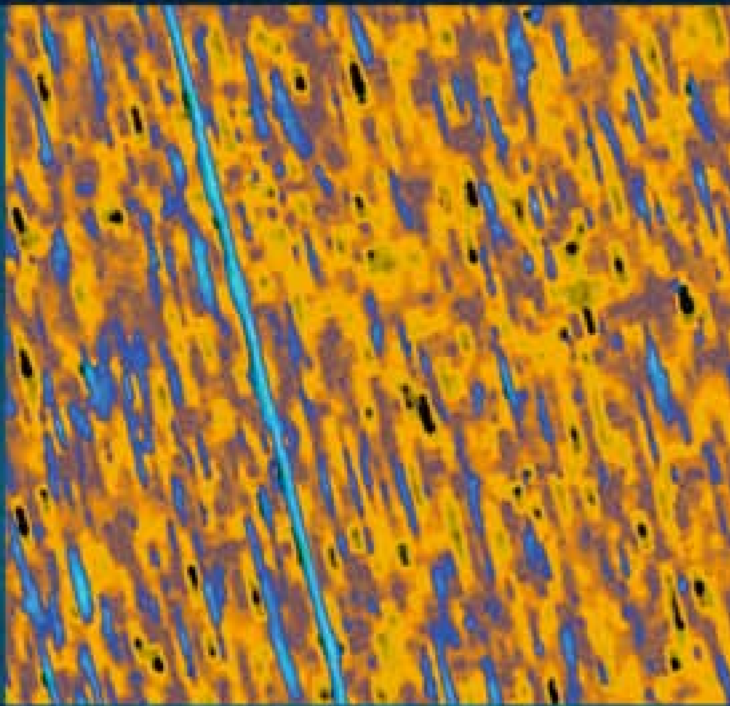
THE FRONTIERS COLLECTION



H. Paul Shuch
*Editor*

# SEARCHING FOR EXTRATERRESTRIAL INTELLIGENCE

## SETI Past, Present, and Future

PRAXIS

Springer

THE FRONTIERS COLLECTION

# THE FRONTIERS COLLECTION



The books in this collection are devoted to challenging and open problems at the forefront of modern science, including related philosophical debates. In contrast to typical research monographs, however, they strive to present their topics in a manner accessible also to scientifically literate non-specialists wishing to gain insight into the deeper implications and fascinating questions involved. Taken as a whole, the series reflects the need for a fundamental and interdisciplinary approach to modern science. Furthermore, it is intended to encourage active scientists in all areas to ponder over important and perhaps controversial issues beyond their own speciality. Extending from quantum physics and relativity to entropy, consciousness and complex systems – the Frontiers Collection will inspire readers to push back the frontiers of their own knowledge.

## *Other Recent Titles*

**Weak Links**
Stabilizers of Complex Systems from Proteins to Social Networks
By P. Csermely

**The Biological Evolution of Religious Mind and Behaviour**
Edited by E. Voland and W. Schiefenhövel
Particle Metaphysics

**A Critical Account of Subatomic Reality**
By B. Falkenburg

**The Physical Basis of the Direction of Time**
By H.D. Zeh

**Mindful Universe**
Quantum Mechanics and the Participating Observer
By H. Stapp

**Decoherence and the Quantum-To-Classical Transition**
By M. Schlosshauer

**The Nonlinear Universe**
Chaos, Emergence, Life
By A. Scott

**Symmetry Rules**
How Science and Nature are Founded on Symmetry
By J. Rosen

**Quantum Superposition**
Counterintuitive Consequences of Coherence, Entanglement, and Interference
By M.P. Silverman

Series home page – springer.com

H. Paul Shuch
Editor

# SEARCHING FOR EXTRATERRESTRIAL INTELLIGENCE:

## SETI PAST, PRESENT, AND FUTURE

Springer

Published in association with
Praxis Publishing
Chichester, UK

PRAXIS

*Editor*

H. Paul Shuch
The SETI League, Inc.
P O Box 555
433 Liberty Street
Little Ferry
New Jersey NJ 07643
USA
paul@setileague.org

*Series Editors:*

Avshalom C. Elitzur
Bar-Ilan University, Unit of Interdisciplinary Studies, 52900 Ramat-Gan, Israel
email: avshalom.elitzur@weizmann.ac.il

Laura Mersini-Houghton
Dept. Physics, University of North Carolina, Chapel Hill, NC27599-3255, USA
email: mersini@physics.unc.edu

Maximilian A. Schlosshauer
Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark
email: schlosshauer@nbi.dk

Mark P.Silverman
Trinity College, Dept. Physics, Hartford CT06106, USA
email: mark.silverman@trincoll.edu

Jack A.Tuszynski
University of Alberta, Dept. Physics, Edmonton ABT6G1Z2, Canada
email:jtus@phys.ualberta.ca

Rüdiger Vaas
University of Giessen, Center for Philosophy and Foundations of Science, 35394 Giessen, Germany
email:ruediger.vaas@t-online.deH.

Dieter Zeh
Gaiberger Straße 38, 69151 Waldhilsbach, Germany
email:zeh@uni-heidelberg.de

ISSN 1612-3018 4
ISBN 978-3-642-13195-0 e-ISBN 978-3-642-13196-7
DOI 10.1007/978-3-642-13196-7
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010936460



Cover design: Jim Wilkie

Cover figure: This cover uses a motif from the SETI League photo http://www.setileague.org/photos/hits/carrier.jpg, used by permission (http://www.setileague.org/).

Copy editor: Rachael Wilkie

Typesetting: David Peduzzi

Printed on acid-free paper

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

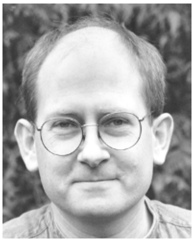
This collection of essays is presented in honor of SETI patriarch Frank D. Drake on the occasion of his 80th birthday. Its publication marks a half-century of observational SETI science.

H. Paul Shuch
Cogan Station, PA, USA
May 2010

# Contents

# About the Authors

**Peter R. Backus** is a radio astronomer who began working in SETI as a post-doctoral fellow in 1982. He joined the SETI Institute in 1985 and was a Principal Investigator in the NASA SETI Program. He continued that work in the Institute's Project Phoenix, and now manages SETI observations at the Allen Telescope Array.

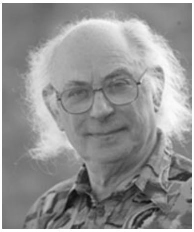


Born in Liverpool, England, **Stephen Baxter** has degrees in mathematics and engineering, has worked as a teacher in mathematics and physics, and since 1987 has published over 40 books, mostly science fiction novels, which have been published internationally and have received many awards. He is a Chartered Engineer, Fellow of the British Interplanetary Society, President of the British Science Fiction Association, and Vice-President of the HG Wells Society.

An Oxford-educated physician, **John Billingham** specialized in Aviation Medicine in the UK Royal Air Force. He held positions in aerospace medicine, exobiology, and SETI at the NASA Ames Research Center, including Chief of the Extraterrestrial Research Division and the SETI Office, receiving the NASA Medal for Outstanding Leadership for Achievements in SETI. Now Senior Scientist and a Member of the Board of Trustees of the SETI Institute, he was inducted into the NASA Ames Hall of Fame for contributions to space medicine, astrobiology, and SETI.

Astronomer **Stuart Bowyer** is recognized as having started the field of extreme ultraviolet astronomy. He became involved in SETI research as a professor at the Space Sciences Laboratory, University of California, Berkeley, where he headed the SERENDIP program.

**David Brin** is a scientist and bestselling novelist, who has won the Hugo and Nebula awards. Much of his writing features provocative thought experiments about alien intelligence and SETI. He lives in San Diego county with his wife, three children, and a hundred very demanding trees.



**Jack Cohen** is an internationally known reproductive biologist. His last position, at Warwick University in the United Kingdom, bridged the Ecosystems Unit of the Biology Department and the Mathematics Institute. His hobbies include boomerang throwing and keeping strange animals: from Hydras to mantis shrimps, and octopuses to llamas.

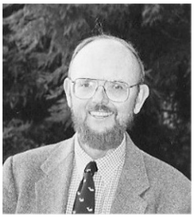
**Kathryn Denning** teaches anthropology and archaeology at York University in Canada, including the anthropology of space exploration. She has been actively researching the scientific culture and discourses of SETI since 2004, and has published social scientific perspectives on SETI subjects ranging from interstellar communication to social evolution to contact. Along with many other contributors to this book, she serves on the SETI Permanent Study Group of the International Academy of Astronautics.

**Steven J. Dick** served as NASA Chief Historian from 2003–2009. He has written on the history of SETI and the extraterrestrial life debate, most notably in *The Biological Universe* (Cambridge University Press, 1996), and *The Living Universe* (Rutgers University Press, 2004). Minor planet 6544 Stevendick is named in his honor.

**Robert Dixon** has served as Assistant and Acting Director of the Ohio State University Radio Observatory, and Chief Research Engineer of the Ohio Academic Resources Network. A Faculty Fellow in the NASA SETI projects *Cyclops* and *OASIS*, he directed the Ohio SETI program, the first full-time search with a large radio telescope, which ran from 1973 to 1997. Bob initiated and popularized the use of the Flag of Earth at SETI worldwide, conceived the Argus omnidirectional radio telescope design, and currently directs its development.

In 1967 Dr John D. Kraus, Director of the Ohio State University Radio Observatory (OSURO), hired physicist **Jerry Ehman** to work as a faculty member in the Ohio State University department of Electrical Engineering, and as a radio astronomer with the "Big Ear" radio telescope. At the OSURO Ehman worked on data analysis for the Ohio Sky Survey (in which over 19,000 radio sources were measured, over half of which had not been seen before). In August 1977, Jerry Ehman wrote the most memorable work of his career: a single word, "Wow!", on a Big Ear computer printout.

**Richard Factor** stepped onto the SETI stage when the U.S. Congress terminated funding for NASA's SETI projects in 1993. Dismayed by that short-sighted political decision, he founded the SETI League to privatize SETI, and continues to serve vigorously as the League's first president. He is a science buff and science fiction fan who wishes he had more time to pursue research and reading.



**Albert A. Harrison** is an emeritus professor of psychology at the University of California, Davis. He has written a number of SETI-related books, and regularly presents papers at various SETI conferences. He is a board member for the annual CONTACT conference, a regent of United Societies in Space, and a member of the National Institute of Discovery Science (NIDS) Science Advisory Board.

**Paul Horowitz** is a Professor of Physics and of Electrical Engineering at Harvard University, where he arrived as an impressionable freshman in 1961 and, well, just never left. He likes to design and build cool stuff, like x-ray microscopes, landmine detectors, cold light stabilizers, and insane new SETIs. He is the co-author, with Winfield Hill, of *The Art of Electronics*.

**Stuart Kingsley** is an expert in fiber optics, with a Ph.D. degree in electronic and electrical engineering. Since 1990, he has vigorously promoted the idea of optical SETI, having created an extensive website and spearheading three optical SETI conferences. Largely due to his efforts, optical SETI has now gained much greater acceptance within the SETI field than it had just a few years ago.

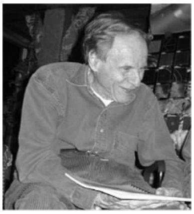
**Eric Korpela** is a research astronomer and the Project Scientist of the SETI@home and Astropulse projects at the University of California, Berkeley. In addition to searching for extraterrestrials, Eric studies interstellar matter at wavelengths ranging from the soft X-ray to the radio. Outside of work, Eric enjoys collecting vintage computers, recreational programming, and bass fishing.

As a graduate student of applied physics at Harvard University, electrical engineer **Darren Leigh** worked closely with Professor Paul Horowitz on the development of the world's most advanced multi-channel spectrum analyzers for SETI data analysis. He designed hardware and software for the Billion-channel ExtraTerrestrial Assay (Project BETA) search for narrow-band carriers.

Italian space scientist **Claudio Maccone**, an active leader and innovator at meetings on interstellar travel and on SETI, studied on scholarships in London, New York, and Turin. He won a 1999 Book Award from the International Academy of Astronautics (IAA). Claudio is a member of IAA and COSPAR, and serves as secretary of the IAA Interstellar Space Exploration Committee and co-chair of the SETI Permanent Study Group.

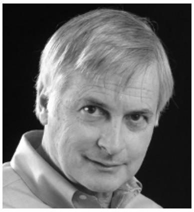
**Stelio Montebugnoli** is working at the National Institute for Astrophisic (INAF) as engineer in charge of the Medicina radiotelescopes (Bologna, Italy). He leads the Seti-Italia program with a Serendip IV system that is connected in piggy back mode to a 32 meter dish antenna. He leads the INAF group involved in Square Kilometer Array (SKA) design and space debris monitoring activities.

SETI patriarch **Philip Morrison** (1915–2005), Institute Professor and Professor of Physics at the Massachusetts Institute of Technology, was a distinguished theoretical astrophysicist and a pioneer in the search for extraterrestrial intelligence through radio communication. He authored scores of books, produced television documentaries, and lectured tirelessly around the world, despite the physical limitations imposed upon him by Post-Polio Syndrome. Phil Morrison co-authored the first scientific SETI paper in 1959, chaired NASA's early study groups on SETI, and was a mentor, friend, or father figure to all of this book's authors.



**Monte Ross**, an electrical engineer, is a pioneer in laser communications and in the concept of optical SETI. He received a Fellow award from IEEE and another from McDonnell-Douglas, both awards for his work in space laser communications. He is currently president and CEO of a company that produces electronic information systems.

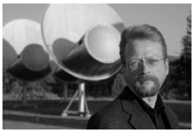

A retired professor of physics, astronomy, and engineering, **H. Paul Shuch** is credited with the design of the first commercial home satellite TV receiver. As executive director (now emeritus) of the grass-roots nonprofit SETI League since 1994, he has been redirecting that technology toward the development of low-cost amateur radio telescopes.  A singer, songwriter, and guitarist of questionable talent, Paul gives technical lecture/concerts around the world under the stage name Dr SETI.



