N)/\partial T$.

Why?

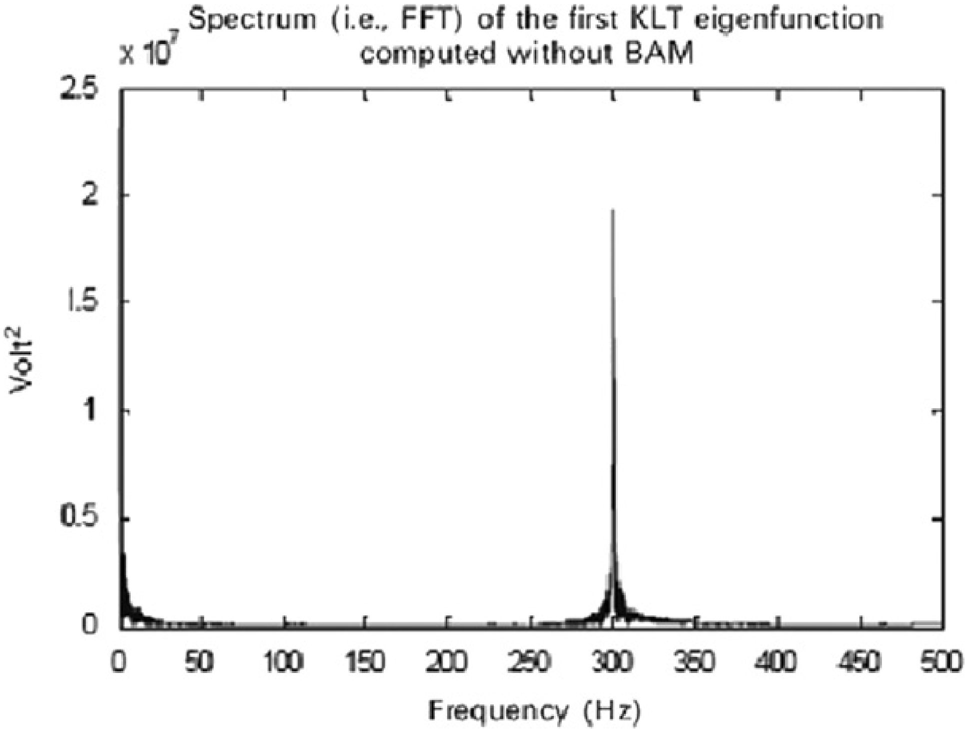


Figure 9.8: The spectrum (i.e. the Fourier Transform) of the first KLT eigenfunction NOT obtained by the BAM, but rather by the very long procedure of solving the N linear algebraic equations corresponding, in discrete time, to the integral Eq. (9.18). Clearly, the result is the same as obtained in Fig.9.7 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 the N^2 calculations requested by the "straight" KLT theory.

No analytic proof of this numerical result was ever found at Medicina. This author, however, had made the first step towards the then missing analytic proof by proving the Final Variance Theorem in May 2007, and he kept talking about this "frontier results" 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 radio- telescope. Young and bright Dr. Sarod Yatawatta of ASTRON then made the next step toward the missing analytic proof: he derived a previously unknown analytic expression for the KLT eigenvalues of the ET sinusoidal carrier [24]. Unfortunately, Dr. Yatawatta made two slight analytical errors in his derivation (described hereafter) that 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 is the proof that the Fourier Transform of the first derivative of the KLT eigenvalues with respect to the final instant T is twice the frequency of the "unknown" ET signal. For stationary processes only, of course.

For non-stationary processes, i.e. for transient phenomena (just as it actually happens in practical SETI, since all celestial bodies move) 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 paper yet! 10. KLT of stationary unitary white noise Before we give the analytic proof that the Fourier Transform of $\partial\lambda_1(T)/\partial T$ is twice the frequency of the unknown ET signal, we must understand what the KLT of stationary unitary white noise is.

Stationary unitary 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 unitary white noise by W(t), one has by definition

$$E\{W(t_1)W(t_2)\} = \delta(t_1 - t_2) \tag{9.32}$$

If one now seeks for the KLT of stationary unitary white noise, one must of course replace the autocorrelation (9.32) into the KLT integral Eq. (9.18), getting

$$\begin{aligned} \lambda_n\phi_n(t_2) &= \int_0^T E\{W(t_1)W(t_2)\}\phi_n(t_1)dt_1 \\ &= \int_0^T \delta(t_1 - t_2)\phi_n(t_1)dt_1 = \phi_n(t_2) \end{aligned} \tag{9.33}$$

This proves that:

1) The KLT eigenvalues of stationary unitary 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 the stationary unitary white noise.

Since **any** set of orthonormal eigenfunctions $\phi_n(t)$ in the Hilbert space is a suitable basis to represent the stationary unitary white noise, from now on we shall adopt the easiest possible such basis, that is the simple Fourier basis made up by orthonormalized sines only over the finite interval $0 \le t \le T$.

$$\phi_n(t) = W_n(t) = \sqrt{\frac{2}{T}} sin\left(\frac{2\pi n}{T}t\right) \tag{9.34}$$

This set of basis functions of course fulfills the orthonomality condition

$$\int_0^T W_m(t)W_n(t)dt = \int_0^T \sqrt{\frac{2}{T}} sin\left(\frac{2\pi m}{T}t\right)\sqrt{\frac{2}{T}} sin\left(\frac{2\pi n}{T}t\right)dt = \delta_{mn} \qquad (9.35)$$

This property will be used in the next section, where we give the proof that the Fourier transform of $\partial\lambda_n(T)/\partial T$ is indeed (twice) the frequency of the unknown ET sinusoidal carrier buried into the white, cosmic noise. We just 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 (9.34). This means that his further results were flawed, even more so since he made a second analytical error in further calculations that we shall not describe here. But the key ideas behind his proof were correct, and we shall describe them in the next section.

9.10 KLT of an ET sinusoidal carrier buried into white, cosmic noise

Consider a new stochastic process S(t) made up by the sum of stationary unitary white noise W(t) plus an alien ET sinusoidal carrier of amplitude a and frequency $\nu = \omega/2\pi$, that is

$$S(t) = W(t) + asin(\omega t) \qquad (9.36)$$

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

$$\begin{aligned} E\{S(t_1)S(t_2)\} &= E\{W(t_1)W(t_2)\} + a^2 sin(\omega t_1)sin(\omega t_2) \\ &+aE\{W(t_1)sin(\omega t_1)\} + aE\{W(t_2 0 sin(\omega t_1)\} \end{aligned} \qquad (9.37)$$

The last two terms in (9.37) 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 (9.37) of the (signal+noise) stochastic process $S(t)$ thus becomes

$$E\{S(t_1)S(t_2)\} = E\{W(t_1)W(t_2)\} + a^2 sin(\omega t_1)sin(\omega t_2) \qquad (9.38)$$

In order to proceed, we now make use of the eigen function expansion of the autocorrelation (16) that, replaced in (38), changes it into

$$\begin{aligned}&\textstyle\sum_{m=1}^{\infty} \lambda_{S_m} S_m(t_1) S_m(t_2)\\&= \textstyle\sum_{m=1}^{\infty} \lambda_{W_m} W_m(t_1) W_m t_2 + a^2 sin(\omega t_1) sin(\omega t_2)\end{aligned} \tag{9.39}$$

In the last equation, $S_m(t)$ clearly are the (unknown) eigenfunctions of the (signal+noise) process $S(t)$, and λ_{S_m} are the (unknown) corresponding eigenvalues. In the right-hand side, λ_{S_m} are the eigenvalues of the stationary unitary white noise that we know to be equal to 1, but, for the sake of clarity, let us keep the symbol λ_{W_m} rather than 1.

To proceed further, we now must get rid of both t 1 and t 2 in (39), and there is only one way to do so: use the orthonormality of the eigenfunctions appearing in (39). We shall do so in a moment. Before, however, let us make the following practical consideration: since the signal is much weaker than the noise (by assumption) (i.e. the signal-to-noise ratio is much smaller than 1, or SNR⋘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⋘1 amounts to the two approximated equations

$$\begin{cases} S_m(t) \approx W_m(t) \\ \\ \lambda_{S_m} \approx \lambda_{W_m} = 1 \end{cases} \tag{9.40}$$

Only the first of these two equations will of course 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 $0 \leq t \leq T$. As a consequence, the new orthonormality condition (nearly) holds

$$\int_0^T S_m(t_1) W_n(t_1) dt_1 \approx \delta_{mn} \tag{9.41}$$

and, similarly,

$$\int_0^T S_k(t_2) W_n(t_2) dt_2 \approx \delta_{kn} \tag{9.42}$$

So, let us now multiply both sides of (39) by $W_n(t_1)$ and integrate with respect to t_1 between 0 and T. Because of (9.41) and (9.35) one has

$$\sum_{n=1}^{\infty} \lambda_{S_n} S_n(t_2) \approx \sum_{n=1}^{\infty} \lambda_{W_m} W_n(t_2) + a^2 sin(\omega t_2) \int_0^T W_n(t_1) sin(\omega t_1) dt_1 \tag{9.43}$$

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 (9.34) and integrating, one gets

$$\begin{aligned} &\textstyle\sum_{k=1}^{\infty} \lambda_{S_k} S_k(t_2) \\ &\approx \textstyle\sum_{k=1}^{\infty} \lambda_{W_k} W_k(t_2) + a^2 sin(\omega t_2) w \sqrt{2} \pi n \frac{\sqrt{T} sin(\omega T)}{\omega^2 T^2 - 4\pi^2 n^2} \end{aligned} \tag{9.44}$$

We next multiply this equation by $W_n(t_2)$ and integrate with respect to t_2 between 0 and T. Because of (9.42) and (9.35), (9.44) becomes

$$\lambda_{S_n} \approx \lambda_{W_n} + a^2 2 \sqrt{2} \pi n \frac{\sqrt{T} sin(\omega T)}{\omega^2 T^2 - 4\pi^2 n^2} \int_0^T W_n(t_2) sin(\omega t_2) dt_2 \tag{9.45}$$

Again, the integral in the last equation can be computed (it is actually the same integral as in (9.43)) and so the conclusion is

$$\lambda_{S_n} \approx \lambda_{W_n} + a^2 \frac{8\pi^2 n^2 T sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2} \tag{9.46}$$

This is Yatawatta's result, as 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

$$\lambda_{S_n}(T) \approx \lambda_{W_n} + a^2 \frac{8\pi^2 n^2 T sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2} \tag{9.47}$$

This equation clearly shows that:

1) For $T \to \infty$, the fraction in the right-hand side approaches zero, and so the eigenvalues of the (signal+noise) approach the pure white noise eigenvalues (as it 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 it is intuitively obvious again). 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

$$\begin{aligned}\lambda_{S_1}(T) \approx \lambda_{W_1} + a^2 \frac{8\pi^2 T sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2)^2} \\ 1 + a^2 \frac{8\pi^2 T sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2)^2}\end{aligned} \tag{9.48}$$

This completes our analysis of the KLT of a sinusoidal carrier buried into white, cosmic noise.

9.11 Analytic proof of the BAM-KLT

We are now ready for the analytic proof of the BAM-KLT method. Let us first re-write (9.47) in the form where the pure white noise eigenvalues are replaced by 1

$$\lambda_{S_n}(T) \approx 1 + a^2 \frac{8\pi^2 n^2 T sin^2(\omega T)}{(\omega^2 T^2 - 4\pi^2 n^2)^2} \tag{9.49}$$

Let us then notice that the final instant T appears three times in the right-hand side of the last equation:

1) Once at the numerator outside the sine;

2) Once at the numerator inside the sine;

3) Once at the denominator.

Therefore, the partial derivative of (49) with respect to T will be made up by the sum of three terms:

1) One term with the derivative of the T at 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

$$sin^2(\omega T) = \frac{1}{2} - \frac{1}{2} cos(2\omega T) \tag{9.50}$$

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

$$2 sin(\omega T) cos(\omega T) = sin(2\omega T) \tag{9.51}$$

3) One term with the derivative of the T at 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 (9.49) with respect to T and just give the final result.

In conclusion, the derivative of (9.49) with respect to T is given the following sum of three terms

$$\frac{\partial \lambda_{S_n}(T)}{\partial T} \approx coeff_1(T) sin(2\omega T) + coeff_2(T) cos(2\omega T) + coeff_3(T) \tag{9.52}$$

where the three coefficients turn out to be (after lengthy calculations)

$$\begin{cases} coeff_1(T) = a^2 \tfrac{8\pi^2 n^2 \omega T}{(\omega^2 T^2 - 4\pi^2 n^2)^2} \\ \\ coeff_2(T) = a^2 \tfrac{4\pi^2 n^2 (3\omega^2 T^2 + 4\pi^2 n^2}{(\omega^2 T^2 - 4\pi^2 n^2)^3} \\ \\ coeff_3(T) = -a^2 \tfrac{4\pi^2 n^2 (3\omega^2 T^2 + 4\pi^2 n^2)}{(\omega^2 T^2 - 4\pi^2 n^2)^2} \end{cases} \tag{9.53}$$

But the right-hand side of (9.53) is nothing but the simple Fourier series expansion of $\partial \lambda_{S_n}(T)/\partial T$. Moreover, (9.53) shows that $\partial \lambda_{S_n}(T)/\partial T$ is a PERIODIC function of T with frequency $2\omega T$. We conclude that: the Fourier transform of $\partial \lambda_{S_n}(T)/\partial T$ equals TWICE the frequency of the buried alien sinusoidal carrier. In other words, the frequency of Alien Signal is a HALF of the frequency found by taking the Fourier transform of $\partial \lambda_{S_n}(T)/\partial T$. And the BAM-KLT method is hence proved analytically.

9.12 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 30 October 2008 that well summarizes the key features of the present scientific paper.

How to eavesdrop on alien chat
30 October 2008
From New Scientist Print Edition
Jessica Griggs

ET, phoney 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 is 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 extra-terrestrial life (SETI) is not listening for such signals, says Claudio Maccone, co-chair of the IAA 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 should not 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, page 14.

9.13 Conclusions

Let us summarize the main results of our paper.

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 (that is the roughest approximation to the full KLT expansion that 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, or

2) (short way= BAM) you compute the derivative of the first eigenvalue with respect to T=N and then Fourier- transform it to get the first eigenfunction.

In practice, numerical simulations of the KLT may be much less time-consuming if option (2) is chosen 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 a noise that has a SNR of about 1 or so, the KLT can, say, filter out signals that have a SNR of, say, 0.001 or so.

This is the superior achievement of the KLT with respect to 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. We provided the results of numerical simulations showing that, by virtue of the BAM, the KLT succeeds in extracting a sinusoidal carrier embedded in lot of noise when the FFT utterly fails.

9.14 Additional reference about non-SETI applications of the KLT

The KLT may of course be used for non-SETI applications also. Just to mention one, in [7] this author pointed out that the range of the planetary radar (sometimes called "asteroidal radar") might be increased just by replacing the KLT to the traditional FFT. In fact, the KLT is capable of retrieving weak sinusoidal signals better than the FFT, as we saw. Thus, without changing the hardware at all, it is possible to improve the performance of the asteroidal radar just by paying for the higher computational burden requested by the KLT.

Acknolwdgments

This author is indebted to many radioastronomers and scientists who helped him over the years to work out what is now the BAM-KLT method. Among them: 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 Dr. Seth Shostak and Dr. Doug Vakoch). 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.

References

1. K. Karhunen, Uber lineare methoden in der wahrscheinlichkeitsrechnung, Ann. Acad. Sci. Fenn., Ser. A 1, Math. Phys. 37 (1946) 3–79.

2. M. Loève, Fonctions aleatoires de second ordre, Rev. Sci. 84 (4) (1946) 195–206.

3. M. Loève, in: Probability Theory: Foundations, Random Sequencies,Van Nostrand, Princeton, NJ, 1955.

4. C. Maccone, Telecommunications, KLT and Relativity, vol. 1, IPI press, Colorado Springs, Colorado, USA, 1994, ISBN: 1-880930-04-8. This book embodies the results of some thirty research papers published by the author about the KLT in the fifteen years 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, M. Poloni, S. Poppi, S. Righini, M. Roma, M. Teodorani, 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 Nanc - ay radio-telescope, Acta Astronaut. 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 Astronaut. 60 (2007) 775–779.

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, Ser. B 100 (1987) 329–342.

Bollettino dell'Unione Matematica Italiana:

9. C. Maccone, Eigenfunctions and Energy for Time-Rescaled Gaussian Processes, Boll. Unione Mat. Ital. Ser. 6 3-A (1984) 213–219.
10. C. Maccone, The time-rescaled Brownian motion B(t2H), Boll. Unione Mat. Ital. Ser. 6 4–C (1985) 363–378.
11. C. Maccone, The Karhunen–Loève expansion of the zero-mean square process of a time-rescaled Gaussian process, Boll. Unione Mat. Ital. Ser. 7 2–A (1988) 221–229.

Journal of the British Interplanetary Society:

12. C. Maccone, Relativistic interstellar flight and genetics, J. Br.Interplanet. Soc. 43 (1990) 569–572.

Acta Astronautica:

13. C. Maccone, Relativistic interstellar flight and Gaussian noise, Acta Astronaut. 17 (9) (1988) 1019–1027.
14. C. Maccone, Relativistic interstellar flight and instantaneous noise energy, Acta Astronaut. 21 (3) (1990) 155–159.

KLT for Data Compression:

15. C. Maccone, The data compression problem for the "GAIA" astrometric satellite of ESA, Acta Astronaut. 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, in: 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, editor), vol. 47, 1993, pp. 128–140.

17. C. Maccone, Karhunen–Loève Versus Fourier Transform for SETI, in: Jean Heidmann and Mike Klein, Eds, Proceedings of the Third Bioastronomy Conference held in Val Cenis, Savoie, France, 18-23 June 1990, Lecture Notes in Physics, vol. 390, Springer-Verlag, 1990, pp. 247–253.

After these seminal works were published, the importance of the KLT for SETI was finally acknowledged by the SETI Institute experts in:

18. SETI 2020, Ron Eckers, Kent Cullers, John Billingham and Lou Scheffer editors, SETI Institute, 2002, pp. 234, note 13. The authors say: "Currently (2002) only the Karhunen Loeve (KL) transform [Mac94] shows potential for recognizing the difference between the incidental radiation technology and white noise. The KL ransform is too computationally intensive for present generation of systems. The capability for using the KL transform should be added to future systems when the 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, J. Br. Interplanet. Soc. 47 (1994) 1.

An early paper about the KLT for SETI-Italia:

20. S. Montebugnoli, C. Maccone, SETI-Italia Status Report 2001, a paper presented at the 2001 IAF Conference held in Toulouse, France, 1–5 October, 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. Schilliro, S. Pluchino, C. Maccone, S. Montebugnoli, Istituto Nazionale di Astrofisica (INAF)—Istituto di Radioastronomia (IRA), Rapporto Tecnico, La KL Transform: considerazioni generali sulle metodologie di analisi ed impiego nel campo della Radioastronomia, Technical Report (in Italian only), January 2007.

23. C. Maccone, Innovative SETI by the KLT, Proceedings of the "Bursts, Pulses and Flickering" Conference held at Kerastari, Greece, June 13–18, 2007 at POS (Proceedings of Science) website[5].

[5] http://pos.sissa.it//archive/conferences/056/034/Dynamic2007_034.pdf

24. Sarod Yatawatta, Personal Communication, 17 June 2008.

A recent paper about the KLT for Relativistic Interstellar Flight:

25. C. Maccone, Relativistic optimized link by KLT, J. Br. Interplanet. Soc. 59 (2006) 94–98.

The "final" 2009 book by the author about the Sun as a Gravitational Lens, the relevant FOCAL space missions, and the KLT, including the relativistic KLT:

26. C. Maccone, "Deep Space Flight and Communications", a 400-pages technical treatise published by Praxis-Springer in 2009, ISBN 978-3-540-72942-6.

Chapter 10

METI: Messaging to Extra-Terrestrial Intelligence

by **Alexander L. Zaitsev**
IRE, Russia alzaitsev@gmail.com

10.1 Introduction

Messaging to Extra-Terrestrial Intelligence (METI) represents cardinally new kind of human activity. Somebody can object that Search for Extra-Terrestrial Intelligence (SETI) also is cardinally new. Yes, new, but not cardinally, because people always passively surveyed heaven in a hope to detect something unknown, whether it's natural or artificial. However, the purposeful efforts directed to converting Terrestrial Civilization into the object of a possible detection by probable Extra-Terrestrial ones, such activity does is substantially new.

The scientific program known as SETI has the main goal to search for any kind of EM-radiation from aliens. In contrast, METI's main goal is to create and to send intelligent messages from humans to aliens. SETI scientists ask only one particular question: "Does Active SETI make sense?" In other words, would it be reasonable, for SETI success, to transmit with the object of attracting ETI's attention?

The goal of METI is much broader: to overcome the Great Silence in the Universe by bringing to ETIs the long-awaited news: "You are not alone!" Since the entire monumental foundation of SETI is based on a tacit assumption of real existence of civilizations transmitting interstellar messages, the scientists who are involved in SETI should unavoidably accept that messaging to ETI is reasonable and strongly motivated by the existence of SETI itself.

10.2 Brief History of METI

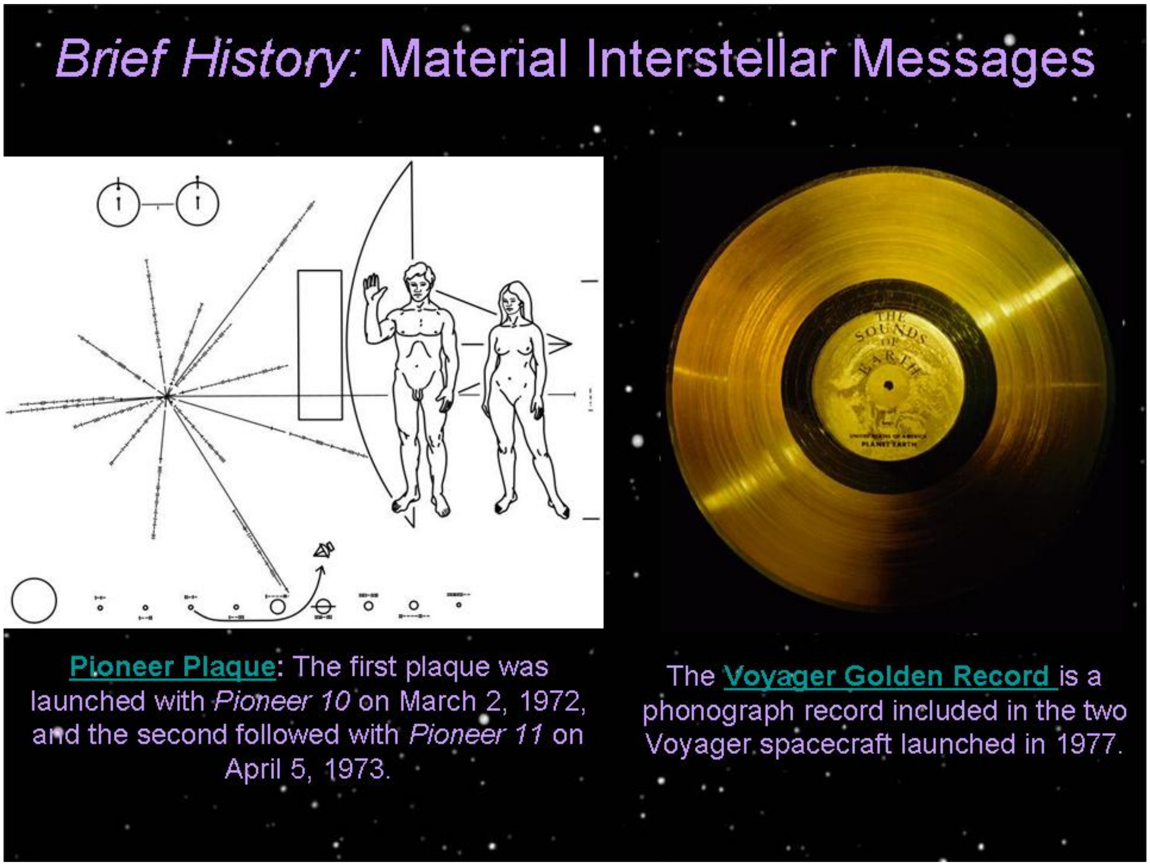
SETI scientists have been listening cosmos for 50 years trying to detect artificial signals in the Universe but, unfortunately, their efforts have brought no results. There are some reasons for this, which were discussed and analysed in corresponding articles and books in more detail. As compared to SETI, METI is in a more advantageous position. Indeed, after composing a dedicated message and sending it to a properly selected star in our galaxy, METI scientists already have some result that makes sense. Specifically, launching the intelligent radio message places a first stone to building a radio bridge between the terrestrial and presumed extraterrestrial civilizations. After this, everything depends only on THEM whether they will discover our Message and send a corresponding response in order to establish the Contact.

Both the first interstellar messages and the first experiments on search for aliens' signals are associated with the name of Frank Drake. In 1972 he together with Carl Sagan and others had made a plate "Pioneer Plaque" [1], and then, in 1977, a disk "Voyager Golden Record" [2], which were placed on the spacecraft flying outside of the Solar system (see Fig.10.1).

"Arecibo Message", the first interstellar radio message, was also created by Drake and Sagan. It sent on November 16, 1974 using a radar telescope with the antenna of diameter of 1000 foots (305 m) and transmitter with the mean power of 500 kW at a wavelength 12.6 cm. Radio messages of four other projects, "Cosmic Call 1999", "Teen Age Message 2001", "Cosmic Call 2003" and "A Message From Earth 2008", have been transmitted into Space using the Evpatoria Planetary Radar, fig.10.2.

Thus, during the entire history of the terrestrial civilization only five interstellar radio messages (IRMs) have been prepared and got out to the space. In Table 10.1 these five IRMs are shown in order of the dates of the first transmission session (notice that the overall number of the sessions is 17). Symbols T and E denote, accordingly, the total duration of all sessions of each IRM and the total energy transmitted. The latter figure correlates to the range of the possible detection.

The Arecibo Message was constructed of 1679 bits. It was sent to a globular cluster M13. The content and structure of the message have been repeatedly described in various books [3] and on the Web [4], and thus do not do not require further discussion. The transmission of radio messages was resumed 25 years later at the Evpatoria Planetary Radar. In 1999 the Evpatoria radar transmitted the message called "Cosmic Call" (CC-1999) to four Sun-like stars [5]. This message represented a kind of encyclopedia of human knowledge about our civilization and the surrounding world written in the special language Lexicon developed by two Canadians: Yvan Dutil and Stephane Dumas. In addition, the structure of CC-1999



Brief History: Material Interstellar Messages

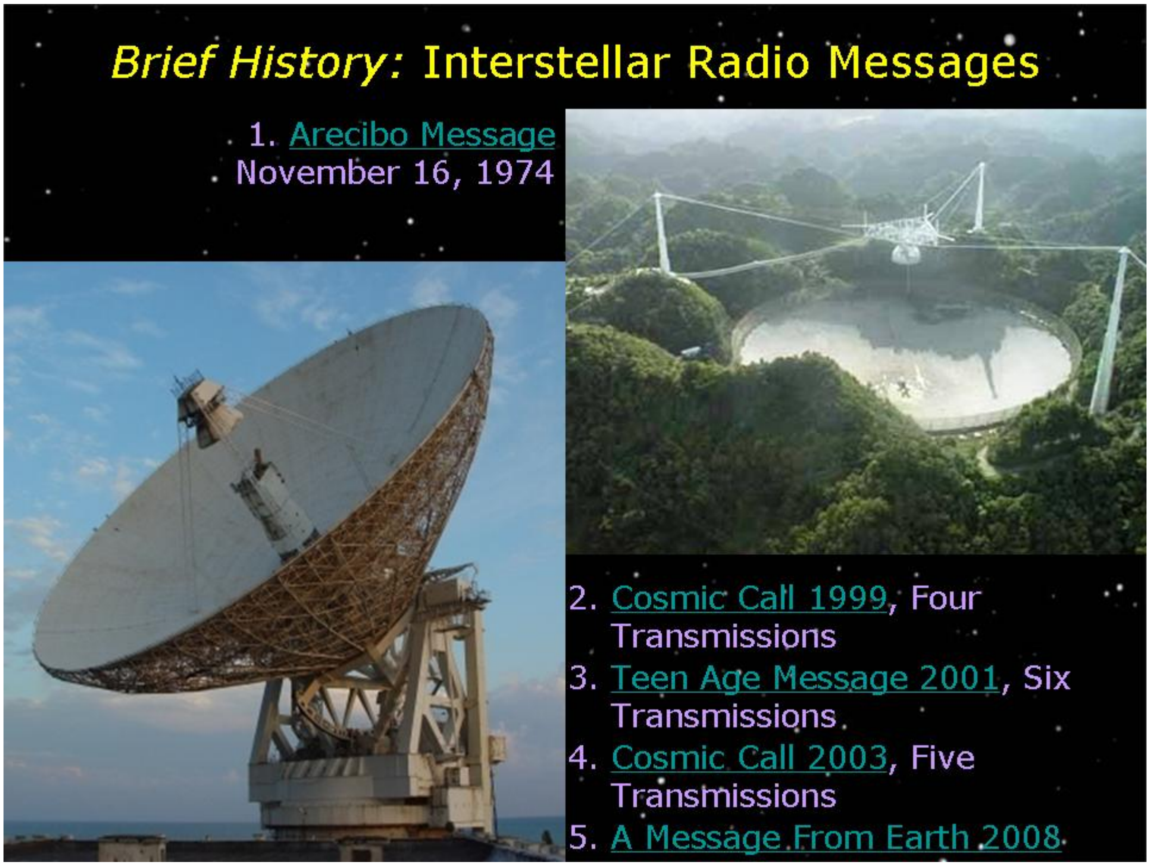
Pioneer Plaque: The first plaque was launched with *Pioneer 10* on March 2, 1972, and the second followed with *Pioneer 11* on April 5, 1973.

The **Voyager Golden Record** is a phonograph record included in the two Voyager spacecraft launched in 1977.

Figure 10.1: The first interstellar messages sent beyond the Solar system onboard the American spacecraft "Pioneer 10 and 11" and "Voyager 1 and 2"

included the technical data of the project, names of its participants and a copy of the Arecibo Message. The size of the "Encyclopedia" was 370967 bits.

In 2001, we were involved in the development and transmission of a Teen Age Message [6] to 6 Sun-like stars. For the first and, regretfully, last time, the transmitted message consisted of three separate sub-messages: 1) monochromatic probing signal, 2) analogue information (music), and 3) digital information. These three elements are described in more detail later in this chapter. As a source of the analogue signal content, we selected a performance on a Theremin ("Termenvox") electronic musical instrument, which generates a quasi-monochromatic signal with a low level of overtones. This significantly facilitates detection of the message and subsequent recognition of its artificial nature over interstellar distances [7]. The digital part consisted of 28 binary images of the total size 648220 bits.



Brief History: Interstellar Radio Messages

1. Arecibo Message November 16, 1974
2. Cosmic Call 1999, Four Transmissions
3. Teen Age Message 2001, Six Transmissions
4. Cosmic Call 2003, Five Transmissions
5. A Message From Earth 2008

Figure 10.2: The first interstellar radio messages.

IRM Cosmic Call 2 was sent to 5 Sun-like stars in 2003 [8]. It was the first truly international IRM, composed by citizens of the USA, Canada and Russia and consisted of a set of fragments of the three previous radio messages. We believe that such a democratic equal-opportunity approach should be applied to all future interstellar messages transmitted from the Earth.

IRM "A Message From Earth" (AMFE) has been prepared and sent from Evpatoria in October 2008 [9]. Its distinctive characteristic was that involvement was opened up, through the Internet, to a great number of participants of the social network Bebo. 501 "best" messages were selected through a web vote for inclusion in the subsequent radio transmission. Initially, the idea of interstellar radio message composition by the general public, through a special website, was suggested in 2002 in article "Project METI@home: Messages to ETI from home" [10].

Standing slightly outside of the main stream, there are two more IRMs: "Across

Table 10.1: Interstellar radio messages transmitted from Earth.

Name	Arecibo Message	Cosmic Call 1	Teen Age Message	Cosmic Call 2	A Message From Earth
Date	16.11.1974	24.05, 30.06, 01.07.1999	29.08, 03.09, 04.09.2001	06.07.2003	09.10.2008
Type	World First IRM (digital)	First multi-page IRM	First digital and analogue	First International IRM	First Collective IRM
Authors	Drake, Sagan, Issacman, et al	Chafer, Dutil, Dumas, Braasta, Zaitsev, et al	Pshenichner, Filippova, Gindilis, Zaitsev, et al	Chafer, Dutil, Dumas, Braastad, Zaitsev, el al.	Madgett, Coombs, Levine, Zaitsev, et al.
Radar	Arecibo	Evpatoria	Evpatoria	Evpatoria	Evpatoria
Sets	1	4	6	5	1
T, min	3	960	366	900	240
E, MJ	83	8640	2200	8100	1440

the Universe 2008" [11] and "Hello From Earth 2009" [12] which were transmitted to the space using 70-m radio dishes of the NASA JPL Deep Space Network and located in Robledo (Spain) and Canberra (Australia). The first of the above-mentioned IRMs was critically discussed in [7]; the second message also had its drawbacks. We consider their main defect to be insufficient scientific and technical justification.

The summary shown as Table 10.2 presents basic data on all 17 transmission sessions for these five terrestrial radio messages. R represents the distance to the target stars, expressed in light years.