**Seth Shostak** is an astronomer with the SETI Institute in Mountain View, California where he conducts research using, among other instruments, the Allen Telescope Array. He is also responsible for much of the Institute's outreach activity, and is currently Chair of the International Academy of Astronautics SETI Permanent Study Group.

**Jill Tarter** is Director of the Center for SETI Research at the SETI Institute in Mountain View, California. Tarter received her Ph.D. in Astronomy from the University of California, Berkeley, and has conducted numerous observational programs at radio observatories worldwide. Currently, she serves on the management board for the Allen Telescope Array, a joint project between the SETI Institute and the UC Berkeley Radio Astronomy Laboratory.

SETI researcher and a futurist **Allen Tough** is a professor emeritus at the University of Toronto. He is a long-term member of the international SETI committee of the International Association of Astronauts, and presented papers at their annual conference for several years. Allen is editor of the book *When SETI Succceeds*, and the founder of the Invitation to ETI.



If we detect extraterrestrial intelligence, how will people around the world respond? Should we reply? If so, what should we say? One scientist who grapples with these questions is **Doug Vakoch**, Director of Interstellar Message Composition at the SETI Institute. Doug coordinates an international team of scholars from the arts, sciences, and humanities to ponder how we might create messages that would represent Earth's diverse cultures.

**Stephen Webb** has worked at several universities in the UK, but now seems to be settled at the University of Portsmouth. He lives not too far away from the LOFAR array at Chilbolton (part of a pan-European facility that will, among other things, contribute to the SETI program by searching for low-frequency radio signals). He is the author of several books, including one on the Fermi paradox, and is currently researching a book about the next generation of astronomical observatories.

**Alexander L. Zaitsev**, Chief Scientist at the Institute of Radio Engineering and Electronics, Russian Academy of Science, supervised the transmission of the 1999 and 2003 Cosmic Calls from the Evpatoria Planetary Radar. Under his scientific leadership, a youth group in Moscow composed and broadcast a "Teen Age Message to ETI". Zaitsev proposed a three-section structure of interstellar radio messages, coined the acronym METI (Messaging to Extra-Terrestrial Intelligence) and the phrase "SETI Paradox", which 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.

# Foreword: Looking Back

Steven J. Dick,
*Former NASA Chief Historian*

In the long history of the extraterrestrial life debate, April 8, 1960 stands out for its sheer audacity. On that date, for the first time in history, the young astronomer Frank Drake, just shy of his 30th birthday, utilized the new technology of radio telescopes to listen for intelligent signals beyond Earth. Turning the 85-foot Tatel radio telescope of the National Radio Astronomy Observatory in Green Bank, West Virginia skyward for this purpose was a brash act of youth, hope and courage. Although in the 1950s astronomers had begun to believe other stars might harbor planets, the subject of extraterrestrials was not yet reputable and was bound to cause controversy. Drake's observations, backed by NRAO Director Otto Struve, did indeed cause a stir, with ripples that continue fifty years later.

Drake was the first to attempt this experiment, but not the first to think of it. On January 20, 1919, the *New York Times* ran a front-page headline "Radio to Stars, Marconi's hope," in which the radio pioneer Guglielmo Marconi expressed the belief that the eternal properties of radio waves "makes me hope for a very big thing in the future … communication with intelligences on other stars … It may some day be possible, and as many of the planets are much older than ours the beings who live there ought to have information for us of enormous value." Suggesting that mathematics might be used as a language of communication, Marconi even claimed he had received unexplained signals that might have originated from the stars. This claim prompted the *Times*, in a lengthy editorial, to advise that humanity should "Let the Stars Alone," lest we receive knowledge "for which we are unprepared precipitated on us by superior intelligences."

That paragraph contains in a nutshell many of the persistent themes of what would become known as the Search for Extraterrestrial Intelligence (SETI). In a famous paper in *Nature* in 1959, physicists Giuseppe Cocconi and Philip Morrison argued that radio waves were indeed the ideal wavelength to search for intelligent interstellar communications. That has ever since been the main overture of the SETI symphony, though optical SETI and other methods have also been tried. For one brief shining moment even NASA had an observational SETI program, before the forces of ridicule prevailed. The idea that other civilizations would be immensely older than us, perhaps by millions or even billions of years, has also held up to our increasing knowledge of the universe, though SETI practitioners have in general failed to take into account the implications of this fact for the nature of intelligence.

Marconi's claim that extraterrestrials might have information of great value, measured against the fears of the *New York Times* that we should leave the stars alone, also foreshadows the hopes of most SETI enthusiasts and the fears of other scientists, including Nobelists Martin Ryle and George Wald and (most recently) Stephen Hawking. The perceived dangers and benefits of SETI give urgency to the debate, actively engaged today by members of the SETI Permanent Study Group of the International Academy of Astronautics, among other stakeholders, as to whether SETI activities should be limited to searching for extraterrestrial signals, or whether we should pro-actively "message" extraterrestrial intelligence (METI).

The fact that no signals have yet been unambiguously detected by Marconi or anyone else gives force to the so-called "Fermi Paradox": given the billions-of-years time scales involved, any extraterrestrials engaged in interstellar travel should have arrived on Earth long ago. Some claim that they have arrived in the form of UFOs, but the extraterrestrial hypothesis for UFOs is an extraordinary claim and, as Carl Sagan cautioned us, extraordinary claims require extraordinary evidence. Alas, such evidence is lacking.

In an attempt to estimate the number of technological civilizations in the Galaxy, the U. S. National Academy of Sciences sponsored a small invitation-only meeting at Green Bank from October 30 to November 3, 1961. Casting about for an agenda, Drake placed on the board an equation with the relevant parameters for this estimate: the rate of star formation, the fraction of Sun-like stars with planets, the fraction of those planets on which life could develop, the fraction of those where life actually does develop, the fraction on which intelligence evolves, and the lifetime of a radio-communicative technological civilization, the all important factor known as L. It has been noted that the so-called Drake Equation, inadequate as it is in yielding a definitive number, serves as an excellent heuristic device – which was, after all, its original intention. The equation encompasses astronomical, biological and cultural evolution, it unites cosmos and culture, and raises more questions than it answers, assuring the continuing vitality of SETI across a spectrum of disciplines.

As we begin the 21st century and the Third Millennium we are indeed increasingly aware of the intimate connections of cosmos and culture, that we are star stuff, as Sagan liked to say. The presence or absence of extraterrestrials has great implications for human destiny, including philosophy, religion and science. Conversely, the history of humanity tells us that where intelligence exists, cultural evolution not only flourishes but rapidly outpaces all other forms of evolution, whether physical or biological. Accordingly, the nature of extraterrestrials and their cultures may exceed our imaginations. The universe may hold in store great surprises.

Steven J. Dick
Former NASA Chief Historian

# Preface

The notion that humanity shares the universe with other sentient beings probably predates recorded history. Since first we realized that the points of light in the night sky are other suns, we have gazed heavenward, and asked "are we alone?" In the mid-20th century, for the first time, we began to develop the technologies with which we might seek a definitive answer to this ages-old question. You hold in your hand a collection of essays by the very scientists and technologists who have led, and continue to lead, the scientific quest known as SETI, the Search for Extra-Terrestrial Intelligence.

This book is written for the educated and informed layperson, one who is technically competent though not necessarily a specialist in the varied disciplines that comprise SETI science. It should appeal to SETI technologists present and future, students of science history, space and astronomy buffs, any radio amateur who ever listened to meteor pings or microwave signals bounced off the moon, and everyone who was ever a kid peering through a backyard telescope, counting the lunar craters and the Jovian moons. It contains overviews basic enough for the technically literate general public, mathematical analyses detailed enough to challenge any academic, and a host of intellectual levels in between. We, the authors, invite you to seek your own level of comfort, and then to challenge yourself, to reach beyond it.

The 26 chapters in this book are grouped into three distinct sections, each of which could well comprise a book in its own right. Not being one to commit blatant acts of trilogy, I have combined these three books under a single cover. Here is how they break down.

*The Spirit of SETI Past*, written by the surviving pioneers of this emerging discipline, documents the first 50 years of SETI science. It begins with a brief overview of SETI terminology and techniques, penned by me, to bring you quickly up to speed on what is to follow. But the true story of observational SETI science starts with Frank Drake, who conducted the very first such experiment a half-century ago. Frank's modesty prevents him from boasting about his own place in history, so that honor falls to me in Chapter 2. Chapter 3, detailing NASA's ambitious Project Cyclops proposal, is penned by Bob Dixon of Ohio State University, one of the (sadly) few surviving members of that study team. There follows a detailed treatment of the most tantalizing

SETI candidate signal ever encountered, by the "Wow!" signal's discoverer, Jerry Ehman. In Chapter 5 Stu Bowyer, who for decades led the various SETI efforts at the University of California, Berkeley (my alma mater), provides an overview of those early efforts. Chapter 6, dealing with a now-defunct U.S. government research project, is penned by John Billingham, who led the short-lived NASA SETI effort. In Chapter 7, Peter Backus, formerly with that same NASA project, relates how it segued into privatized research. The *SETI Past* volume concludes with a description by Harvard professor Paul Horowitz and his former graduate student, Darren Leigh, of that institution's landmark SETI research, including the development of the world's most powerful million (and later, billion) channel receivers.

*The Spirit of SETI Present* summarizes the state of the SETI art, circa 2010, and provides technical details of several contemporary SETI instruments and experiments. We begin with a description of what is currently the world's most advanced SETI array, and the research it is presently conducting, presented by Jill Tarter of the SETI Institute, arguably the world's best known radio astronomer and SETI scientist. In Chapter 10, photonics engineers Stuart Kingsley and Monte Ross tell about the expansion of the SETI search space into the optical spectrum. Next, we hear from the University of California's Eric Korpela about the hugely popular SETI@home distributed SETI experiment, which has attracted millions of participants from all over this planet. In Chapter 12, I review my own current SETI project, a global network of small radio telescopes built and operated by dedicated amateurs, coordinated through the internet. There follows a chapter by Richard Factor, founder and president of the grass-roots SETI League, showing how distant massive objects can focus weak signals, for likely detection by even modest receiving stations on Earth. Mathematician Claudio Maccone then compares and contrasts the two most popular algorithms for SETI digital signal processing. In Chapter 15, engineer Stelio Mongebugnoli and his colleagues at Italy's Istituto di Radioastronomia show how they are implementing the most computationally demanding of these algorithms at the SETI Italia facility. Next we hear from Bob Dixon again, this time in a discussion of a new radio telescope design concept that seeks to map the whole sky at once, from a single instrument. The *Present* section concludes with British scholar Stephen Webb's contemporary treatment of perhaps the oldest question in SETI science: "Where are they?"

*The Spirit of SETI Future* looks forward to the next 50 years of SETI activities, extrapolating our technological prowess, and ultimately leading (dare we hope?) to that long-anticipated communication from, and perhaps even dialog with, our distant cosmic companions. In Chapter 18, we hear again from Claudio Maccone on how gravitational lensing can focus and propel the interstellar internet. Then science fiction author Stephen Baxter presents a thorough review of how SETI science is depicted in the literature, and how those depictions will both inform and inspire future science. Next, psychologist Doug Vakoch delves into some of the deeper questions of the social consequences of communicating with the Other. This is followed in Chapter 21 by Russian radar scientist Alexander Zaitsev's treatment of the challenges of beaming deliberate messages into space. Next, physicist, science fiction author, and self-proclaimed contrarian David Brin cautions about the potential hazards associated with the transmission of messages to extraterrestrial intelligence. In Chapter 23, well-known SETI Institute spokesman Seth Shostak contemplates the longevity of extraterrestrial civilizations. British reproductive biologist Jack Cohen next takes a speculative stab at the question of ETI's physical form and characteristics. Canadian anthropologist Kathryn Denning tries to pin down just what constitutes technology, in an effort to narrow our search parameters. And finally, retired psychology professor Al Harrison wrestles with the impact which SETI success may have on human society.

In all, these three books in a single volume cover the multitude of topics which constitute SETI, perhaps the most highly interdisciplinary of scientific fields. The various chapters are written by SETI's shining stars, past, present, and future. The quest for contact is herein revealed as an exciting multigenerational journey, in which you, the reader, are invited to participate.

H. Paul Shuch, Ph.D.
Executive Director Emeritus
The SETI League, Inc.

# Part I: The Spirit of SETI Past

# 1

# A Half-century of SETI Science

H. Paul Shuch,
*Executive Director Emeritus, The SETI League, Inc.*

We begin our journey with a brief review of half a century of SETI science. The material in this introductory chapter is offered for the benefit of those educated laypersons whose enthusiasm for the Search for Extraterrestrial Intelligence exceeds their detailed knowledge of the relevant technologies. It is my hope that readers of this volume will better appreciate the material which follows if they first have a basic understanding of SETI concepts. Hence, I offer an overview, which is intended not to be exhaustive, but rather representative. Together, we will explore the nature of radio telescopes, experimental design strategies, SETI instrumentation, signal analysis, and the hallmarks of artificiality that allow us to differentiate between natural astrophysical emissions and intelligent interstellar transmissions. If you are already a technical specialist in these areas, feel free to bypass this introduction, and proceed directly to the subsequent chapters.

## 1.1 Birth of Radio Astronomy

Are we alone, the sole sentient species in the vast cosmos, or might there be others out there, with which we may some day hope to communicate? This is a fundamental question, which has haunted humankind since first we realized that the points of light in the night sky are other suns. Now, for perhaps the first time in human history, we have the technology to seek a definitive answer.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_1, 

That technology derives largely from radio astronomy, a relatively young science which was born quite accidentally in the 1930s, with the chance discovery that stars emit electromagnetic radiation in the radio spectrum. At Bell Laboratories in New Jersey, USA, a young radio engineer, Karl Jansky, was tasked with tracking down a source of interference that was plaguing transatlantic radiotelephone communications. Building a large, steerable directional antenna, he tracked the noise source across the sky to determine its periodicity. The interference did indeed repeat, on a 23 hour, 56 minute cycle. From this observation, Jansky concluded that the emissions were not originating on Earth or from the Sun, but rather from interstellar space. Today, we know that Jansky was detecting radio emissions from the center of the Milky Way galaxy. Thus was radio astronomy born.