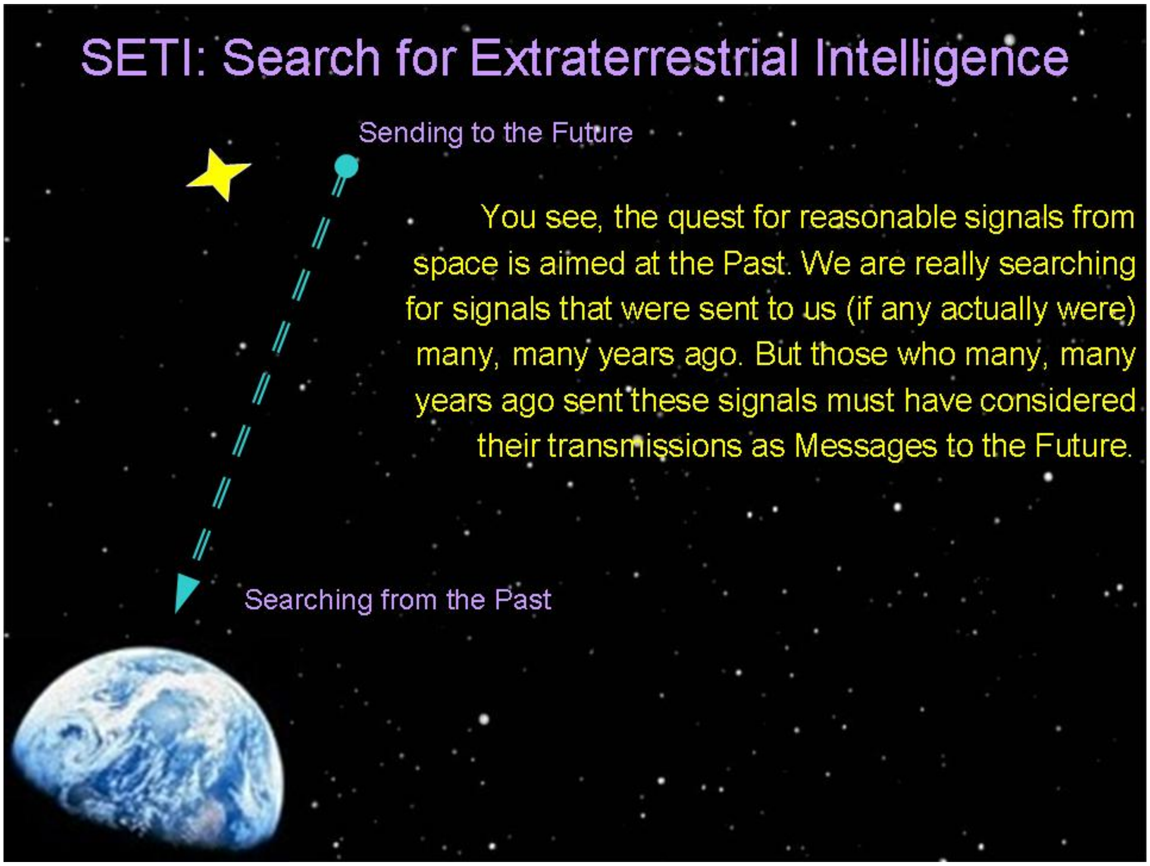
The last column of Table 10.2 predicts the time when the "Great Silence of the Universe" can potentially come to an end, reaching any aliens who happen to exist at the receiving side of the communication link, on the highly optimistic chance that they are capable of detecting our radio messages. If they do detect our signals, they will likely perceive that they now live in a drastically different habitable Universe. Thus a scientific revolution may start in one alien's civilization, hopefully to propagate through the entire Universe, being passed from one civilization to another, once they realize that they are not alone. This fundamental transformation of the Universe through interstellar messaging can be triggered by us, by our intellect and our good will. We consider triggering such a communications revolution to be one of the most worthy applications of the united intellect of human civilization!

Table 10.2: Details of the 17 sessions of the conducted IRM transmission

Message	Target	Constellation	Date sent	R, LY	Arrival date
AM	NGC 6205	Hercules	Nov 16, 1974	≈ 25000	≈ 26974
CC-1	HD 186408	Cygnus	May 24, 1999	70.5	Nov 2069
CC-1	HD 190406	Sagitta	Jun 30, 1999	57.6	Feb 2057
CC-1	HD 178428	Sagitta	Jun 30, 1999	68.3	Oct 2067
CC-1	HD 190360	Cygnus	Jul 1, 1999	51.8	Apr 2051
TAM	HD 197076	Delphinus	Aug 29, 2001	68.5	Feb 2070
TAM	HD 95128	Ursa Major	Sep 3, 2001	45.9	Jul 2047
TAM	HD 50692	Gemini	Sep 3, 2001	56.3	Dec 2057
TAM	HD 126053	Virgo	Sep 3, 2001	57.4	Jan 2059
TAM	HD 76151	Hydra	Sep 4, 2001	55.7	May 2057
TAM	HD 193664	Draco	Sep 4, 2001	57.4	Jan 2059
CC-2	HIP 4872	Cassiopeia	Jul 6, 2003	32.8	Apr 2036
CC-2	HD 245409	Orion	Jul 6, 2003	37.1	Aug 2040
CC-2	HD 75732	Cancer	Jul 6, 2003	40.9	May 2044
CC-2	HD 10307	Andromeda	Jul 6, 2003	41.2	Sep 2044
CC-2	HD 95128	Ursa Major	Jul 6, 2003	45.9	May 2049
AMFE	HIP 74995	Libra	Oct 9, 2008	20.3	Feb 2029

10.3 Interdependence SETI and METI

Our present SETI activity, a quest for reasonable signals from space, is directed to the past, as we are searching for signals that were presumably sent to us many, many years ago. We search in the locations where known exoplanets were at the time they might transmit signals to us (Figure 10.3). In fact, the currently observed starry sky is an image of the past, in the sense that we see celestial objects where they were when they emitted the light now reaching the Earth. Actually, each observed celestial body is now in a slightly different place. This slight difference in the angular position of an object on the sky is related to PM, the proper motion of the celestial body, and is defined as the product of PM [in arc sec per year] and distance D [in Light Years] to the given body.

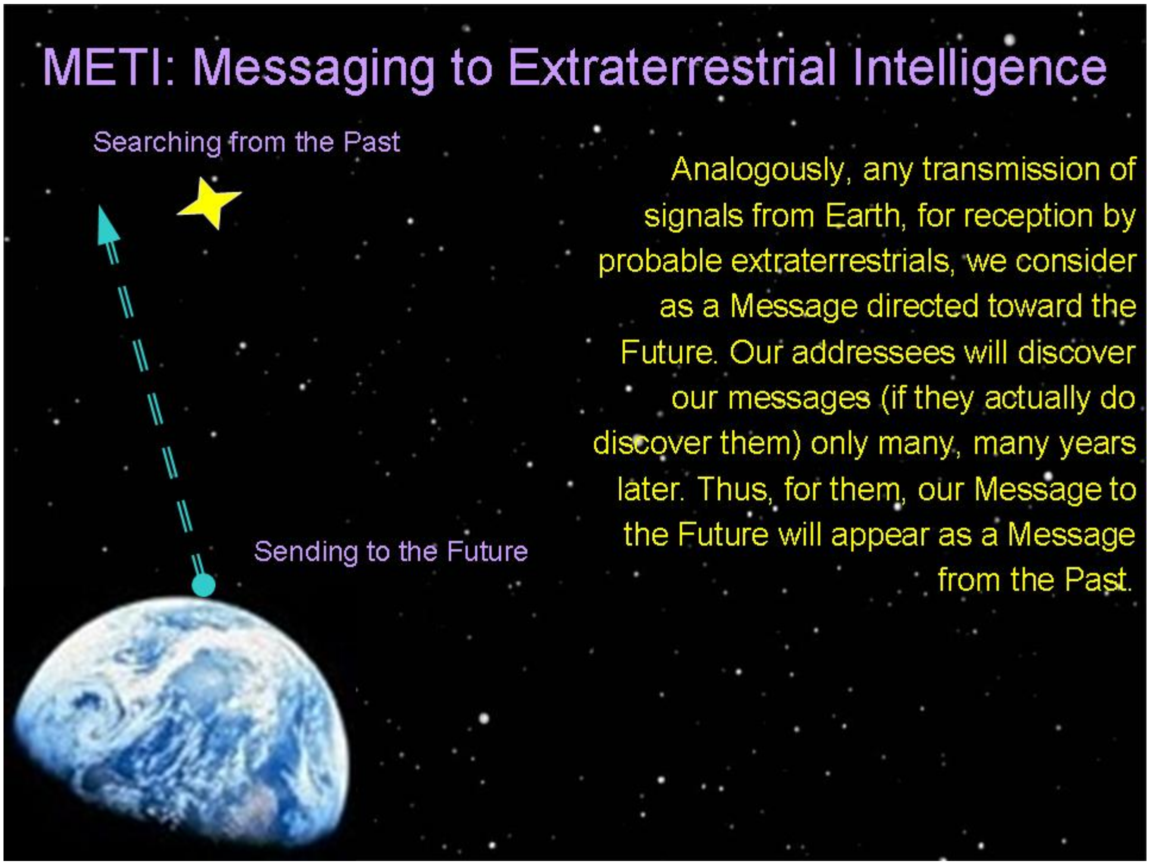


SETI: Search for Extraterrestrial Intelligence

You see, the quest for reasonable signals from space is aimed at the Past. We are really searching for signals that were sent to us (if any actually were) many, many years ago. But those who many, many years ago sent these signals must have considered their transmissions as Messages to the Future.

Figure 10.3: Searching for intelligent signals from the cosmos that come from the past, in both time and position.

In contrast, any METI transmission of signals from Earth, for detection by an extraterrestrial civilization, must be considered directed toward the future. Our addressees will discover our messages in many years to come, and not at the location where they are now, but rather where they will be at the moment our signal reaches them. One needs to count on the finite speed of light in locating the potential recipient of our message in the directional diagram of our transmitting antenna, because the target star will move between "now" and "then" positions in the sky. This effect is similar to that of the proper motion accounted for SETI, but with the opposite sign (Figure 10.4).



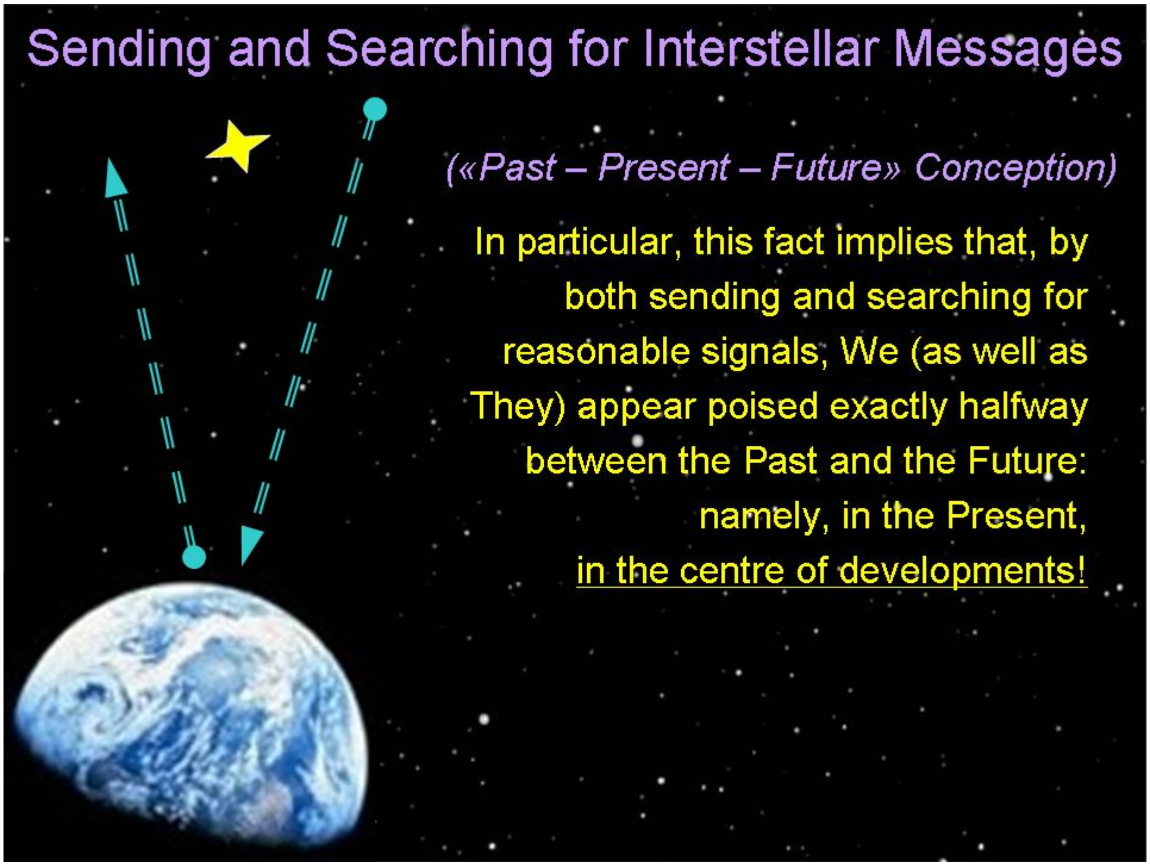
METI: Messaging to Extraterrestrial Intelligence

Analogously, any transmission of signals from Earth, for reception by probable extraterrestrials, we consider as a Message directed toward the Future. Our addressees will discover our messages (if they actually do discover them) only many, many years later. Thus, for them, our Message to the Future will appear as a Message from the Past.

Figure 10.4: Any transmission of intelligent signals to prospective extraterrestrial civilizations is directed to the future, in both time and position.

Thus, we have good reasons to say that conducting both search (SETI) and transmission (METI) of intelligent signals, we find ourselves just halfway between the past and the future, i.e., in the present! It is rather symbolic that in Russian the word "nastoiashchee" has two different meanings: "present" and "genuine".

An advanced civilization, terrestrial or extraterrestrial, may at some point attain such a high level of intellectual and technological development that it starts feeling the need to engage itself in both searching for (SETI) and transmission of (METI) intelligent radio signals (Figure 10.5). The latter can be thought of as a purely altruistic, unselfish activity, seeking to help our neighbors to learn that they are not alone in the vastness of the Universe. Such a socially mature civilization is worthy to be called "genuine". In implementing both SETI and METI, this civilization overcomes the passage of time, positioning itself directly between the past and the future – in the present. Acting unselfishly, not expecting any direct benefit, seeking only the goal of helping other civilizations to realize that they are not alone, such a civilization performs a chivalrous, unsurpassed deed!



Sending and Searching for Interstellar Messages

(«Past – Present – Future» Conception)

In particular, this fact implies that, by both sending and searching for reasonable signals, We (as well as They) appear poised exactly halfway between the Past and the Future: namely, in the Present, in the centre of developments!

Figure 10.5: A truly advance civilization will perform both sending and searching for artificial radio signals.

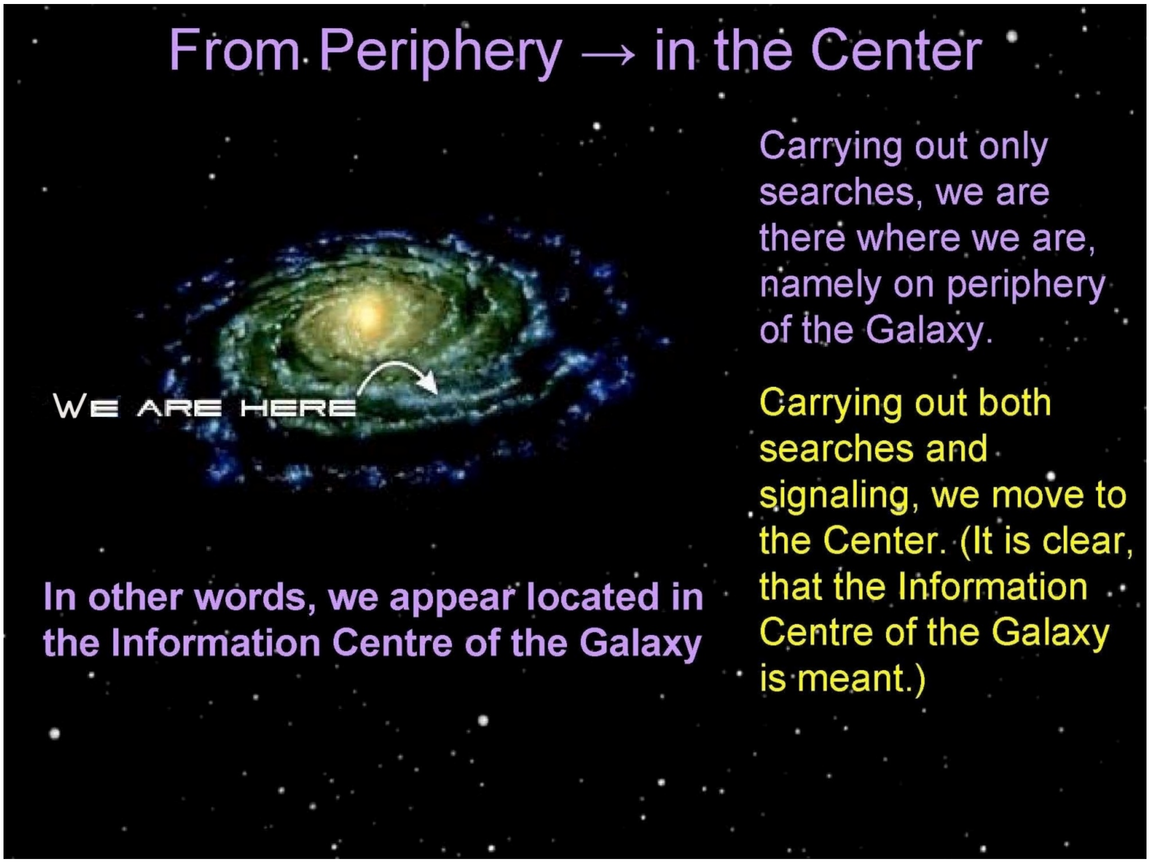
Let us consider the case of optical telescopes used for both transmitting and searching for artificial signals in the Universe. Such a telescope would have lenses and/or mirrors that focus either a beam of a powerful laser (during transmission), or the radiation coming in from the space (when searching). Accounting for the proper motion of the target celestial body must be carried out at both ends of the optical link. In the "Search" mode, we have to direct the telescope to the visible ("past") position of the presumably transmitting celestial body, while in the "Transmission" mode we have to enter a correction by pointing the telescope to that point in the sky where our target body is supposed to be at the time of arrival of our signal. It is important to emphasize that, when transmitting, the correction applied to the orientation of the telescope correction must equal twice of the product of the target proper motion (PM) and distance to the target in light years (D).

In the case of transmission of radio signals, all of the above mentioned considerations will hold, except perhaps for the necessity of redirecting the antenna between the "Transmission" and "Search" modes. We can expect that the angular width of the emitting beam is significantly larger than the angle of the proper motion correction, even for the largest radio antennas. Therefore, it seems reasonable to keep the radio antenna directed continuously to the present position of the target body, i.e., toward the point in the space that is midway between the "past" and the "future", where the celestial body is now ("at present")

French philosopher Blaise Pascal, who lived in the 17th century, expressed his emotions by saying: "The eternal silence of these infinite spaces fills me with dread". The mature planetary consciousness, through sharing Pascal's feeling and perceiving that this "silence of the spaces" should perhaps frighten not only us but also other intelligent inhabitants of the Universe, is coming to the understanding that our mission is to do whatever we can in order to break out the silence of space.

An important point here is that any civilization engaged in both transmitting and receiving becomes an "information centre", as it appears at the centre of the events that move it in the information space from the periphery of our Galaxy to its centre (Figure 10.6).

In the interrelated processes of sending and searching for intelligent signals in the Universe, it is necessary to see distinctly that in the case of sending we create and transmit the messages that would not exist in Nature without human intellect. In this sense, the creation of messages is a kind of art, a creative process of composing something new, with the intention that it be transferred and understood by intellectual beings elsewhere in the Universe. To create and transmit such messages, we must solve purely scientific and technical questions. However, the main issue here is the creative process of producing new information that is intended for propagation



From Periphery → in the Center

Carrying out only searches, we are there where we are, namely on periphery of the Galaxy.

Carrying out both searches and signaling, we move to the Center. (It is clear, that the Information Centre of the Galaxy is meant.)

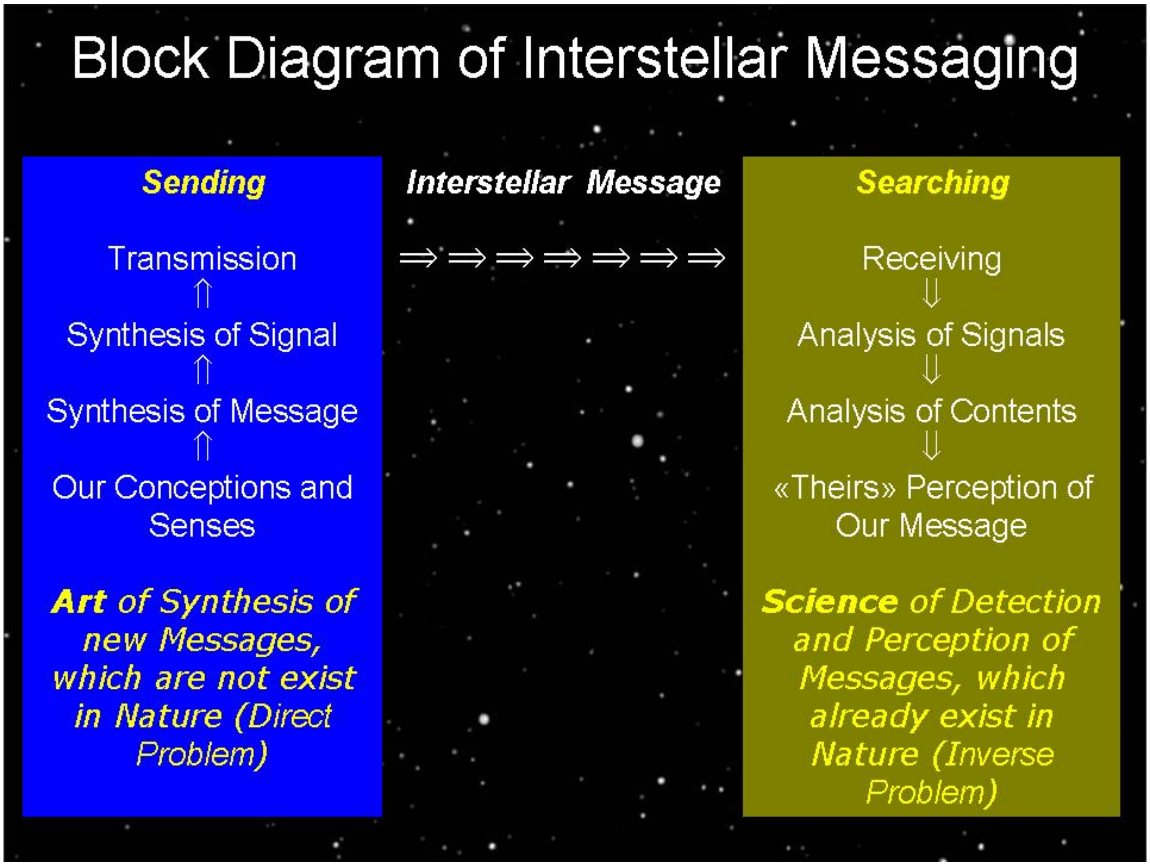
In other words, we appear located in the Information Centre of the Galaxy

Figure 10.6: The concept of the "Information Centre of the Galaxy" is applied to a socially mature, advanced civilization that performs transmitting as well as searching for interstellar radio messages.

to other, yet unknown, intellectual beings.

Searching is a very different matter, an example of a typical inverse problem: we search for what is not yet known to us, but presumably already exists in Nature (Figure 10.7). In other words, by searching, we are solving a scientific problem of the signal detection, its decoding, and extracting information from it. Thus, the specific of the inverse problem is that, in searching, we are looking not for a natural regularity, but rather the opposite – for intelligent messages, signals of mind, not Nature!

Here it is important to notice that the civilization which is carrying out only searching is in a less advantageous position than the civilization which conducts both sending and searching for intelligent signals. For a civilization that both trans-



Block Diagram of Interstellar Messaging

Sending

Transmission
⇑
Synthesis of Signal
⇑
Synthesis of Message
⇑
Our Conceptions and Senses

Art of Synthesis of new Messages, which are not exist in Nature (Direct Problem)

Interstellar Message

⇒⇒⇒⇒⇒⇒⇒

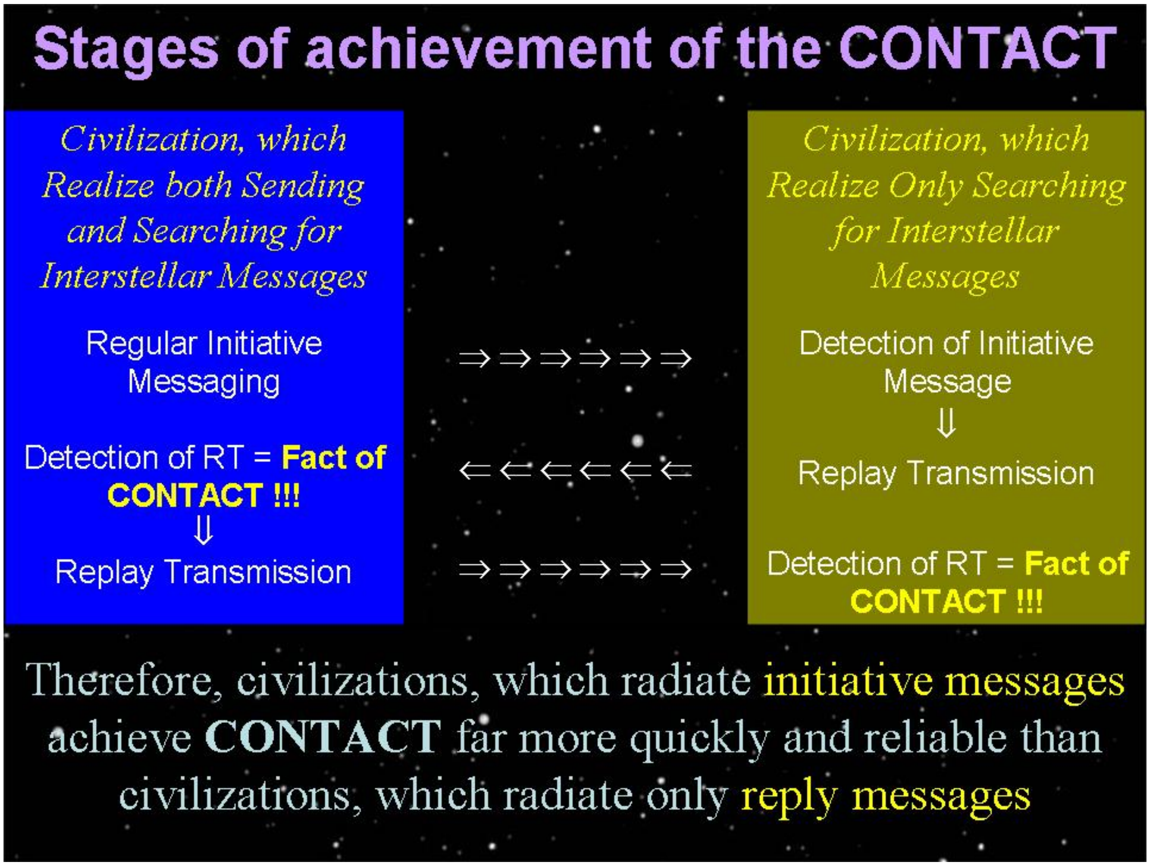
Searching

Receiving
⇓
Analysis of Signals
⇓
Analysis of Contents
⇓
«Theirs» Perception of Our Message

Science of Detection and Perception of Messages, which already exist in Nature (Inverse Problem)

Figure 10.7: Block diagram of sending and searching for intelligent signals in the Universe.

mits and receives, to confirm the fact of establishing contact, it will be sufficient to receive a replay signal from another civilization. Success in searching conducted by a receiving-only civilization requires at least twice the time and effort. Indeed, after signals are detected, it is necessary to send a response, and then to wait for its acknowledgement. Only after such an acknowledgement is received, it will be possible to say that contact has been established (Figure 10.8).



Stages of achievement of the CONTACT

Civilization, which Realize both Sending and Searching for Interstellar Messages

Regular Initiative Messaging

Detection of RT = **Fact of CONTACT !!!**

⇓

Replay Transmission

⇒⇒⇒⇒⇒⇒

⇐⇐⇐⇐⇐⇐

⇒⇒⇒⇒⇒⇒

Civilization, which Realize Only Searching for Interstellar Messages

Detection of Initiative Message

⇓

Replay Transmission

Detection of RT = **Fact of CONTACT !!!**

Therefore, civilizations, which radiate initiative messages achieve **CONTACT** far more quickly and reliable than civilizations, which radiate only reply messages

Figure 10.8: Comparison of two civilizations, one of which is both transmitting and receiving" and one which is only receiving.

10.4 Sending and Searching for Interstellar Messages: Ten Questions

In order to underscore the complexity of a SETI program that entails an extremely large uncertainty and, as a consequence, a necessity to process a large volume of data to find an artificial signal from space, Jill Tarter used the figurative concept of a Cosmic Haystack [13]. She showed the space of unknown parameters to search (SETI search space) to encompass eight dimensions, but we believe two more questions are reasonable and should be added. Thus, we must consider all of the following:

1. Where to search?

2. When to search?

3. At what wavelength?

4. Type of polarization?

5. Power of a receiving signal?

6. How to demodulate a detected signal?

7. How to decode the received information?

8. How to understand the sense of the message?

9. (Why should they send messages?)

10. (Do they consider IRM's transmitting dangerous?)

The first eight questions have been formulated by Tarter, and questions 9 and 10 are suggested by us. We believe that the last two questions should inevitably be posed by aliens who do SETI, assuming that their reasoning is similar to our logic. The resulting list of questions can be applied to solving the inverse problem (which as a matter of fact is a direct problem) – that is, transmitting from the Earth our own signals to the presumably existing extraterrestrial civilizations. In a more general sense, replacement of SETI with METI represents a transition from the "science of search and perception" of something which already exists in Nature but has not yet been known to us, to the "art of synthesis" [14] of information that is not originally present in Nature, and is intended for comprehension by the aliens (about whom we can make only quite a general assumption that they are sufficiently intelligent).

When we consider the problem of METI, we formulate the same type of questions that were considered in SETI, [15]:

1. Where to transmit?

2. When to transmit?

3. At what wavelength?

4. What polarization to use?

5. What should be the energy of the transmitted radio signals?

6. What modulation to apply?

7. What is the optimum structure of the transmitted messages?

8. What content should the message bring to aliens?

9. Why try to transmit interstellar messages?

10. Will METI jeopardize the safety of our own civilization?

In the previous section we discussed the interrelationship between the two programs (sending to vs. searching for intelligent signals in the Universe). We have formulated two sets of questions, each of which referred to only one of the programs. It is not unreasonable to propose that METI and SETI be combined in a single project. From this perspective, the two sets of questions can be unified into a single block of problems to solve [16], namely:

1. Presumable targets for sending and searching.

2. Synchronization of sending and searching.

3. Optimum frequency bands

4. Polarization.

5. Power of transmitting and receiving signals.

6. Type of modulation.

7. Structure and methods of encoding of messages.

8. Content of the messages.

9. Do METI and SETI make any sense?

10. Potential danger of sending and receiving messages.

Next, we present our vision of how to solve these novel and controversial problems. We understand that our approach should be critically analysed and discussed. As a result, new – perhaps more adequate – solutions can be found. Nonetheless, we suggest that our answers may be fairly close to the optimum solution. So, let us explore possible answers to these above ten issues:

10.5 Presumable targets for sending and searching

The identification of celestial coordinates of a presumably existing extraterrestrial civilization is a non-trivial problem. Fortunately, it has become much easier since 1995, when Swiss astronomers Michel Mayor and his graduate Didier Queloz made the remarkable discovery of the first known planet orbiting another sun-like star, 51 Pegasus [17]. The discovery of this first known exoplanet made it clear that, just as stars are ubiquitous in the Universe, so should planets probably exist everywhere. Our Galaxy alone contains about 100 billion stars, with 1% of those stars (or about a billion) being of solar or nearly-solar type. This remarkable figure places an upper limit on the number of stars to which our interstellar radio messages can be sent. Of course, much careful study should be done to select, among this billion, those stars that presumably have planets with intelligent life. These planets are the main targets of our interstellar messaging program.

We do not propose restricting our targets by only the solar-type stars, but they should be our main goal, defined by our present understanding of astrophysics, biophysics, chemistry, etc. We recognize that the problem of identification of the life sites in our Galaxy has not been yet resolved, and that there remain enormous opportunities for further discoveries and research. Our present list of requirements for candidate stars includes the following characteristics:

- the star must be on the Main Sequence;
- it must have relatively constant luminosity;
- its age must be 4-7 billion years;
- spectral class of the star must be close to the solar type;
- position of the star in the sky must be close to some "preferable direction" – ecliptic, remarkable astronomical object, the centre or anticancer of the Galaxy, etc.;
- it is desirable that, as viewed from the target star, our Solar System is also visible in a direction that is close to some remarkable astronomical object, so that aliens might find us in the course of their routine astronomical observations;
- in the case of targets representing known planetary systems, it is desirable that the orbits of these exoplanets have low eccentricity, as such planetary systems are more stable and there is no significant temperature fluctuation preventing the formation of life;

- it is desirable to choose stars located inside the "Life Belt", [18] – the "greenhouse" of our Galaxy – where stars and spiral arms co-rotate, thus making conditions for the origin and long development of a life less hostile.

As our knowledge about the origin of life in the Universe grows, other criteria for identification of possible targets for METI and SETI programs may be recognized.

10.6 Synchronization of sending and searching

The problem of time synchronization between our transmission and an alien civilisation's searches or, in the case of SETI, between an alien's transmission and our searches, is vitally important. Peter Makovetsky estimated [19] that proper synchronization can allow us to increase the probability of establishing of radio contact by a factor of ten. A possible method of establishing this synchronization is to associate the moment of transmission ("over here") and searching ("over there") with some astronomical event which is observable by both parties. Perhaps novae and supernovae explosions are the best candidates for such synchronizing events. Using simple geometrical relationships, Makovetsky has calculated a "schedule" of transmitting/receiving sessions for neighbour stars. One example of such a synchronizing event was a nova explosion in the constellation Cygnus, which was observed on the Earth on August 29, 1975. Using modern, large optical telescopes, it is now possible to register the events of supernovae explosions in neighbouring galaxies. These can also be used for the time synchronization of messaging and searching.

10.7 Optimal frequency band

It seems to us that an ideal frequency band for transmitting IRMs would coincide with that spectral range frequently used for SETI covering, at wavelengths from 20 cm to 1 cm. This is because the propagation range of radio communications in this band covers almost the entire Galaxy. We define the energy potential of a space radio link as the product of the power of transmitter and the combined gains of the transmitting and receiving antennas, divided by the noise temperature of the receiving system. Currently, the state of the art in terrestrial technology is such that this energy level (signal-to-noise ratio) is maximized at wavelengths between about 1 and 10 centimetres. We recognize and accept that, in the course of development of space communication technology, the spectral segment for maximum signal power (hence range) may shift to the infrared or optical wavelengths. If this happens, the optimum wavelength of sending/receiving will of course change as well.

The exact value of the optimum wavelength may even take one of the "magic" values. For example, it is likely that the wavelength 3.36 cm = 21 cm / 2π or its multiple, is to be known to all technological civilizations as the ratio of two universal constants, one physical (the radio emission line of interstellar neutral hydrogen) and the other mathematical (number π) [20].

10.8 What polarization?

Specifically chosen parameters of polarization of the transmitted signal is one of the possible indicators of its artificial origin. In addition, discrete or continuous modulation of the polarization parameters, such as direction of rotation of circular polarization or orientation of the plane of linear polarization, can be used for encoding intelligent messages. By the way, in Carl Sagan's remarkable science-fiction novel Contact, the radio message from Vega indeed had the polarization modulation

10.9 Power of transmitted radio signals

Should we desire to build a transmitter for the purpose of continuous METI transmission, we would have to evaluate its presumed power. This evaluation is not difficult, and can be readily accomplished when required. However, if we are interested in doing METI today, with existing radio dishes and transmitters, then it is more relevant to replace the question about the power of transmitters with another one: the specific energy of the radio emission which is required for sending each bit of information. The answer to this question will determine a detectable data rate for the transmitted information.

The following summary [21] shows the rates of transmission of information in METI experiments conducted with the three most powerful transmitting radio systems currently available on Earth. The numbers in parentheses represent the diameter of the transmitting dish, the average power, and the wavelength, respectively:

1. Radar Telescope in Arecibo, Puerto Rico (300 m; 1000 kW; 12.5 cm) – 1000 bits per second;

2. Solar System Planetary Radar in Goldstone, California (70 m; 480 kW; 3.5 cm) – 550 bits per second;

3. Planetary Radar near Evpatoria, Crimea (70 m; 150 kW; 6.0 cm) – 60 bits per second.

In these calculations, we have conservatively assumed that the distance to a presumed alien recipient is about 70 light years, that the alien receiving antenna has an effective capture area of 1 million square meters, and that the ratio of the effective area to the system's noise temperature equals to 50,000 m2/K. These parameters are similar to those of the Square Kilometer Array (SKA) now under development on Earth, which we expect to be built and commissioned within the next decade [22].

10.10 Type of Modulation

We still know nothing about our message's intended recipients, except that we presume them to be intelligent. Therefore, while trying to synthesize an IRM, we should bear in mind that its recipients will first deal with a physical phenomenon, and only after that perceive the information. At first, their receiving system will detect the radio signal. Only then will the issue of extraction of the received information and comprehension of the obtained message arise. Therefore, above all, the designer of an IRM should be concerned about the ease of signal determination. In other words, the signal should have maximum openness, which is understood here as an antonym of the term security. This branch of signal synthesis can be named anticryptography. A possible variant of such a synthesis is presented below. The variant is based on spectral representation, [7].

During 50 years of nearly continuous searches for intelligent signals from presumed existing extraterrestrial civilizations, the overwhelming number of studies have employed surprisingly similar detection algorithms. It is commonly accepted practice to apply digital spectral analysis with the number of parallel channels reaching from hundreds of millions up to several billions. For example, Project Phoenix at the SETI Institute used a digital spectral analyser consisting of two million channels with a bin width 1 Hz. This allowed scientists to analyse a bandwidth on the order of 2 MHz in a real time, on-line mode, and on the order of 2 GHz of spectrum in post-processing (off-line) mode [23].

If we assume that the optimum receiver has parameters similar to those used in current SETI projects, and that we intend not only to search for radio signals from other civilizations, but also to transmit to ETIs, we will inevitably come to the conclusion that modulation of the transmitted signals should have a distinctive spectral signature, allowing anybody to decipher it with minimum ambiguity using the above-mentioned parallel spectral analysers, [24]. Such a modulation scheme, with a format well known and widely used on Earth, is frequency modulation (FM).

10.11 Structure and methods of encoding of messages

If one agrees that a radio message can be synthesized on the basis of a spectral approach, it is logical to propose the following possible spectral compositions of a message that is based on the temporal behaviour of frequency of the radiated signal:

- the frequency is constant over time;
- the frequency jumps chaotically between several fixed values; or
- the frequency drifts smoothly over time.

Transmitting a constant frequency assumes that the signal is monochromatic. The idea behind radiation of a monochromatic wave with a constant frequency is that such a signal is optimum for detection by the receiver described above, because it can be integrated over long timeframes, maximizing receiver sensitivity. Therefore, a monochromatic signal is a natural choice for radiating at the beginning of a longer message, as it plays the role of a call sign. Besides, such a signal, which contains zero initial information, can still be identified as being of intelligent origin, even if received by aliens having a different type of reasoning and logic which may prevent them from recognizing our more complicated informative messages.

We emphasize that a monochromatic signal contains no semantic information. However, during the journey from the Earth to another civilization, it will be influenced by the interstellar medium and other possible factors, and thus will gradually acquire physical information about the processes going on along its way. Such monochromatic signals are used in space radio science to study planetary atmospheres, solar corona, and interplanetary space. Applied to METI, this method, called radio sounding [25] is extended to study the interstellar medium.

Let us assume that aliens indeed receive our signal which was originally monochromatic, but has now been affected by interaction with the interstellar medium. In this case, they will need to determine that the signal was indeed initially monochromatic. This means that they have to eliminate the distortions of the received signal produced by their atmosphere, or by the propagation path, as well as Doppler drift due to rotation of their planet and orbital motion around their central star. These considerations should be also applied to our search for extraterrestrial signals.

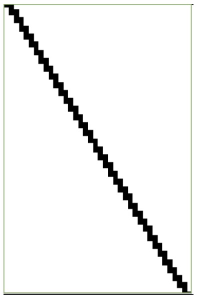
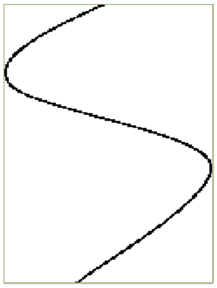
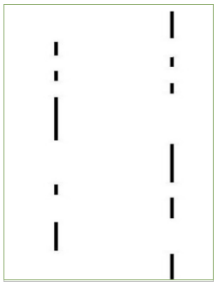
Interestingly, the accuracy of estimation of frequency, and, hence, radial velocity, even for existing radar systems, is very good. For example, let us assume that we emit radio signals from Evpatoria to aliens who are located at the distance of 70 light years from Earth, and who possess an Evpatorian-type radio dish and receiver. In this case, the received signal-to-noise ratio will be 16 dB, assuming a receive signal filter with a bandwidth of 0.1 Hz. The estimated error of the Doppler frequency shift will not be worse than 0.015 Hz, and the accuracy of any resulting measurement of radial velocity will be 0.9 mm/sec. If the aliens have an Arecibo-like antenna, the error of a single measurement in a 10-second interval will decrease to 0.2 mm/sec. Besides the frequency, we can also estimate other possible measured parameters of radio signals, such as polarization, amplitude, and phase variations. We notice also that the interference associated with the presence of the terrestrial ionosphere and interplanetary plasma is significantly lower if one is sending radio signals in the direction opposite to the direction to our Sun...

We propose that the structure of an ideal radio message have three distinct parts, corresponding to the three types of temporal behaviour of frequency: "Constant", "Continuous", and "Discrete." The monochromatic part of the transmission becomes modulated by physical processes that occur in Nature, and thus imparts scientific data. The modulation of the other two parts of the transmission is done by people. I call these different types of the modulation "the Language of Nature", "the Language of Emotions", and "the Language of Logic" respectively. Table 10.3 explains these modulations; the term "Sonogram" designates two-dimensional visualization of the spectral structure of the signal in coordinates X-frequency, and Y-time.

Here we can apply an analogy to the threefold structure of human way of thinking, which is split in intuitive, emotional, and logical components [26]. The first part of the radio message is designed by radio engineers, and represents a coherent electromagnetic wave with monochromatic or periodic linear frequency modulation (LFM). The slow tuning of the message's frequency is required to compensate the variable Doppler shift due to the orbital motion of the Earth with respect to the barycentre of the Solar System (or to the centre of the Galaxy, if we transmit at one of the "magic frequencies"), calculated so that a constant carrier frequency is perceived by our intended recipients. If aliens are sufficiently intelligent and intuitive, they definitely will be able to figure out that they have received a radio message of artificial origin.

We believe that the second part of the message should be created by composers, artists, architects, etc., and represent the analogue variation of the message frequency associated with our emotions and artistic sensibilities. An elementary example of such analogue modulation would be the melodies of classical music compositions. From psychology, we know that human emotions are transitive, i.e., they propagate from one individual to another by various expressive means. Here, we are extending the concept of the transitivity of emotions to interstellar broadcasting.

Table 10.3: Spectral Languages for Messaging to ETI

Parameter	Three types of modulation		
Type	1. Constant	2. Continuous	3. Discrete
Author (Earth site)	Radio Engineer	Artist	Scientist
Language	"Nature"	"Emotion"	"Logic"
Information	Absent	Analogue	Digital
Sonogram of transmitting signals (X axis is horizontal, Y axis is vertical)			
Analyst (Alien)	Astrophysicist	Art Critic	Linguist

The third part of the message consists of discrete frequency shifts, digital data flow showing constituents of our logic (algorithms, theories, etc.) and representing cumulative knowledge about ourselves and the world around us.

In Table 10.3, the row "Analyst" represents our vision of how the message to aliens can be explored by the recipients. The first part of the message is optimized for astrophysical analysis with the purpose of revealing the effects of the interstellar environment, and supporting diagnostics of the propagation channel. The second part of the message is analysed by art critics; the third part by linguists, logicians, and behavioural scientists. As time goes on, the meaning of the radio message gradually penetrates to the alien's mind, and becomes an integral part of the recipient's culture.

10.12 Content of Radio Message

The content of the radio messages (more exactly, their digital parts) that have already been transmitted has a common feature. Specifically, in all five previous IRMs, a binary code has been used, under the implicit assumption that the concept of prime numbers is a universal and known not only to us, but also to extra-terrestrial recipients of our messages. In the Arecibo Message (AM), Cosmic Call 1 (CC-

1), and Teen Age Message (TAM), the transmitted sequence of binary information represented components of a two-dimensional matrix with elements equal to the product of two prime numbers. We imagined that, upon receiving these binary codes, aliens will be able to arrange the numbers in a proper way to convert them back to the original two-dimensional matrix. In the radio message Cosmic Call 2 (CC-2), the generated structure again had this same form, of a two-dimensional matrix representing an image. Each row of the matrix had a length equal to a prime number, with the first and last elements being identical. Frames (columns) in the two-dimensional image were separated from each other by the same symbols in each row. Thus, all these messages assume that the alien recipients are able to perceive two-dimensional information in the form of images, like those used by terrestrial oculists for testing human sight.

Each transmitted IRM also contained an introductory (educational) part. In the AM this part was short and contained only the concept of binary representation of numbers, while the CC-1 and CC-2 included a whole introductory chapter written in a language that is methodologically similar to an artificial language LINCOS, first described in 1960 by Hans Freudenthal [27]. The originality of the radio message TAM consists in the structure of its prologue, which is bilingual and constructed on the basis of a concept of BIG = Bilingual Image Glossary (Russian-English), a dictionary with image recognition.

The bodies of each of the IRMs transmitted to date are unique, essentially different from one another both in terms of the representation of the information and in its volume. Detailed descriptions of these radio messages can be found in corresponding publications [4-8]. We would like to emphasize that, at present, neither has a standard procedure of the synthesis of IRMs been developed, nor is there general agreement on what should be included in the content of such messages.

A widely held opinion begs clarification, that it is necessary to transmit knowledge about ourselves and the world surrounding us. It is highly plausible that the most essential part of our own knowledge is already known to advanced extraterrestrials, and therefore we should be much more selective in choosing the information to be included in our messages. For example, aliens might be interested in exact values of the coordinates and proper motions of the stars available for measurement from our Solar System. Comparing this information with their own measurements, the aliens would learn much more precise distances to stars and their dynamics, due to a newly found ability to perform parallax measurements with the baseline of the radio link reaching distances of tens to hundreds of light years. Others suggest that we send information about terrestrial social life and culture, because it is very unlikely that the alien civilization has the same principles of social organization and art.

We recall here the opinion of Academician Vladimir Vernadsky: "I believe that for deeper understanding of the world, music and the feelings experienced by people in the process of creative work are most essential" [The Diary, 1932]. We share this opinion, and believe that broadcasting our music and our art will help Them to achieve a really deeper understanding of the world.

10.13 Why should we transmit interstellar radio messages?

Accepting for now the point of view that the goal of the search for intelligent signals from space is intuitively clear, let us try to answer the question on whether METI makes any sense. Here, we walk on the shaky ground of fuzzy and insufficiently precise reasoning and assumptions. Straightforward justification of the necessity and practicality of METI is impossible, at least, for now. Emotional and ethical reasons like "we bring to aliens the long-awaited news that they are not alone in the Universe" is not scientifically justified, and can convince only a few people. For this reason, there are voices saying that METI makes no sense. Such METI critics should understand a simple thing: if all civilizations in the Universe are only recipients, and none are message-sending civilizations, then SETI searches make no sense either. We emphasize that all terrestrial programs that search for intelligent signals in the Universe start with the implicit assumption that aliens exist, and that some of them send interstellar radio messages. Accepting this assumption, we see that METI programs stand on exactly the same ground as SETI, and should not cause doubts.

In 2006, I published the paper "The SETI Paradox" [28], which sought an answer to the question as to whether METI makes sense. There, we analysed the terrestrial situation of the paradoxical co-existence of two opposite tendencies: a persevering aspiration to searches for intelligent signals from other civilizations, and a strong aversion to any attempt of sending similar signals from the Earth to presumably existing extraterrestrials. If we accept that such situation is typical for any civilization in our Universe, then SETI would make no sense at all, [29].

The paper was extensively discussed in blogs [30, 31] where more than 90 comments were posted, and in the SETI League's site, [32]. If we, ourselves, do not have a need to pass over information to extraterrestrials, how is it possible to justify that such need is experienced by them? If they have no such need and do not send radio messages to other civilization, what can we expect to find with SETI? The answer is clear: nothing. Discussion of the SETI Paradox leads us to an inevitable conclusion: either we do both METI and SETI, or we do nothing. Later, an anonymous author of a Wikipedia article on SETI proposed a slightly different version: "SETI's Paradox refers to an apparent 'paradox' where two distant civilizations capable of interstellar communication will always remain silent unless one of them contacts the other first, resulting in a deadlock of silence."

At present, one can judge the existence of intellect in our Universe based on only one case: our own terrestrial civilization. We are interested in estimating the likelihood of a transfer of our information to other civilizations. For a numerical evaluation of this likelihood, and how it affects the estimate of the number of communicative civilizations in our Galaxy, we suggest using the Drake equation with an additional parameter, the so called "METI-factor" fm. After taking into account this factor, Drake's classical formula now assumes the following form:

$$N = R_* \times f_p \times n_e \times f_l \times f_i \times f_c \times f_m \times L \tag{10.1}$$

where f_m is the fraction of the communicative civilizations indeed conducting systematic transmission of purposeful interstellar messages, [33].

We note that to be in a communicative state of the development, and to actually emit METI messages, are not the same thing. For example, we terrestrials have apparently reached the communicative state, but can not yet consider ourselves a communicative civilization, because we do not practice such activities as a purposeful and systematic transmission of interstellar messages.

We may try to estimate the METI-coefficient f_m for the only known, terrestrial, civilization. As we pointed out above, our civilization is in the communicative phase and does conduct SETI activities. However, our METI/SETI ratio is less than one percent: these data follow from the review of Jill Tarter published in the recently released "SETI-2020" collection of papers [34]. It lists 100 various SETI programs starting from the first Ozma project and going until the present time. The total time of the search for extraterrestrials is several years, whereas the total transmission time is only 41 hours. This characterizes the present attitude of researchers. However, we must also take into account the effect of the general reluctance to support METI activities. Thus, if we make an estimate of the f_m coefficient based on the only known civilization, we find that it is fairly close to zero and, consequently, the same should be true for the number of potentially detectable extraterrestrial civilizations, as we do not expect the presumed aliens to be significantly more likely to transmit than are we ourselves.

Hence, we can formulate **the SETI Paradox** also in this form: "The search for intelligent life is meaningless if no one feels the need to transmit..." In other words: "SETI makes sense only in a Universe that creates an Intellect which realizes the need, not only to search for another Intellect, but also for transmitting intelligent signals.

It should become possible to establish contact, only if one of the distinguishing features of the Intellect in our Universe is a mission to carry out to aliens the good news that they are not alone in space. Given such enormous distances and, consequently, long signal propagation time, communications can be expected to be mostly one-way – our addressees will receive our messages, and we, in turn, will detect messages from those who had chosen us as their addressees. This is how the Universe, at a certain stage of its development, allows observers to discover its habitability. Unless this process is triggered and is ongoing, intelligent life in different parts of the Universe will remain lonely, isolated, and inclined to extinction.

10.14 Is it dangerous to receive and transmit interstellar messages?

A comparison of the total number of transmissions generated by conventional radar astronomy, to those having been sent to extraterrestrial civilizations, reveals that the probability of detection of radio signals deliberately sent to extraterrestrials (ETs) is about one million times smaller than that of the radar signals used to study planets and asteroids in the Solar System.

There are three large-dish instruments in the world that are currently employed for doing radar investigations of planets, asteroids and comets [35]: ART (Arecibo Radar Telescope), GSSR (Goldstone Solar System Radar), and EPR (Evpatoria Planetary Radar). The radiating power and directional coverage of these instruments is so outstanding that it allows us to emit radio messages to extraterrestrials, which are detectable practically everywhere in the Milky Way.

Recently, some scientists and scientific-fiction authors have expressed concerns [36] that sending messages to those stars in our Galaxy which may have a habitable life, jeopardizes the very existence of our own civilization. There is a fear that our transmitted signals might help ETIs to pin down the location of our Solar System in the Milky Way. If the aliens reached the level of a super-civilization, some argue, they might send a space fleet to the Earth to either destroy us or to convert us to slaves.

The goal of this section is to estimate the probability of detection of terrestrial radio signals by a presumably hostile super-civilization existing somewhere in our Galaxy. Our calculation starts by noting that, over all of our radar astronomy history, about 1400 sets of radio transmissions were produced. Their distribution all over the sky is shown in Figure 10.9 in the ecliptic plane [37].

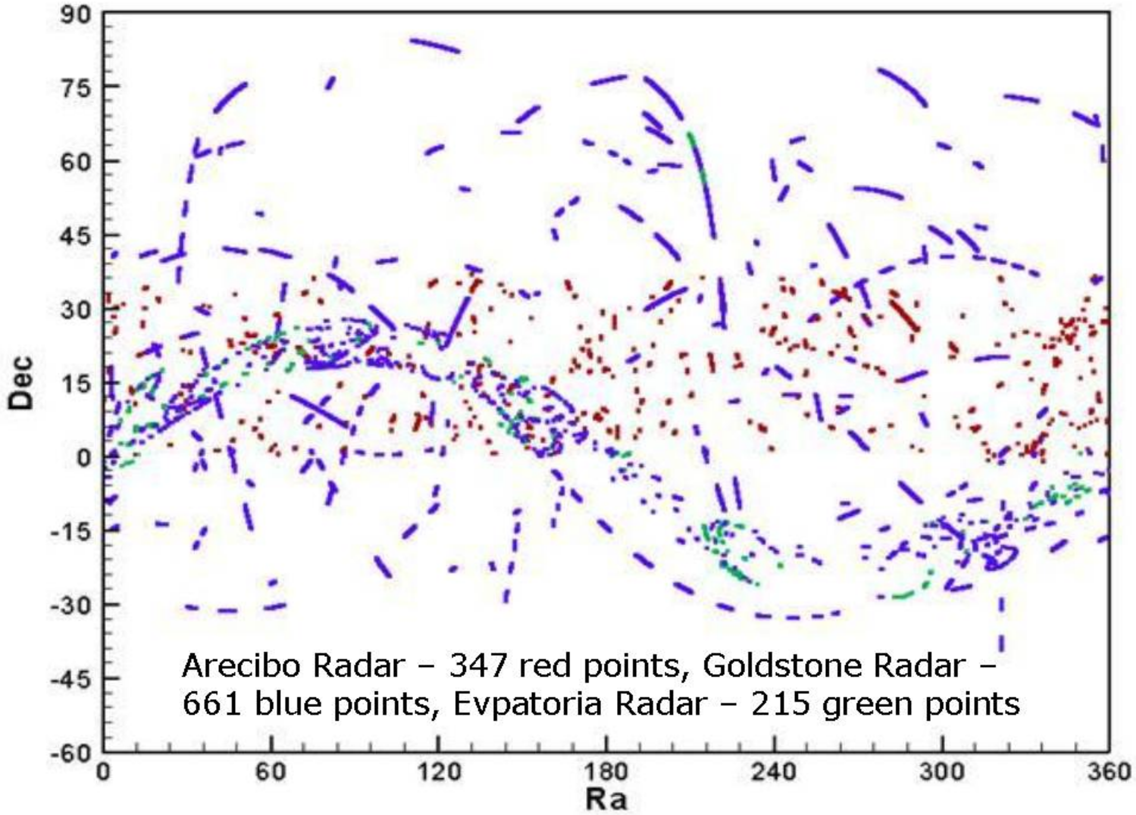


Figure 10.9: Illumination of the sky by the overall radiation emitted during radar observations of celestial bodies.

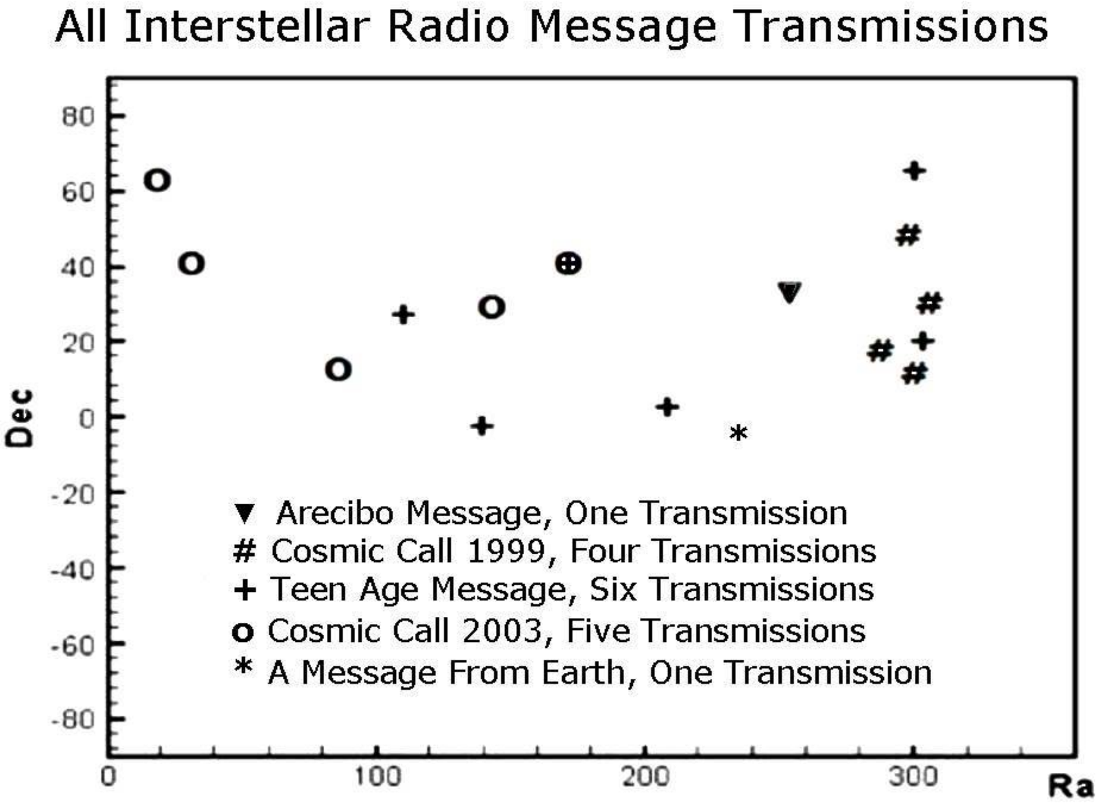


Figure 10.10: 17 sessions of radiation of interstellar radio messages.

The total area of the sky illuminated by these transmissions is about 0.022 steradians (sr), or about 2×10^{-3} (two parts in a thousand) of the whole sky. The total number of METI transmissions to date is only 17 sets, and the total area of sky, illuminated by the METI transmissions, is about 10^{-5} sr, or 2000 times less than that covered by radar astronomy transmissions (see Figure 10.10).

The total duration of our combined planetary radar transmissions exceeds the overall time interval of the METI transmissions by a factor of 450. Therefore, we can conclude [38] that the probability of detecting the radar astronomy transmissions by a hostile super-civilization is $(2000 \times 450) \sim$ a million times higher than that of the METI transmissions!

So, if someone is concerned about the chances of our possible detection by an aggressive and paranoid super-civilization so-called METI-phobia, [38], he or she would have to prohibit, first of all, not METI, but rather radar astronomy. However, nobody is going to ban radar astronomy, it is an important and indispensable component of both our asteroid hazard detection programs (planetary defense) and Earth's various national security defense systems, [39]. For this reason, we conclude that all the on-going conversations about the dangers posed to our civilization by METI activity are meaningless, and that the radar astronomy instruments in Arecibo, Goldstone, and Evpatoria should remain open for further exploration of interstellar space and our galaxy through METI transmissions.

Regarding the sources of the METI fear firstly mentioned in England by radio astronomer Martin Ryle [3], if such hostile super-civilizations really exist, then our civilization is already doomed to extinction or slavery. Such mighty and ruthless super-civilizations inevitably will find us because the anomalously high percentage of molecular oxygen contained in terrestrial atmosphere definitely indicates the presence on Earth of some organic matter. After having detected the indirect signs of life, the aliens will establish a program for continuous monitoring of our planet, in order to detect the activities associated with intelligent life. No doubt, they will eventually find this activity, which includes the isotropic radiation of the broadcasting radio stations, TV centres, and the anisotropic emission of our radar astronomy transmitters.

At the same time, the probability of detection of our civilization by aggressive super-civilizations through our METI activity is almost negligible. Therefore, all conversations regarding the danger that METI imposes through the transmission of interstellar radio messages are rather superficial, specious, emotional, and entirely non-scientific. METI-phobia is nothing more than a consequence of paranoiac self-agitation based on fantasy, superstition, and a prejudice. There are always people who are illogical, who can not comprehend scientific arguments, and who rely not on knowledge but rather on pseudo-science. Those who perceive that METI puts Earth in dire jeopardy are no different from those who anticipated the end of the world following the activation of the Large Hadron Collider.

It appears that some terrestrials express similar fears regarding perceived dangers of a SETI listening program, [40]. They consider that any messages received by us through SETI searches are dangerous also, as they may contain super-refined computer viruses, or any unknown, extremely reactionary or extremist doctrine which can destroy us, either individually, or societally. I consider such fears as extreme, and no valid reason to abandon either SETI or METI science.

10.15 Conclusion

Finally, let us return to the original reference, and give a classic quotation from the seminal SETI paper by Cocconi and Morrison: "The probability of success is difficult to estimate, but if we never search the chance of success is zero".

The above argument is certainly true. However, the incidental detection of extraterrestrials as a result of routine astronomical observations is also possible. This may happen if and only if there exist extraterrestrial civilizations that actually send interstellar messages. Therefore, in regard to METI, the Cocconi-Morrison statement can be reformulated as follows: "The probability of success is difficult to estimate, but if nobody transmits the chance of success is zero in principle".

So, we can formulate the following two versions of the thesis, implied by the SETI Paradox: "Only the Universe, which spawns a sociable type of Intellect, acquires, with the lapse of time, its own Voice" and "Only those who are trying to overcome the Great Silence, will hear the Voice of the Universe"...

10.16 Acknowledgements

The author is sincerely grateful to numerous colleagues from Russia, U.S.A., Ukraine, and Western Europe together with whom he worked on the development and transmission of the interstellar radio messages. My Russian colleagues, who now work in the USA Sergei Kopeikin (University of Missouri-Columbia) and Ludmilla Kolokolova (University of Maryland), provided a great improvement to the text by editing my initial English translation. I highly appreciate their help. Comments to the text by Sergey Kopeikin and Stephane Dumas (Canada) also were a valuable source for the improvement of this chapter. Many thanks to my numerous colleagues from Russia, U.S.A., Canada, and Western Europe, electronic correspondence with whom and pleasure of a direct dialogue at seminars and conferences undoubtedly have contributed to the creation of this chapter.

References

1. Pioneer Plaque; `http://grin.hq.nasa.gov/ABSTRACTS/GPN-2000-001623.html`

2. Voyager Record; `http://voyager.jpl.nasa.gov/spacecraft/goldenrec.html`

3. Murmurs of Earth, by Carl Sagan, Frank Drake, Ann Druyan, Timothy Ferris, Jon Lomberg, Linda Salzman Sagan. Random House, 1978.

4. Arecibo Message; `http://en.wikipedia.org/wiki/Arecibo_message`

5. Alexander L. Zaitsev and Sergey P. Ignatov. Report on Cosmic Call 1999; `http://www.cplire.ru/html/ra&sr/irm/report-1999.html`

6. Л. М. Гиндилис, С. Е. Гурьянов, А. Л. Зайцев, С. П. Игнатов, Е. В. Казаков, Н. Т. Петрович, Б. Г. Пшеничнер, И. А. Феодулова, Л. Н. Филиппова, С. П. Яценко. Сигнал отправлен: 1-е Детское радиопослание внеземным цивилизациям. Вестник SETI, № 3/20, НС РАН "Астрономия", М., 2002, `http://lnfm1.sai.msu.ru/SETI/koi/bulletin/20/articles/1.html`

7. Alexander L. Zaitsev. The First Musical Interstellar Radio Message. Journal of Communications Technology and Electronics, vol. 53, No 9, pp. 1107-1113, `http://springerlink.com/content/m62151781m500p16/?p=a9f198a3a40a488dbe3d3e84bbfbbda`
   `pi=11`

8. Richard Braastad and Alexander Zaitsev. Synthesis and Transmission of Cosmic Call 2003 Interstellar Radio Message; `http://www.cplire.ru/html/ra&sr/irm/CosmicCall-2003/index.html`

9. A Message From Earth, `http://en.wikipedia.org/wiki/A_Message_From_Earth`

10. Alexander L. Zaitsev. Project METI@home: Messages to ETI from home, `http://www.cplire.ru/html/ra&sr/irm/METI@home.html`

11. NASA Beatles Transmission, `http://www.nasa.gov/home/hqnews/2008/jan/HQ_08032_NASA_Beatles.html`

12. Hello From Earth, `http://en.wikipedia.org/wiki/HELLO_FROM_EARTH`

13. Jill C. Tarter. "The Cosmic Haystack and Recent U.S. SETI Programs."Presented at SETI - 81 Symposium in Tallinn, Estonia, December 1981, published in Problema Poiska Jeeznee vo Vselennoi (Russian) (1986).

14. Alexander Zaitsev and Richard Braastad. METI Art, `http://www.cplire.ru/html/ra&sr/irm/METI_Art.html`

15. Alexander L. Zaitsev. Transforming SETI to METI, `http://www.cplire.ru/html/ra&sr/irm/metitran.html`

16. Alexander L. Zaitsev. Sending and Searching for Interstellar Messages. Acta Astronautica, vol. 63, Issues 5-6, Sept 2008, pp. 614-617, http://arxiv.org/abs/0711.2368

17. Michel Mayor and Didier Queloz. A Jupiter-mass companion to a solar-type star. Nature 378, 355-359, 23 November 1995; http://www.nature.com/nature/journal/v378/n6555/abs/378355a0.html

18. Л. С. Марочник и Л. М. Мухин. Галактический "пояс жизни". В сборнике "Проблема поиска жизни во Вселенной", М.: Наука, 1986, стр. 41-46.

19. П. В. Маковецкий. Новая Лебедя – синхросигнал для внеземных цивилизаций. АЖ, 1977, т. 54, No 2, с. 449-451.

20. П. В. Маковецкий. О структуре позывных внеземных цивилизаций. АЖ, 1976, т. 53, No 1, стр. 222-224.

21. Alexander L. Zaitsev. Limitations on Volume of Interstellar Radio Messages; http://www.cplire.ru/html/ra&sr/irm/limitations.html

22. SKA – Square Kilometer Array; http://www.skatelescope.org

23. Project Phoenix General Overview; http://www.seti.org/Page.aspx?pid=583

24. А. Л. Зайцев. Язык радиопосланий к другим цивилизациям. Доклад на конференции "Джордано Бруно и современность", февраль 2000, ГАИШ МГУ. Вестник SETI, No 2/19, 2002, стр. 73-82; http://lnfm1.sai.msu.ru/SETI/koi/bulletin/19/articles/1.html

25. R. A. Phinney and D. L. Anderson. On the Radio Occultation Method for Studing Planetary Atmoshheres. J. Gephys. Res., v. 73, 1968, pp. 1819-1927.

26. Г. М. Идлис. В поисках истины. М. Издательство "Агар", 2004.