Jansky's report, published in a radio journal, was read with considerable interest by another radio engineer, Grote Reber, in Wheaton, IL, USA. It was Reber, an accomplished amateur radio experimenter, who built the first modern radio telescope, a 10-meter diameter parabolic reflector, and in 1937 used it to produce the first known radio maps of the Milky Way.

Although in hiatus during the Second World War (during which most of the world's physicists were otherwise occupied with matters of weaponry), radio astronomy emerged as an observational science in 1951, with the first detection (by Harold Ewen, a graduate student at Harvard University, and his research advisor, Edward Purcell) of the 21-cm hyperfine emissions from interstellar Hydrogen, the most abundant element in space.

## 1.2 Radio Telescope Modalities

The three primary operating modes for modern radio telescopes include radiometry, spectroscopy, and interferometry. Each mode requires unique hardware and a specific experimental design.

The early observations of Jansky and Reber are examples of total-power radiometry, a time-domain measurement in which the thermal blackbody emissions from astrophysical sources are plotted against antenna aiming coordinates. Aiming can be either dynamic (i.e., actively varying the antenna in azimuth or elevation) or drift-scan (in which the Earth's rotation causes the antenna to sweep varying right ascensions over time). Radiometers are the simplest of radio telescopes, requiring only that the incoming signal be sufficiently amplified, and then applied to a square-law detector.

Spectroscopy is a frequency-domain mode, used to observe the molecular absorption or emission lines of the source being monitored. Ewen's pioneering hydrogen emission detection was an early example of astrophysical radio spectroscopy. In its most common implementation, spectroscopes involve downconversion of a portion of the electromagnetic spectrum, using a fixed intermediate frequency and a swept local oscillator.

Interferometry uses the interference fringes from multiple antennas to generate a spatial-domain image of an area of space. Interferometers require complex digital correlators along with well-matched antennas and receivers. Examples of advanced interfometers include the 27-dish Very Large Array (VLA) in Socorro, New Mexico, USA, and the 30-dish Giant Meterwave Radio Telescope (GMRT) in Khodad, India.

## 1.3 Early SETI Science

The notion that existing radio telescopes were capable of receiving purported artificial transmissions from distant, technologically advanced civilizations was first articulated by Cocconi and Morrison exactly a half-century ago. Their short paper "Searching for Interstellar Communications" in the journal *Nature* (1959) is generally regarded as the blueprint for the modern Search for Extra-Terrestrial Intelligence (SETI).

Even as that paper was in press, Frank Drake, then a young radio astronomer at the newly formed National Radio Astronomy Observatory (NRAO) in Green Bank, West Virginia, USA, was quietly preparing to perform the very experiment which the two Cornell professors were proposing. Drake and Coconni/Morrison had independently arrived at similar search strategies, and independently derived nearly identical experimental designs. Clearly, SETI science was ready to be born.

Project Ozma, Drake's Green Bank effort, observed two nearby sun-like stars (Tau Ceti and Epsilon Eridani) for a few weeks in the spring of 1960, scanning a narrow band of frequencies surrounding the Hydrogen emission line, using an 85-foot diameter parabolic reflector. Drake's receiver effectively combined radiometry and spectroscopy, in that it employed a square-law detector, but scanned a range of frequencies related to a known astrophysical emission line. Although he detected no promising candidate signals, Drake's Project Ozma served as a model for hundreds of SETI searches to follow.

The year after conducting Project Ozma, Drake convened at Green Bank the world's first scientific conference devoted to SETI. The agenda for that week-long meeting consisted of seven topics, touching upon various astrophysical, biological, technological, and societal aspects of the emergence of potential communications partners in the cosmos. Stringing those seven topics together into a multiplicative model, Drake created the now-famous Drake Equation, a tool widely used for estimating the number of communicative civilizations which might exist in our Milky Way galaxy.

## 1.4 The NASA Years

In the summer 1971, a landmark study was conducted at the NASA Ames research center in Mountain View, California, USA. Chaired by Dr Bernard M. Oliver, then vice-president of engineering for the Hewlett-Packard company, Project Cyclops sought to design (on paper) the ultimate SETI receiving system which could be conceived if money were no object.

A proposed interferometer array, consisting ultimately of 900 large parabolic reflector antennas coupled to an advanced optical computer for multi-channel spectral analysis, would have cost in the tens of billions of US dollars. It was never seriously considered for funding, and hence never built. However, the resulting publication (which was reprinted in 1996, and is still available) serves to this day as a blueprint for how large-scale SETI hardware and software might be developed.

A modestly funded NASA SETI program followed in the US. In other countries, parallel studies ensued, each receiving limited government or institutional financial support. Most borrowed limited observing time on existing radio telescopes.

The NASA SETI office, headquartered at the Ames Research Center, where the Project Cyclops studies had been conducted, expended significant effort on the development of advanced Multi-Channel Spectrum Analyzers (MCSAs) capable of scanning hundreds to thousands of MHz of the electromagnetic spectrum in real time. This was in marked contrast to earlier SETI efforts, whose receivers had been restricted to merely a few tens to hundreds of kHz of bandwidth.

NASA SETI launched a 10-year, two-pronged search in October 1992, significantly the 500th anniversary of Columbus' first voyage of discovery. Its two complementary strategies involved a targeted search of nearby sun-like stars, and a methodical sweep of the entire sky for signals emanating from the vicinity of stars not specifically known to us. Each search strategy required a different instrumentation approach. Targeted searches are conducted by tracking known stars with the largest, highest gain antennas available for hours on end. Sky surveys, on the other hand, tend to be operated in meridian transit (drift scan) mode, employing smaller antennas (to provide increased spatial coverage) and limiting observing time in any one direction, in favor of maximizing sky coverage. As an analogy to the differing equipment capabilities required for each of these two modalities, you can mentally contrast optically viewing the night sky through a toilet paper roll vs. a soda straw.

Although budgeted at a mere five cents per US citizen per year, NASA SETI's $12.6 million annual budget proved an easy target for legislators. The US Congress cancelled the NASA SETI program in 1993, after just one year of observations (and, in the process, reduced the federal deficit by

0.0006 percent). Some in the SETI community have taken this as definitive proof that there exists no intelligent life in Washington.

## 1.5 Privatization of SETI

With the demise of NASA SETI, two nonprofit organizations in the US stepped up to privatize and continue SETI research. The California-based SETI Institute revived the targeted-search prong of the late NASA SETI program in 1995, under the guise of Project Phoenix (symbolically named for having risen from the ashes of its predecessor search). The SETI Institute secured for its Project Phoenix observations the spectral analysis equipment which it had previously developed on contract in support of NASA SETI.

In 10 years of observations involving renting time on large radio telescopes at Parkes and Mopra in Australia, Green Bank in West Virginia, Woodbury in Georgia (US), Arecibo in Puerto Rico, and at Jodrell Bank, UK, Project Phoenix monitored 1,000 nearby sun-like stars across a substantial portion of the microwave spectrum. The project employed sophisticated follow-up detection procedures to validate candidate signals and eliminate terrestrial interference. Project Phoenix achieved a null result, in that none of its candidate signals passed the follow-up detection test.

Beginning in 1996 from donated office space in New Jersey, the nonprofit SETI League's Project Argus is an ongoing attempt to resurrect the all-sky survey component of the NASA SETI effort. Project Argus is named for the mythical Greek guard-beast who had 100 eyes and could see in all directions at once. It seeks to see in all directions at once, in real time, by bringing online thousands of small radio telescopes around the world, built and operated by dedicated amateur radio astronomers, whose efforts are coordinated through the internet. To date, Project Argus has in operation 144 stations in 27 different countries on all seven continents. It too has yet to detect conclusive evidence of ETI.

## 1.6 Dedicated SETI Instruments

While in recent years dozens of SETI experiments have begged, borrowed, or bought observing time on various radio telescopes in the UK, Germany, Russia, Argentina, Italy, Japan, Australia, Puerto Rico, and elsewhere, ever since the Project Cyclops study the dream of a full-time SETI observatory, optimized for the detection of intelligently generated extraterrestrial signals, has remained foremost in the minds of SETI scientists.

In 1999 the nonprofit, membership-supported SETI League began planning its *Array2k* observatory. Receiver prototypes were constructed and tested, key technologies developed and patented, land acquired in New Jersey, and a prototype *Very Small Array* (VSA) constructed in Pennsylvania, before global economic conditions brought an end to the project's private funding. It is hoped that work on this instrument can resume should new funding materialize.

In California, the nonprofit SETI Institute has been a little more fortunate in securing funding for its Allen Telescope Array (ATA). Planned as a phased array of 350 Gregorian antennas with cryogenically cooled broadband front-ends and fiber-optic links to its advanced MCSA, the ATA is currently online with 42 dishes operational.

When fully implemented, it is expected that the ATA will permit fulltime SETI observations of the entire sky which can be seen from its location in Hat Creek in Northern California, and will be used to survey upwards of one million stars across the entire microwave spectrum.

## 1.7 Signal Analysis Techniques

A challenge facing all SETI observatories, extant and planned, is differentiation between candidate signals and the ever-present natural background noise of the cosmos. Whereas natural thermal emissions are broadband, extending across the entire electromagnetic spectrum, it is expected that signals used for deliberate electronic communication over interstellar distances will likely have narrow-band components. Thus, minimizing receiver bandwidth is an accepted technique for pulling such weak signals out of the noise.

However, the actual frequency of transmission is not known to us *a priori*. The quietest portion of the electromagnetic spectrum that could efficiently support interstellar contact is extremely wide, spanning from a few hundred MHz to tens of GHz. Assuming a sufficiently narrowband receiver, there are billions of possible frequencies to which it might be tuned. Here is where digital signal processing (DSP) techniques become an important part of SETI research.

It is common SETI practice to receive, amplify, filter, and digitize extremely wide portions of the electromagnetic spectrum. DSP is then employed to subdivide the broad received spectrum into a multitude of contiguous, vanishingly narrow frequency bins, each of which excludes much of the broad spectrum of noise so as to maximize signal to noise ratio. Most frequently, the fast Fourier transform (FFT) is employed to produce these narrow bins. At 1 Hz of bin width, for example, a 1 billion point FFT can simultaneously monitor all 1 billion such 1 Hz channels within 1 GHz of bandwidth.

A limitation of the FFT is that it is optimized for the detection of sinusoidal signal components. If the nature of the incoming signal is unknown, one would desire an adaptive transform to detect it. One such tool, which makes no *a priori* assumptions as to the characteristics of the signals hidden in the noise, is the Karhunen-Leove transform (KLT). SETI scientists have for years sought, with mixed success, to implement the KLT on radio telescopes in Ohio, Italy, and elsewhere.

Unfortunately, the KLT is extremely demanding of computer power. As the state of the art in computer technology continues to experience Moore's Law exponential growth, it is felt that the KLT will eventually displace the FFT as the SETI signal analysis algorithm of choice.

## 1.8 Hallmarks of Artificiality

Most SETI scientists hold that detection of artificially generated electromagnetic waves remains the most likely mechanism of contact between humans and ETI, at least at our present state of technological development, and excluding from consideration any laws of nature not presently in evidence. The photon is, after all, the fastest spaceship known to man. It travels relatively unimpeded through the interstellar medium, at the fastest speed which our understanding of physics would allow.

Based upon the primitive state of Earth's communications technology, such contact is most likely to occur in the microwave spectrum, although optical SETI is becoming more viable. We would have a high confidence level that such contact had taken place upon simultaneous detection (at widely separated terrestrial coordinates) of signals of sufficient duration or periodicity to allow multiple independent observations. In addition, such signals must exhibit some reasonable combination of the following hallmarks of artificiality:

- spatial/temporal characteristics consistent with sidereal motion;
- coherence not achievable by known natural emission mechanisms;
- Doppler signatures indicative of planetary motion;
- frequency selection which exhibits a knowledge of one or more universal constants; and information content suggestive of a mathematically based culture.

## 1.9 Standards of Proof

We now address the issue of what constitutes incontrovertible proof of ETI contact. The question is complicated by the fact that the general public may make only a vague distinction between fact and faith. The spectrum of human skepticism vs. gullibility encompasses a wide range of extremes, characterized by diverse viewpoints ranging from "of course they exist - we couldn't possibly be alone!" to "I'll believe in the existence of intelligent extraterrestrials only when one walks up and shakes my hand." We must take pains to prevent such declarations of faith from clouding the judgment of our SETIzens.

We start by acknowledging that one can never conclusively prove the negative, but that it takes only one counter-example to disprove it. Conservative experimental design demands that we frame our research hypothesis in the null form: "resolved that there are no civilizations in the cosmos which could be recognized by their radio emissions." Now a single, unambiguous signal is all it takes to disprove the null hypothesis, and negate the notion of humankind's uniqueness.

What exactly constitutes an unambiguous signal? A popular definition holds it to be one which could not have been produced by any naturally occurring mechanism which we know and understand. But this is an insufficient condition. The first pulsars, after all, fitted that definition. They were first labeled "LGM" for Little Green Man, and their intelligent extraterrestrial origin seriously considered for several months, until our knowledge of the mechanics of rapidly rotating, dense neutron stars became more complete. There is the risk that any signal which cannot be produced by any known natural mechanism could well have been generated by an astrophysical phenomenon which we have yet to discover. So, we need an additional metric.

We listed above several of the hallmarks of artificiality, which we can expect to be exhibited by an electromagnetic emission of intelligent origin. The common denominator of all these characteristics, in fact of all human (and we anticipate, alien) existence, is that they are anti-entropic. Any emission which appears (at least in the short term) to defy entropy is a likely candidate for an intelligently generated artifact. In that regard, periodicity is a necessary, though not a sufficient, condition for artificiality (remembering once again the pulsar).

Ideally, we would hope to receive communication rich in information content, signals which convey otherwise unknown information about the culture which generated them. Unless we are blessed with such a message, we are unlikely to ever achieve absolute certainty that what we have received is indeed the existence proof we seek. Multiple independent observations, however, can do much to dispel the obvious alternative hypotheses of equipment malfunction, statistical anomaly, human-made interference, and deliberate hoax. In that respect the development of well coordinated signal verification protocols can do much to narrow our search space.