27. Lincos, http://en.wikipedia.org/wiki/Lincos_(language)

28. Alexander L. Zaitsev. The SETI Paradox. http://arxiv.org/abs/physics/0611283

29. А. Л. Зайцев. Парадокс SETI. Бюллетень САО, т. 60-61, стр.. 226-229; http://fire.relarn.ru/126/paradox.htm

30. SETI's Paradox and the Great Silence; http://www.centauri-dreams.org/?p=928

31. Overflow Thread: SETI's Paradox; http://www.centauri-dreams.org/?p=933

32. Paul Gilster. SETI's Paradox and the Great Silence; http://www.setileague.org/editor/silence.htm

33. Alexander Zaitsev. The Drake Equation: Adding a METI Factor, http://www.cplire.ru/html/ra&sr/irm/Drake_equation.html

34. "SETI 2020: A Roadmap for the Search for Extraterrestrial Intelligence". Eds.: Ekers R. D., Billingham J., Cullers D. K., Schefer L. K., Zajdel T. T. SETI Press, 2003.

35. Radar Astronomy, http://en.wikipedia.org/wiki/Radar_astronomy

36. The San Marino Scale, http://avsport.org/IAA/smiscale.htm

37. Д. А. Чураков. Анализ работы планетных радаров применительно к SETI и METI. Журнал радиоэлектроники, No 3, 2009, http://jre.cplire.ru/jre/mar09/index.html

38. Alexander L. Zaitsev. Detection Probability of Terrestrial Radio Signals by a Hostile Super-civilization; http://arxiv.org/abs/0804.2754

39. Asteroid and Comet Impact Hazard, http://impact.arc.nasa.gov/index.cfm

40. Richard A. Carrigan, Jr. Do Potential SETI Signals Need To Be Decontaminated?, Acta Astronautica, vol. 58, Issue 2, Jan 2006, pp 112-117.

Chapter 11

Principal component analysis and applications

by **Stephane Dumas**
The SETI League, inc.
jgsdumas@gmail.com

Abstract

Principal Component Analysis (PCA) is a powerfull tool of factorial analysis. It can be used to classified objects, to reduce the size of a database and extract information from a noisy signal. This paper will briefly introduce the PCA and present applications related to astrobiology and SETI. It will also discuss how to use the algorithm Lanczos to compute eigenvalues of huge matrices (N=1,000,000).

11.1 Introduction

Principal Component Analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations (i.e. variables) into its eigenspace. It is a way of identifying patterns in data or reduce the dimension of the original data set.

The core of the technique is to find the principal components (i.e. eigenvectors) of the data set. This set is made of vectors composed of variables independent of the other.

Given a problem is N vectors X of M variables, then the first step of the process is to compute the covariance matrix **A** of the vectors. The covariance of two vector X and Y is given by equation 11.1 where X_i is the i-th element of X.

$$cov(X,Y) = \frac{\sum_{i=1}^{M}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{11.1}$$

Each element of **A** (e.g. $A(i,j)$) would then be value of covariance of the i-th and j-th vector X. An $N \times N$ matrix **A** is said to have an eigenvector $\vec{x}$ and corresponding eigenvalues λ if equation 11.2 is satisfied and which hold if and only if equation 11.3 is satisfied also.

$$\mathbf{A} \cdot \vec{x} = \lambda \vec{x} \tag{11.2}$$

$$\det |\mathbf{A} - \lambda \mathbf{I}| = 0 \tag{11.3}$$

Equation 11.3 provides a series of N equations of N unknowns (i.e. the eigenvalues) to be solved. Several techniques are known to deal with this problem (some will be described in this paper).

There is no direct meaning of the eigenvectors, nor eigenvalues, of a data set. There are Mathemetical objects used to described the information contained in the data.

This paper will provide examples of usage of PCA in the field of astrobiology and SETI. PCA is also names the discrete *Karhunen-Loeve transform* (KLT), the *Hotelling transform* or *Singular Decomposition Values* (SVD).

11.2 Data compression

One of the first application of the PCA is to reduce the dimension of the original data set. Once the original vectors have been transfer to the eigenspace, they can be represented with less dimensions. The number of dimensions required to represent the original information depend only on the magnitude of their respective eigenvalues. By plotting them from the largest to the smallest, one can select the q largest needed to describe the original information. Figure 11.1 shows a typical representation of the first largest eigenvalues. Most of the time, the first eigenvalues (λ_1) is very large and the amplitude drops very fast. Only a few λ could be used to represent the original data.

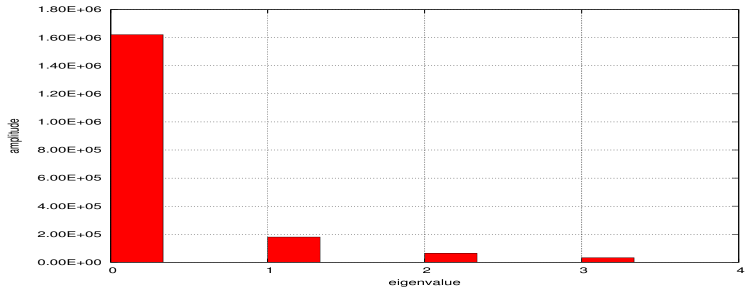


Figure 11.1: Eigenvalues plotted by their amplitude.

Figure 11.2 provides a example of data compression. In this case, a sinusoidal curve can be represented by only 3 eigenvalues. The original data can be approximated by summing the first three eigenvectors into a single vector. Each eigenvectors can be built by using their corresponding eigenvalues.

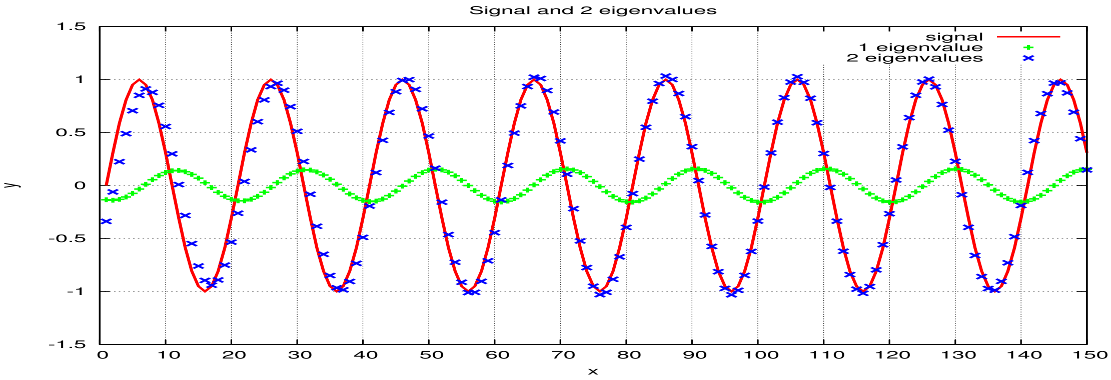


Figure 11.2: Illustration of data compression of $\sin(x)$. The whole curve can be approximated using only 2 eigenvalues.

Figure 11.3 illustrates a more complex scenario where the original data set is a Brownian motion. The data being more complex than a single sin, more eigenvectors are required to approximated it.

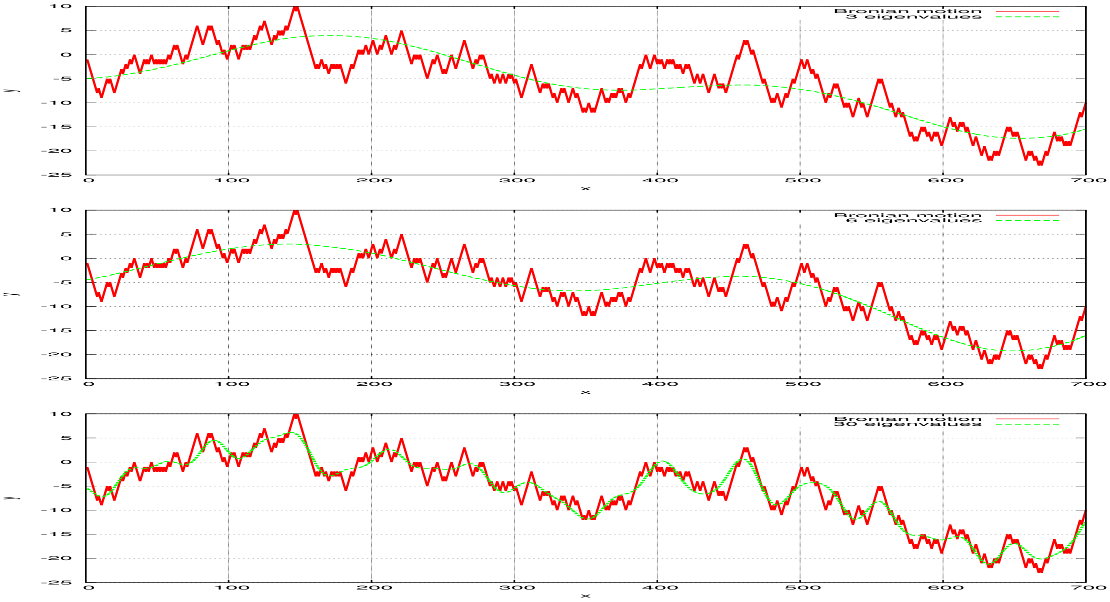


Figure 11.3: Illustration of data compression of a Brownian motion. The top figure shows how 3 eigenvalues can roughly approximate the data. The middle figure used 6 eigenvalues and the bottom 30.

11.3 Classification of spectra

Singular Value Decomposition (SVD) is a variant of PCA that can be used to compare unknown data set to known data and classify them. SVD is summarised by equation 11.4 where **D** the matrix containing the original data set and **U** is the matrix containing the data set represented in the eigenspace. **W** is a square matrix with the eigenvalues on its diagonal and **V** is matrix containing the eigenvectors. [1] provides a good algorithm of this process.

$$\mathbf{D} = \mathbf{U}\mathbf{W}\mathbf{V}^T \tag{11.4}$$

Given a group of infrared spectra taken on rock samples containing life form (i.e. endolithes) and mineral. The problem is to classify the spectra and isolate the biological signatures vs. the mineral ones.

The data set is composed of N spectra of M wavelength (taken in the IR range of the light spectrum). All the spectra are joined together in a single matrix **D** of M columns by N rows. The mean value of each wavelength is computed and subtracted

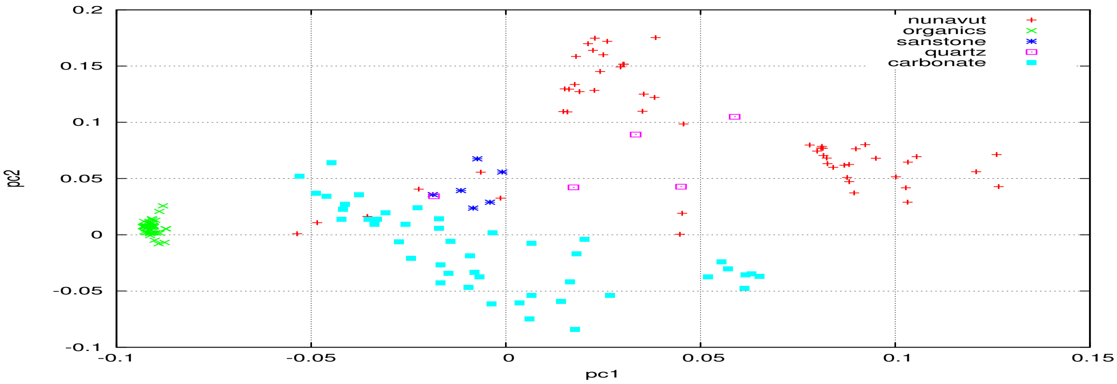


Figure 11.4: Spectral classification of biomarkers and other samples using SVD. The plot shows how the corresponding spectra could be represented with only 2 eigenvalues (or Principal Component, pc1 and pc2). The grouping indicated the main composition of each unknown spectra (i.e. Nunavut) relative to known ones.

from the spectra. The analysis is performed on the variation, from a mean value, of each wavelength. The SVD algorithm computes the covariance matrix of **D** and produce the eigenvalues and eigenvectors from it.

The resulting matrix **U** contains a series of N vectors of M variables (which are not wavelength, but rather projections on the eigenvectors). By virtue of the data compression nature of PCA, it is possible to shorten the length of the new vector by considering only the most important eigenvalues. Instead of N vectors of M elements, the new data set has N vectors of size P ($P < M$).

Clustering techniques can be used on the reduced data set to group similar vectors. By adding spectra, of known mineral and biological signatures, to the matrix **D**, the result of SVD and clustering can be used to classify the unknown spectra and identify them. This technique was used in [2] to classify spectral signatures of rocks and identify possible biomarkers.

It is possible to transpose spectrum to the eigenspace without solving the whole system. The matrices WV^T are similar to a rotation matrix and can be used to transfer from one space to other once they have be properly computed with known spectra.

11.4 Signal analysis

Signal analysis refers to the analysis of a time dependent signal, defined by $X(t)$. Any signal can be represented using a infinite summation of sin functions with increasing frequencies (equ. 11.5). Among those frequencies, some will give information on the signal and the other will be noise. The classical approach is to work in the frequency domain (e.g. Fourier), remove the noise through filtering then transfer back to the time domain and obtain a signal with less noise.

$$X(t) = \int_{-\infty}^{\infty} H(f) e^{2\pi i f t} df \tag{11.5}$$

KLT (Karhunen-Loeve transform) can be used to do a similar analysis (e.g. removing noise). The core of the process is to compute the eigenvectors of the autocorrelation matrix of $X(t)$, defined by equation 11.6

$$R(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \tag{11.6}$$

The eigenvalues and eigenvectors are calculated from $\mathbf{R}$ using an algorithm such as *Jacobi* (or any other similar algorithm). Projection of the $X(t)$ into the eigenspace is computed using equ. 11.7 where $\Phi_n(t)$ are the eigenvectors and T is the number of element in $X(t)$ (i.e. the KL Expansion). $X(t)$ can than be reconstructed using the Z_n vectors by simple adding enough (equ. 11.8, where K is the number of eigenvectors to sum) of them until the approximation is good enough.

$$Z_n = \int_{1}^{T} X(t) \Phi_n(t) dt \tag{11.7}$$

$$X(t) = \sum_{i=1}^{K} Z_i \Phi_i(t) \tag{11.8}$$

11.4.1 Comparison between FFT and KLT

FFT, or Fast Fourier Transform, is the numerical (and discrete) implementation of the Fourier Transform. There are several implementation of this algorithm some slow and some very fast. While FFT is faster than KLT, KLT performs better when the SNR[1] is very low.

The first eigenvector of the autocorrelation matrix of $X(t)$ is sufficient in order to extract the main frequencies of a signal using KLT. Those frequencies can be found by performing a FFT on the eigenvector vector λ_1.

Figure 11.5 illustrates a case with a noisy sinus wave (i.e. SNR=-23dB). The Fourier Transform cannot isolate the main frequency and display a lot of different

[1] Signal to Noise ratio : defined, in dB, as $SNR = 10log(A/A_0)$ for amplitudes.

frequencies. While KLT did not show only the main frequency, the strongest peak is the main one.

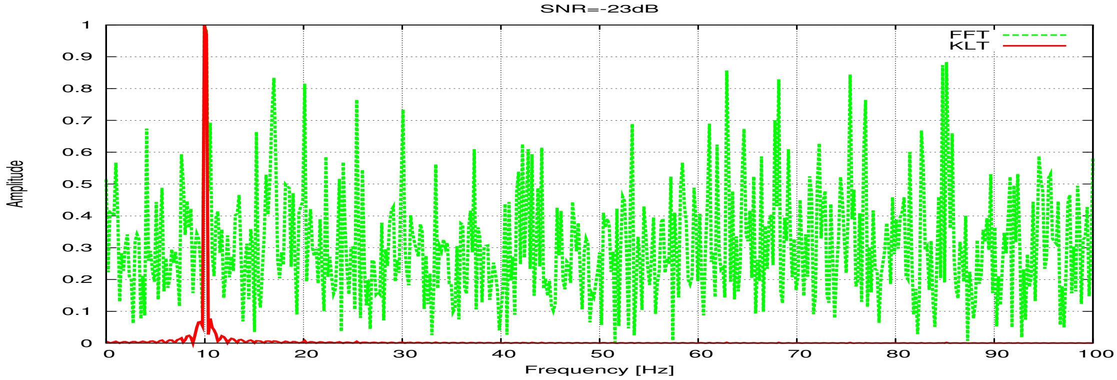


Figure 11.5: Comparison between FFT and KLT with a very noisy signal (SNR=-23dB). The main frequency of the signal is located at x=10 on the frequency unit axis.

11.4.2 Brownian motion

When the times series, $X(t)$, is a Brownian motion (or Wienner process), then the autocorrelation matrix $\mathbf{R}$ is defined by equation 11.9. [3] explains this special case.

$$R(t_1, t_2) = \alpha \min(t_1, t_2) \tag{11.9}$$

11.5 Lanczos algorithm

Solving N equations with N variables can be very computer intentive when N is great. There are several computer algorithms to perform this task. Most are based on variants of *Jacobi*, *Householder* and *LU Decomposition* algorithms. They all requires the content of the matrix $\mathbf{A}$ (equations 11.2 and 11.3) to be completely stored in memory and fully accessible.

The Lanczos algorithm [4] is a technique that can be used to solve certain large, sparse, symmetric eigenproblems. The technique is very fast and appropriate to large matrices. It involves partial tridiagonalisation of a given matrix $\mathbf{A}$. However, unlike the previous techniques (i.e. Jacobi, Householder and LU Decomposition), no intermediate, full submatrices are generated.

The author wrote a computer program implementing this algorithm, that includes some High Performance Calculation mathematical libraries (i.e. BLAS[2]). Modifications were performed to support the autocorrelation matrix of a Brownian Motion. Similar changes can be performed to support autocorrelation of normal time series.

These modifications and improvements could not be done on programs based on the other algorithms.

Figure 11.6 shows the results of the algorithm when applied to a data set of 1 millions points of a Brownian motion (involving an autocorrelation matrix of 1M x 1M). The calculations took 309 minutes on a simple workstation (e.g. quad core CPU and 8 Gb of RAM).

A 1,000,000 × 1,000,000 matrix requires 7.62×10^6 Mb of memory on a computer. This is far more than any standard computer could handle. The modification performed to the algorithm to process huge matrices is to avoid storing them completely in memory and compute only the part needed (e.g. $R(t_1, t_2)$). This is the strength of modified Lanczos algorithm which does not require the full matrix. This process slows down the calculation but permit to considere matrix too big to be processed.

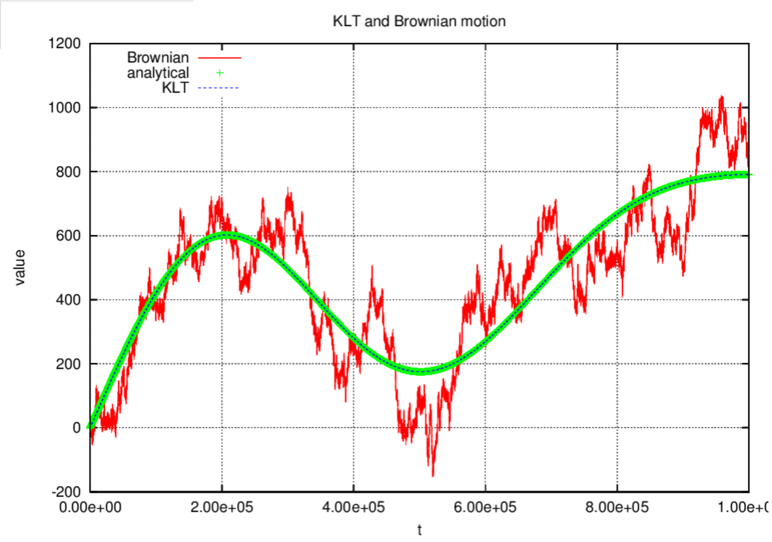


Figure 11.6: KTL applied to a Brownian motion data set of 1 millions points.

[2] http://www.netlib.org/blas/

11.6 BAM

The Bordered Autocorrelation Method (BAM) has its mathematical foundation in the Final Variance theorem described in [5] and summarised by equation 11.10. It stated that the first-order partial derivative of all the λ_n with respect to T for stationnary processes are just constant. The Modified Lanczos algorithm was used to compute the values of eigenvalues up to N=6000 and provide a validation for the BAM. Figure 11.7 illustrates this result. However, figure 11.8 shows that the linear relation cannot be used to find higher order.

$$\lambda_n(T) \propto T \tag{11.10}$$

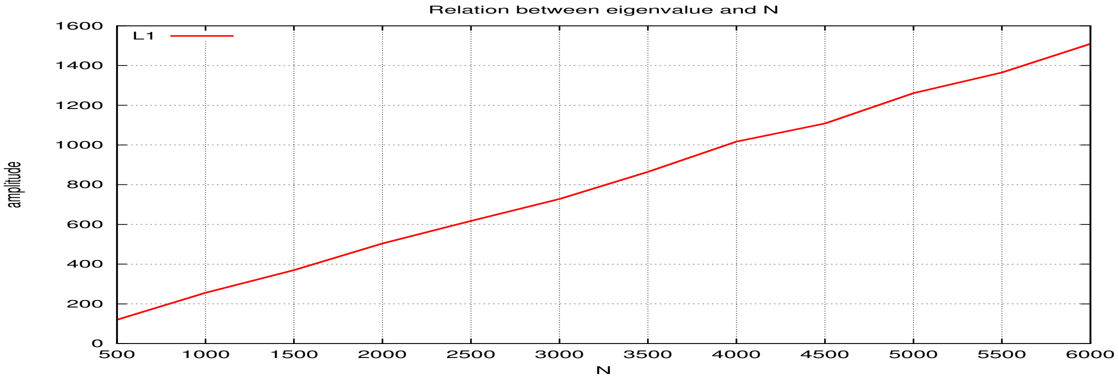


Figure 11.7: Plot of λ_1 vs. N. The relation is linear as the theory predicted.

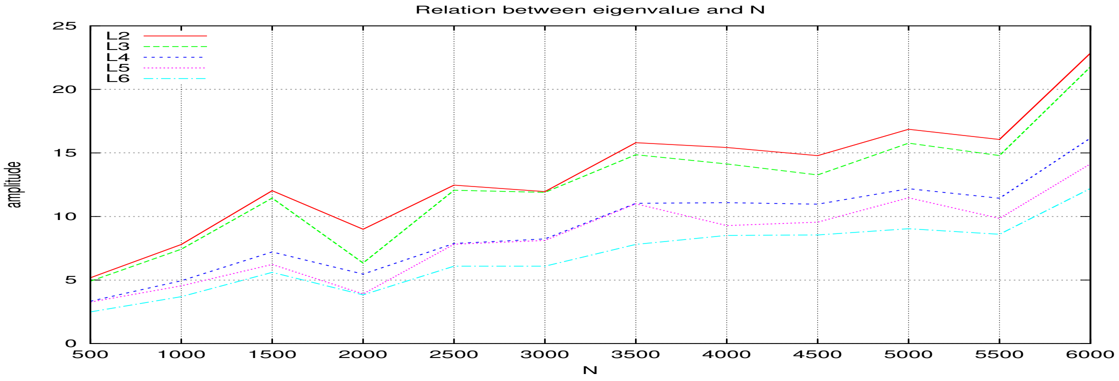


Figure 11.8: Plot of λ_2 to λ_6 vs. N. There is no clear relation with N.

The BAM theory gives a quick way to find the value of λ_1 when the problem is related to finding the eigenvalues of the autocorrelation of matrix $\mathbf{R}$ for a Brownian motion.

11.7 Using PCA with data set larger than the rank of R

The application of PCA to time series (i.e. $X(t)$) is only linked to the autocorrelation matrix $\mathbf{R}$. Once the eigenvectors are found, a difference series (i.e. $Y(t)$) can be process as long as it has the same length. Figure 11.9 illustrates this by applying the same PCA results of $\mathbf{R}$ to different vector $X(t)$. Here the algorithm does not try to reproduce the signal but to pass a single $\sin(x)$ through the points.

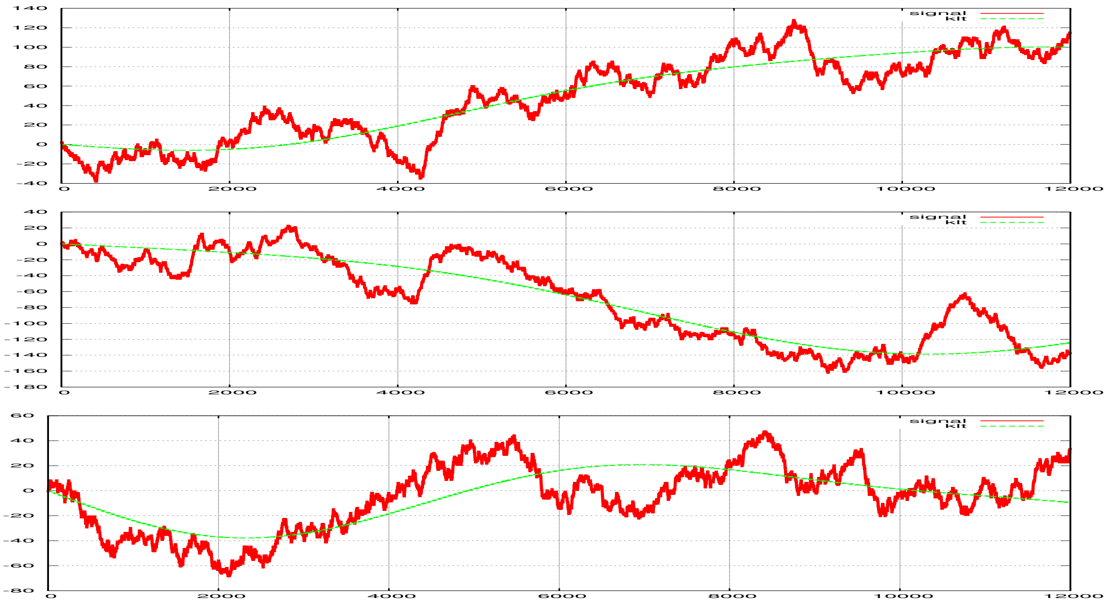


Figure 11.9: The same eigenvectors solution applies to different Brownian Motion.

This idea can be pushed further by solving the PCA problem for a matrix $\mathbf{R}$ of rank N and applying it to a vector $X(t)$ with M elements where $M > N$. Figure 11.10 illustrates the results of the algorithm. The data set $X(t)$ is a Brownian motion of 3,600 points (in red). PCA is solved using an autocorrelation matrix $\mathbf{R}$ of rank N=200. The green curve (i.e. PCA) follows well the path described by $X(t)$. This example also illustrate the capacity of PCA to reproduce the curve.

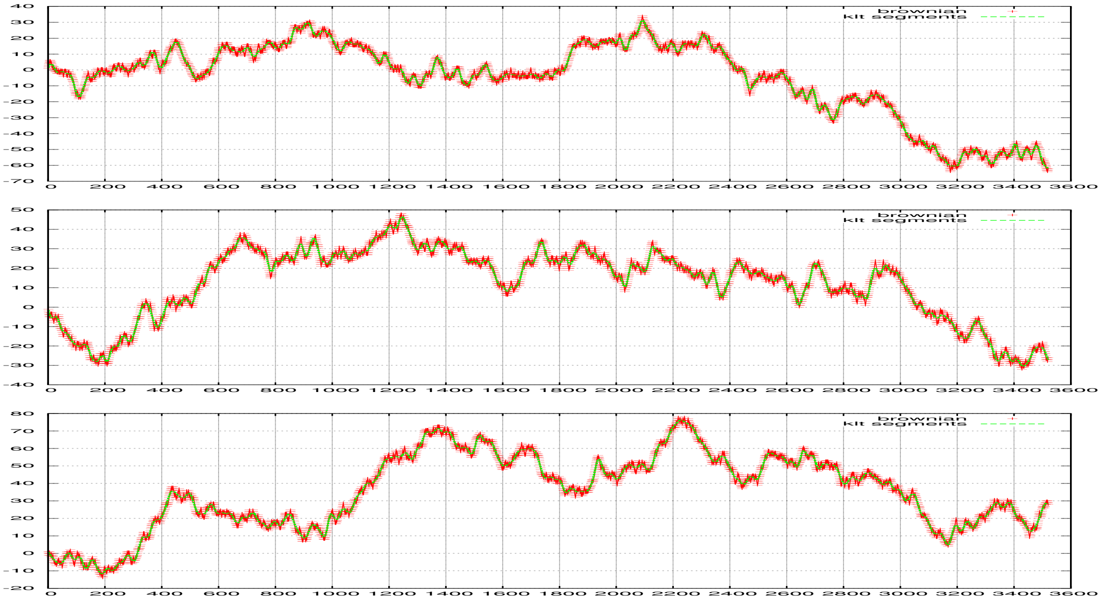


Figure 11.10: An example of using PCA with a data set larger then the rank of **R**. This sample uses a matrix of rank N=200, find the eigenvectors and applied the result to several segments of $X(t)$ that has 3,600 points.

11.8 Conclusion

Principal Component Analysis (PCA) is a powerful mathematical technique that can be applied to data compression, classification and noise filtering.

While its requirement on computer time is far more than the Fourier Analysis (or FFT), it can perform better in most cases. The simple case demonstrates in this paper shows that a SNR of -23dB could defeat the FFT.

When applied to particular type of problem (i.e. Brownian Motion), the solution of PCA could be pre-computed and use on different data set.

Bibliography

[1] W.H. Press, S.A. Teukolsky, W.T. Vettering, and B.P. Flannery. *Numerical Recipes in Fortran.* Cambridge University Press, 1988.

[2] S. Dumas, Y. Dutil, and G. Joncas. Detection of biomarkers using infrared spectroscopy. *Acta Astronautica*, 67:1356–1360, 2010.

[3] A. Papoulis and S.U. Pillai. *Probability, Random Variables and Stochastic Processes",.* Graw Hill, 2002.

[4] G.H. Golub and C.F. van Loan. *Matrix computations.* John Hopkins Press, Baltimore, 1983.

[5] C. Maccone. *Deep Space Flight and Communications.* Praxis-Springer, 2009.

Chapter 12

Writing a letter to ET

by **Stephane Dumas**
The SETI League, inc.
jgsdumas@gmail.com

Abstract

The SETI experiences are searching for messages from outer space. The format and shape of those messages are unknown but it is expected they will be easy to detect from the background noise. Technology gives us the mean to broadcast our own messages in the direction of our other habitable worlds. The content of such messages would be very important. Transmissions cannot last very long and there are so many potential targets that the message must be short but more elaborated than a simple *Hi there* while less complex than the whole Encyclopedia Britannica. The message should have enough information to be interesting and provide a starting point to establish a conversation. This paper will discuss those points and provide a structure, based on Mathematics, for the Message's Primer. Furthermore, learning how to write efficient messages may help us in the deciphering of potential extraterrestrial ones.

12.1 Introduction

Sending message to other worlds may be another way to establish contact with an alien civilisation.

The content of the message is very important. Transmissions cannot last very long and there are so many targets that the message must be short in order to sent it more then once. It most be more elaborated than a simple *Hi there* and less than the whole Encyclopedia Britannica. The message should have enough information to be interesting and provide a starter for the next transmission.

The message must not be culturally link. It must however reflect the planet as a whole and not just a single culture. The Human must be treated as a single group without politic, religion and other ethical distinctions. Nevertheless, those concepts would too complex to explain in the first message.

The message must not be temporally link. Sending a message reflecting the 21st Century would be meaningless. If it is received in an hundred years, it may no longer reflect the state of the Human civilisation and it will no longer describe us. However, the first part of the message, the primer, may contain a lot of information that will be tighten to this contemporary period and some that may be already obsolete. This part of the message serves as a introduction and should contain simple things.

Telling about our technologies would not be directly useful to an Alien civilisation since they may already possess the knowledge. But the real piece of information would be in the context of this technology. The information contains in the message can be perceived from multiple point of view. For example, telling ET about our knowledge of the proton, and even the quarks, implies that we have achieved that level of technology, that we know about the constituent of the universe. They might already be much advanced in that field and the information would not be useful for them. However, they will certainly understand where we are relative to them. Science and knowledge are like a tree. Knowing about protons implies knowledge of particle Physics and a fair amount of Mathematics.

It is similar to archaeologist that found some piece of pottery or bronze near an encampment. The amount of contextual information they generate may be more important than the piece it-self. Even if they will never meet in person those who used those artefacts. However, in the case of Human Archaeologist, they have the Human culture as reference.

Sending the value of π, the boiling point of water, or the age of the universe to an alien civilisation, serve no real purpose outside the fact to establish reference points.

The precision of the details relative to our knowledge also carries information about us. With our technology, we can measure electromagnetic properties with a very high precision, but we face extreme difficulties when trying to measure the gravitational constant. The same is true for properties of the universe, for which our knowledge is quite crude at best.

12.2 An interstellar language

12.2.1 Metalanguage

In order to communicate with an alien civilisation, a common language must be use and known from both parties. A language is a communication protocol used to exchange information between two interlocutors. It is based on common references known by both parties. All Human languages share those references since we live on the same planet. A glass of water is recognised by everyone even if the name changed.

Everything from our interplanetary neighbours is unknown. It will be quite surprising if they know any human language. They live in the same universe as us and more specifically, in the same neighbourhood of the Milky Way. We assume that the laws of Physics are the same everywhere in the Universe. Those laws could be used as a starting point for discussions. Mathematics are more fundamentals than Physics and may be more common. Any civilisation capable of building devices to listen to radio waves must know some sort of engineering. Starting from Mathematics and building up to Physics may seem a good approach to begin our conversation.

12.2.2 Artificial Languages

Synthetic, or auxiliary languages, are not a new concept and a few were created during the last century. They have been created by design and are not natural evolution of human language. While keeping most of the cultural fingerprint, they are a good basis to start.

Those languages could be the starting point of a more elaborated communication protocol with the alien civilisation once the first step (of contact) would be achieved.

Latino Sine Flexione

This artificial language has been proposed by Giovanni Peano [1] in 1903. It is similar to latin but without the heavy grammar, inflexion, double root, and so forth. This would be closest to a real human language but still depend on social context. It is also referred to as *Interlingua*.

It is composed of elements of seven major European languages (e.g. English, French, German, Italian, Portuguese, Spanish and Russian). As such, its vocabulary is mostly of Greco-Latin origin. Its grammar is very simple.

This language focus on word families built around root like 'currer' (run), 'prender' (take), 'caper' (grasp) and a common set of prototypic affixes (ad-, pre-, pro-,

-ion, -ive, -ura, etc.). Table 12.1 presents some examples of this language.

The Peano's Interlingua is not to be confused with the IALA's Interlingua. The former favoured the abolition of grammar while the later reduced it to a minimum. They both uses a common Latin base.

Table 12.1: Example of Latino Sine Flexione. (a) in latin, (b) in latino sine flexion and (c) in english

(a)	Vox populi, vox Dei
(b)	Voce de populo, voce de Deo
(c)	The voice of the people is the voice of God
(a)	Hodie mihi, cras tibi.
(b)	Hodie ad me, cras ad te.
(c)	It is my lot today, yours to-morrow.
(a)	In medio stat virtus.
(b)	Virtute sta in medio.
(c)	Virtue stands in the middle.

Langua cosmica

Lincos was created by Hans Freudenthal in 1960 as a language to be used in extraterrestrial communication. The draw back of Lincos is that its primitives must be defined prior to being used. It is not a self-taught communication system. Preceding the Lincos text with preamble part to teach it could be solution.

Table 12.2 presents an example of Lincos. In summary, Ha asks Hb what is x for 10x=101. Hb answer 101/10 and Ha says it is good. Ha and Hb can be interpreted as humans A and B. inq means inquit (say) and ben mean bebe (good). The dialog form is another example of structure to present the reader the information without entering into too complexe descriptions of the concept.

Table 12.2: Example of Lincos - dialogue about the value of x for a given equation.

Ha inq Hb ?x 10x=101
Hb inq Ha 101/10
Ha inq Hb ben

However, in practice most experts rule out purely logical messages like those proposed by Freudenthal. In place, a hybrid communication scheme including logical propositions and graphical representations of ideas and objects is preferred [2, 3, 4, 5, 6, 7]. In hybrid representations, the ideal is to have a new character for each proposition. In such cases, we shall speak of ideograms and not characters, like Chinese, Egyptian, and Mayan hieroglyphs.

Lambda Calculus

Created by Alonzo Church in 1930, *Lambda Calculus* relies heavily on the use of function (e.g. Mathematical functions). It is considered as the first computer language. From this a pseudo computer language can be created to describe processes and concepts.

Tables 12.3 and 12.4 shows examples of the syntax. It is somehow a cumbersome notation.

Table 12.3: Example of Lambda Calculus - enumeration of numbers

$1 \equiv \lambda sz.s(z)$
$2 \equiv \lambda sz.s(s(z))$
$3 \equiv \lambda sz.s(s(s(z)))$

Table 12.4: Example of Lambda Calculus - additions of numbers (2+3).

$2S3 \equiv (\lambda sz.s(sz))(\lambda wyx.y(wyx))(\lambda uv.u(u(uv))$

12.3 A Basic Lexicon

The process of writing a message to an Extraterrestrial Civilisation (EC) is similar to the process of analysis a message received from them. The tools (i.e. information theory and statistics) that we are going to use in decoding the received message, should be the very similar to the tools an EC would used upon reception of our message.

An Interstellar Rosetta Stone should be used as a prefix, or an introduction, chapter of the whole message. Something to teach the reader how to read the rest of the communication. This first part of the message will be based on Mathematics. The following sections will be related to elementary notions of Physics, Chemistry and Biology. The idea is to communicate enough information so the receiver understand the concepts and thus creating a common reference for discussion.

The choice of science as the language of the preamble chapter is dictated by the fact the alien civilisation receiving it most have built some sort of device (i.e. radio-telescope) to detect such transmissions. This device implies knowledge of engineering and therefore Mathematics and Physics. It is also expected that the message itself will be analysed, upon reception, by a group of people with the appropriate knowledge.

While Mathematics and Physics served to introduce reference points in the communication, they are not adequate to communicate social and human concepts, that is to talk about our civilisation. This will be the subject of further chapters of the message.

12.3.1 Introducing Numbers

The first step is to establish, or introduce, a set of symbols representing the numbers. The choice of the base 10 numerical system is to simplify the coding of the message. There is no evidence of an EC using a similar numeral system and they could use any other system of counting. In fact, during the History of humanity, different civilisations have used different system (i.e. Babylonian used base 60, the Roman, the Chinese, etc.).

Numbers would be introduced by association with binary representation and a list of N items (i.e. 2 = 0010 = xx). The Mathematical aspect of the message could be written using binary representation but the space require to write huge binary numbers is not practicable. Also, the base 10 system offers the possibility of writing floating point (i.e. 1.23456) which it rather difficult in binary.

The introduction of numbers is a little tricky but can be achieve through repetition. Numbers could be used in other context such as arithmetic. Displaying arithmetic expression could increase the level of information concerning the 10 symbols and serve also to introduce Mathematical operators.

Arithmetic expressions can be introduced by listing a series of operations using the value numbers (i.e. 0+1=1, 1+1=2, 2+1=3). Permutation of numbers (i.e. 2+3=3+2) reinforces the definition of numbers and help teaching them the alien reader. Operations such as subtraction, multiplication and division will be also presented through a list of examples (e.g. using the same sequences of numbers, 2+3=5, 2-3=-1, 2*3=6 and 2/3=0.6666). Negatives numbers will be introduced through the subtraction examples. The notion of division by 0 would be used to introduce the notion of infinity.

Division of two numbers could be used to introduce the notion of real number (or floating point notation) using examples such as 1/2=0.5, 1/4=0.25 and

1/3=0.33333(periodic). The last example introduce the notion of periodicity which is useful to indicate a pattern.

12.3.2 Power Notation

Exponents are used to reinforce the notion of fraction and also to introduce a compact form of writing large values. Large values will be needed when describing the real world (e.g. mass, distance, etc.). Using the same series of number as for the previous operations (i.e. 0+1, 1+1, 2+1), exponent can be introduced via the same examples (i.e. 0⇑1=1, 1⇑1=1, 2⇑2=2)

A typical way to write large number using exponent notation is 1.2×10^2 (also known as the scientific notation). There is also the engineering notation (i.e. 1.2E+2). The engineering notation would introduce a second way to write large number and may be confusing. For the sake of reducing the size of the symbols, the proposed notation would be the scientific one. A series of examples on the use of the power of 10 and how to write small or large numbers using a short expression would also be in the message (2,300,000,000 can be written as 2.3*10⇑9).

12.3.3 Mathematical Logic

Logic statements are introduced by listing examples, like the others operators. Logics can be used to indicate that a situation is good or bad and be used in the same kind of syntax as in Lingua Cosmica.

12.3.4 Group Theory

Group theory is used to introduce the notion of grouping elements or setting precedence in reading statements. Let define a group A composed of elements a, b et c as $A = \{a, b, c\}$. We can says that $a \in A$ (a is an element of set A) and that $d \notin A$ (d is not part of A). Let define another group B, $B = \{d, e, f\}$. Now $d \in B$ and $a \notin B$. The union of A and B is written as $A \cup B = \{a, b, c, d, e, f\}$ and their interception as $A \cap B = \{\}$.

It is then possible to write $Earth = \{6.7 * 10 \Uparrow 9 \ human\}$ (assuming that Human has been defined previously) that could interpreted as the set Earth has 6.7 billions humans. Longevity of human could be expressed with an expression like this: $human + 100 \ years = \{\}$. The concept of parentheses follows easily and could be introduced via a list of expressions (e.g. 1+2*3=7 vs. (1+2)*3=9).

12.4 Points of Reference

Communication with an alien civilisation requires more than doing arithmetic operations. The numbers are important to be able to quantify concepts but those concept must be introduce somehow.

Sending a large amount of text (i.e. the internet) may solve problems related to the exchange of knowledge. Information theory shows us that it is possible to rebuilt the grammar of a language using a large amount of data. The receiver of such message would be able, through information theory and statistics, to understand the message from a grammatical point of view. But it is doubtful that he will understand the meaning of it. It would like knowing the grammar without the meaning of each word. We have a similar problem with the dolphin language [8, 9].

12.4.1 Physics

Physics is the study of the real world, from the atoms to the galaxies. The Laws of Physics are universal and can be used as reference to communicate between us and an EC. It is assumed that the alien civilisation would know and understand science. They have built devices to receive radio wave and must known a minimum of engineering.

Building from the first section of the message, where Mathematics are introduced, notions of Physics can be used to introduce length, time and mass. In the 1999 Evpatoria Message, the hydrogen atom was used to illustrate those notions. The hydrogen atom was schematised using the Rutherford Model. While the use of drawings may be good or not (depending on the interpretation of the Extraterrestrial readers), the symbols (i.e. glyphs) beside the atom represent the masses of proton, neutron and electron. The glyphs representing numbers are easily read if the Mathematics section was understood. Once the proton, neutron and electron have been defined, their masses can be introduced. The hydrogen can be also used to introduce notion of length through its energy level (e.g. Balmer, Lyman spectral lines).

With the discovery of exoplanets by orbital observatories (i.e. Kepler, Corot), it will be possible to observe other solar systems. It is quite possible that an EC will have similar capabilities and therefore can observe our own Solar System. Therefore a diagram of our Solar System with masses of Jupiter and the Sun can be added in the message. Another possible way to illustrate our Solar System without the use of drawing could be to list all planets with their mass and period of revolution around the sun.

Notion of Mass

The Mass is a measure of how much matter an object possessed. The proton's mass relative to the electron is a fix ratio regardless of the unit used. The electron's charge relative to the proton's is a also fix quantity. Table 12.5 uses the notions taught so far regarding the notation of large and small numbers. It also introduces the unit of mass, the **kg**.

In table 12.5, mass and charge are written using a capital letter follow by the particle name. This type of notation is short and practice. The usual way to write the proton's mass would be M_{proton}. However, this implies the used of half-line when writing the message. In the message, the words proton, neutron and electron would be replaced by the appropriated symbols identifying the particles.

Table 12.5: Atom used to defined mass. Here a comparison between the mass and charge of proton, neutron and electron

M proton = $1.67262158 \times 10^{-27}$ [kg]
M neutron = $1.67262158 \times 10^{-27}$ [kg]
M electron = $9.1093812 \times 10^{-31}$ [kg]
M proton = **M** neutron
M proton = 1836 * **M** electron
C proton = -**C** electron
C neutron = 0

Atoms can be described a group of proton and neutron. The notion of union (∪) is more suited than the notion of addition. Hydrogen can be seen as a group of 1 proton, helium as a group of 2 protons and 2 neutrons, and so on.

The definition of proton and neutron is reinforced by combining tables 12.5 and 12.6.

Table 12.6: Introducing some atoms.

Hygrogen = $1 \times \mathrm{P} \cup 0 \times \mathrm{N}$
Helium = $2 \times \mathrm{P} \cup 2 \times \mathrm{N}$
Carbon = $6 \times \mathrm{P} \cup 6 \times \mathrm{N}$
Nitrogen = $7 \times \mathrm{P} \cup 7 \times \mathrm{N}$
Oxygen = $8 \times \mathrm{P} \cup 8 \times \mathrm{N}$

Notion of Length

The length is the measure of the distance between two objects, or one of the object size. Atoms were introduced in the previous section to define the mass. They can be used to defined length by showing their dimensions. However, it could be unbiguous. The Hydrogen spectrum produces could be also used to introduce length. The speed of light is related to the frequency and wavelength by equ. 12.1.

$$c = \lambda\nu \tag{12.1}$$

Table 12.7 lists some spectral lines form the Lyman and Balmer's series. The same information could be written as a list of attributes of hydrogen such as : H=$\{122\times10^{-9}$ m, 103×10^{-9} m, 97.2×10^{-9} m, 94.9×10^{-9} m$\}$. This notation use the notion of the group theory which has been introduced earlier. The list of values followed by the symbol **m** will not be understood at first. But the group of aliens experts would find it soon or later. This type of work is very similar to those deciphering ancient human languages.

Table 12.7: List of Lyman and Balmer spectral lines from Hydrogen.

n	Lyman	Balmer
	λ (nm)	λ (nm)
2	122	
3	103	656
4	97.2	486
5	94.9	410

Notion of Time

Time is a difficult concept. It could be presented as a extension of the 3D world through the relation of space and time. Time is a component of velocity (L/T), acceleration (L/T^2) and frequency ($1/T$). One may proposed to use any of those three concepts, or all of them, and illustrate the concept of time.

Frequency is illustrated by showing a sinusoid wave pattern. It is linked to periodicity of event. Wavelength of photon can be used to introduce the frequency by equ. 12.1. One has just to find a way to introduce the wavelength of photon. Maybe through the emission spectrum of some element such as hydrogen.

Notion of Temperature

Temperature may be introduced by listing the previously known elements (i.e. H, He, Li, O, 2...) and their temperature of melting and fusion.

Table 12.8: Temperatures of atoms

hydrogen	14.01K	20.28K
helium	0.95K	4.22K
lithium	453.69K	1,615K
nitrogen	63.153K	77.36K
oxygen	54.36K	90.20K

12.4.2 Chemistry and Biology

Notions of Chemistry and Biology can be introduced by using atoms and group of atoms (e.g. molecules). Contrary to Mathematical expression, the use of = in Chemical expression could be confusion. The equality symbol indicates that both part of the expression are the same. A Chemical equation may involve energy, or a catalyser. In a sense they are equal but the meaning in not exactly the same. Lets use ⇒ instead of = to indicate a process. This could then be used further in the message for other purposes.

A example of Chemical expression could be illustrated by equation 12.2. The indices refer to the number of atom in the structure (N_2 means 2 Nitrogens). Referring to the discussion in section 12.4.1 regarding the usage of indices, one can see the potential conflict of meaning.

$$N_2 + 3H_2 \Rightarrow 2NH_3 \tag{12.2}$$

In section 12.4.1, the notation M_{proton} referred to the mass of the proton while in equation 12.2 N_2 indicates 2 nitrogen atoms linked together. Normally, both notations are not a problem since they are used in different occasions. However, if we want to use them in a single message, uniformity must prevail. An expression like equation 12.3 could then be used instead. Note that everything is now explicit (i.e. instead of $3H$, we used $3 * H$)

$$2 \times N + 6 \times H \Rightarrow 2 \times (N + 3 \times H) \tag{12.3}$$

12.5 Talking of Human

Describing the Humans with simple terms is not an easy task. Ask a thousand persons to do it and you will get a thousand different answers. However, there are some physical aspects that could be generalised.

A physical description of the subject is often the first step to define it. This would be the external appearance of an Human. The mean value, around the world, of a human's height is 1.737 meters for male and 1.605 for female. Using the prefix notation (previously) discussed, this information can be written as *HEIGHT HUMAN = 1.671 METER*.

Before using the notion of male and female, it must be introduced some how. This could be difficult without using complex notion of biology involving reproduction.

The mean mass of human may be written using the same technique : MASS HUMAN = 70 KG. And the temperature is given as follow TEMPERATURE HUMAN = 310.15 KELVINS.

The Biological aspect of a human require the use of chemistry and biology. A human is a carbon-base live form. Listing a few typical chemical reaction involving sugar may be a start. Using the set theory and the notion of time, it is possible to describe the longevity of a human HUMAN + 100 years = { }.

Once the physical aspect of a human have been introduced, the next step would be to talk about group of human (our society). This section of the message will treat the living style of Human. We tend to live in group and form communities. This information can be best describe using the set theory.

Following a rough description of our civilisation, talking about politic may bring a complement of information. This section of the message would treat the political aspect of the human society. Human tends to prefer democracy as a form of government but tolerate dictatorship, oligarchy and other non-democracy. This information can be best describe using the game theory and decision making processes.

12.6 Conclusion

The SETI program is looking for extraterrestrial messages. They are look for modulations of radio and optical waves. The search is centred on the structure of the message. Until we receive a real one, it would not be possible to guest its content. Building messages for an alien civilisation help understand the kind of message we may one day received.

Our attempts to send message revealed that its could not be written in any human language. The first part of it should be written in some universal language that an alien civilisation may understand. Since they have build some sort of device to receive radio transmission, it is reasonable to assumed they known science. Mathematics, Physics and Chemistry should be part of the first part of the transmission.

The format is also important in the construction of a message. A linear message (i.e. Morse) would be too susceptible to noise. A series of images would offer a better resistance to noise and degradation. Several statistical tools may be used to find some structure and rebuilt the images.

Bibliography

[1] G. Peano. De Latino Sine Flexione. *Revues de Mathematiques*, 8(3):74–83, 1903.

[2] M. A. Arbib. The psychology of Interstellar Communication. *Cosmic Search*, 1:21–24, 1979.

[3] C. De Vito. Language Based on Scienc. *Acta Astronautica*, 26:268, 1992.

[4] C. DeVito and R. Oehrle. A Language Based on the Fundamental Facts of Science. *Journal of the British Interplanetary Society*, 43:561–568, 1990.

[5] D. Vakoch. Possible pictorial messages for communication with extra-terrestrial intelligence. *Journal of the Minnesota Academy of Science*, 44:23–25, 1978.

[6] D. Vakoch. Constructing Message to Extraterrestrial: An Exosemiotic Perspective. *Acta Astronautica*, 42:697–704, 1998.

[7] D. Vakoch. The Dialogic Model: Representing Human Diversity in Messages to Extraterrestrials. *Acta Astronautica*, 42(705-710), 1998.

[8] R. Ferrer and B. McGowan. A law of Word Meaning in Dolphin Whistle Types. *Entropy*, 11:688–701, 2009.

[9] B. McGowan, S.F. Hanser, and L.R. Doyle. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle. 57:409–419, 1999.

Chapter 13

Is it dangerous or not to transmit signals?

by **L.M. Gindilis**
Sternberg Astronomical Institute, Moscow
lgindilis@mail.ru

Abstract

In recent years discussions on whether it is dangerous or not to send signals to extraterrestrial civilizations have been spreading in the Internet, in the scientific and quasi-scientific community. Such mentalities are probably provoked by fantastic movies that fill television screens. However, we must recognize that some scientists also express concern. See, for instance, discussions at the conference in Kavli, October 2010. Ethical and technical aspects of METI are examined. Pro and contra arguments are discussed.

With the emergence and formation of SETI in the middle of the past century, both in the scientific community and public opinion focus has been on the outstanding value of the fact of finding intelligent life beyond the Earth, let alone making contact (if at all possible). This problem has troubled thinkers of mankind for centuries. And when the practical detection capability appeared, it aroused widespread enthusiasm. At that time no worries about transmissions arose. Dangers of another kind were discussed: philosophical, social, and political implications of detecting ETI signals. Discussions that took place at the time were well summarized and clearly presented in Carl Sagan's novel, Contact [1]. Now there is apprehension about transmissions of signals up to their complete prohibition. Under the influence of public opinion, fuelled by the media, some scientists begin to express concern. See, for instance, discussions at the conference in Kavli, October 2010 [2]. Opponents of the transmission charge their opponents with levity and ideological motives. We believe that these accusations are unfounded. Supporters of the transmission rely on deep thinking and serious analysis, which has nothing to do with political or ideological motives. Let us remind some ideas discussed for decades of the development of the SETI problem.

13.1 A retrospective review

In 1964, Semyon Emmanuilovich Khaikin, outstanding Russian physicist and radio astronomer, developed a very interesting and very meaningful concept of establishing contact with extraterrestrial civilizations based on mutual search [3]. According to this concept, a less advanced civilization, having reached a certain level, sends a signal of readiness; having received it, a more advanced civilization that has created a system of signal detection determines the direction to the signal source, estimates the distance to it, and immediately starts transmission of information at the frequency of the signal of readiness. Khaikin emphasized that by refusing to transmit a signal of readiness, the civilization risks to put themselves outside the organized Galactic communication. The signal of readiness is the contribution that a civilization must make to join the Galactic Club. According to Khaikin, in the community of galactic civilizations each civilization must, in accordance with their level of development, to expend some effort. Not doing their share of problems, a civilization may remain outside the community. The need for transmission, together with their search, was pointed out by V.S. Troitskii [4].

Another Soviet scientist who strongly supported the need for signal transmission was Andrei Sakharov. In 1971, in response to a CETI questionnaire, he wrote: "In this case, I would like to stress the importance of design work for sending signals that are brought to the concrete implementation of some projects – the only way to understand the subtle aspects of the contact. Here, as in other cases, egoists, in the end lose out" [5].

In developing these ideas, A.L. Zaitsev said that if all civilizations observed the ban on the transmission of signals, the search would be nothing, and the problem of SETI would lose sense. Hence it follows that SETI has meaning only in a universe where there is awareness of the need and willingness for the transmission of interstellar messages [6].

In recent years, original and very profound ideas in this domain have been developed by A.D. Panov [7]. He drew attention to the fact that our civilization has entered, or is very close to, the state of the information crisis S. Lem warned about as early as in the middle of the twentieth century. A civilization that has reached the state close to the information crisis needs access to a new source of knowledge, different from the source of modern scientific knowledge. Panov noted that if we cannot resolve the problem of access to a new source of knowledge in some other way, such a source could be information obtained from other civilizations. This means that the SETI-contact can be vitally important for post-singular (i.e., those which have overcome the internal crisis) civilizations. Developing these ideas, Panov came to the hypothesis of a Galactic cultural field. He suggested that the cultural field in the Galaxy appears when each of exohumanitarian (post-technological) civilizations implements search and transmission of information to other civilizations. At the same time, it rebroadcasts transmissions received from other civilizations. As a result, the amount of information circulating in the Galaxy increases as an avalanche, and the Galaxy becomes a unified cultural field. The model of the Galactic cultural field leads to the concept of an exobank of knowledge. By its character the study of the exobank content resembles the process of studying Nature (conceptual model-test-new model). Panov calls the process of studying (comprehension) of the exobank exoscience. In his opinion, exoscience must take over the leadership in methods of cognition after the information crisis. Thus, overcoming the information crisis and creating a new source (exobank of knowledge) is associated with the search and transmission of information to other civilizations. Currently this activity does not play an important role, but it is a factor of excessive variety, which is probably destined to play a key role in overcoming the information crisis and forming the Galactic cultural field, i.e., transition to a completely new stage of evolution.

Similar views were expressed by A.L. Zaitsev in 1999. He noted the specifics of the communication to outside as selfless and messianic activity, carrying to alleged brothers in mind the good news "You are not alone." Zaitsev emphasized that awareness of the need for broadcasting to extraterrestrial civilizations is a sign of transition to a qualitatively new and higher level of intellectual and technological development. Targeted transmission of information to extraterrestrial civilizations, he said, “can serve to justify our existence, becoming one of the guarantees of the future sustainable development because among the causes of extinction of a civilization there is also “loss of interest”. It is also important that the development of methodology of broadcast to extraterrestrial civilizations allows us to better understand SETI strategies and tactics” [8].

Russian scientists who discussed the problem of SETI at the conference “Horizons of Astronomy and SETI” (Special Astrophysical Observatory of the Russian Academy of Sciences, September 2005) concluded that, in addition to searching, we also need the transmission of signals. The memorandum contained an item on the conference to support the efforts of METI [9].

13.2 Two aspects of the problem: ethical and technical

Is it dangerous or not to transmit signals? This problem has two aspects: ethical and technical. We consider first the ethical side. Ethics of advanced civilizations.

Returning to Khaikin's idea about the signal of readiness, we note that this idea, which arose in connection with the development of radio search strategy, has a wider philosophical sound. Every contact includes the desire and efforts from both sides. In this sense, the signal of readiness can be interpreted as the internal psychological and moral commitment to human contact. But isn't it dangerous?

Historical experience teaches us that until now (at least in the last millennium) development on the Earth was such that stronger individuals and civilizations aimed to subdue (and subdued) weaker ones. The system of socio-cultural restraints prevented complete destruction of opposing parties. But now the mankind for the first time has come to a point, has reached a level, where further increase in aggression and destruction of the enemy will inevitably lead to self-destruction of the human civilization, and, perhaps, to the death of the entire terrestrial biosphere. Therefore, the historical correction should lead to a change in consciousness: the human race should move from hostility to cooperation. If it fails to do this step it will perish in the flames of self-destruction or as a result of the complete destruction of the environment. It seems that people are becoming aware of it, and the idea of cooperation, in spite of a fierce resistance of the opposing forces, becomes increasingly stronger among people. We may think that the same applies to the extraterrestrial civilizations whose development includes some element of aggression. Either the spirit of cooperation will win or they will end in self-destruction. Therefore, advanced civilizations that have passed the crucible of crisis should possess high ethics and high culture. The evolution of civilizations produces a law requiring that high knowledge must not be given to evil hands. It seems that in this respect the mankind has reached the limit. Next, either a change of the way (change in consciousness, soul-searching) or self-destruction. From these positions, we can agree with Tsiolkovsky, when he wrote that the Universe is filled with a highly conscious, perfect life, in it the Supreme Mind and perfect public relations dominate.

In spite of the convincingness of this argument, it has a weak link. We still know too little about the laws of evolution of cosmic civilizations and, therefore, may be mistaken in our conclusions. And the risk is too great to ignore them, even at a low probability. Let us refer to the technical side of the problem.

13.3 Technical detection capabilities

From the technical point of view, it is clear that highly developed extraterrestrial civilizations can detect us (and probably have detected long ago) by the radio emission of planetary radar and television transmitters. A TV signal is weaker than a directional METI signal, but it can be detected using the technology similar to the ours of recent years at a distance of tens of light years. Moreover, since the TV signal is directed to all sides, it is easier to detect it.

Planetary radars, which are used for the detection of small bodies in the vicinity of the Earth and prevent the comet and asteroid hazard, have a longer duration compared to a METI signal, and they illuminate a much larger area of the sky. The probability of detecting such signals is a million times higher than the detection of a METI signal. Consequently, the risk of detection does not depend on our SETI/METI activity.

An advanced civilization can detect signs of life on the Earth also from observation of oxygen lines in the Earth's atmosphere. After all, it is thus that we are going to search for habitable planets around other stars.

There is no doubt that an extraterrestrial civilization that has reached such a level that for it aggression on interstellar scale becomes possible, possesses means of detection of less advanced civilizations of interest to it. Trying to hide from these civilizations, abandoning the transmission of signals is similar to the position of an ostrich burying his head in the sand. The position of scientists supporting METI is not levity; it is based on a realistic assessment of the situation.

In our opinion, the people that build up fears about possible aggression of extraterrestrial civilizations execute, consciously or unconsciously, a "social order" on the separation of our civilization from the rest of the Cosmos. This trend has deep historical roots, and it is quite fallacious.

13.4 Evaluation of arguments

Let us consider arguments of opponents of signal transmission.

1. Since the problem concerns the entire human race, the widest possible strata of the mankind must be involved in the discussion. This boundless democracy is in fact reduced to endless demagoguery. It is impossible to solve scientific problems

at meetings such as the Novgorod Veche, or the Athenian People Assembly. These problems cannot be resolved with the help of mass media, especially given the current technology to manipulate the public opinion.

History offers us vivid examples. During the heyday of Athenian democracy, the People Assembly expelled Anaxagoras; removed Pericles, Phidias; Socrates was sentenced to drink a cup of poison. Add the prosecution of the Pythagoras and Plato.

Nowadays, mass media have launched an unprecedented campaign of intimidation in connection with the commissioning of the Large Hadron Collider in Switzerland. The most incredible versions have been put forward. And all this was presented as from supposedly scientific grounds.

We believe that such matters affecting the destiny of the mankind must be addressed by competent and responsible people on the basis of serious analysis, not emotions.

2. METI opponents point to the danger of the arrival of aliens, but possibilities of such direct contacts are not analyzed at all. It is assumed that they are easy to implement, as in Hollywood films. You can send unmanned probes that, hundreds of thousands of years later, will reach other stellar systems. Our Pioneers and Voyagers are examples of such probes. But what is the danger for anyone they represent?

3. Concerns and conclusions about the behavior of "aliens" are considered from the standpoint of the history of our civilization. Is it rightful? Civilizations that are similar to us in their development, firstly, are quite rare and, secondly, do not pose any danger. Civilizations that have outperformed us by millions and billions of years are like Gods for us. Assessing their behavior in terms of the current instant of the human history is absurd. You may not accept arguments about the high ethics of such civilizations, but hiding from them is stupid.

4. Possibilities of detecting our civilization by facilities we possess even today are completely disregarded, not to mention the detection abilities of more advanced civilizations. People simply do not understand the problem.

5. At the beginning of SETI, when it engaged a small community of scientists, there were no concerns about the risk of transmissions. Other issues, philosophical, social and political implications of the fact of signal detection were discussed.

13.5 Conclusion

There is no reason to believe that extraterrestrial civilizations are highly aggressive and aim at the conquest of other planets. There is no reason to assume that the conditions of the Earth may be of interest to them, because the nature of their life may be different. Our concepts about the high ethics of developed extraterrestrial civilizations, in spite of convincing arguments, are neither conclusive. However, the analysis of technical capabilities shows that such civilizations can detect us by various methods, regardless of our SETI/METI activity.

Now with the help of the mass media and with the active participation of intellectuals, such as Michael Michaud and others, hysteria has been raised against METI, ready at any moment to spill over at SETI. One gets an impression that a kind of a "social order" is being executed, perhaps obscure for the performers themselves. The history of mankind is the history of the struggle between good and evil. We can also note two philosophical trends going through the history of the human thought: cosmism and isolationism. The anti-METI hysteria is an expression of the position of isolationism.

However, it should be emphasized once again that the position of METI supporters is not based on ideological arguments (they play a supporting role) but on a serious analysis of technical possibilities of the detection.

References

1. Sagan Carl. Contact. Simon and Schuster, 1985

2. See `http://royalsociety.org/extra-terrestrial-life/` and `https://royalsociety.org/General_WF.aspx?pageid=4294977022`
   See also `http://www.centauri-dreams.org/?p=17861`

3. Khaikin S.E. On The problem of communication with extraterrestrial civilizations // Extraterrestrial civilizations. Trudy soveschaniya. Byurakan, May 20-23 1964. Erevan, 1965/ P. 83-94. – In Russian.

4. Troitskii V.S. Some thoughts about the search for intelligent signals from the Universe // Extraterrestrial civilizations. Trudy soveschaniya. Byurakan, May 20-23 1964. Erevan, 1965/ P. 97-112. – In Russian.

5. Gindilis L.M. Andrei Dmitrievich Sakharov and the Search for Extraterrestrial Intelligence //Third Decennial US-USSR Conference on SETI. Santa Cruz, California, August 5-9, 1991 /Edited by G.Seth Shostak. Astronomical Society of the Pacific. San Francisco, 1993. p. 27-33.

6. Zaitsev A.L. The SETI Paradox, `http://arxiv.org/abs/physics/0611283`

7. Panov A.D. Evolution and the SETI problem. – In Russian. http://lnfm1.sai.msu.ru/SETI/koi/articles/EvolAndSETI.pdf

8. Zaitsev A.L. Broadcasting for extraterrestrial civilizations // SETI Information Bulletin, 1999. No 15. P. 31-47. – In Russian. See also http://lnfm1.sasi.msu.ru/SETI/roi/articles/beti-2.html These ideas had got the farther development in Zaitsev's recent paper "Rationale for METI" http://arxiv.org/abs/1105.0910

9. Final Memorandum //Astrophysical Bulletin of Special Astrophysical Observatory of Russian Academy of Sciences. 2007, 60-61. P. 5.– In Russian.

Chapter 14

Realistic targets at 1000 AU for interstellar precursor missions

by **Claudio Maccone**
International Academy of Astronautics
Via Martorelli, 43, Torino (Turin) 10155, Italy

Abstract

The nearest stellar system, the Alpha Centauri three star system, is located at about 4.40 light-years away. This amounts to 278,261 AU. But at only 550 AU, or, more generally, at only about 1000 AU, the focus of the gravitational lens of the Sun is found, that is then 278 times closer than our nearest interstellar target. In other words, assuming equal engineering problems, the trip to the Sun focus takes 278 times less than the time to the nearest stellar target. This makes the Sun focus a reasonable target for our probes to reach within this century. It also plainly appears that, before we send a probe towards anyone of the nearest stellar systems, we will need a detailed radio map of that stellar system. In other words, we need a huge radio magnification of all objects located in that neighbourhood, and nothing is better than the huge magnification provided by the gravitational lens of the Sun. Thus, sending a preliminary probe to 1000 AU in the direction opposite to the target stellar system clearly must be done before any interstellar flight to that stellar system is designed, not to say attempted. In this paper, a status review is presented about the "FOCAL" probe to 550 or 1000 AU. The relevant scientific, propulsion and telecommunication issues are briefly summarized and updated.

14.1 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 1420 MHz (21 cm wavelength). This is the frequency that all SETI radio-astronomers 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, by this author. The relevant Proceedings were published in 1994 in the Journal of the British Interplanetary Society [8]. Meanwhile, on May 20, 1993 this 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 one 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.

In the third edition of his book "The Sun as a Gravitational Lens: Proposed Space Missions" [12], the author summarized all knowledge available as of 2002 about the FOCAL space mission to 550 AU and beyond to 1000 AU. On October 3, 1999, this book had already been awarded the Engineering Science Book Award by the International Academy of Astronautics (IAA).

Finally, in March 2009, the new and comprehensive 400-pages book by the author, entitled "Deep Space Flight and Communications-Exploiting the Sun as a Gravitational Lens" [19], was published. This book embodies all the previous material published about the FOCAL space mission and updates it. On November 25, 2009, this book was presented in a talk that the author gave at the SETI Institute in Mountain View, CA, USA (You Tube site[1]).

The FOCAL name may also be regarded as an acronym for "Fast Outgoing Cyclopean Astronomical Lens", summarizing the mission's main features.

14.2 Why 550 AU is the minimal distance that "FOCAL" must reach

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

$$\alpha(r) = \frac{4GM_{Sun}}{c^2 r} \tag{14.1}$$

Fig. 14.1 shows the basic geometry of the Sun gravitational lens with the various parameters in the game.