Once again, in signal verification activities, it is the null hypothesis we should be attempting to verify. We thus expect that we will continue to rule out most candidate signals. There may eventually come a signal, however, which simply cannot be explained away. At that point, we may dare to conclude that *we are not alone*.

## 1.10 Conclusion

Now that we've summarized where we've been in a half-century of SETI activity, you, the reader, are encouraged to pursue the balance of the submitted chapters in this single-volume trilogy, to gain a better sense of where we are in our pursuit of SETI science, where we have been, and where we are going.

**Acknowledgment**

This chapter is based upon an introductory SETI lecture by the author, presented at the 60th International Astronautical Congress, Daejon, Korea, on October 1, 2009.

# 2

# Project Ozma: The Birth of Observational SETI

H. Paul Shuch,
*Executive Director Emeritus, The SETI League, Inc.*

It was an idea whose time had come, but nobody dared admit that out loud. Frank Drake, in particular, was keeping silent. Like many of his generation, he had long speculated about the existence of extraterrestrial life, and pondered how we humans might probe for direct evidence of our cosmic companions. Now, in 1959, the young astronomer was finally in a position to do more than ponder. At 29, he had just completed graduate school, the ink on his Harvard diploma as wet as he was behind the ears. As the new kid on the block at the National Radio Astronomy Observatory, he had access to the tools necessary to mount a credible search for radio evidence of distant technological civilizations. Drake knew enough to tread lightly; a publicly announced hunt for Little Green Men would be tantamount to professional suicide, so he approached his superior with understandable trepidation.

Fortunately, NRAO director Otto Struve was sympathetic, even as he counseled caution. Having theorized that the slowed rotation rate of certain stars suggested that their angular momentum had been dissipated in the formation of planets, Struve himself speculated on the probable existence of extraterrestrial civilizations. So, he authorized Drake to use the 85 foot diameter Howard Tatel telescope (Figure 2.1) in his off-duty time, to conduct what was to become the world's first observational SETI experiment. Only, do so quietly, Struve warned; we don't want the word getting out that we're using a government facility to hunt for aliens.

Drake had already run the numbers. He knew the most likely frequency on which to search, and the best receiver circuitry to employ. He had picked

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_2, © Springer-Verlag Berlin Heidelberg 2011 

**Fig. 2.1** The 85-foot diameter Howard Tatel Telescope at the National Radio Astronomy Observatory, Green Bank WV, used by Frank Drake for his *Project Ozma* observations in April and May of 1960.

his candidate stars, two nearby sunlike ones which he reasoned were likely to harbor habitable planets. He had selected his research methodology, and proceeded (very quietly) to assemble his listening station.

And then, the *Nature* article hit the newsstands. "Searching for Interstellar Communications" was written by two Cornell University professors, Giuseppe Cocconi and Philip Morrison, and it proposed, in brief but clear detail, the very experiment which Drake was preparing to conduct! This very first scientific article in the not-yet-named discipline of SETI was complete, down to the selection of frequencies and target stars – and it paralleled Drake's work exactly. Neither the team of Morrison and Cocconi, nor that of Drake and Struve, knew anything about the others' interest in this esoteric study. Both groups had arrived at the same crossroads in history, completely independently, in an elegant example of what I like to call the Parenthood Principle: when a great idea is ready to be born, it goes out in search of a parent. Sometimes, it finds more than one.

Now Schrodinger's Cat was out of the bag, and Drake had no choice but to go public. The publicity he received was widespread, and generally enthusiastic; the scientific community, it appeared, was ready to embrace the notion of SETI. Struve began writing about the possibility of extraterrestrial life: “An intrinsically improbable event may become highly probable if the number of events is very great… it is probable that a good many of the billions of planets in the Milky Way support intelligent forms of life. To me this conclusion is of great philosophical interest. I believe that science has reached the point where it is necessary to take into account the action of intelligent beings, in addition to the classical laws of physics.”

His cover now blown, Drake soon found himself in the company of other open-minded scientists and technologists, who collectively found themselves unwitting parents to a newly-emerging scientific discipline. Among those contacting Drake after reading about his nascent experiment were: microwave communications expert Bernard M. Oliver, then vice-president of engineering at Hewlett-Packard (and, later, president of the Institute of Electrical and Electronic Engineering); Dana Atchley, president of Microwave Associates in Massachusetts; and a young planetary scientist, Berkeley post-doctoral researcher Carl Sagan. These individuals, as well as Struve, Morrison, and a handful of others, were ultimately to become SETI’s patriarchs. (Cocconi, though having co-authored the seminal SETI article with Morrison, went on to distinguish himself in particle physics research at CERN, never to return to the SETI fold.)

Drake named his search Project Ozma, after the princess of Oz in the L. Frank Baum books, as he saw his efforts leading humans to a far-off and exotic land. Launched in April 1960, and running only through May of that year, *Ozma* searched only two stars, on a single frequency, for mere dozens of hours, but established the protocols and laid the groundwork for all subsequent SETI experiments. It was a paradigm-shifting endeavor, successful for its audacity, if not for its discoveries.

And yet, for one brief moment early on, Frank Drake thought he had hit paydirt. As he slewed his antenna off Tau Ceti and onto Epsilon Eridani, he was greeted with a strong, periodic, pulsed signal on 1420 MHz, the hyperfine transition emission line of interstellar hydrogen atoms proposed by Cocconi and Morrison, and still favored as a promising hailing frequency for interstellar communications. “My god,” Frank mused, “can it really be this easy?”

The next day, when the signal reappeared, Drake was ready with a second, low-gain antenna. The pulses were there as well, sadly disproving their extraterrestrial origin. But they were not exactly terrestrial interference, either. The rate at which the phantom signal traversed the sky suggested that it was emanating from an aircraft cruising at unprecedented altitude – perhaps 80,000 feet! Of course, in April 1960, no known aircraft could reach the stratosphere. Such an aircraft, as it happened, didn’t “come into existence” until the following month, when Francis Gary Powers was shot down over the Soviet Union. (Frank wisely decided to withhold publication of this positive result, so he never did receive proper credit for "discovering" the U-2.)

A year after Project Ozma's brief tenure, Drake convened at Green Bank the first scientific conference devoted to modern SETI. He gathered together 10 scientists from disparate disciplines to spend a week contemplating areas from the physical, biological, and social sciences which had relevance to the question of extraterrestrial technological civilizations, and how to communicate with them. The assembly included the six SETI patriarchs already mentioned, along with J. Peter Pearman of the National Academy of Sciences' space science board, Su Shu Huang of NASA, University of California chemist Melvin Calvin (whose Nobel prize was to be announced during the Green Bank meeting), and neuroscientist John C. Lilly, who was then studying the language of dolphins, and attempting to communicate with these intelligent Earth mammals. The group called themselves the Order of the Dolphin, a tribute to Lilly's studies into human -dolphin communication, which they deemed a worthy metaphor for the challenge of interspecies communications on a grander, cosmic scale.

Drake chalked on a blackboard seven topics for discussion, which would comprise the agenda for the week-long meeting. They included stellar formation, planetary formation, the existence of planets within habitable zones, the emergence of life, the evolution of intelligence, communications technology, and the longevity of technological civilizations.

Having established that the emerging discipline of SETI was to encompass fields as diverse as stellar evolution, planetary astronomy, environmental science, biology, anthropology, engineering, and sociology, Drake next did something almost whimsical, which assured his lasting fame: he strung these seven factors together into an equation (Figure 2.2).

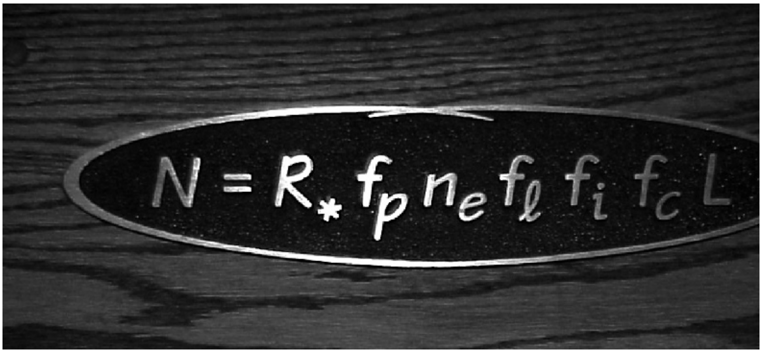



**Fig. 2.2** The seven agenda items for the Order of the Dolphin meeting. Strung together, they form the famous Drake Equation. This plaque graces the very wall on which the equation was originally scrawled on a chalkboard at NRAO Green Bank in 1961.

The idea was to multiply seven unknowns together and, in so doing, to estimate N, the number of communicative civilizations in our Milky Way galaxy. The Drake Equation, as it is now called, appears in every modern astronomy textbook. It is a marvelous tool for quantifying our ignorance: never intended for quantification, but quite useful in narrowing the search parameters. We still use it, not to seek a numerical solution, but rather to help us to focus our thinking in designing our searches for life.

Drake's seven factors are cleverly ordered, from solid to speculative. Today's astrobiology meetings are similarly structured. When the equation was first published, only the first factor (the rate of stellar formation) was known to any degree of certainty.In the intervening decades, Drake's equation has guided our research in an orderly manner, from left to right, so that today we have a pretty good handle on Drake factors two and three (planetary formation, and habitable zones). The remaining four factors are still anybody's guess, and it may well take decades more before our research begins to quantify those areas of our ignorance. But the Drake Equation is most valuable in guiding our research, because it asks the important questions. It is still up to us to answer them.

The lessons learned during the brief course of Project Ozma, amplified and expanded at the Order of the Dolphin meeting, have informed and enriched every subsequent SETI experiment. The interdisciplinary nature of the science now known as SETI was articulated at the outset. Drake's work clearly showed that Earth's technology was at last approaching the level at which a disciplined search for extraterrestrial microwave emissions was becoming feasible. The quietest part of the electromagnetic spectrum was explored then, as now. Highly directional, high gain parabolic antennas, coupled to very low noise microwave preamplifiers, remain our preferred observational tools. Although the advent of multi-channel spectrum analyzers means we no longer have to select a single channel to scan, SETI scientists continue to speculate as to universal calling frequencies that alien civilizations might employ to make their presence known. Concentrating our efforts on known, nearby sun-like stars remains an accepted technique for planning targeted searches, one of the two primary search modalities still practiced. (The other popular SETI research strategy, the all-sky survey, was long employed at the Ohio State University "Big Ear" radio telescope, and more recently forms the basis of The SETI League's Project Argus search, as discussed in subsequent chapters of this book.)

Most important, Frank Drake's early efforts began to lend legitimacy to an endeavor previously considered fringe science. Today, the preponderance of informed opinion holds that we inhabit a universe teeming with life. The only matter for speculation is whether we yet possess the technology necessary to detect it. The emphasis here is on *yet*. Most of us contemplating such a detection no longer argue "if" but rather "when."

Drake subsequently distinguished himself as Director of the famed Arecibo Observatory, from which he orchestrated the Arecibo Message, humankind's first deliberate microwave transmission to the stars, as detailed elsewhere in this volume. His astronomical research has led to important discoveries about pulsars and Jovian radio emissions. Now retired, he is today recognized as the godfather of observational SETI. Much in demand as a speaker at scientific meetings (Figure 2.3), Frank remains deeply involved in SETI science fully half a century after Project Ozma, serving as a Director of the SETI Institute in California and on the scientific advisory board of the nonprofit SETI League.

This, then, is Project Ozma's legacy: it, and Frank Drake, have turned science fiction into credible, respectable science.


**Fig. 2.3** In his rightful role as SETI Elder Statesman, Frank Drake still serves as keynote speaker at SETI conferences all over the world.

# 3

# Project Cyclops: The Greatest Radio Telescope Never Built

Robert Dixon,
*Acting Director, Ohio State University Radio Observatory*

## 3.1 Introduction to Project Cyclops

Each summer NASA sponsors a number of research and development projects at their various research centers across the country, often in cooperation with a nearby university. Selected groups of university faculty and professionals are brought together to study some research problem of interest to NASA, and to provide continuing education for the participants. The great advantage of these summer research programs is that NASA gains the experience of talented people who can look at problems with fresh eyes and no preconceived solutions. The participants are freed from their normal day-to-day responsibilities, and can let their imaginations run wild and be totally dedicated to the problem at hand. These programs are exhilarating, wonderful and can even be career-changing experiences.

In 1971 the NASA-Ames Research Center, Stanford University and the American Society for Engineering Education organized one of these studies, called Project Cyclops. Twenty Faculty Fellows from many universities across the country and fields of study worked together for 11 weeks on this specific objective:

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_3, © Springer-Verlag Berlin Heidelberg 2011

"To assess what would be required in hardware, manpower, time and funding to mount a realistic effort, using present (or near-term future) state-of-the-art techniques, aimed at detecting the existence of extraterrestrial (extrasolar system) intelligent life."

The Fellows were from electrical engineering, mathematics, management science, civil engineering, space science, astronomy, and mechanical engineering. But it was clear from the beginning that this was no ordinary summer study project. Two superstars co-directed the program: John Billingham, Chief of the Life Sciences Division at NASA-Ames Research Center (Figure 3.1), and Barney Oliver, Vice President of Research at Hewlett-Packard corporation (Figure 3.2). And many other famous people came to NASA-Ames to give presentations and advice to the study group, including Philip Morrison (MIT), Ronald Bracewell (Stanford), Sebastian von Hoerner (National Radio Astronomy Observatory), Richard Goldstein (Jet Propulsion Laboratory), Gordon Pettingill (MIT), Martin Rees (Cambridge University, UK), David Heeschen (Director of the National Radio Astronomy Observatory) and others. These people all realized that Project Cyclops was a very important milestone in the progress of SETI, and they wanted to be part of it.

All these summer projects produce a final report for internal NASA purposes.They are in the public domain, but usually little known afterwards.

**Figure 3.1** John Billingham, Chief of the Life Sciences Division at NASA-Ames Research Center..

But NASA chose to give the Project Cyclops report much wider circulation and importance, under the designation Contractor Report (CR) 114445. Ten thousand copies were printed and distributed.

*Cyclops* was the first large-scale effort to study this problem from all aspects in a coordinated way. It was orders of magnitude beyond what had been done before. But of course, it was understood that even more orders of magnitude of study would be needed to actually construct the system envisioned by the study.

The Cyclops report was initially misunderstood by some people to be an "all or nothing" plan, and hence impractical. In reality, the plan was to start with a small, relatively inexpensive system using a single antenna, and then enlarge it as needed until success was achieved.

Some notable firsts for *Cyclops* included:

1 Determination that the region of the radio spectrum between 1.4 and 1.7 GHz has the lowest natural radio noise and ease of reception, and is hence the logical radio band of choice for inter-species communication.

2 Recognition that the hydrogen (H) radio spectral line at 1.4 GHz and the hydroxyl (OH) radio spectral line at 1.7 GHz bound this lowest-noise band, in a cosmic coincidence (Figure 3.3).

**Figure 3.2** Barney Oliver, Vice President of Research at Hewlett-Packard.

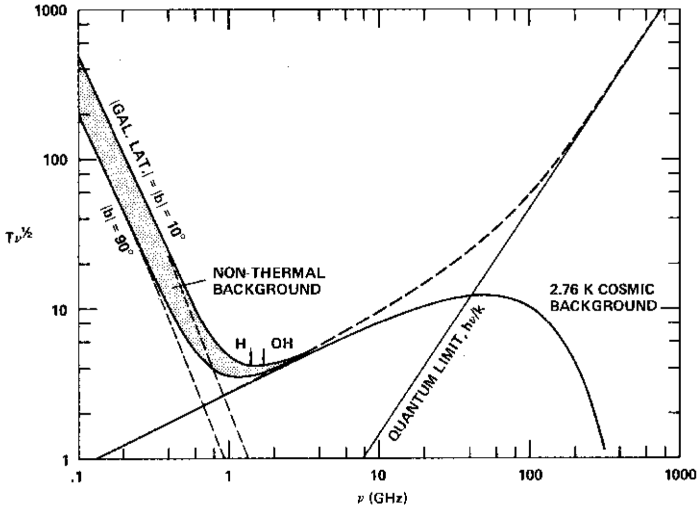



**Figure 3.3** The difficulty of reception of radio signals at various frequencies is determined by the natural noise from the galaxy, the 3 degree blackbody noise left over from the beginning of the universe, the quantum noise of radio photons, and the fact that drifting signals require greater bandwidth or smaller integration time. All these effects are added up here, showing that the best frequency range is the "Water Hole" between H and OH.

3 Recognition that these two molecules combine to form water ($H+OH=H_20$), and then creating the name "Water Hole" for this band

4 Making the philosophical connection between the radio Water Hole, and the wilderness water holes found on Earth, which are the meeting places of different species

5 Recognizing that the senders of any signals intended for other civilizations would design their signal in such a way as to make it as easy as possible for the recipients to decipher it. That is the whole purpose of sending it, and designing it this way maximizes the probability of successfully establishing contact. The term "Principle of Anticryptography" was assigned to this idea

6 Determination of the optimum size of each antenna to be used in an array of required total size, so as to minimize cost

7 The system of upconverting each received band to a fixed frequency maser amplifier

8 The system of inverting the intermediate frequency signals periodically along their way from the dishes to the central control building, to remove variations in cable loss at different frequencies.

9 The system of sending a common local oscillator signal out from the control building to all the dishes, and then back to the control building, to cancel out phase variations caused by temperature changes in the cables.

10 Determination that the Golay detector for narrowband signals is no more sensitive than a conventional linear signal detector.

11 Determination that the Drake Ensemble method of detecting a group of narrowband signals is no more sensitive than conventional methods, for the examples of the television and FM broadcast bands in the USA.

12 Design of an optical spectrum analyzer capable of searching a 100 MHz band with a resolution of 0.1 Hz.

In addition to its technical findings and recommendations, *Cyclops* brought together for the first time in one place a large collection of relevant technical information from many fields of science and engineering. This information is presented as tutorial introductions and appendices, leading up to and serving as the starting point for the *Cyclops* work. Here are some examples.

On the topic of Life in the Universe, there are explanations of:

1 Origin of the universe

2 Evolution of galaxies and stars

3 Evolution of planets

4 Evolution of planetary atmospheres

5 Origin and evolution of life on Earth

6 Development of intelligent life on Earth

7 Life on other planets: the Drake equation

8 Probability of interstellar communications

On the topic of Radio Communications, there are explanations of:

1 Antenna design

2 Space communications

3 Receiver design

4 Signal detection

5 Radio interference

6 Doppler shift

## 3.2 Findings of the *Cyclops* Project Team

### 3.2.1 Antenna system and physical arrangements

A total collecting area of around 10 $km^2$ is required. It is not practical at the current state of the art to construct such an antenna in space, so it must be ground based. A single dish antenna of that size is not practical, so an array of many smaller dishes must be used. Analysis of dish size vs. cost shows that using the largest dish that can be reasonably built with today's technology has the lowest total cost. That is about 100 meters in diameter, ultimately requiring a total of about 1000 dishes. The cost could be significantly reduced

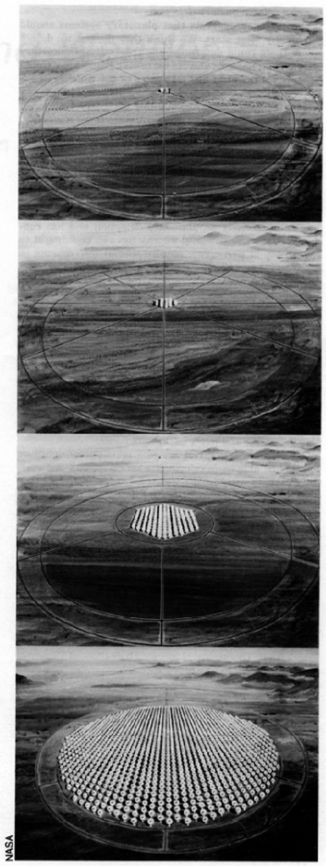



**Figure 3.4** The *Cyclops* array grows with time as needed.

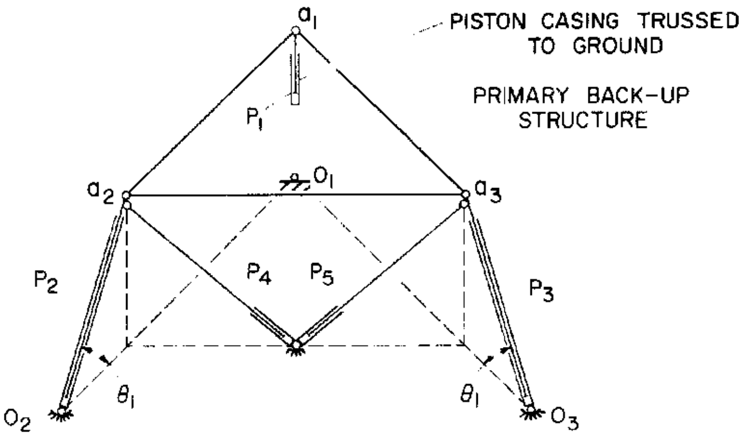



**Figure 3.5** Piston mounting system considered for the *Cyclops* dishes.

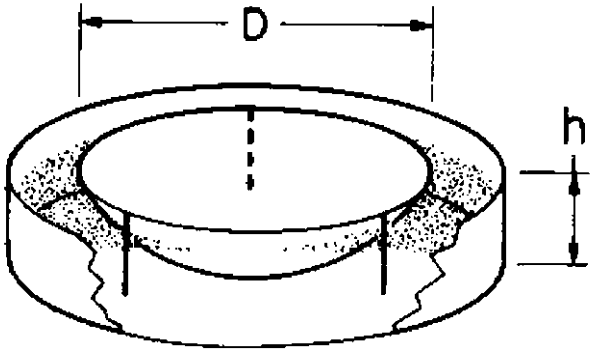



**Figure 3.6** Tethered floating mounting system considered for the *Cyclops* dishes.

through the use of mass production techniques. (Informally it was stated that a factory could be built to stamp these things out.) The array would start with a single element, and then grow as needed until signal detection was achieved. (Figure 3.4).

Various types of antenna mounting systems were investigated, including the usual azimuth-elevation and equatorial mounts, plus novel ideas such a piston mount and a floating mount (Figures 3.5 and 3.6). Azimuth-elevation was chosen as being the least expensive, and easiest to implement with current technology.

A system of underground service tunnels links all of the elements to a data processing center, located at the center of the array. These tunnels carry the power, control and signal cables needed to operate the array. Various tunnel arrangements were investigated, and the one shown in Figure 3.7 was found to be least expensive.

A small community called Cyclopolis is nearby, but far enough away to not cause radio interference. That provides housing for the system staff, offices, laboratories and other buildings needed.

The array should be sited at a location not subject to earthquakes, and with low humidity, low wind, mild climate, flat terrain and be remote from habitation and air routes.

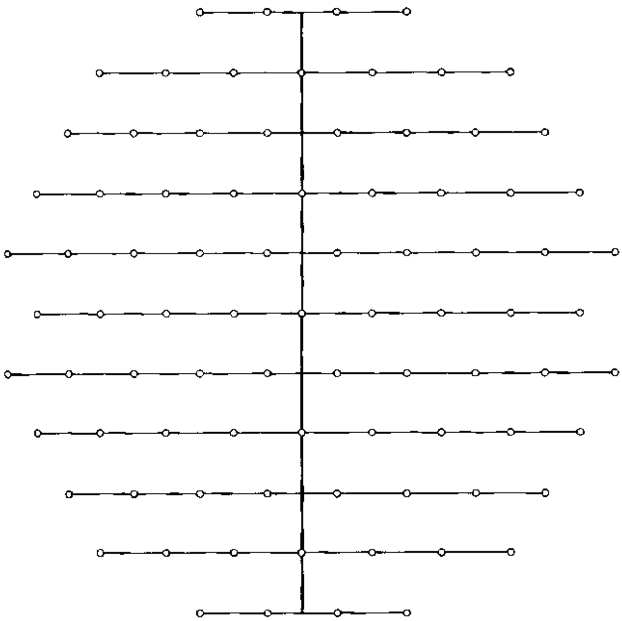


**Figure 3.7** The least expensive tunnel arrangement for connecting the *Cyclops* dishes to the central control building.

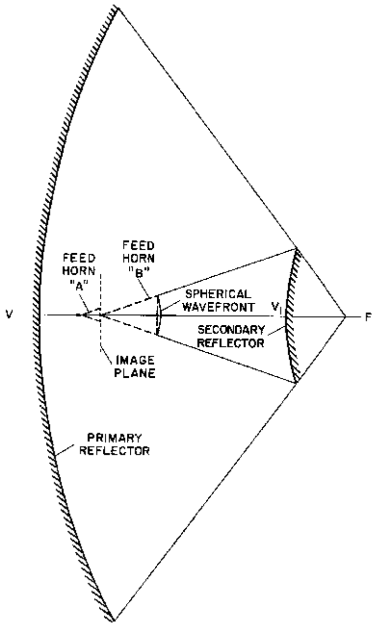



**Figure 3.8** The Cassegrainian dish feed system.

.

### 3.2.2 Dish feed system

A Cassegrainian feed system is used, because that allows the receivers and feed horns to be located at the center of the dish, and be shielded from the signal path by being in the shadow of the secondary reflector Figure 3.8). This allows the feed horns to be as large as desired, without causing signal blockage and introducing additional antenna noise. To maximize efficiency, dielectric loaded corrugated feed horns are used (Figure 3.9). To cover the frequency range of interest (0.5 to 3 GHz), six separate horns are needed, and they are mounted on a rotating turret (Figure 3.10). Each horn receives dual circular polarization.

### 3.2.3 Receiver system

The desired frequency band is 0.5 to 3 GHz, with optional extension to 10 GHz. The band of interest is divided into six subbands for maximum efficiency. Each subband has its own cooled upconverter, which sends the signals to a fixed frequency (10 GHz) cooled maser amplifier. After further amplification, the signals are downconverted to a 75–175 MHz Intermediate Frequency (total bandwidth 100 MHz) for transmission to the central control building. There are two such receiver systems for each dish, to handle

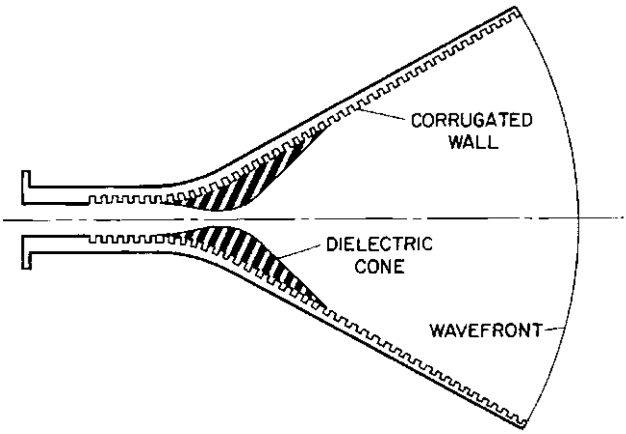



**Figure 3.9** The dielectric loaded corrugated feed horn.

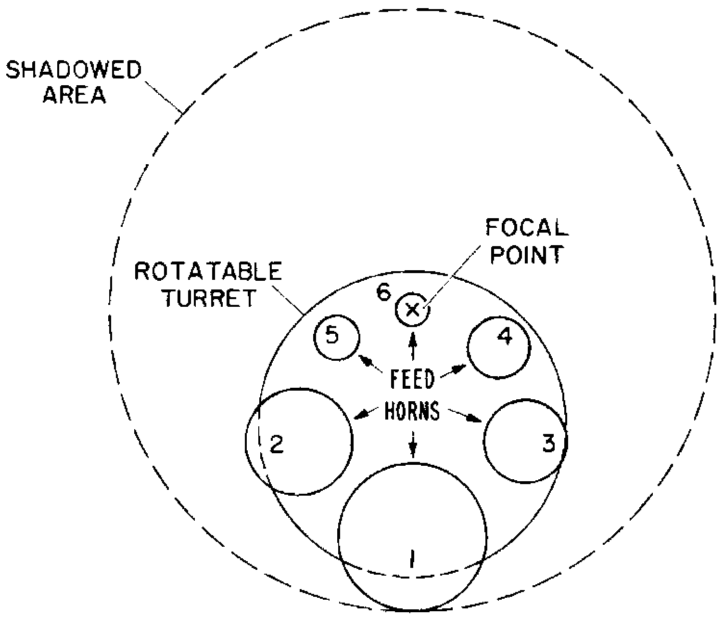



**Figure 3.10** The rotating turret to hold the six feed horns.

the two polarizations coming from the feed horns. A system temperature of 20 degrees Kelvin is achieved by this combination of feed horns and receivers.