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 a occurs for those rays just grazing the Sun surface, i.e. for $r = r_{Sun}$. This yields the inequality

$$\alpha(r_{Sun}) > \alpha(r) \tag{14.2}$$

[1] http://www.youtube.com/watch?v=ObvKVe5H8pc

with

$$\alpha(r_{Sun}) = \frac{4GM_{Sun}}{c^2 r_{Sun}} \tag{14.3}$$

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

$$\tan(\alpha(r_{Sun})) = \frac{r_{Sun}}{d_{focal}} \tag{14.4}$$

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

$$\alpha(r_{Sun}) \approx \frac{r_{Sun}}{d_{focal}} \tag{14.5}$$

Eliminating the angle $\alpha(r_{Sun})$ between Eqs. 14.3 and 14.5, and then solving for the minimal focal distance d_{focal}, one gets

$$d_{focal} \approx \frac{r_{Sun}}{\alpha(r_{Sun})} = \frac{r_{Sun}}{4GM_{Sun}/c^2 r_{Sun}} = \frac{c^2 r_{Sun}^2}{4GM_{Sun}} \tag{14.6}$$

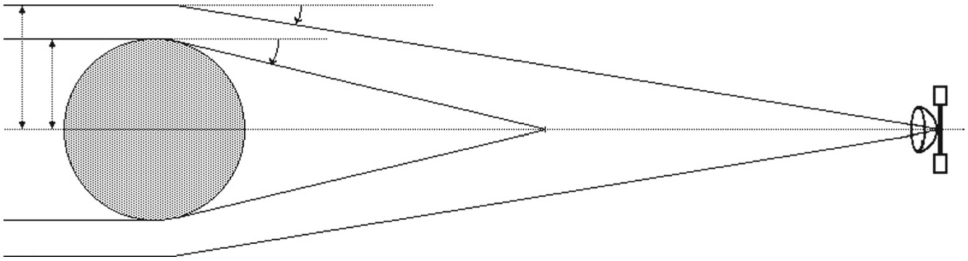


Figure 14.1: Geometry of the Sun gravitational lens with the minimal focal length of 550 AU (= 3.17 light days = 13.75 times beyond Pluto's orbit) and the FOCAL spacecraft position beyond the minimal focal length.

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

$$r_{Schwarzschild} = \frac{2GM_{Sun}}{c^2} \tag{14.7}$$

yielding

$$d_{focal} \approx \frac{r_{Sun}}{\alpha(r_{Sun})} = \frac{c^2 r_{Sun}^2}{4GM_{Sun}} = \frac{r_{Sun}^2}{2r_{Schwarzschild}} \tag{14.8}$$

Numerically, one finds

$$d_{focal} \cong 542AU \approx 550AU \approx 3.171 \ light \ years \tag{14.9}$$

This is the fundamental formula yielding the minimal focal distance of the gravitational lens of the Sun, i.e. 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 further 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 further it goes beyond 550 AU the less distorted the collected radio waves by the Sun Corona fluctuations.

The important and difficult problems of the plasma fluctuations in the Sun's Corona (electrons) were studiedby Von Eshleman at Stanford and by Slava Turyshev at JPL (please refer to Chapter 6 of Ref. [19]).

Just to summarize how the electrons in the Sun's Corona "push the true focus out" (i.e. they create a "diverging lens effect" that opposes the "converging lens effect of gravity") Fig.14.2 (taken from Ref. [19, p. 143]) shows that the true minimal focal distance F, that the FOCAL spacecraft must reach, is higher for lower frequencies of the source's electromagnetic waves crossing the Corona, and lower for higher frequencies. For instance, at 500 GHz the true focus falls at about 650 AU, while at 160 GHz (the cosmic microwave background (CMB) peak frequency) the true focus is found at 763 AU, and finally at 60 GHz the true focus is located at 1000 AU or beyond. Thus, we might briefly say that the FOCAL spacecraft must actually get beyond 550 AU to get rid of the problems caused by the Sun's Corona. Also, it must be stressed that these numbers were obtained on the basis of the so-called Baumbach-Allen model for the Solar Corona (described in Chapter 8 of Ref. [19]), while other and more up-to-date models might yield different results. In conclusion, more research work about the plasma in the Sun's Corona has to be made in order to find how much the "true" focal distance might be larger than just 550 AU.

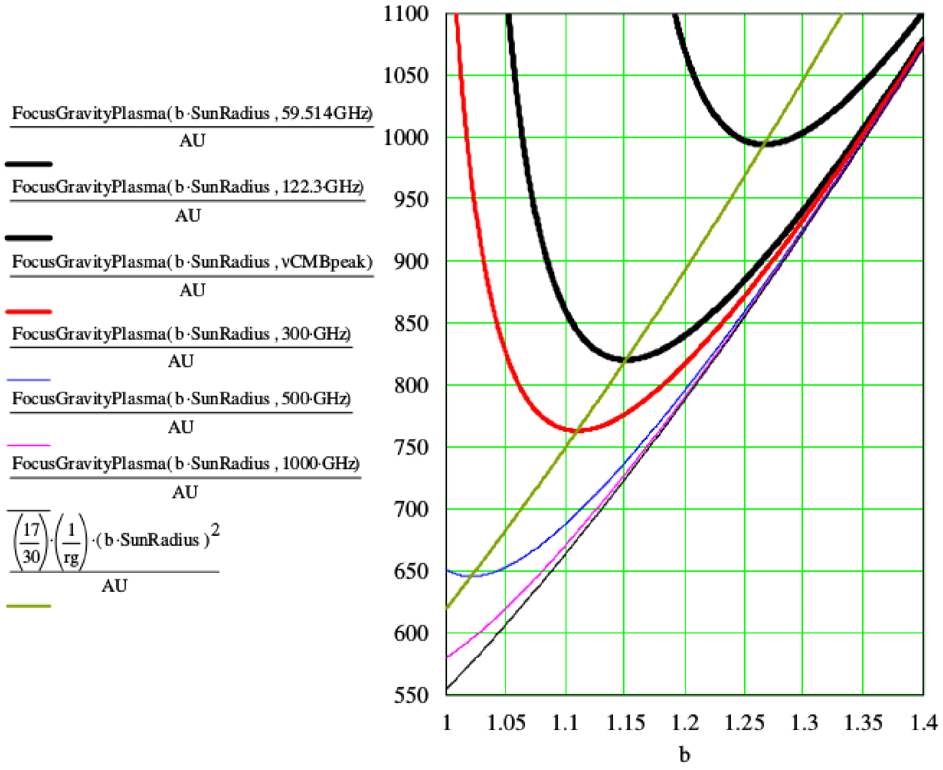


Figure 14.2: How the electrons in the Sun's Corona "push the true focus out". In this plot we see the (gravity + plasma) Sun's lens focal distance F (in AU) as a function of the impact parameter b (in units of the Sun radius) for all radio waves focused by the Sun above 120 GHz (plot of Eq. (8.6-8) of [19]). 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 the distance F from the Sun. Each curve corresponds to radio waves of a different frequency increasing from 59.514 GHz (top, thick curve) to 1000 GHz (bottom thin curve). The shifting of the minimum of obvious, and the "locus" of all these minima is the just the parabola of equation $F(b) = (17/30) * (1/r_{Schwarsschild}) * (b)^2$ (see Eqs. (8.7-9) and (9.4-4) of Ref. [19]).

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 will do without demonstration; that can be found on p. 55 of Ref. [13]. Consider a spherical star with radius r_{star} and mass M_{star}, that 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 focussing star center? The answer is given by the formula

$$d_{focal} = \frac{r_{star}^2}{(4GM_{star}/c^2) - (r_{star}^2/D_{source})} \tag{14.10}$$

This is the key to gravitational focussing 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 further out than 550 AU with respect to the focussing, intermediate star (see Chapter 6 of Ref. [19] for more details about this topic). Finally, notice that Eq. 14.10 reduces to Eq. 14.6 in the limit $D_{source} \to \infty$, i.e. (6) is the special case of (10) for light rays approaching the focussing star from an infinite distance.

14.3 The huge (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, i.e. how much focussing 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", that stems out of 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 (refer, for instance, to Ref. [14], in particular p. 6-117, Eq. (6-241)):

$$G_{antenna} = \frac{4\pi A_{effective}}{\lambda^2} \tag{14.11}$$

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

$$A_{effective} = \frac{A_{physical}}{2} = \frac{\pi r_{antenna}^2}{2} \tag{14.12}$$

Substituting this back into (14.11) yields the antenna gain as a function of the antenna radius and of the observed frequency:

$$G_{antenna} = \frac{4\pi A_{effective}}{\lambda^2} = \frac{2\pi A_{physical}}{\lambda^2} = \frac{2\pi^2 r_{antenna}^2}{\lambda^2} = \frac{2\pi^2 r_{antenna}^2}{c^2}\nu^2 \tag{14.13}$$

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

$$G_{Sun} = 4\pi^2 \frac{r_{Schwarzschild}}{\lambda} \tag{14.14}$$

or, invoking the expression (14.7) of the Schwarzschild radius

$$G_{Sun} = \frac{8\pi^2 GM_{Sun}}{c^2}\frac{1}{\lambda} = \frac{8\pi^2 GM_{Sun}}{c^3}\nu \tag{14.15}$$

The mathematical proof of Eq. (14.14) is difficult to achieve. The author, unsatisfied with the treatment of this key topic given in Refs. [1,3,14], 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 (14.14), but also of the focal distance for rays originated from a source at finite distance, Eq. 14.10. Their proof is fully described in Ref. [13], 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 (Fig. 1). 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 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".

Table 14.1: The GAIN of the Sun's lens alone, the gain of a 12-m spacecraft (S/C) antenna and the combined gain of the Sun +S/C Antenna system the at five selected frequencies important in radioastronomy.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	1420 MHZ	327 MHz	1.6 GHz	5 GHz	22 GHz
Wavelength λ (cm)	21	92	18	6	1.35
S/C antenna beamwidth (deg)	1.231	5.348	1.092	0.350	0.080
Sun gain (dB)	57.4	51.0	57.9	62.9	69.3
12-m antenna S/C gain (dB)	42.0	29.3	43.1	53.0	65.8
Combined Sun+S/C gain (dB)	**99.5**	**80.3**	**101.0**	**115.9**	**135.1**

Table 14.2: The image sizes for a 12 m FOCAL spacecraft antenna that has reached the distances of 550, 800 and 1000 AU from the Sun for each of the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	1420 MHz	327 MHz	1.6 GHz	5 GHz	22 GHz
Wavelength λ (cm)	21	92	18	6	1.35
Image size (down 6 dB) at 500 AU	2.498	10.847	2.217	0.709	0.161
Image size (down 6 dB) at 800 AU	3.033	13.169	2.691	0.861	0.196
Image size (down 6 dB) at 1000 AU	3.391	14.724	3.009	0.963	0.219

One is now able to compute the total gain of the Antenna+Sun system, that is simply obtained by multiplying equations the two equations yielding the spacecraft gain proportional to ν^2 and the Sun gain proportional to ν :

$$G_{Total} = G_{Sun} G_{antenna} = \frac{16\pi^4 G M_{Sun} r^2_{antenna}}{c^5} \nu^3 \tag{14.16}$$

Since the total gain increases with the cube of the observed frequency, it favors electromagnetic radiation in the microwave region of the spectrum. The table in Fig. 15.2 shows the numerical data provided by the last equation for five selected frequencies: the hydrogen line at 1420 MHz and the four frequencies that the Quasat radio astronomy satellite planned to observe (Table **??**), had it been built jointly by ESA and NASA as planned before 1988, but Quasat was abandoned by 1990 due to lack of funding. The definition of dB is of course:

$$N\mathtt{dB} = 10 log_{10} N = \frac{10 ln N}{ln 10} \tag{14.17}$$

14.4 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 > 4550 AU). We may define such an image size as the distance from the focal axis (i.e. from the spacecraft straight trajectory) at which the gain is down 6 dB. The formula for this (proven in Ref. [8]) is

$$r_{6dB} = \frac{\lambda}{\pi^2}\sqrt{\frac{z}{2r_S}} = \frac{c}{2\pi^2\sqrt{GM_{Sun}}}\lambda\sqrt{z} = \frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{\sqrt{z}}{\nu} \tag{14.18}$$

Thus the image size increases with the spacecraft distance z from the Sun. Table 14.2 shows how the image size increases with the spacecraft distance from the Sun in between the distances of 550 AU (minimal distance) and 1000 AU (maximal distance regarded as useful).

It is clear that these image size values are very small compared to 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) would enable us to gradually see much of the galactic center at the huge resolution provided by the gravitational lens of the Sun.

14.5 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 optimally:

(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} , or, mathematically,

$$r_{antenna} \lll r_{6dB} = \frac{c}{2\pi^2\sqrt{GM_{Sun}}}\lambda\sqrt{z} = \frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{\sqrt{z}}{\nu} \tag{14.19}$$

(2) Angle requirement: The impact-radius circle around the Sun within which electromagnetic waves are focussed towards the FOCAL spacecraft must fall well within antenna beamwidth of the FOCAL spacecraft. In a little more technical terms, the half-power beam width (= HPBW, i.e. 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

$$HPBW \ggg 2\alpha(r) = \frac{8GM_{sun}}{c^2 r} \tag{14.20}$$

Tables 14.3 and 14.4 show that both these conditions are fulfilled at the three FOCAL distances from the Sun and for our five selected frequencies.

14.6 The angular resolution at the spacecraft distance z

The notion of angular resolution of the Sun lens is very relevant to the discussion that will follow in the second half of this paper, inasmuch as we will want to know the linear resolution provided by a FOCAL spacecraft (with a 12-m antenna) at three very different distances from the Sun:

(1) The Galactic Center, distant about 30,000 AU or so from the Sun jointly with its huge, Galactic Black Hole (Sagittarius A*). Such a FOCAL space mission would be especially appealing to Astrophysicists and Cosmologists.

(2) The Alpha Centauri system of three stars (Alpha Cen A, B and Proxima), located just 4.37 light-years away from the Sun. This FOCAL mission would be of special interest to the many, different scientists that regard the Alpha Centauri system as the first real target for the future Interstellar Flights.

(3) Finally, we will provide an example of FOCAL space mission applied to one of the over 400 extrasolar planets that have been discovered since 1995. We selected the small one (above 1.9 Earth radius in radius) Gliese 581 e (or Gl 581 e). This is the fourth extrasolar planet found around Gliese 581, an M3V red dwarf star approximately 20.5 light-years away from Earth in the constellation of Libra. As described at the site[2], this planet was discovered by an Observatory of Geneva team lead by Michel Mayor, using the HARPS instrument on the European Southern Observatory 3.6 m (140 in) telescope in La Silla, Chile. The discovery was announced on 21 April 2009. Mayor's team employed the radial velocity technique, in which the

[2] http://en.wikipedia.org/wiki/Gliese_581_e

Table 14.3: The image sizes vs. the antenna radius for a 12 m antenna located at various distances from the Sun for the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	1420 MHz	327 MHz	1.6 GHz	5 GHz	22 GHz
Wavelength λ (cm)	21	92	18	6	1.35
Image size at 550 AU vs. antenna radius	2.498 km $\ggg$ 6 m	10.85 km $\ggg$ 6 m	2.22 km $\ggg$ 6 m	0.71 km $\ggg$ 6 m	0.16 km $\ggg$ 6 m
Image size at 800 AU vs. antenna radius	3.03 km $\ggg$ 6 m	13.17 km $\ggg$ 6 m	2.69 km $\ggg$ 6 m	0.86 km $\ggg$ 6 m	0.20 km $\ggg$ 6 m
Image size at 1000 AU vs. antenna radius	3.39 km $\ggg$ 6 m	14.72 km $\ggg$ 6 m	3.01 km $\ggg$ 6 m	0.96 km $\ggg$ 6 m	0.22 km $\ggg$ 6 m

Table 14.4: The half-power beam width (HPBW) vs. the aspect angle of the Sun for a 12 m antenna located at various distances from the Sun for the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	1420 MHz	327 MHz	1.6 GHz	5 GHz	22 GHz
Wavelength λ (cm)	21	92	18	6	1.35
HPBW at 550 AU vs. 2α	$1.23154^\circ \ggg 1.5 \times 10^{-7\circ}$	$5.34798^\circ \ggg 1.5 \times 10^{-7\circ}$	$1.09299^\circ \ggg 1.5 \times 10^{-7\circ}$	$0.34976^\circ \ggg 1.5 \times 10^{-7\circ}$	$0.07949^\circ \ggg 1.5 \times 10^{-7\circ}$
HPBW at 800 AU vs. 2α	$1.23154^\circ \ggg 1.5 \times 10^{-7\circ}$	$5.34798^\circ \ggg 1.5 \times 10^{-7\circ}$	$1.09299^\circ \ggg 1.5 \times 10^{-7\circ}$	$0.34976^\circ \ggg 1.5 \times 10^{-7\circ}$	$0.07949^\circ \ggg 1.5 \times 10^{-7\circ}$
HPBW at 1000 AU vs. 2α	$1.23154^\circ \ggg 1.5 \times 10^{-7\circ}$	$5.34798^\circ \ggg 1.5 \times 10^{-7\circ}$	$1.09299^\circ \ggg 1.5 \times 10^{-7\circ}$	$0.34976^\circ \ggg 1.5 \times 10^{-7\circ}$	$0.07949^\circ \ggg 1.5 \times 10^{-7\circ}$

orbit size and mass of a planet are determined based on the small perturbations it induces in its parent star's orbit via gravity. At a minimum of 1.9 Earth masses, it is the smallest extrasolar planet discovered around a normal star, and the closest in mass to Earth. At an orbital distance of just 0.03 AU from its parent star, however, it is outside the habitable zone. It is unlikely to possess an atmosphere due to its high temperature and strong radiation from the star. Although scientists think it probably has a rocky surface similar to Earth, it is also likely to experience intense tidal heating similar to (and likely more intense than) that affecting Jupiter's moon Io. Gliese 581e completes an orbit of its sun in approximately 3.15 days.

Having so described three different targets for three different FOCAL space missions, we complete this section by pointing out that the angular resolution provided by FOCAL simply is defined as the ratio of the image size (at the spacecraft distance z from the Sun) to that distance z. From (18), we thus get

$$\theta_{resolution}(z) = \frac{r_{6dB}}{z} = \frac{c^2}{2\pi^2\sqrt{GM_{Sun}}} \frac{1}{\nu\sqrt{z}} \tag{14.21}$$

Clearly the angular resolution also depends on the spacecraft distance z from the Sun, and it actually improves (i.e. it gets smaller) as long as the distance increases beyond 550 AU. This is an advantage, of course, and one more reason (apart from avoiding the Corona effects) to let FOCAL reach distances above 550 AU and up to 1000 AU.

Table 14.5 gives the angular resolutions for the same three FOCAL spacecraft distances of 550, 800 and 1000 AU from the Sun, at the same five selected frequencies.

Table 14.5: Angular resolution at spacecraft distances of 550, 800 and 1000 AU, at the five selected frequencies.

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	1420 MHz	327 MHz	1.6 GHz	5 GHz	22 GHz
Wavelength λ (cm)	21	92	18	6	1.35
Angular resolution at 500 AU S/C distance (arcsec)	6.3458×10^{-6}	2.7557×10^{-5}	5.6319×10^{-6}	1.8022×10^{-6}	4.0959×10^{-7}
Angular resolution at 850 AU S/C distance (arcsec)	5.2267×10^{-6}	2.2697×10^{-5}	4.6387×10^{-6}	1.4844×10^{-6}	3.3736×10^{-7}
Angular resolution at 1000 AU S/C distance (arcsec)	4.6749×10^{-6}	2.0301×10^{-5}	4.1490×10^{-6}	1.3277×10^{-6}	3.0174×10^{-7}

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 did not fire, forcing technicians to take 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 ms of arc precision. Checking this figure against the above table, one can see that the gravitational lens of the Sun plus a (modest) 12-m antenna would improve the angular resolution by about three orders of magnitude (at radio frequencies).

14.7 The spatial resolution at the spacecraft distance z

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

$$R_{object} = d_{Sun-Object}\Theta_{resolution} = d_{Sun-Object}\frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{1}{\nu\sqrt{z}} \tag{14.22}$$

Again, beyond 550 AU the resolution improves (i.e. the angle gets smaller) slowly with the increasing spacecraft distance from the Sun. Table 14.6 shows the spatial resolutions for a very wide range of object distances, from the Oort Cloud to cosmological objects like quasars. 8.

14.8 The 2009 new book by the author about the "FOCAL" space mission

In March 2009, the new, 400-pages and comprehensive book by the author, entitled "Deep Space Flight and Communications – Exploiting the Sun as a Gravitational Lens" [19], was published. This book embodies all the previous material published about the FOCAL space mission and updated it in view of submitting a formal Proposal to NASA about FOCAL. The front and back covers of this book are reproduced in Fig. 15.3.

14.9 Using two antennae and a tether to get a much larger field of view for FOCAL

The goal of this section is to put forward the new notion of a TETHERED SYSTEM tying up TWO ANTENNAE for the FOCAL spacecraft. We are going to show that the length of this tether system does not need to be very long: actually, just a couple of km or so is sufficient to get a radio picture of the big Galactic Black Hole, 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 the system stability. To build a 2-km long truss in space, however, is a difficult engineering task.

We thus prefer to speak about a tethered system rather than a truss system, leaving the actual design to expert engineers. We start by pointing out the problem of the Sun corona plasma fluctuations with the relevant disturbances caused upon the radio waves passing through the corona itself, as described in Chapters 8 and 9 of the author's 2009 book [19]. Finding a solution to this problem is vital for the success of the FOCAL space mission. We now claim that the best way to solve the corona problem is by doing interferometry 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, additionally, introduces the new and fruitful idea of a tether tying up each antenna to the main cylindrical body of the FOCAL spacecraft, as shown in Fig. 14.4.

Thus, 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, 400-500 AU, all "engines" (solar

sails? Nuclear-electric? Antimatter?) are turned off, so we can assume that, at least beyond 550 AU, the Sun-speed of the whole system is uniform.

(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 tense because of the angular rotation of the whole system.

(4) The two antennas are inflated at the same time just after they have reached the minimal safety distance from the spacecraft.

(5) The two antennas are oriented and pointed each toward the Sun. This means that the two antenna axes are parallel or nearly parallel to each other. In practice, a huge isosceles triangle is created in space, having as basis the distance between the two antennas and as apex the center of the Sun (at any distance higher than 550 AU).

(6) Slowly, both tethers are deployed by the same amount of length on each side of FOCAL. Because of the uniform angular rotation of the whole system, this means that the end-points of the tether, i.e. the center of each antenna, is made to describe an Archimedean spiral (i.e. a spiral with polar equation $\rho(\theta) = const\ \theta$) 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... On the other side of the Sun, at the distance of the galactic center (i.e. about 32,000 light-years 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 Fig. 15.5 hereafter. This gigantic black hole we call the Galactic Black Hole, as described in the next section.

14.10 Observing the galactic black hole magnified by virtue of FOCAL

(7) On the other side of the Sun, at the distance of the Galactic Bulge (i.e. some 26,000-32,000 light-years 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 Fig. 5 hereafter. This gigantic black hole we call the Galactic Black Hole, and provisionally assign to it the estimated mass of a million times the mass of the Sun (as of 2009, its estimated mass is actually 4.31±0.06 Sun masses). Consequently, the Schwarzschild radius of the galactic black hole is a

Table 14.6

Line	Neutral hydrogen		OH radical		H_2O
Frequency ν	1420 MHz	327 MHz	1.6 GHz	5 GHz	22 GHz
Wavelength λ (cm)	21	92	18	6	1.35
Resolution at 0.5 ly Oort Cloud (km)	145	632	129	41	9
Resolution at 4.29 ly α Centauri (km)	1248	5422	1108	355	81
Resolution at 10 pc=32.6 ly (km)	9576	41,583	8499	2719	618
Resolution at 100 pc=326 ly (km)	95,759	415,833	84.98	27.19	6180
Resolution at 1 kpc=3,260 ly (km)	957.58 km	4,158,330 km	849,861 km	271,955 km	61,808 km
	= 0.006AU	= 0.028 AU	= 0.005 AU	= 0.001AU	= 0.0004 AU
Resolution at 10 kpc=32,600 ly (km)	9,575,870 km	41,583,000 km	8,498,610 km	2,719,550 km	618,082 km
Galactic Center	= 0.06401 AU	= 0.27797 AU	= 0.05681 AU	= 0.01818 AU	= 0.00413 AU
Resolution at 50 kpc = 160,000 ly	4.78794×10^7km	2.07917×10^8km	4.2493×10^7 km	1.3597×10^7 km	3.0903×10^6km
Magellanic Cloud	=0.32006 AU	1.38984AU	= 0.28405 AU	= 0.0909 AU	=0.02066AU
Resolution at 612 kpc=2 millions ly	5.82123×10^8 km	2.52788×10^9km	5.16631×10^8km	1.65322×10^8km	3.75732×10^7km
Andromeda Galay M31	= 3.89125 AU	=16.8978 AU	=3.45349AU	=1.10512AU	=0.25166AU
Resolution at 18,406 pc = 60 millions ly	1.74636×10^{10}km	7.5836×10^{10}km	1.5499×10^{10}km	3.95968×10^{10}km	1.1272×10^9km
"Jet" Galaxy M87 in Virgo	=116.738 AU	=506.934 AU	=103.605AU	=33.1535AU	=7.53488AU
Resolution at 3.07 millions kpc	2.91059×10^{12}km	1.26393×10^{13}km	2.58316×10^{12}km	8.2661×10^{11}km	1.8786×10^{11}km
10 billions ly Radius of the Universe	=19,456 AU	=84,489 AU	=17,267AU	=5525.58AU	=1255.81AU
	=0.30765ly	1.33598 ly	=0.27304ly	=0.08737ly	=0.01985ly

million times larger than the Sun Schwarzschild radius, i.e. it equals $\sim 2.95 \times 10^9$ km $\sim$ 0.01976 AU. This linearity between mass and Schwarzschild radius obviously appears in the definition (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 two FOCAL Archimedean spirals. Fig. 5 clearly shows the 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 the proportion:

$$\frac{Minimal\ Tether\ Length}{550AU} = \frac{2r_{Schwarzschild\ of\ Galactic\ Black\ Hole}}{32,000 lightyears} \tag{14.23}$$

But the Galactic Black Hole Schwarzschild radius is simply given by the Schwarzschild radius formula (14.7)

$$r_{Schwarzschild\ of\ Galactic\ Black\ Hole} = \frac{2GM_{Galactic\ Black\ Hole}}{c^2} \tag{14.24}$$

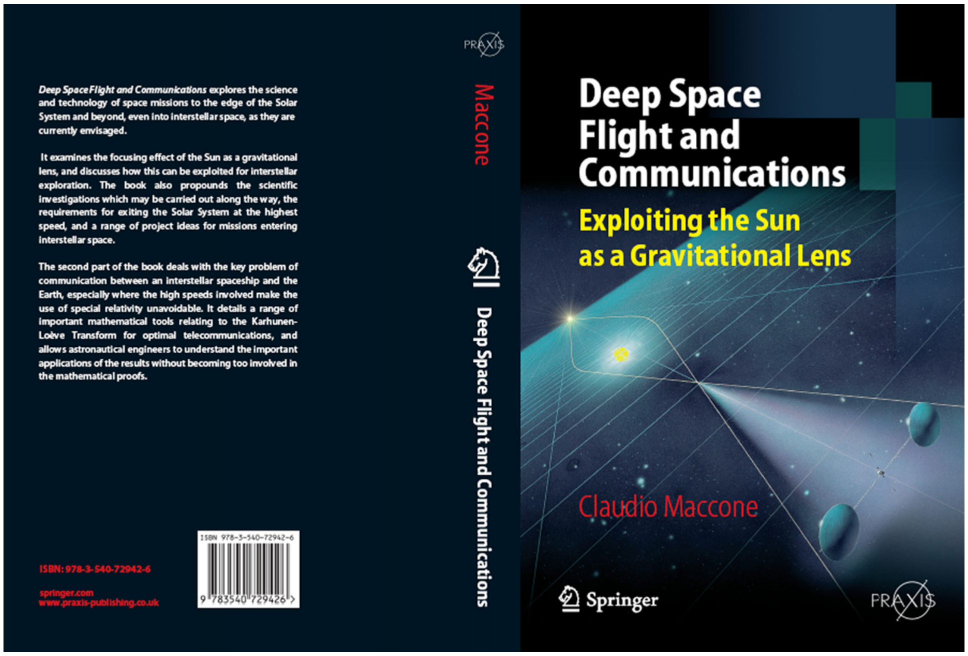
Astronomers have recently estimated the mass of the Galactic Black Hole to equal some four million solar masses. This is described at the Wikipedia site[3]. There one finds that, monitoring stellar orbits around Sagittarius A* for 16 years, the following conclusion was announced in 2008 by Reinhard Genzel, team leader of the research

[3] http://en.wikipedia.org/wiki/Sagittarius_A*

study: "The stellar orbits in the galactic centre show that the central mass concentration of four million solar masses must be a black hole, beyond any reasonable doubt." Thus, substituting (14.24) into (14.23) and solving for the minimum tether length, one gets

$$minimum\ tether\ length = 1.6km \tag{14.25}$$

for the basic case of 1 million solar masses for Galactic Black Hole. If the real value is four times as much, we must multiply (14.25) by four, getting 6.4 km. Also, the distance of the galactic center was changed by astronomers in recent years, letting it get down from 32,000 light-years to about 26,000 light-years. And, since the actual tether length must be higher than this minimal tether length, we reach the conclusion that a tether about 10 km long would certainly allow us to see not just the Galactic Black Hole, but also a host or astrophysical phenomena taking place around it, like the "swallowing" of stars by the Galactic Black Hole. In fact, from Table 14.6, row 9, we see that the linear resolution provided by FOCAL at the Galactic Center ranges between 1/10 and 1/100 AU.



Deep Space Flight and Communications explores the science and technology of space missions to the edge of the Solar System and beyond, even into interstellar space, as they are currently envisaged.

It examines the focusing effect of the Sun as a gravitational lens, and discusses how this can be exploited for interstellar exploration. The book also propounds the scientific investigations which may be carried out along the way, the requirements for exiting the Solar System at the highest speed, and a range of project ideas for missions entering interstellar space.

The second part of the book deals with the key problem of communication between an interstellar spaceship and the Earth, especially where the high speeds involved make the use of special relativity unavoidable. It details a range of important mathematical tools relating to the Karhunen-Loève Transform for optimal telecommunications, and allows astronautical engineers to understand the important applications of the results without becoming too involved in the mathematical proofs.

ISBN: 978-3-540-72942-6

springer.com
www.praxis-publishing.co.uk

ISBN 978-3-540-72942-6
9 783540 729426 >

PRAXIS

Maccone

Deep Space Flight and Communications

Deep Space
Flight and
Communications

Exploiting the Sun
as a Gravitational Lens

Claudio Maccone

Springer

PRAXIS

Figure 14.3: Front and back covers of the author's new book entitled "Deep Space Flight and Communications – Exploiting the Sun as a Gravitational Lens" published by Springer-Praxis in March 2009 (see [19]).

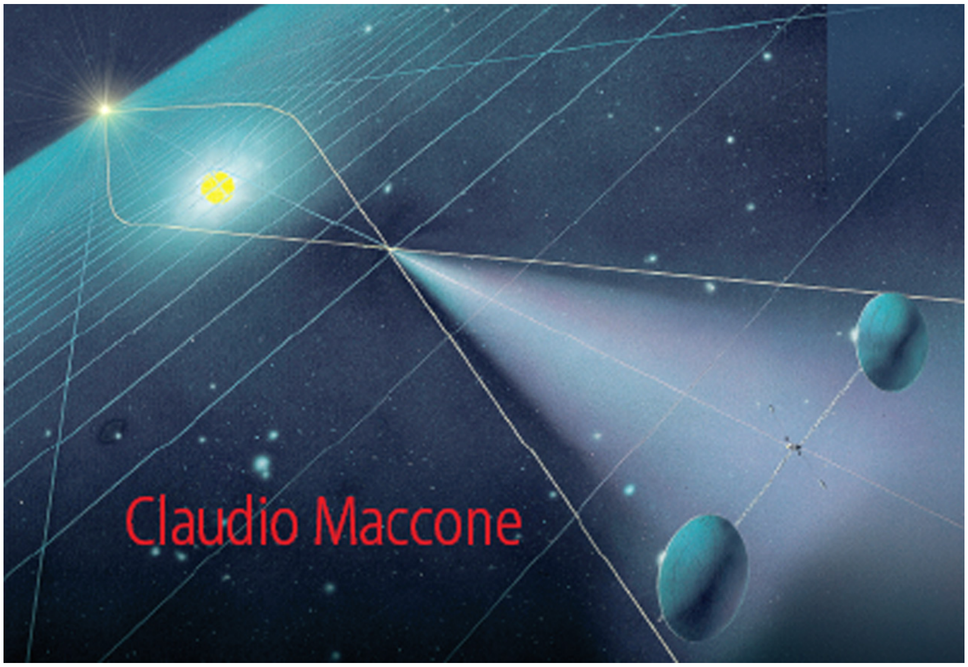


Figure 14.4: Enlarged part of the front cover of the author's 2009 book [19] showing: (1) the bright radio source at infinity (i.e. the horizon); (2) its radio waves flying by the Sun and made to focus at 550 AU; (3) the FOCAL spacecraft made up by two (say) 12-m antennae tied to each other by a tether and revolving in the orthogonal plane to the spacecraft's velocity vector. The same is shown here, with the two Archimedean Spirals covered by the antennae.

To sum up, 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 stellar target and thus launched along a different direction out of the solar system. And the guess is made here that all of them will use the tethered system as described in this section to avoid, by virtue of interferometry, all the problems caused by random fluctuations occurring within the Sun's corona.

14.11 Observing the 3 Alpha Centauri stars magnified by virtue of FOCAL

Alpha Centauri (a Centauri/ a Cen) is a triple star system and is the brightest star system in the southern constellation of Centaurus. Alpha Centauri AB (α Cen AB) is a close binary system revolving in 79.91 years. To the unaided eye it appears as a single star, whose total visual magnitude would identify it as the third brightest star in the night sky. As we all know, the triple Alpha Centauri system is the closest star system to the Solar System, the center of gravity of a AB Cen being only 1.34 parsecs, or 4.37 light-years away from our Sun. Because of this, the very first truly interstellar space mission will very likely be aimed at reaching the Alpha Centauri system, rather than any other nearby star system in the Galaxy.

From site[4] (the Alpha Centauri Wikipedia site), we learn that the star called Alpha Centauri A is the principal member or primary of the binary system, being slightly larger and more luminous than our Sun. It is a solar-like main sequence star with a similar yellowish-white color, whose stellar classification is spectral type G2V. From the determined mutual orbital parameters, a Cen A is about 10% more massive than our Sun, with a radius about 23% larger. The projected rotational velocity (v sin i) of this star is 2.7±0.7 km s^{-1}, resulting in an estimated rotational period of 22 days, which gives it a slightly faster rotational period than our Sun's 25 days.

The star Alpha Centauri B is the companion star or secondary, slightly smaller and less luminous than our Sun. This main sequence star is of spectral type of K1V, making it more an orangish-yellow color than the whiter primary star. a Cen B is about 90% the mass of the Sun and 14% smaller in radius. The projected rotational velocity (v sin i) is 1.1±0.8 km s^{-1}, resulting in an estimated rotational period of 41 days (an earlier estimate gave a similar rotation period of 36.8 days). Although it has a lower luminosity than component A, star B's spectrum emits higher energies in X-rays. The light curve of B varies on a short time scale and there has been at least one observed flare.

Finally, Alpha Centauri C, also known as Proxima Centauri, is of spectral class M5Ve or M5VIe, suggesting that this is either a small main sequence star (Type V) or sub-dwarf (VI) with emission lines, whose B-V color index is +1.81. Its mass is about 0.12 times the Sun mass. Proxima is approximately 12,000 or 13,000 AU away from Alpha Cen AB and its orbital period around them is of the order of 100,000-500,000 years or more (its orbit might even be hyperbolic). Because of this situation, Proxima is indeed the closest star to us of all, its distance being 4.243±0.002 light-years (1.3009±0.0005 pc).

[4] http://en.wikipedia.org/Alpha_centauri

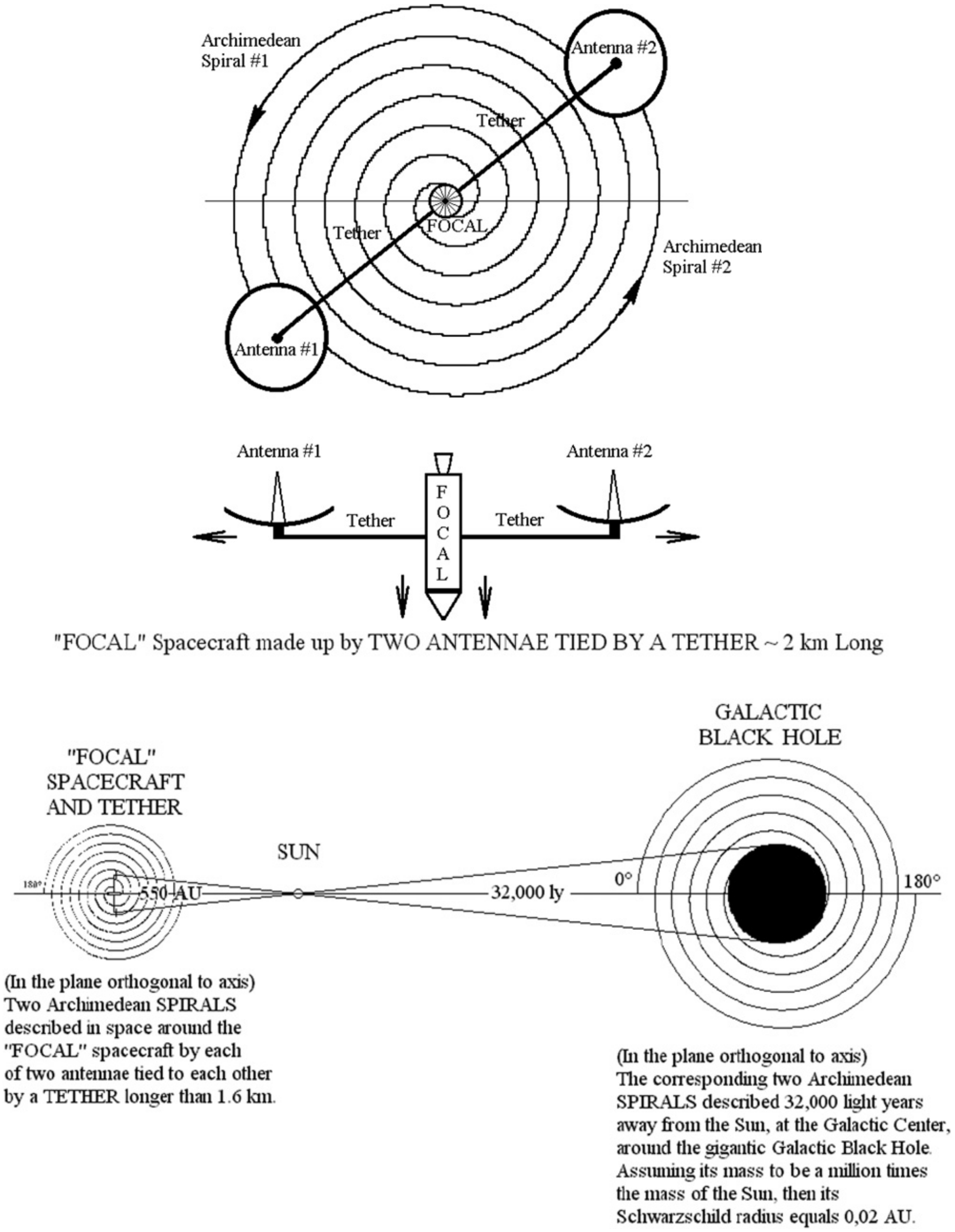

Archimedean
Spiral #1
Antenna #2
Tether
Tether
FOCAL
Archimedean
Spiral #2
Antenna #1
Antenna #1
Antenna #2
Tether
F
O
C
A
L
Tether
"FOCAL" Spacecraft made up by TWO ANTENNAE TIED BY A TETHER ~ 2 km Long
GALACTIC
BLACK HOLE
"FOCAL"
SPACECRAFT
AND TETHER
SUN
180°
550 AU
32,000 ly
0°
180°
(In the plane orthogonal to axis)
Two Archimedean SPIRALS
described in space around the
"FOCAL" spacecraft by each
of two antennae tied to each other
by a TETHER longer than 1.6 km.
(In the plane orthogonal to axis)
The corresponding two Archimedean
SPIRALS described 32,000 light years
away from the Sun, at the Galactic Center,
around the gigantic Galactic Black Hole.
Assuming its mass to be a million times
the mass of the Sun, then its
Schwarzschild radius equals 0,02 AU.

Figure 14.5: Imagine the above two Archimedean spirals in parallel planes both ORTHOGONAL to the axis FOCAL, Sun Center, Galactic Center. Then, two SIMILAR TRIANGLES relate the FOCAL tether length, the FOCAL spacecraft distance from the Sun, the size of the Galactic Black Hole and its distance from the Sun. They allow us to compute the minimal tether length.

We now want to clarify the notion of position angle, usually abbreviated PA and defined as the angular offset in degrees of the secondary star to the primary, relative to the north celestial pole. This is visually described in Fig. 14.6, taken from the relevant Wikipedia site.

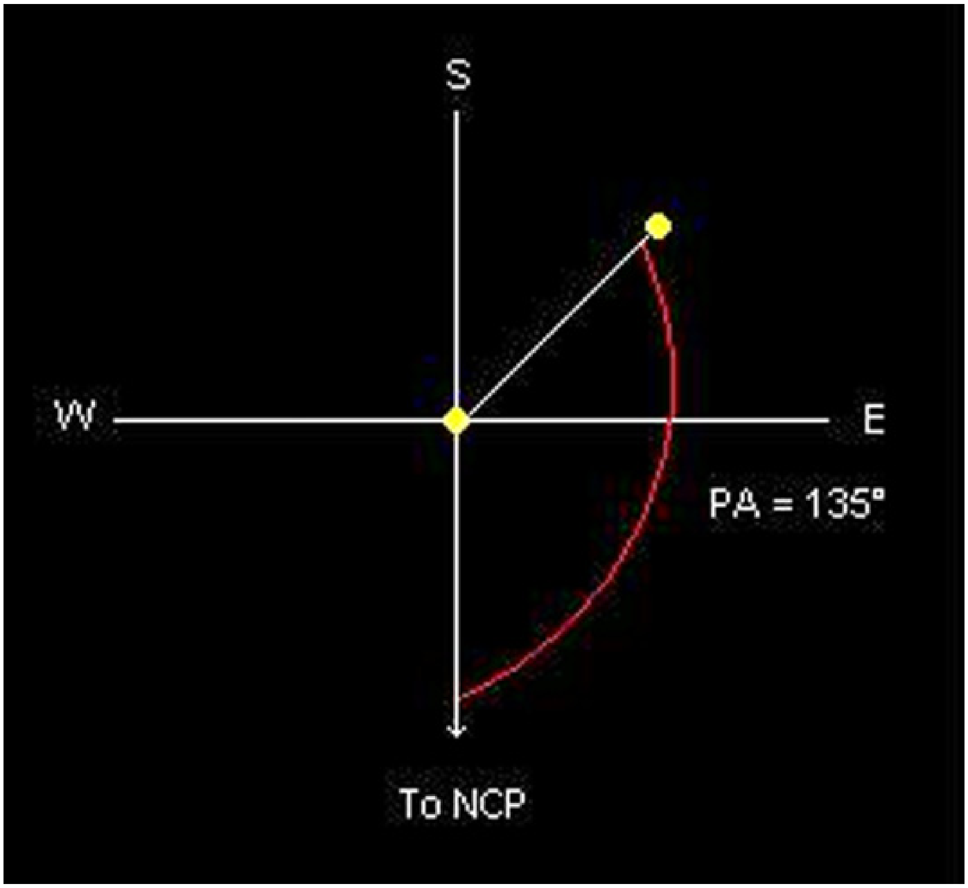


Figure 14.6: How the position angle PA is estimated through a telescope eyepiece. The primary star is at center. If one were observing a hypothetical binary star with a PA of 1351, that means an imaginary line in the eyepiece drawn from the north celestial pole (NCP) to the primary (P) would be offset from the secondary (S) such that the NCP-P-S angle would be 1351. The NCP line is traditionally drawn downward—that is, with north at bottom—and PA is measured counterclockwise, from 01 to 3591 (from site http://en.wikipedia.org/wiki/Position_angle).

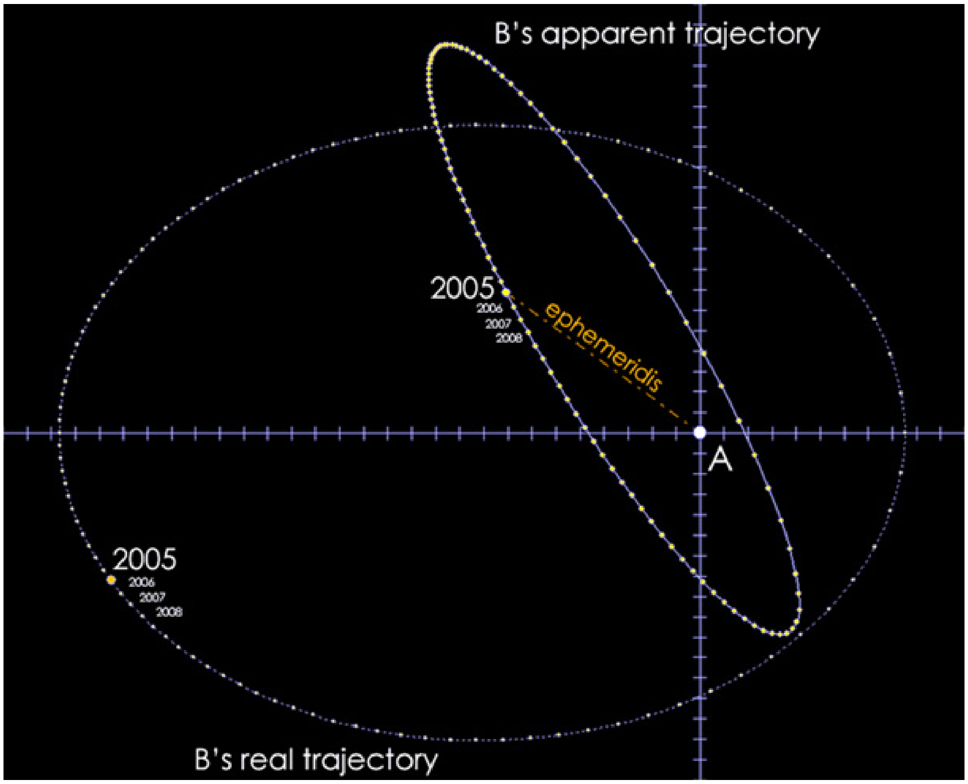


Figure 14.7: Apparent and True Orbits of Alpha Centauri B (the secondary) around Alpha Centauri A (the primary). Motion is shown from the A component against the relative orbital motion of B component. The Apparent Orbit (thin ellipse) is the shape of the orbit as seen by the observer on Earth. The True Orbit is the shape of the orbit viewed perpendicular to the plane of the orbital motion (taken from site http://en.wikipedia.org/wiki/Alpha_Centauri).

Figure 14.8: The building of the Guggenheim Museum in New York City shown tilted by 901, as if it was lying horizontally on the ground rather than vertically! Then, from right to left in this picture, the profile of this building is a conical helix of increasing radius.

Let us now go back to the Alpha Centauri system.

Viewed from Earth, the apparent orbit of this binary star system means that the separation and the position angle are in continuous change throughout the projected orbit. Observed stellar positions in 2008 are separated by 8.29 arcsec through a P.A. of 237°, reducing to 7.53 arcsec through 241° in 2009. Next closest approach will be in February 2016, at 4.0 arcsec through 300°. Observed maximum separation of these stars is about 22 arcsec, while the minimum distance is a little less than 2 arcsec. Widest separation occurred during February 1976 and the next will be in January 2056 (see Fig. 14.7 for B's apparent trajectory with respect to A).

In the true orbit, closest approach or periastron was in August 1955; and next in May 2035. Furthest orbital separation at apastron last occurred in May 1995 and the next will be in 2075. Thus, the apparent distance between the two stars is presently decreasing.

Going now back to the FOCAL space mission, the first question we wish to answer is: can we use a tethered system of two antennae to watch the Alpha Centauri system (as it is possible to do in order to observe the Galactic Black Hole? Unfortunately, the answer is "no", since the tether length would be far too long: of the order of millions of km!

To realize why it is so, just compute the expression

$$550\ AU\ tan\ \left(\frac{22arcsec}{2}\right) = 4.388 \times 10^6 km \tag{14.26}$$

yielding the tether length requested to encompass the view of both A and B at their maximum visual separation of 22 arcsec. Even if we consider the minimal visual separation of 4 arcsec, this is no better:

$$550\ AU\ tan\ \left(\frac{4arcsec}{2}\right) = 7.978 \times 10^5 km \tag{14.27}$$

So, a tethered system to encompass the whole A-B system is out of consideration. But this does not mean that a tethered system can be adopted to visualize each of the two stars separately: simply, it would provide too a narrow field of view because the whole system is just too close to the Sun.

So, we must resort to some other idea.

The new idea proposed here for the first time is to change the FOCAL orbit beyond 550 AU from a just an outgoing straight line to a conical helix of increasing radius.

To understand immediately what this means, consider the famous building of the Guggenheim Museum in New York City, shown in Fig. 14.8 as if it was tilted from right to left.

The profile of this building is a conical helix, and if you look at it from right to left, you have just the conical helix, i.e. the helix in on the surface of a cone with apex at just 550 AU and then higher and higher radius. This is the modified orbit we propose for FOCAL after 550 AU.

But how to achieve such an orbit in space?

Well, we need a small but continuous thrust, like those used in electric propulsion and called FEEP, an acronym for field emission electric propulsion: see, the relevant Wikipedia site and references[5].

Actually, the acceleration produced by these FEEPs is so small, and the times involved in having the FOCAL spacecraft moving along its conical helix trajectory are so large (decades), that one might well add a tethered system revolving orthogonal

[5] http://en.wikipedia.org/wiki/Field_Emission_Electric_Propulsion

to the speed vector, as described for the observation of the Galactic Black Hole. The radio image of the Alpha Cen system would then appear more and more detailed over the years while FOCAL would travel from 550 to 1000 AU along its conical helix trajectory.

We stop our description at this point, for the next step would require an accurate engineering design and an excellent astrodynamical calculation of the conical helix, both of which we do not have the time to compute now. But the idea of the conical helix plus a tether is a good one, and will have to be developed in further papers by this or other authors.

14.12 Observing extrasolar planets magnified by virtue of FOCAL

The most important discovery in Astronomy after 1995 is probably the discovery of extrasolar planets. As of August 2009 ([6]), 373 exoplanets are listed in the Extrasolar Planets Encyclopaedia. We thus wish to conclude this paper by providing an example of how the FOCAL space mission would be able to provide largely magnified images of extrasolar planets. For instance, consider Gliese 581e (or Gl581e), the fourth extrasolar planet found around Gliese 581, an M3V red dwarf star approximately 20.5 light-years away from Earth in the constellation of Libra. The planet was discovered by an Observatory of Geneva team lead by

Michel Mayor, using the HARPS instrument on the European Southern Observatory 3.6 m (140 in) telescope in La Silla, Chile. The discovery was announced on 21 April 2009.

Going now back to FOCAL, consider (14.22) again:

$$R_{Object} = d_{Sun-Object}\Theta_{resolution} = d_{Sun-Object}\frac{c^2}{2\pi^2\sqrt{GM_{Sun}}}\frac{1}{\nu\sqrt{z}} \tag{14.28}$$

This is the linear resolution of our extrasolar planet radio-pictures provided by FOCAL. In (22) we know

(1) The distance between the Sun and the target star, $d_{Sun-Object}$ given by the Hipparcos Catalogue.

(2) The distance z between the Sun and the FOCAL spacecraft after it reached at least 550 AU away from the Sun.

[6] http://en.wikipedia.org/w/index.php?title= Extrasolar_planet&action=edit

(3) The observing frequency n that we can choose at will (with many technological constraints) when we design the FOCAL spacecraft dedicated to observe that particular extrasolar planet only.

So, the key variable is the frequency, of course, and (14.22) clearly shows that the higher the frequency, the smaller (i.e. the better) is the linear resolution provided by FOCAL.

We just wanted to point this clearly out.

The interested reader may wish to read more by consulting the author's recent book [19], especially Chapter 9 and Section 9.4, where the Sun's Coronal Effects are taken into account also. Thanks very much.

14.13 Conclusion

In these few pages we could just sketch the FOCAL space mission to 550 AU and beyond to 1000 AU.

A number of issues still have to be investigated in:

(1) the many scientific aspects related to the mission,

(2) in the propulsion tradeoffs to get there in the least possible time and

(3) the optimization of the telecommunication link.

Yet, it plainly appears that the Sun focus at 550 AU is the next most important milestone that Humankind must reach in order to be prepared for the following and more difficult task of achieving the interstellar flight.

14.14 Additional references

We would like to further help the reader with a few additional references to papers about the FOCAL space mission. Ref. [14] is the standard textbook about radioastronomy by the late John D. Kraus of the Ohio State Radio Observatory. Ref. [15] suggests a hypothetical propulsion system based on the exploitation of suitable radioactive materials according to the lines of Nobel-laureate Carlo Rubbia's "interstellar propulsion". Ref.[16] describes the telecommunication link between the FOCAL probe and the Earth optimized by virtue of the KLT (Karhunen-Loève transform). The KLT is a more powerful tool than the classical FFT, but implies a higher computational burden. This topic also is dealt with in ref. [17]. Finally, Ref. [18] is a review paper about the propulsion tradeoffs for a space general mission to Alpha Centauri.

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: it's potential for observations and communications over interstellar distances, Science 205 (1979) 1133-1135.

4. F. Drake, Stars as gravitational lenses, in: G. Marx (Ed.), Proceedings of the Bioastronomy International Conference, Balatonfured, Hungary, June 22-27, 1987, pp. 391-394.

5. N. Cohen, The pro's and con's of gravitational lenses in ceti, in: G.Marx (Ed.), Proceedings of the Bioastronomy International Conference, Balatonfured, Hungary, June 22-27, 1987, p. 395.

6. F. Drake, D. Sobel, Is Anyone Out There? Delacorte Press, New York, 1992, pp. 230-234 (in particular).

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: C. Maccone (Ed.), Proceedings of the International Conference on Space Missions and Astrodynamics, Turin, Italy, June 18, 1992, 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, C. Maccone, AstroSail and FOCAL: two extra solar system missions to the Sun's gravitational focuses, Acta Astronautica 35 (1994) 409-410.

11. C. Maccone, The SETISAIL project, in: G. Seth Shostak (Ed.), Progress in the Search for Extraterrestrial Life, Proceedings of the 1993 Bioastronomy Symposium, University of California, Santa Cruz, 16-20 August 1993, Astronomical Society of the Pacific Conference Series 74 (1995) 407-417.

12. C. Maccone, The Sun as a Gravitational Lens: Proposed Space Missions, 3rd ed., IPI Press, Colorado Springs, Colorado, USA, ISBN 1-880930-13-7, 2002.

13. R. Orta, P. Savi, R. Tascone, Analysis of gravitational lens antennas, in: C. Maccone (Ed.), Proceedings of the International Conference on Space Missions and Astrodynamics, Turin, Italy, June 18, 1992, Journal of the British Interplanetary Society 47 (1994) 53-56.

14. J.D. Kraus, Radio Astronomy, 2nd ed., Cygnus-Quasar Books, Powell, Ohio, 1966, pp. 6-115-6-118.

15. C. Maccone, Radioactive decay to propel relativistic interstellar probes along a rectilinear hyperbolic motion (Rindler spacetime), Acta Astronautica 57 (2005) 59-64.

16. C. Maccone, Telecommunications, KLT and Relativity, vol. 1, IPI Press, Colorado Springs, CO, USA, ISBN 1-880930-04-8, 1994.

17. C. Maccone, Relativistic optimized link by KLT, JBIS 59 (2006) 94-98.

18. L. Derosa, C. Maccone, Propulsion tradeoffs for a mission to Alpha Centauri, Acta Astronautica 60 (2007) 711-718.

19. C. Maccone, Deep space flight and communications—exploiting the Sun as a gravitational lens, a 400-pages treatise about the FOCAL space mission that embodies and updates all previously published material about FOCAL. ISBN 978-3-540-72942-6 Springer, Berlin, Heidelberg, New York, 2009, Library of Congress Control Number: 2007939976, & Praxis Publishing Ltd., Chichester, UK, 2009.

Chapter 15

Protected antipode circle on the Farside of the Moon

by **Claudio Maccone**
International Academy of Astronautics
Via Martorelli, 43, Torino (Turin) 10155, Italy

Abstract

The international scientific community, and especially the IAA (International Academy of Astronautics) have long been discussing the need to keep the Farside of the Moon free from man-made RFI (radio frequency interference). In fact, the center of the Farside, specifically crater Daedalus, is ideal to set up a future radiotelescope (or phased array) to detect radio waves of all kinds that are impossible to detect on Earth because of the ever-growing RFI.

Nobody, however, seems to have established a precise border for the circular region around the antipode of the Earth (i.e. zero latitude and 180° longitude both East and West) that should be Protected from wild human exploitation when several nations will have reached the capability of easy travel to the Moon.

In this paper we propose the creation of PAC, the Protected Antipode Circle, centered around the antipode on the Farside and spanning an angle of 30° in longitude, in latitude and in all radial directions from the antipode.

There are sound scientific reasons for this:

(1) PAC is the only area of the Farside that will never be reached by the radiation emitted by future human space bases located at the L4 and L5 Lagrangian points of the Earth–Moon system;

(2) PAC is the most shielded area of the Farside, with an expected attenuation of man-made RFI of 100 dB or higher;

(3) PAC does not overlap with other areas of interest to human activity except for a minor common area with the Aitken Basin, the southern depression supposed to have been created 3.8 billion years ago during the “big wham” between the Earth and the Moon.

In view of these unique features, we propose PAC to be officially recognized by the United Nations as an International Protected Area, where no radio contamination by humans will possibly take place now and in the future for the benefit of all humankind.

15.1 Introduction

The need to keep the Farside of the Moon free from man-made RFI (radio frequency interference) has long been discussed by the international scientific community. In particular, in 2005 this author reported to the IAA (International Academy of Astronautics) the results of an IAA "Cosmic Study" that had been started back in 1994 by the late French radio astronomer Jean Heidmann (1920–2000) and had been completed by this author after Heidmann’s death (see, for instance, [1,2]).

The center of the Farside, specifically crater Daedalus, is ideal to set up a future radiotelescope (or phased array) to detect radio waves of all kinds that it is impossible to detect on Earth because of the ever-growing RFI.

Nobody, however, seems to have established a precise border for the circular region around the antipode of the Earth (i.e. zero latitude and 180° longitude both East and West) that should be Protected from wild human exploitation when several nations will have reached the capability of easy travel to the Moon.

In this paper we propose the creation of PAC, the Protected Antipode Circle. This is a large circular piece of land about 1820 km in diameter, centered around the Antipode on the Farside and spanning an angle of 30° in longitude, in latitude and in all radial directions from the antipode, i.e. a total angle of 60° at the cone vertex right at the center of the Moon.

There are three sound scientific reasons for defining PAC this way:

(1) PAC is the only area of the Farside that will never be reached by the radiation emitted by future human space bases located at the L4 and L5 Lagrangian points of the Earth–Moon system (the geometric proof of this fact is trivial);

(2) PAC is the most shielded area of the Farside, with an expected attenuation of man-made RFI ranging from 15 to 100 dB or higher;

(3) PAC does not overlap with other areas of interest to human activity except for a minor common area with the Aitken Basin, the southern depression supposed to have been created 3.8 billion years ago during the "big wham" between the Earth and the Moon.

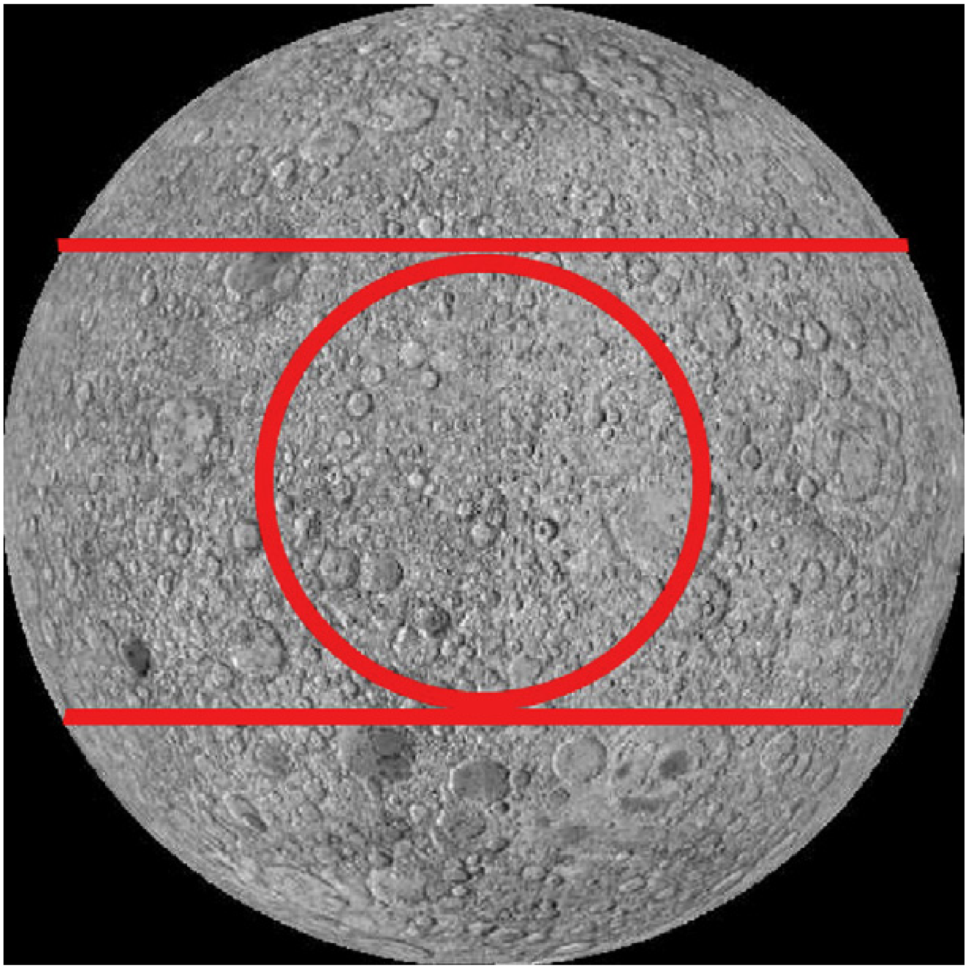


Figure 15.1: PAC, the Protected Antipode Circle, is the circular piece of land (1820 km in diameter along the Moon surface) that we propose to be reserved for scientific purposes only on the Farside of the Moon. At the center of PAC is the antipode of the Earth (on the equator and at 180° in longitude) and near to the antipode is crater Daedalus, an 80 km crater proposed by the author in 2005 as the best location for the future Lunar Farside Radio Lab. Inside Daedalus, the expected attenuation of the man-made RFI (radio frequency interference) coming from the Earth is of the order of 100 dB or higher.

Fig. 15.1 shows a photo of the Farside of the Moon, the two parallels at plus and minus 30° drawn by solid red lines, and PAC shown as the red, solid circle centered at the antipode and tangent to the above two parallels at plus and minus 30° .

In view of these unique features, we propose PAC to be officially recognized by the United Nations as an International Protected Area, where no radio contamination by humans will possibly take place now and in the future.

This will be for the benefit of all humankind.

15.2 Urgent need for RFI-free radioastronomy

In order to detect radio signals of all kinds, as radio astronomers do, it is mandatory to firstly reject all RFI. But RFI is produced in ever-increasing amounts by the technological growth of civilization on Earth, and has now reached the point where large bands of the spectrum are blinded by legal or illegal transmitters of all kinds.

Since 1994, the late French radio astronomer Jean Heidmann pointed out that radio astronomy from the surface of the Earth is doomed to die in a few decades if uncontrolled growth of RFI continues. Heidmann also made it clear, however, that advances in modern space technology could bring radio astronomy to a new life, if radio astronomy was done from the Farside of the Moon, obviously shielded by the Moon spherical body from all RFI produced on Earth.

In view of the following developments in this paper, we present now a short review about the five Lagrangian points of the Earth–Moon system, shown in Fig. 15.2.

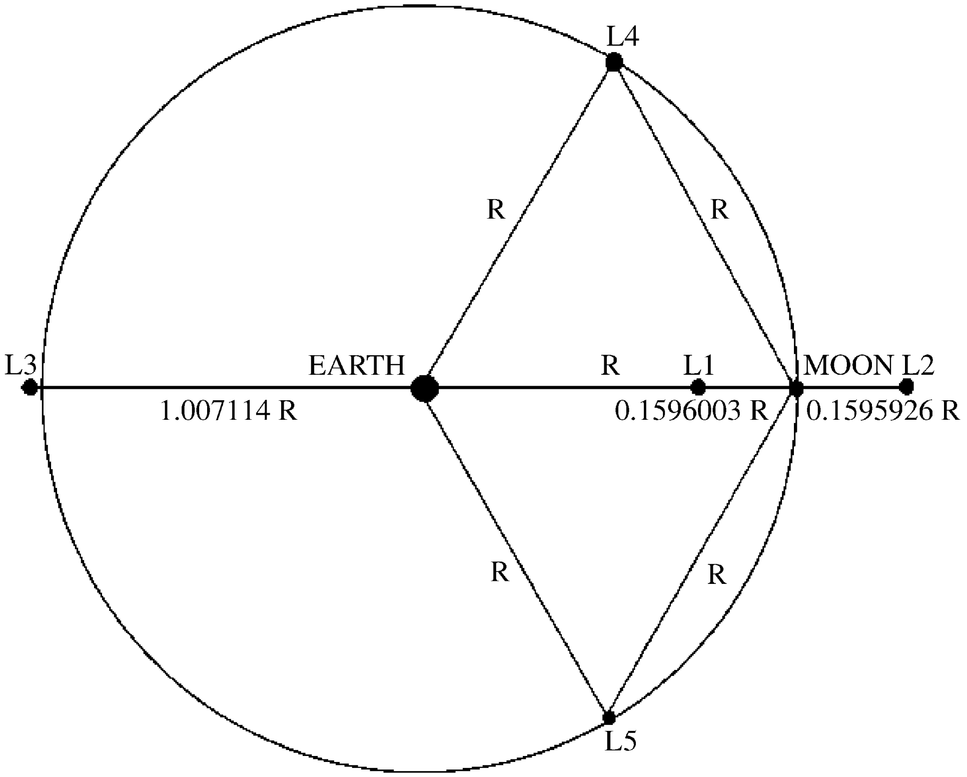


Figure 15.2: The five Earth–Moon Lagrangian Points (i.e. the points where the Earth and Moon gravitational pulls on a spacecraft cancel out!): (1) Let R denote the Earth–Moon distance that is 384 400 km. Then, the distance between the Moon and the Lagrangian point L1 equals 0.1596003*R, that is 61 350 km. Consequently the Earth-to-L1 distance equals 0.8403997*R, that is 323 050 km. (2) The distance between the Moon and the Lagrangian point L2 equals 0.1595926*R, that is 61 347 km. (3) The distance between the Earth and the Lagrangian point L3 equals 1.007114*R, that is 387 135 km. (4) The two "triangular" Lagrangian Points L4 and L5 are just at the same distance R from the Earth and Moon.

15.3 Terminal longitude λ on the Moon Farside for radio waves emitted by telecommunication satellites in orbit around the Earth

In this section we prove an important mathematical formula, vital to select any RFI-free Moon Farside Base.

We wish to compute the small angle α beyond the limb (the limb is the meridian having longitude 90° E on the Moon) where the radio waves coming from telecommunications satellites in circular orbit around the Earth still reach, i.e. they become tangent to the Moon's spherical body. The new angle $\lambda = \alpha + 90°$ is called the "terminal longitude" of these radio waves. In practice, no radio wave from telecom satellites can hit the Moon surface at longitudes higher than this terminal longitude λ.