### 3.2.4 Signal transmission to the central control building

Coaxial cable of about 1 inch in diameter is used to carry the Intermediate Frequency signals from the individual dishes to the central control building, through the service tunnels. Other transmission methods such as waveguide, lasers, optical fibers and microwave links were considered, but found to be too expensive, too unstable or too lossy. It is noted that this choice could well change in the future as these technologies develop further.

Coaxial cables are not without their own problems, primarily variation of loss and delay with frequency and temperature. These problems are solved by first upconverting one of the two polarization signals to 324 -425 MHz, thereby creating a double sideband spectrum centered at 250 MHz (Figure 3.11). This spectrum is inverted at each ¼ distance point along the cable from the dish to the central control building, so the upper sideband signal (from one of the polarizations) becomes the lower sideband signal, and vice versa. This technique cancels out the variations in signal loss at different frequencies. The 250 MHz local oscillator signal is generated at the central building, and sent out to all the dishes, so they all have exactly the same intermediate frequencies. The phase of the return oscillator signal is compared to the phase of the original oscillator signal, and any difference found is removed by a variable delay unit. This cancels out delay changes caused by temperature variations.

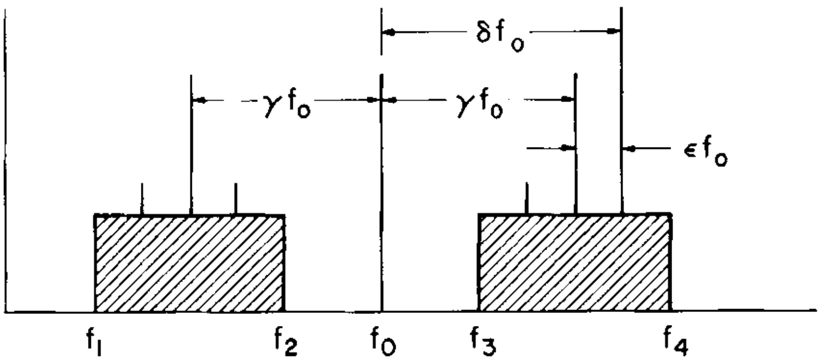



**Figure 3.11** The double sideband intermediate frequency spectrum, used to transmit the data from both polarizations of the dishes to the central control building.

### 3.2.5 Signal processing

The most likely signals of intelligent origin are very narrowband, with bandwidths on the order of 0.1 to 1 Hz. To find such signals buried within the 100 MHz bandwidth IF signal requires calculation of the Fourier transform to change the data from the time domain to the frequency domain. Then each of the millions of frequency data points can be examined to see if its amplitude exceeds the noise level, thereby indicating detection of a narrowband signal.

1 **Digital spectrum analyzer**: To do this electronically, the IF signal has to be digitized with an analog-to-digital converter, and then the Cooley-Tukey fast fourier transform technique (FFT) is applied. At the current state of the art, using hardwired FFT systems, 18,000 of them are needed, at a total cost of $4.5 billion. That would dwarf all other portions of the *Cyclops* system, and is completely impractical.

2 **Optical spectrum analyzer**: A simple optical lens has the ability to automatically transform an optical signal from the time domain to the frequency domain, with almost unlimited speed and bandwidth. This transformation is done in analog fashion, with no analog-to-digital converter needed. The signal must first be recorded continuously on photographic film in raster form, as shown in Figure 3.12.

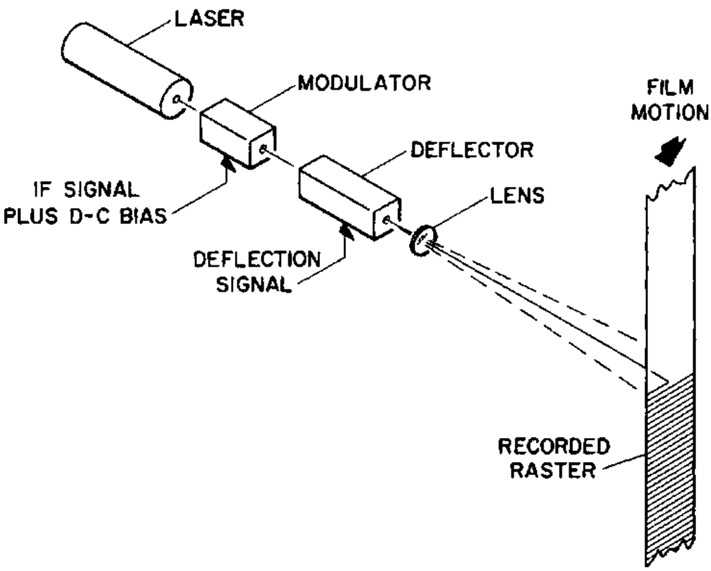



**Figure 3.12** Recording the *Cyclops* data onto photographic film to serve as input to the optical spectrum analyzer.

In a second step, light is shown through the film, through the Fourier transforming lens, and the resulting frequency domain display is recorded on a second film strip (Figure 3.13). A single system like this can transform 100 MHz bandwidth into 10 Hz bandwidth channels. Since that is not narrow enough, we need 20 or 200 optical spectrum analyzers (for both polarizations) to achieve the desired narrow bandwidths. The film used is 35 mm wide, and runs at about 9 cm/sec total, costing $47/hr, which is reasonable. The frequency display is imaged on a high-resolution vidicon tube, scanned and converted to a video signal, and recorded onto magnetic disks.

3 **Doppler shift compensation**: The frequency of any extraterrestrial signal will change with time, since there are many time-varying motions of both the remote transmitter and our local receiver, which cause Doppler shift. Some of these motions are rotation of the planets, revolution of the planets around their stars, motions of the stars through the galaxy, and rotation of the galaxies. It is highly desirable to average (integrate) the data at each narrow frequency. This causes the noise to average out, and the signal to remain. Unfortunately, this does not work if the signal is changing from one frequency to another with time. So it is necessary to search not for stationary signals, but instead for moving signals. (Figure 3.14 is an example of a

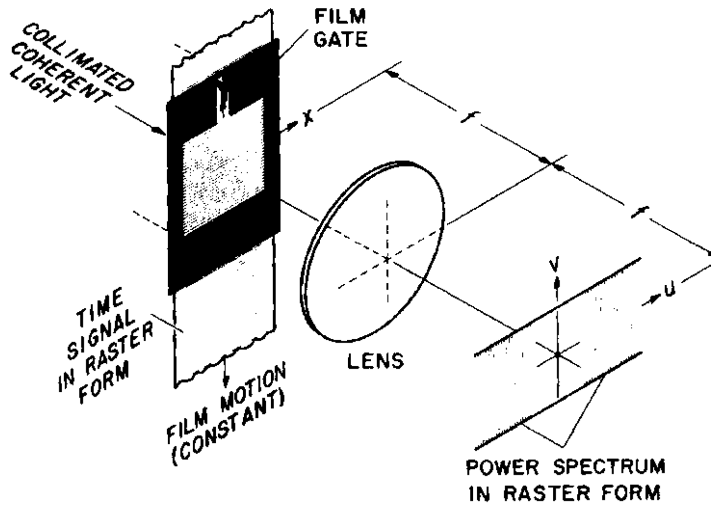



**Figure 3.13** The optical spectrum analyzer.The frequency display is imaged on a high-resolution vidicon tube, scanned and converted to a video signal, and recorded onto magnetic disks.

moving signal.) It is necessary to search for about 100 different Doppler shifts, (which would correspond to 100 different line slopes in the figure). This is done by playing back the data from the magnetic disks into a series of video delay lines that apply varying amounts of delay to each signal. This "straightens out" the sloping lines and makes them vertical. Then they can be averaged in the normal way.

4 **Using *Cyclops* as a camera**: Since a fully-deployed *Cyclops* is orders of magnitude larger than any existing radio telescope, it is a magnificent tool for studying Natural radio signals, apart from its SETI capabilities. It is capable of forming a detailed broadband image of the portion of sky visible to a single element. However, the amount of data that must be processed to accomplish this is daunting, and no really good method has been found. Since the state of the art does not allow digital processing, an analog method must be used. This requires construction of a large shielded building (550 feet long), in which the signals would be re-radiated at the original frequency, from one end by a scaled-down model of the array (using elements such as those shown in Figure 3.15), to another array at the other end which forms the desired image.

### 3.2.6 Cost of *Cyclops*

The costs of the various components of the *Cyclops* system are estimated in Figure 3.16). As can be seen, the cost of the dish antennas is much larger than

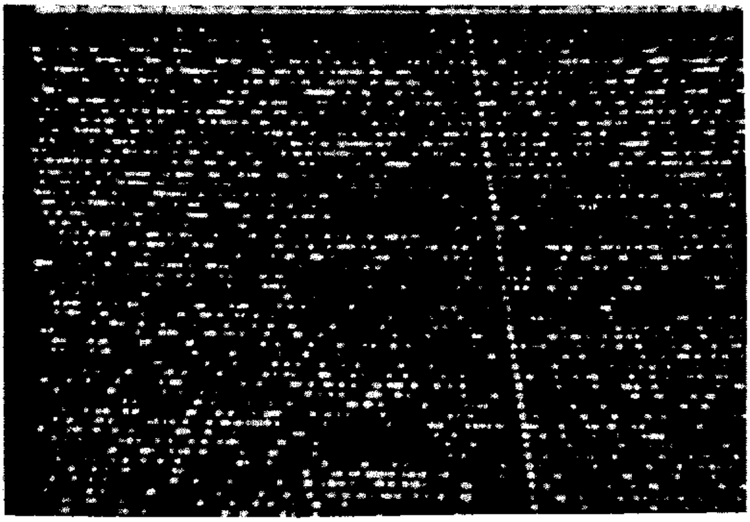


**Figure 3.14** A signal whose frequency drifts with time, as a result of Doppler shift caused by the motions of the transmitter and receiver.

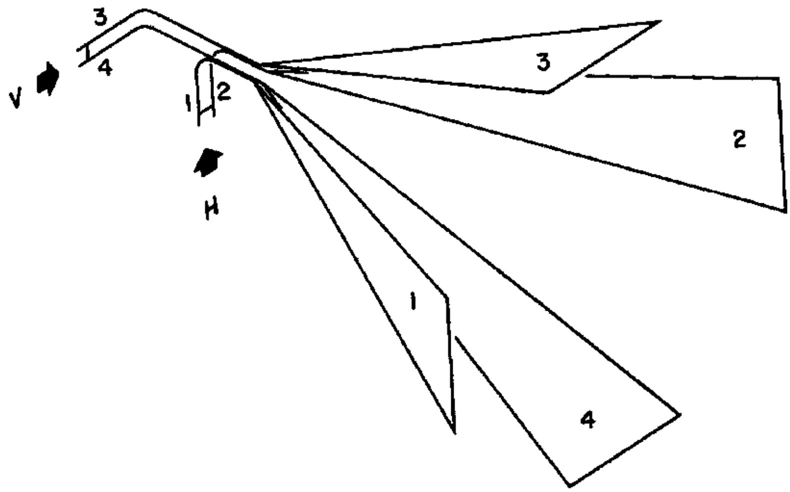



**Figure 3.15** Antenna elements which could be used inside a chamber to create a radio image of the sky as seen by the *Cyclops* telescope.

any of the other component costs. Therefore it is important to investigate further the use of mass production in building the antennas. Since the *Cyclops* system is built gradually as needed, there is no total cost but, rather, there is an estimated $600 million annual cost, over several decades.

### 3.2.7 Conclusions

The conclusions of the *Cyclops* study group are stated here verbatim, as they stand well for themselves, without editing or further comment.

1 It is vastly less expensive to look for and to send signals than to attempt contact by spaceship or by probes. This conclusion is based not on the present state of our technological prowess but on our present knowledge of physical law.

2 The order-of-magnitude uncertainty in the average distance between communicative civilizations in the galaxy strongly argues for an expandable search system. The search can be begun with the minimum system that would be effective for nearby stars. The system is then expanded and the search carried farther into space until success is achieved or a new search strategy is initiated.