To find α (see Fig. 15.3) we draw the straight line tangent to the Moon's sphere from G, the point tangent to the circular orbit having radius R. This straight line forms a right-angled triangle with the Earth–Moon axis, EM, with right angle at G. Next, consider the straight line parallel to the one above but from the Moon center M, intersecting the EG segment at a point P. Once again, the triangle EPM is right angled in P, and it is similar to the previous triangle. So, the angle is now equal to the EMP angle. The latter can be found, since:

(1) The Earth–Moon distance $\overline{EM} = D_{Earth-Moon}$ is known and we assume its worst case (Moon at perigee): Earth–Moon distance equal to 356 410 km.

(2) The $\overline{EP}$ segment equals the $\overline{EG} = R$ segment minus the Moon radius, R_{Moon}.

(3) Using Pythagoras' theorem one finds $\overline{PM} = \sqrt{\overline{EM}^2 - \overline{EP}^2}$.

(4) The tangent of the requested angle is then given by

$$\tan\alpha = \frac{\overline{EP}}{\overline{PM}} = \frac{\overline{EP}}{\sqrt{\overline{EM}^2 - \overline{EP}^2}} \tag{15.1}$$

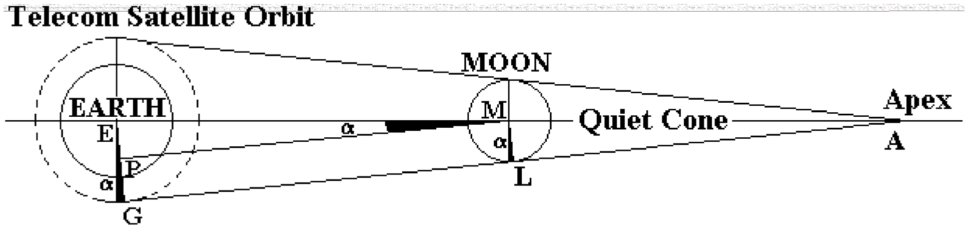


Figure 15.3: The simple geometry defining the "Terminal Longitude", on the Farside of the Moon, where radio waves emitted by telecom satellites circling the Earth at a radius R are grazing the Moon surface.

Inverting the last equation and making the substitutions described in points (1), (2) and (4), one gets the terminal longitude of radio waves on the Moon Farside (between 90° E and 180° E) emitted by a telecom satellite circling around the Earth at a distance R:

$$\lambda = atan\left(\frac{R - R_{Moon}}{\sqrt{D^2_{Earth-Moon} - (R - R_{Moon})^2}}\right) + \pi/2 \qquad (15.2)$$

Here the independent variable R can range only between 0 and the maximum value that does not make the above radical become negative, that is $0 \leq$ R $\leq D_{Earth-Moon} + R_{Moon}$. The equation above for λ shows that the $\lambda(R)$ curve becomes vertical for $R \rightarrow (D_{Earth-Moon} + R_{Moon})$ and $\lambda = 180°$

15.4 Selecting crater Daedalus near the Farside center

This author claims that the time will come when commercial wars among the big industrial trusts running the telecommunications business by satellites will lead them to grab more and more space around the Earth, pushing their satellites into orbits with apogee much higher than the geostationary one. A "safe" crater must be selected East along the Moon equator. How much further East? The answer is given by the diagram in Fig. 15.4, based on the above equation for .

The vertical trait predicted by our equation for shows up in Fig. 15.4 as the "upgoing right branch". This shows that, if we only take the equation for into account, the maximum distance from the Earth's center for these telecom satellites is about 8.479 times the geostationary radius, corresponding to a circular orbital radius of 358 148 km. If a telecom satellite were put in such a circular orbit around the Earth, its radio waves would flood Moon longitudes as high as about ≈175° or more. However we did not consider the Lagrangian points yet!

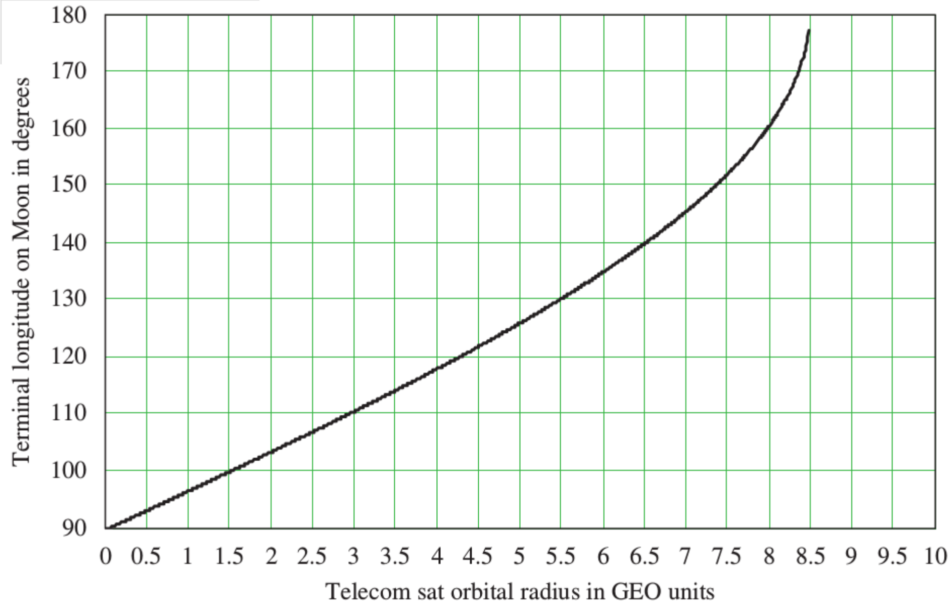


Figure 15.4: Terminal longitude λ (vertical axis) on the Moon Farside versus the telecom satellites orbital radius R around the Earth (horizontal axis) expressed in units of the Earth's geostationary radius (42 241.096 km).

So, it will never be possible to put a satellite into a circular orbit around the Earth at a distance of 358 148 km, simply because this distance already lies beyond the distance of the Lagrangian point L1 nearest to the Earth that is located at 323 050 km (Lagrangian points are, by definition, the points of zero orbital velocity in the two body problem!).

So we are now led to wonder: what is the Moon Farside terminal longitude corresponding to the distance of the nearest Lagrangian point, L1? The answer is given by the above equation for upon replacing R=323 050 km, and the result is λ=154.359°. In words, this means the following: *the Moon Farside Sector in between 154.359 E and 154.359 W will never be blinded by RFI coming from satellites orbiting the Earth alone.*

In other words, the limit of the blinded longitude as a function of the satellite's orbital radius around the Earth is 180° (E and W longitudes just coincide at this meridian, corresponding to the "change-of-date line" on Earth). But this is the antipode to Earth on the Moon surface that is the point exactly opposite to the Earth direction on the other side of the Moon. And our theorem simply proves that the antipode is the most shielded point on the Moon surface from radio waves coming from the Earth; an intuitive and obvious result, really.

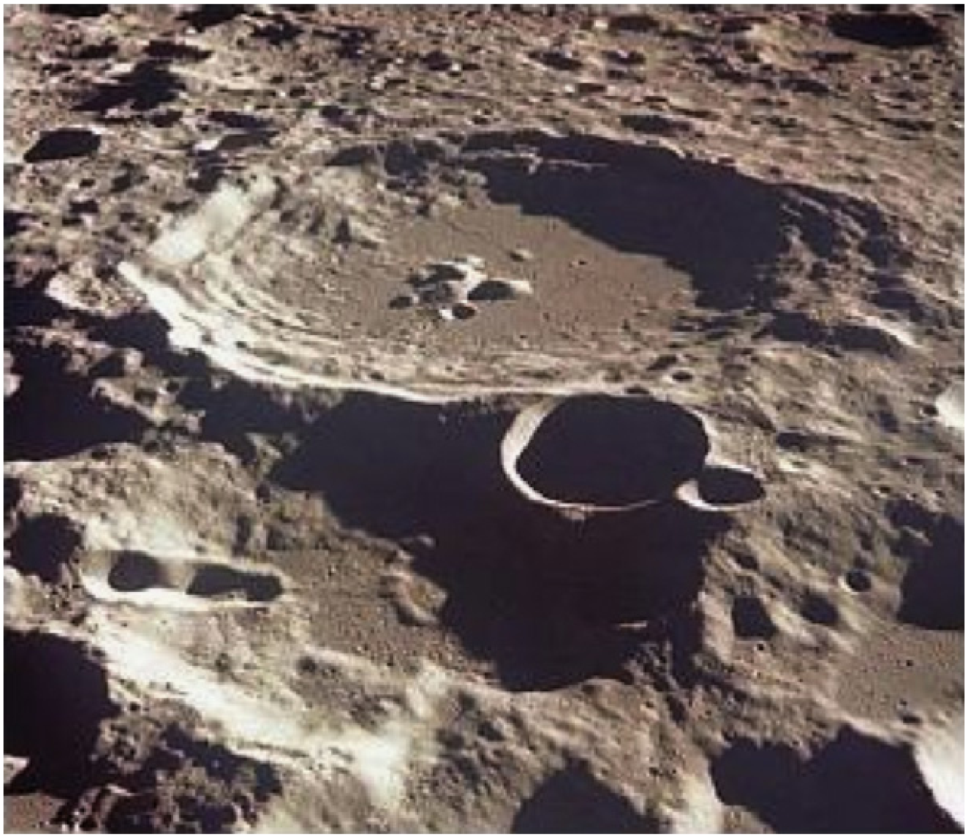


Figure 15.5: AS11-44-6609 (July 1969) – An oblique view of the Crater Daedalus on the Lunar Farside as seen from the Apollo 11 spacecraft in lunar orbit. The view looks southwest. Daedalus (formerly referred to as I.A.U. Crater No. 308) is located at 179° east longitude and 5.5° south latitude. Daedalus has a diameter of about 50 statute miles ($\approx$ 80 km). This is a typical scene showing the rugged terrain on the Farside of the Moon, downloaded from the web site: http://spaceflight.nasa.gov/gallery/images/apollo/apollo11/html/as11_44_6609.html.

So, where are we going to locate our SETI Farside Moon base? Just take a map of the Moon Farside and look. One notices that the antipode's region (at the crossing of the central meridian and of the top parallel in Fig. 15.5) is a too rugged region to establish a Moon base. Just about 5° South along the 180° meridian, however, one finds a large crater about 80 km in diameter, just like Saha. This crater is called Daedalus. So, this author proposes to establish the first RFI-free base on the Moon just inside crater Daedalus, the most shielded crater of all on the Moon from Earth-made radio pollution!

15.5 Our vision of the Moon Farside for RFI-free science

Let us replace the value of = 154.359° with the simpler value of = 150° . This matches perfectly with the need for having the borders of the Pristine Sector making angles orthogonal to the directions of L4 and L5. The result is this author's vision of the Farside of the Moon, shown in Fig. 15.6.

Fig. 15.6 shows a diagram of the Moon as seen from above its North Pole with the different "colonization regimes" proposed by this author. One sees that:

(1) The near side of the Moon is left totally free to activities of all kinds: scientific, commercial and industrial.

(2) The Farside of the Moon is divided into three thirds, namely three sectors covering 60° in longitude each, out of which:

(a) The Eastern Sector, in between 90° E and 150° E, can be used for installation of radio devices, but only under the control of the International Telecommunications Union (ITU-regime).

(b) The Central Sector, in between 150° E and 150° W, must be kept totally free from human exploitation, namely it is kept in its "pristine" radio environment totally free from man-made RFI. This sector is where crater Daedalus is, a $\approx$100 km crater located in between 177° E and 179° W and around 5° of latitude South. At the moment, this author is not aware of how high the circular rim surrounding Daedalus is.

(c) The Western Sector, in between 90° W and 150° W, can be used for installation of radio devices, but only under the control of the ITU-regime. Also:

(d) The Eastern Sector is exactly opposite to the direction of the Lagrangian point L4, and so the body of the Moon completely shields the Eastern Sector from RFI produced at L4. Thus, L4 is fully "colonizable".

(e) The Western Sector is exactly opposite to the direction of the Lagrangian point L5, and so the body of the Moon completely shields the Western Sector from RFI produced at L5. Thus, L5 is fully "colonizable" in this author's vision. In other words, this author's vision achieves the full bilateral symmetry around the plane passing through the Earth–Moon axis and orthogonal to the Moon's orbital plane.

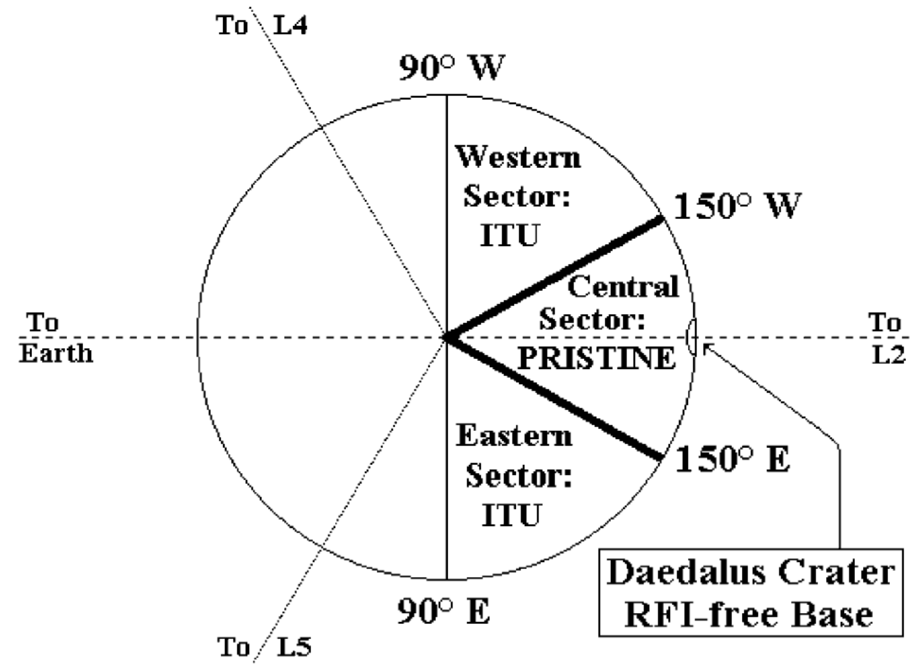


Figure 15.6: This author's vision of the Moon Farside Protection. The Central Sector of the Farside extends from the North to the South Pole in between the two meridians at 150° E and 150° W. The protected antipode circle (PAC) is the circular piece of land tangent to these two meridians at the Moon's equator. It also is tangent to the two parallels at +30° and -30°. We claim that the PAC should be kept free from man-made electromagnetic pollution at all times in the future, i.e. that it should become an INTERNATIONALLY PROTECTED LAND, just as Antarctica is on Earth.

(f) Of course, L2 may not be utilized at all, since it faces crater Daedalus just at the latter's zenith. Any RFI-producing device located at L2 would flood the whole of the Farside, and must be ruled out. L2, however, is the only Lagrangian point to be kept free, out of the five located in the Earth–Moon system. Finally, L2 is not directly visible from the Earth since it is shielded by the Moon's body, which calls for "leaving L2 alone"!

15.6 The further two Lagrangian points L1 and L2 of the Sun–Earth system: their "polluting" action on the Farside of the Moon

There still is an unavoidable drawback, though.

This is coming from the further two Lagrangian points L1 and L2 of the Sun–Earth system, located along the Sun–Earth axis and outside the sphere of influence of the Earth that has a radius of about 924 646 km around the Earth. Precisely, the Sun–Earth L1 point is located at a distance of 1 496 557.035 km from the Earth toward the Sun, and the L2 point at the (virtually identical) distance of 1 496 557.034 km from the Earth in the direction away from the Sun, which is toward the outer solar system. These two points have the "nice" property of moving around the Sun just with the same angular velocity as the Earth does, while also keeping at the same distance from the Earth at all times. Thus, they are ideal places for scientific satellites.

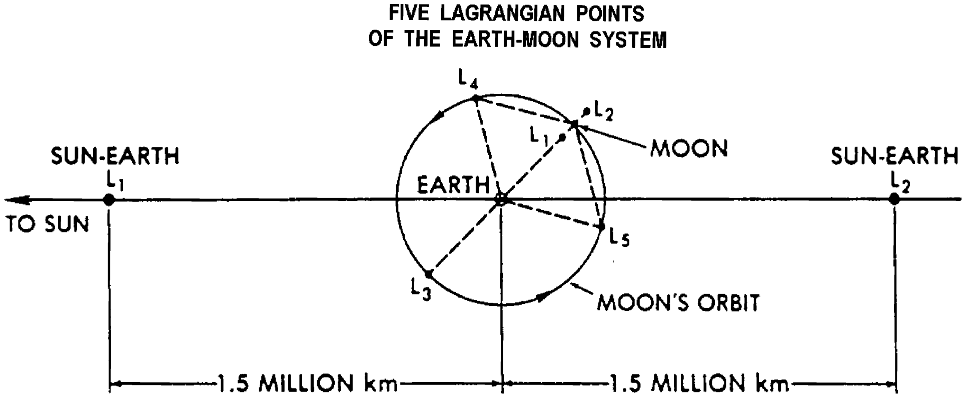


Figure 15.7: (Courtesy of Dr. Robert “Bob” Farquhar, Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA). In addition to the five Lagrangian Points of the Earth–Moon system (already described in Fig. 15.1) the next two closest Lagrangian Points to the Earth are the Lagrangian Points L1 and L2 of the Sun–Earth system. These are located along the Sun–Earth axis at the distances of about 1.5 million kilometers from the Earth toward the Sun (L1) and outward (L2). Unfortunately, spacecrafts located in the neighborhood of these L1 and L2 Sun-Earth Points do send electromagnetic waves to the Farside of the Moon. Examples are the ISEE-III and Soho spacecrafts, already orbiting around L1, and more spacecrafts will do so in the future around both L1 and L2.

Table 15.1: Radio waves attenuation in the lunar equatorial plane and at lunar longitude λ=180° (i.e. near the Daedalus crater) for radio sources emitting at 100 kHz, 100 MHz and 100 GHz, respectively.

Frequency or radio waves	f = 100 kHz	f = 100 MHz	f = 100 GHz
Source in GEO (dB)	-42.62	-72.62	-102.62
Source in an orbit passing through the L1 point (dB)	-30.32	-60.32	-90.32
Source still at L4 or L5 Lagrangian points (dB)	-29.15	-59.15	-89.15

All attenuation values are in dB.

Actually, the Sun–Earth L1 Point has already been in use for scientific satellite location since the NASA ISEE III spacecraft was launched on August 12, 1978 and reached the Sun–Earth L1 region in about a month.

On December 2, 1995, the ESA-NASA "Soho" spacecraft for the exploration of the Solar Corona was launched. On February 14, 1996, Soho was inserted into a halo orbit around the Sun-Earth L1 point, where it is still librating now (2007) (Fig. 15.7).

As for the Sun–Earth L2 point, there are plans to let the NASA's SIM (Space Interferometry Mission) satellite be placed there, as will be ESA's GAIA astrometric satellite as well.

So, all these satellites do "POLLUTE" the otherwise RFI-free Farside of the Moon when the Farside is facing them. Unfortunately, the Moon Farside is facing the Sun–Earth L1 point for half of the Moon's synodic period, about 14.75 days, and it is facing the Sun–Earth L2 point for the next 14.75 days. Really all the time!

This radio pollution of the Moon Farside by scientific satellites located at the Lagrangian Points L1 and L2 of the Sun-Earth system is, unfortunately, Unavoidable. We can only hope that telecom satellites will never be put there. As for the scientific satellites already there or on the way, the radio frequencies they use are well known and usually narrow band. This should help the Fourier transform of the future spectrum analyzers to be located on the Moon Farside to get rid of these transmissions completely.

15.7 Attenuation of man-made RFI on the Moon Farside

In a recent paper presented by this author at the International Astronautical Congress held in Valencia in October 2006, his co-worker Salvo Pluchino succeeded in computing the RFI attenuation on the Farside of the Moon [3]. Basic results proven there

Table 15.2

Origin of radio wave	Radio frequency f	Source in GEO (dB)	Source in orbit distance (dB)	Source still at L4 or L5 (dB)
ELF	0.003 MHz	-27	-15	-14
VLF	0.030 MHz	-37	-25	-23
Jupiter's storm	20 MHz	-53	-53	-23
Deuterium	327.384 MHz	-77	-65	-64
Hydrogen	1420.406 MHz	-84	-71	-70
Hydroxyl radical	1612.231 MHz	-84	-72	-71
Formaldehyde	4829.660 MHz	-89	-77	-75
Methanol	6668.518 MHz	-90	-78	-77
Water vapor	22.235 GHz	-96	-83	-82
Silicon monoxide	42.519 GHz	-98	-86	-85
Carbon monoxide	109.782 GHz	-103	-90	-90
Water vapor	183.310 GHz	-105	-92	-91

are the RFI attenuation values shown in Table 15.1 hereafter. Perhaps even more important than the "generic" frequency values listed in Table 15.1 are the following precise line frequencies high scientific importance again taken from the paper [3]. In practice, these are the attenuations of man-made RFI to be expected at crater Daedalus and within the PAC. It should also be stated that these are the attenuation values assuming that the Moon is not surrounded by a very thin ionosphere. Since a very tiny Lunar Ionosphere might possibly exist, however, the values below might be slightly incorrect (Table 15.2).

15.8 Legal issues: a possible strategy to have the PAC approved by the United Nations (COPUOS)

Having thus made clear what the PAC is and why its creation is urgently needed for the benefit of all humankind, our next issue is legal: how can we have the PAC become legally approved by the United Nations?

To answer this question, let us first review the function of the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS).

As described at the relevant official web site[1] COPUOS was set up by the United

[1] http://www.unoosa.org/oosa/COPUOS/copuos.html

Nations General Assembly in 1959 (resolution 1472 (XIV)) to review the scope of COPUOS, to devise programs in this field to be undertaken under United Nations auspices, to encourage continued research and the dissemination of information on outer space matters and to study legal problems arising from the exploration of outer space. As of 2007, the number of member states of COPUOS is 67, and the relevant list can be found at the above web site.

The Committee has two standing Subcommittees of the whole:

- the Scientific and Technical Subcommittee and
- the Legal Subcommittee.

The important points that may lead to the approval of the PAC by COPUOS seem to be the following two:

(1) The Committee and its two Subcommittees meet annually to consider questions put before them by the General Assembly, reports submitted to them and issues raised by the member states.

(2) The Committee and the Subcommittees, working on the basis of consensus, make recommendations to the General Assembly. Detailed information on the work of the Committee and the Subcommittees is contained in their annual reports.

Because of point (1) it is pretty clear that there must be at least one member state of the United Nations raising the issue of the PAC creation so that COPUOS can later make any recommendation to the General Assembly.

Now, it happens that the IAA is only an observer at the United Nations level, and so it cannot raise issues. Issues can be raised at the United Nations only by member states, and so the steps to follow seem to be:

(1) Step one: let at least one member state of the United Nations raise the issue of the PAC creation at the United Nations. After Step one has been accomplished, COPUOS may well ask the IAA and the IISL (the International Institute of Space Law) about technical clarifications regarding the PAC. Thus, in preparation of this request for information by COPUOS to the IAA and the IISL, step two may be started now.

(2) Step two: the IAA and the IISL will provide COPUOS with all technical information and clarifications about PAC. In view of this, the author, as an IAA member, has asked for permission to lead an IAA Study Group to study the Moon Farside Protection and the PAC. When this phase has been accomplished too, the COPUOS may finally take step three.

(3) Step three: the COPUOS may finally either:

(a) submit the PAC issue to the General Assembly of United Nations for a vote or

(b) include the PAC issue in the "New Moon Treatise" that has long been envisaged by some COPUOS members as the "good replacement" to the "bad Moon Treatise of 1979".

So, the story now goes over to the already existing Outer Space Treatise of 1967 (sometimes called "the good Moon treatise" in the legal space jargon) and the (failed, in practice) Moon Treatise of 1979 (sometimes called "the bad Moon treatise" in the legal space jargon), as described briefly in the next section.

15.9 The existing "Outer Space Treatise" of 1967 and the (failed) "Moon Treatise" of 1979

In order to understand the issue of the two key space treatises of 1967 and 1979, the reader may wish to consult the relevant two Wikipedia sites:

(1) For the Outer Space Treatise of 1967[2].

(2) For the 1979 Moon Treatise (proposed, but never ratified by all the major space-faring nations)[3].

Figs. 15.8 and 15.9 are reproduced from the above two web sites, respectively. We shall now express our opinion about these two treatises, an opinion that may be of course debatable

but points out the great state of uncertainty about "who owns what" on the Moon surface today and tomorrow.

(1) The Outer Space Treatise signed in 1967 by the USA, Great Britain and the Soviet Union was "wise" in that it forbade any country from "colonizing" the Moon and planets and asteroids in the name of one country only. As such, no wonder it was later signed and ratified by virtually all the countries in the world. Back in 1967, however, it was hard to envisage a time when private investors could reach the Moon and other bodies by virtue of their own private means only. By 1980 the situation had already evolved so much in this regard that...

(2) When the Moon Treatise of 1979 was finalized and proposed, the United States Senate finally failed to approve it because it forbade both private property and the capability of terraforming (i.e. altering the atmosphere so as to "make it breathable"). In this author's view, the terraforming of the Moon (if at all feasible!) and Mars will be necessary in the long run. So, it was good that the USA and the

[2] http://en.wikipedia.org/wiki/Outer_Space_Treaty

[3] http://en.wikipedia.org/wiki/Moon_Treaty

rest of the world did not sign the 1979 Treaty. By doing so, however, they left the door open to those who claim that private property on the Moon will be a good thing.

(3) And this is precisely the issue: can single individuals own pieces of land on the Moon? The vast majority of legal experts would probably answer "no" to this question, but the matter is not settled at all yet, as one may see at the Wikipedia site about "Extraterrestrial Real Estate"[4]

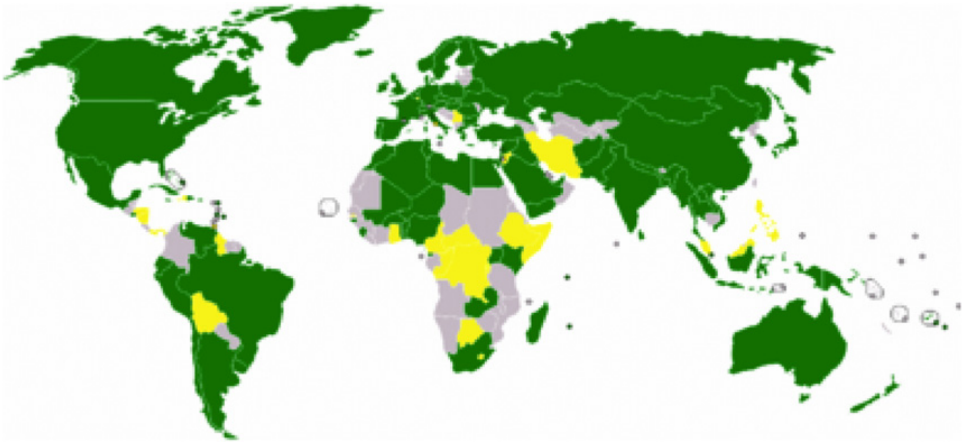


Figure 15.8: Countries that signed and ratified the Outer Space Treatise of 1967 (green) and countries that signed it but did not ratify it (yellow) (taken from the Wikipedia site http://en.wikipedia.org/wiki/Outer_Space_Treaty).

[4] http://en.wikipedia.org/wiki/Extraterrestrial_real_estate.

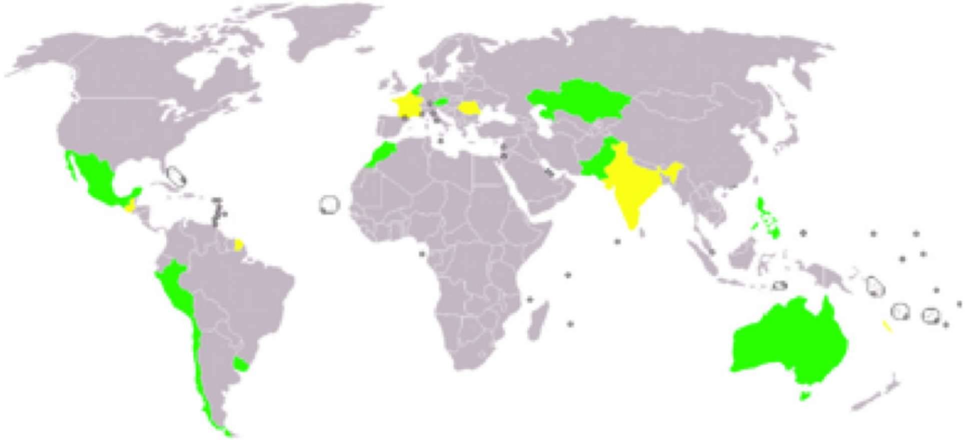


Figure 15.9: Countries that signed and ratified the Moon Treatise of 1979 (green) and countries that signed it but did not ratify it (yellow) (taken from the Wikipedia site http://en.wikipedia.org/wiki/Moon_Treaty)

In the USA there have unfortunately been private companies like "The Lunar Embassy" selling real estate on the Moon and they have made quite a bit of money out of this sale! Please see the web site[5] .

Yet, nobody has done anything in the USA to prevent this illegal sale. Only in China the "Lunar Embassy" firm was recently prevented from doing so by the government: see the web site[6].

This author does not know whether the Lunar Embassy already sold any piece of land within the borders of the PAC, but the issue of having the PAC recognized as a sort of "Natural Preserve" on the Farside of the Moon is urgent indeed.

15.10 Conclusions

The goal of this paper was to make the readers sensitive to the importance of protecting the Central Farside of the Moon from any future wild, anti-scientific exploitation.

In particular, we gave sound scientific reasons why the PAC, should be declared an international land under the Protection of the United Nations, or, in absence of that institution, by direct agreement among the governments of the space-faring nations.

[5] http://www.lunarlandowner.com/

[6] http://www.moondaily.com/reports/China_Bans_Firm_From_Selling_Land_On_The_Moon_999.html

The Farside of the Moon is a unique place for us in the whole universe: it is close to the Earth, but protected from the radio garbage that we ourselves are creating in an ever-increasing amount that is making our radio telescopes blinder and blinder.

The Farside cannot be left in the realtors' hands! And this is an urgent matter!

Some international agreement must be taken for the benefit of all humankind.

Acknowledgments

The author would like to thank the International Academy of Astronautics (IAA) for allowing him to serve as Coordinator of the "Lunar Farside Radio Lab" IAA Cosmic Study in the years 2000–2005.

Interest in the international role that the United Nations might play in the creation of the PAC also came from several members of the International Institute of Space Law (IISL), in particular Prof. Vladimir Kopal, Chair of the IAA "Scientific-Legal Liaison Committee". Hopefully, these and other far-sighted minds will unite their efforts to save the Farside of the Moon from new man-made RFI pollution right over there!

Finally, the support of the author's co-workers Salvo Pluchino and Nicolò Antonietti is gratefully acknowledged in the study of the mathematical problems to quantify the RFI on the Moon Farside.

References

1. C. Maccone, The quiet cone above the farside of the moon, Acta Astronautica 53 (2003) 65–70.

2. C. Maccone, Moon farside radio lab, Acta Astronautica 56 (2005) 629–639.

3. S. Pluchino, N. Antonietti, C. Maccone, Protecting the moon farside radiotelescopes from RFI produced at future Lagrangian-points space stations, Paper IAC-06-D4.1.01 presented at the International Astronautical Congress held in Valencia, Spain,October 2–6, 2006.

Prologue

The conference venue, The Institute of Applied Astronomy.

The conference room. Claudio Maccone listening to the English translation for a discussion between A.Finkelstein and M.Y. Marov (in Russian).

A. Finkelstein and A. Panov

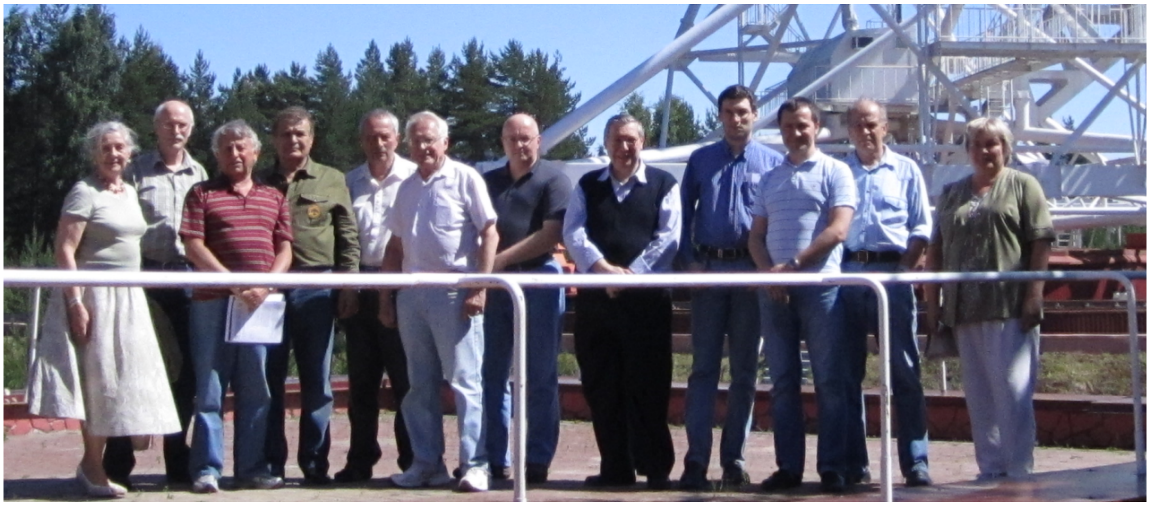
A. Zaitzev and S. Dumas.

M.Ya. Marov and L.M. Gindilis

N. V. Sokulina and Y.N. Efremov

C. Maccone and G. Gontcharov

S.A. Yasev



Group picture at the Svetloe observatory, including people from the conference and observatory.