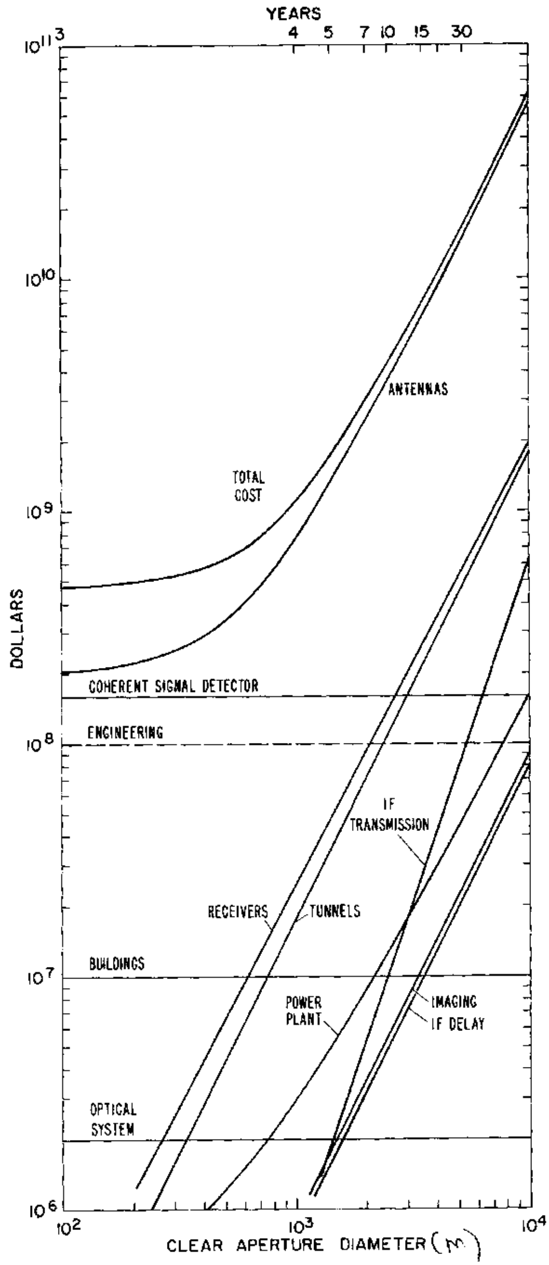



**Figure 3.16** Costs of the various components of the *Cyclops* system, over its initial period of construction and operation.

3 Of all the communication means at our disposal, microwaves are the best. They are also the best for other races and for the same reasons. The energy required at these wavelengths is least and the necessary stabilities and collecting areas are fundamentally easier to realize and cheaper than at shorter wavelengths.

4 The best part of the microwave region is the low frequency end of the "microwave window" - frequencies from about 1 to 2 or 3 GHz. Again, this is because greater absolute frequency stability is possible there, the Doppler rates are lower, beamwidths are broader for a given gain, and collecting area is cheaper than at the high end of the window.

5 Nature has provided us with a rather narrow quiet band in this best part of the spectrum that seems especially marked for interstellar contact. It lies between the spectral lines of hydrogen (1420 MHz) and the hydroxyl radical (1662 MHz). Standing like the Om and the Um on either side of a gate, these two emissions of the disassociation products of water beckon all water-based life to search for its kind at the age-old meeting place of all species: the water hole.

6 It is technologically feasible today to build phased antenna arrays operable in the 1–3 GHz region with total collecting areas of 100 or more square kilometers. The *Cyclops* system is not nearly this large, but we see no *technological* limits that would prevent its expansion to such a size.

7 With antenna arrays equivalent to a single antenna a few kilometers in diameter at both the transmitting and receiving end, microwave communication is possible over intergalactic distances, and high-speed communication is possible over large interstellar distances. Thus rapid information transmission can occur once contact has been confirmed between two civilizations.

8 In the search phase we cannot count on receiving signals beamed at us by directive antennas. Neither can we afford to overlook this possibility. Beamed signals may be radiated at relatively low powers by communicative races to as many as a thousand nearby likely stars and for very long times. Long range beacons, intended to be detectable at any of the million or so likely stars within 1000 light-years, will probably be omnidirectional and very high powered ( $> 10^9$W).

9 Beacons will very likely be circularly polarized and will surely be highly monochromatic. Spectral widths of 1 Hz or less are probable. They will convey information at a slow rate and in a manner that does not seriously degrade their detectability. How best to respond will be contained in this information.

10 The efficient detection of beacons involves searching in the frequency domain with very high resolution (1 Hz or less). One of the major contributions of the *Cyclops* study is a data processing method that permits a 100 MHz frequency band to be searched simultaneously with a resolution of 0.1 Hz. The *Cyclops* system provides a receiver with a billion simultaneous narrow channel outputs. Although the *Cyclops* system bandwidth is 100 MHz, no very great technological barriers prevent widening it to 200 MHz. This would permit searching the entire "Water Hole" simultaneously. *If our conclusion as to the appropriateness of this band is correct, the problem posed by the frequency dimension of the search can be considered solved.*

11 The cost of a system capable of making an effective search, using the techniques we have considered, is on the order of 6 to 10 billion dollars, and this sum would be spent over a period of 10 to 15 years. If contact were achieved early in this period, we might either stop expanding the system or be encouraged to go on to make further contacts. The principal cost in the *Cyclops* design is in the antenna structures. Adopting an upper frequency limit of 3 GHz rather than 10 GHz could reduce the antenna cost by a factor of two.

12 The search will almost certainly take years, perhaps decades and possibly centuries. To undertake so enduring a program requires not only that the search be highly automated, it requires a long-term funding commitment. This in turn requires faith. Faith that the quest is worth the effort, faith that man will survive to reap the benefits of success, and faith that other races are, and have been, equally curious and determined to expand their horizons. We are almost certainly not the first intelligent species to undertake the search. The first races to do so undoubtedly followed their listening phase with long transmission epochs, and so have later races to enter the search. Their perseverance will be our greatest asset in our beginning listening phase.

13 The search for extraterrestrial intelligent life is a legitimate scientific undertaking and should be included as part of a comprehensive and balanced space program. We believe that the exploration of the solar system was and is a proper initial step in the space program but should not be considered its only ultimate goal. The quest for other intelligent life fires the popular imagination and might receive support from those critics who now question the value of landings on "dead" planets and moons.

14 A great deal more study of the problem and of the optimum system design should precede the commitment to fund the search program. However, *it is not too early to fund these studies*. Out of such studies

would undoubtedly emerge a system with a greater capability-to-cost ratio than the first *Cyclops* design we have proposed.

15 The existence of more than one *Cyclops*-like system has such great value in providing complete sky coverage, continuous reception of detected signals, and in long base-line studies, that international cooperation should be solicited and encouraged by complete dissemination of information. The search should, after all, represent an effort of all mankind, not just of one country.

### 3.2.8 Recommendations

The recommendations of the *Cyclops* study group are stated here verbatim, as with the Conclusions.

1 Establish the search for extraterrestrial intelligent life as an ongoing part of the total NASA space program, with its own budget and funding.

2 Establish an office of research and development in techniques for communication with extraterrestrial intelligence. Appoint a director and a small initial staff.

3 Take steps to protect the "Water Hole." through the FCC and corresponding agencies. Use of the spectrum from 1.4 to 1.7 GHz should be limited to interstellar communications purposes. The hydrogen line is already protected. All that is needed to extend this protection upward in frequency is to include the hydroxyl line.

4 Establish, perhaps through the National Academies of Science and Engineering, an advisory committee consisting of interested astronomers, radio astronomers, engineers, physicists, exobiologists and appropriate specialists. The advisory committee should have the initial responsibility for reviewing the available material on the subject, including this report, and of recommending an appropriate course of action. Assuming the committee concurs that further investigations should be undertaken, it should have the responsibility to see that the necessary preliminary scientific studies and engineering design and development are carried out in an orderly manner over a 3 to 5 year period.

5 Make use of outside study contracts initially, but gradually build up internal design and system analysis teams to provide competent contract review and creative in-house (NASA) leadership.

6 As the various systematic and strategic problems of the search yield

to continued study and the overall feasibility approaches general acceptability, begin a series of releases to the scientific community and to the general public to stimulate interest in, and appreciation of, the value of the search.

7 Establish at the outset a policy of open liaison with comparable groups in other countries, that there be no classified information and that all reports be publicly available.

8 When all systemic and strategic problems have been solved, a go-no-go decision must be made. If "go", then political support must be obtained for the funding. The funding must be on a long-term basis so that construction, once started, is not interrupted and can proceed in an orderly way.

9 Make it clear that the system will be available for a certain fraction of the time during the search phase for radio astronomy research and other space programs.

10 Establish the policy of reporting publicly all advances made through the use of the facility. Produce educational films on its capabilities and its mission, and conduct tours of the facility for the public, to sustain interest

**Figure 3.17** An artist's conception of a portion of a fully-expanded *Cyclops* system.

and develop a popular sense of participation in, and identification with, the search.

An artist's conception of a portion of a fully-expanded *Cyclops* system is shown in Figure 3.17.

**Personal Note**

I was one of the Fellows in the *Cyclops* study, in charge of the Signal Processing group. My experiences there were responsible for my starting the first full-time large-scale observational SETI program, using the large radio telescope at Ohio State University (as described elsewhere in this book by Jerry Ehman), and many subsequent years of activity in the field. I shall always be grateful for that experience.

## 3.3 Project Oasis: The Sequel to Project Cyclops

In 1979 NASA sponsored a summer study program at NASA-Ames, called Project Oasis, which was the follow-on to Project Cyclops. Its goal was to design a signal detector to analyze the output of the 8 million channel multi-channel digital spectrum analyzer (MCSA) that had been developed by NASA since Project Cyclops, made possible by the rapid advance of digital electronics in the 1970s. The MCSA replaced the optical spectrum analyzer of *Cyclops*, and it could be used in conjunction with any radio telescope, *Cyclops* or otherwise. This study was a continuation of the Signal Processing portion of the *Cyclops* study. Its title was "The Design of a Signal Detector for the Search for Extraterrestrial Intelligence".

The name *OASIS* was chosen in allusion to the "water hole" region of the electromagnetic spectrum between 1400 and 1700 MHz that is bounded by the hydrogen and hydroxyl molecule spectral lines, that are the disassociation products of water, and where nature passes radio signals with relative ease.

Project OASIS was jointly sponsored by NASA, The University of Santa Clara, and the American Society for Engineering Education. Timothy Healy of the University of Santa Clara and Mark Stull of NASA-Ames served as co-directors. Twenty-four faculty and professionals from across the country, in the fields of Physics, Astronomy, Psychology, Computer Science, Geology, Electrical Engineering, Statistics, Mathematics and Space Sciences were brought together for the summer study program. As it turned out, I was the only Fellow who had also been in Project Cyclops.

Presentations and advice to the group were provided by a number of SETI luminaries including Barney Oliver and John Billingham (directors of

Project Cyclops), Jack Welch (Director of the Hat Creek Radio Observatory at the University of California at Berkeley), Jill Tarter of UC Berkeley (now director of SETI Research at the SETI Institute), Charles Seeger of NASA-Ames, Ron Bracewell (Director of the Stanford University Radio Observatory) and many others.

The specific problems presented to the group were:

1 How does one process half a terabit of data in 1000 seconds?

2 How does one detect a completely unspecified signal with acceptable sensitivity?

### 3.3.1 The OASIS Signal Detector

Since:

1 the characteristics of the signal to be detected are completely unspecified (by the definition of the problem to be solved), and

2 computing resources are now abundantly available,

there is no reason to use just a single detection algorithm. In fact the opposite is true; we should use a battery of different detectors all at once (the shotgun approach) in the hopes that at least one of them will find the signal. The detectors are grouped into five categories:

1 The narrowband signal detector

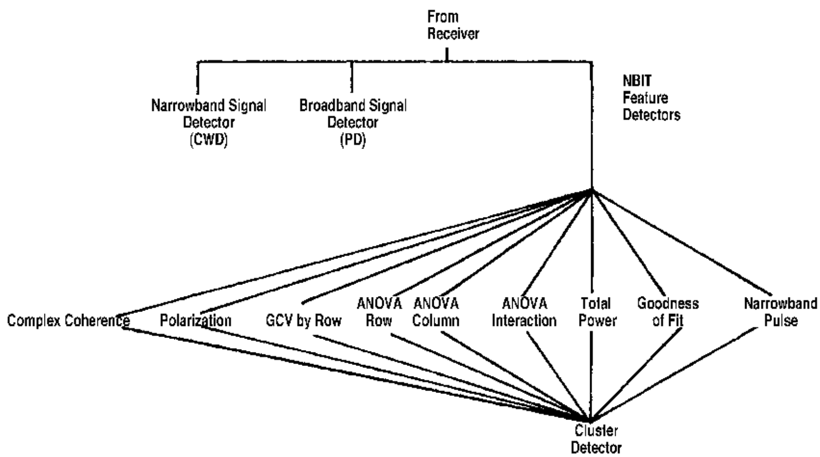



**Figure 3.18** Overall design of the OASIS signal detector.

2 The broadband (pulsed) signal detector

3 A battery of nine independent “feature detection” algorithms, called NBIT, that make no assumptions as to the form of the signal

4 A pattern recognition algorithm (cluster analysis) that looks for global “features” across all of the above nine algorithms taken together

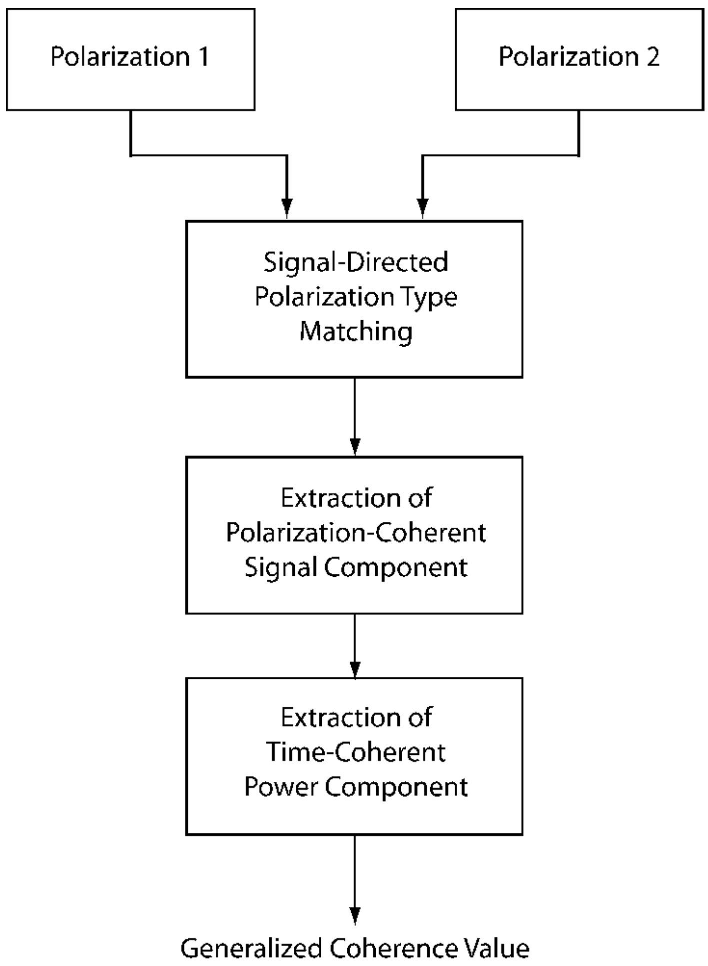



**Figure 3.19** Calculation of the Generalized Coherence Value (GCV).

5 Use of the unique characteristics of the human eye–brain system as a detector of unusual events and patterns that might not be detected by any of the computer-based detectors

Figure 3.18 illustrates how the various computer-based algorithms are organized.

### 3.3.2 Generalized Coherence

Coherence is the property of an intelligent signal that distinguishes it from random noise. There is Time Coherence, which measures the consistency of the amplitude and phase of a signal with respect to a time-delayed version of itself. And there is Polarization Coherence, which measures the consistency of the amplitude and phase between the signals received by two orthogonally polarized antennas. Both are important in detecting intelligent signals. Both are typically determined in different ways, but it is possible to combine them into a more general quantity called Generalized Coherence.

It is possible to process the output signals of two orthogonally polarized antennas so as to create a single synthetic antenna whose polarization is automatically matched to the polarization of the incoming signal, regardless of what its polarization might be. This is called polarization matching, and it is an optimum first step in signal detection since it maximizes the signal-to-noise ratio and provides a single signal for further analysis, avoiding the doubling of cost and complexity otherwise required to process both polarizations separately.

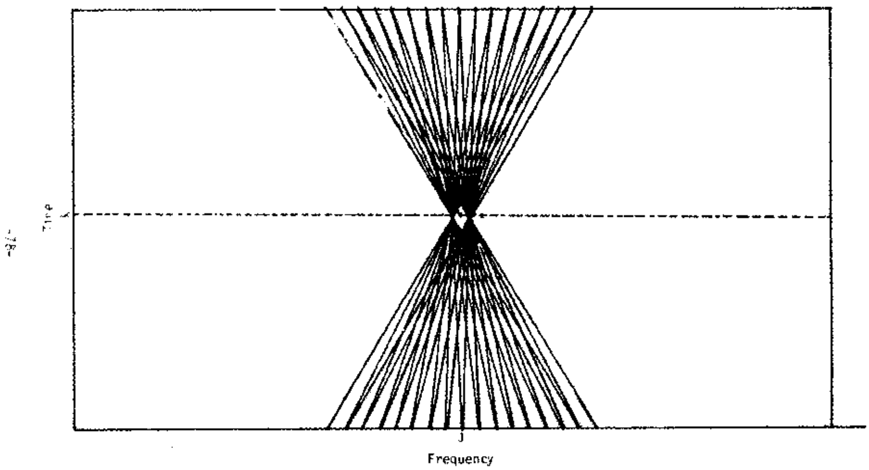



**Figure 3.20** Time-frequency masks used to detect drifting signals in the narrowband signal detector.

Not all signals are polarized. Many signals of natural origin, such as from stars and galaxies, are not polarized at all, and some are only partially polarized. But any signal radiated from an actual antenna is always completely polarized. So this is another important criterion that can be used to distinguish intelligent signals from natural ones. It is possible to process the output signals of two orthogonally polarized antennas to calculate the Degree of Polarization, which ranges from 0 to 1. This factor can be used to assist in decisions related to signal detection.

All of the above calculations can be combined into a single quantity called the Generalized Coherence Value (GCV), as illustrated in Figure 3.19. The GCV is used in several of the Feature Detectors.

### 3.3.3 Narrowband signal detector

The MCSA produces an 8 million point spectrum once per second. The output of the MCSA is overlaid with a mask of 20 rays, each of which corresponds to a different drift rate (Figure 3.20 shows two such masks). A mask is centered at each of the MCSA frequency output bins exactly once, and selects a ray of 20 MCSA data points (20 seconds of data). The GCV is calculated and compared to a threshold. If it exceeds, that is noted. In addition, all GCVs are accumulated for 1000 seconds and compared against another threshold, to look for more extended drifting features.

### 3.3.4 Broadband pulse detector

The MCSA outputs are coarsely binned together to create wider frequency channels, more suitable for detecting pulsed signals. The binning is done with adjacent bins repeatedly, creating first 2 Hz bins, then 4 Hz bins, etc. for a total of 23 progressively wider bandwidths, ending up with the the final bin being the total power in the 8 million channels all averaged together. All the various bandwidths are compared against a threshold, and any that exceed are noted.

### 3.3.5 Numerical Battery of Independent Tests (NBIT)

The MCSA output data is subdivided into blocks of 40 Hz by 20 seconds in frequency -time space. Nine different tests are applied to each block, and any which exceed a threshold set a single bit in a 9 bit number. The nine tests are:

1 Total power in two polarizations

2 Degree of polarization

3 Complex amplitude coherence detection

4 Broadband coherent pulse detection for each polarization using generalized coherent values

5 8 Hz pulse detection for each polarization

6 Analysis of variance (ANOVA) for rows (frequency)

7 ANOVA for columns (time)

8 ANOVA interaction (frequency-time), for each polarization

9 Probability distribution of the data compared with the expected distribution in the presence of just noise

A two-stage cluster detector then examines all the 9 bit numbers. In the first stage, the total number of set bits is compared against a threshold and, if exceeded, that is noted. The second stage is a single linkage cluster-seeking algorithm that searches the entire block of 9 bit numbers for subtle signal patterns, and notes if any are found. The *OASIS* report explains all this in far greater detail. Calculation of quantities such as GCV and ANOVA are of course affected by the presence of noise, and these effects were calculated and numerically simulated, with the results shown in appendices.

### 3.3.6 The human signal detector

The human eye–brain system can do things that a computer cannot, since it has no preconceived ideas ("programs") that tell it what to look for. The tradeoff for accepting some degree of inconsistency and low precision is to add a "wild card" that might just see something - maybe a wavy line instead of a straight line, or maybe a "hunch" about something that seems to fade in and out, or change shape and size. Maybe a funny-looking blob among otherwise random noise. Human performance is known to:

- degrade gracefully under less than optimal conditions
- extract interesting and unusual events from a large body of information
- process information on several levels, both conscious and subconscious; and
- benefit from training and experience

Clearly a human cannot cope with the entire raw data coming out of the MCSA, but an appropriate role may be at the cluster analysis stage, where the data are already greatly filtered.

## 3.4 Obtaining the *Cyclops* and *OASIS* Reports

The original edition of the *Cyclops* report is long out of print, although used copies can be found on Amazon.com. Both the *Cyclops* and *OASIS* documents can be obtained from the NASA Technical Reports Server, in printed form or as a downloadable pdf.

*Cyclops* pdf document http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730010095_1973010095.pdf (14.5 MB)

*OASIS* pdf Document http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19820016111_1982016111.pdf (29.5 MB)

To purchase printed-to-order copies, go to

https://www.sti.nasa.gov/cgi-bin/ordersti.pl

and provide the document ID numbers:

*Cyclops* Document ID 19730010095

*OASIS* Document ID 19820016111

In 1995 the SETI League and the SETI Institute collaborated to reprint the Cyclops report in its original form. It can be purchased from the SETI League at http://www.setileague.org/photos/premiums.htm

The reprint contains a preface by Barney Oliver in which he comments on the relevant technological changes that have occurred during the ensuing 25 years. Chief among those are the digital revolution that rendered the optical spectrum analyzer obsolete and made possible much more powerful signal analysis. Additional introductory material by SETI League president Richard Factor and executive director H. Paul Shuch, as well as a concluding tribute to Barney Oliver penned by John Billingham, further place this remarkable publication in its proper historical context, as a blueprint for the greatest radio telescope never built.

# 4

# “Wow!” - A Tantalizing Candidate

Jerry R. Ehman,
*Radio Astronomer, Ohio State University Radio Observatory (retired)*

This chapter discusses the “Wow! signal” detected on August 15, 1977 by the Ohio State University Radio Observatory (OSURO) radio telescope (often called the “Big Ear”). Let’s start with some history of OSURO prior to that detection.

## 4.1 Brief History of the Ohio State University Radio Observatory

Before the “Big Ear” radio telescope was constructed, Dr John D. Kraus (the Director of OSURO), designed and built a radio telescope that eventually consisted of 96 11-turn helical elements as shown in Figure 4.1. Each helix was about 2 meters long and 30 centimeters in diameter. These elements were perpendicular to a tiltable wire mesh plane. This helix array was operated at a frequency of 250 MHz (megahertz). The single beam from the array pointed along the north -south meridian, depending on the tilt of the plane. The Earth’s rotation was used to scan each strip of sky (each strip covered a constant span of declinations). This instrument had a moderately large collecting area and worked well. A map of the sky visible to that telescope was made. However, John Kraus realized that the frequency span was only about 2:1 and he wanted to design a new instrument that had a much wider frequency range (at least 10:1).

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_4, © Springer-Verlag Berlin Heidelberg 2011

**Figure 4.1** The first radio telescope designed and constructed by John Kraus.

By 1956 he had completed the design of his new radio telescope and construction began. By 1961 it became operational. Figure 4.2 shows (from left to right): a fixed standing paraboloidal (curved) reflector at the south end; a flat aluminum-covered ground plane; two feed horns, and a tiltable flat reflector at the north end. Underground, below the feed horns, was a room that contained the receivers and other electronics, and a computer.

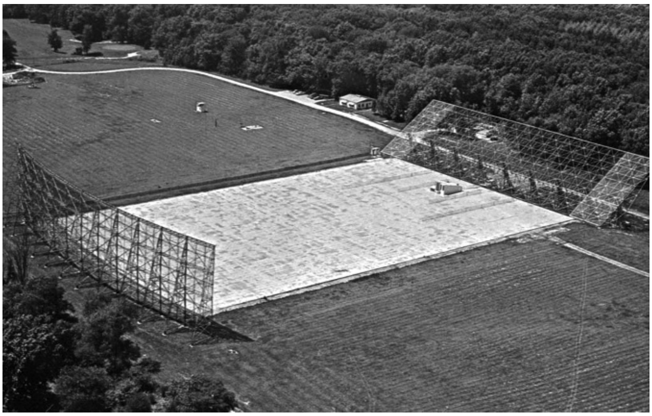


**Figure 4.2** John Kraus's new radio telescope, 1961.

This telescope was also a meridian-transit instrument that used the earth's rotation to bring radio sources into view. Figure 4.3 shows how a radio wave (photons) reflects off the flat reflector, travels to and reflects off the paraboloidal (curved) reflector and is collected in turn by each of the feed (collecting) horns. The aperture (physical collecting area) was 340 feet (104 meters) in the east -west direction by 70 feet (21 meters) in height. Although the term "a big ear" was used by a local newspaper reporter for the first 6 helical elements of what would become the 96-element helix array, the name "Big Ear" was applied to this new radio telescope. Because of its unusual design, it also came to be called a "Kraus-type radio telescope." A telescope of similar design was built in Nancay, France.

For the first several years, the main frequency band of operation of the "Big Ear" was 1411 -1419 MHz (or wavelengths of about 21 cm), although observations were also conducted at 612 MHz and 2650 MHz (wavelengths of 49 cm and 11 cm, respectively). For the frequency band of 1411–1419 MHz, the angular size (HPBW=half-power beamwidth) of each beam (and there were two beams) was 8 arcminutes in right ascension by 40 arcminutes in declination. The two feed horns were separated by about 1.5 meters in the east -west direction. The receiver switched between the two horns at a rate of 79 Hz (cycles per second) and amplified the difference between the two signals. This switching had the effect of significantly increasing the sensitivity of the receiver to discrete (small angular diameter) radio sources. The other effect of having two horns was that two responses on a detectable radio source were received, one after the other, with the first response (from the "negative horn") dropping below the baseline and the second response (from the "positive horn") rising above the baseline.

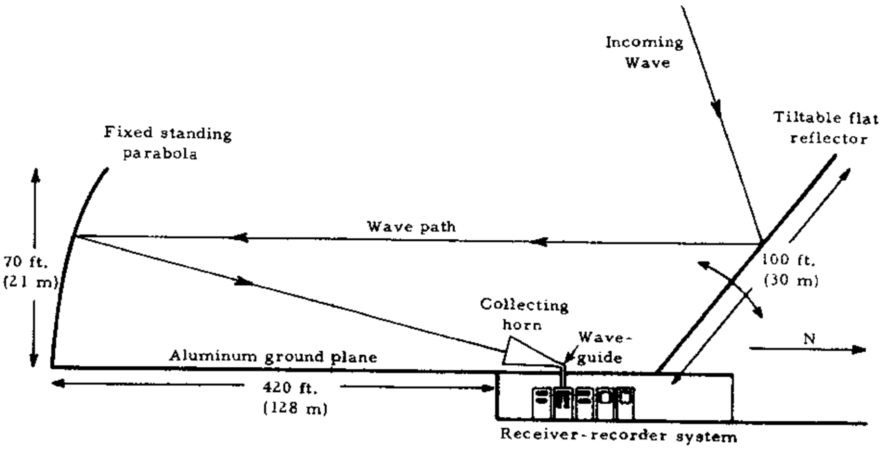



**Figure 4.3** Radio wave reflects off the flat reflector, travels to and reflects off the paraboidal (curved) reflector and is collected in turn by each of the feed (collecting) horns.

Until the mid 1970s the Ohio Sky Survey was the major project at OSURO. (I joined OSURO as a radio astronomer in 1967; the Ohio Sky Survey was already in progress when I joined.)

In the early 1960s Robert S. (Bob) Dixon came to OSURO as a graduate student in electrical engineering. He worked on the Ohio Sky Survey developing the procedures for analyzing the data. He wrote much of the computer software for analysis of the “Big Ear” continuum (wideband) data for the Ohio Sky Survey. He coordinated the team that worked on that analysis and wrote his PhD dissertation on those procedures. Later, Bob became the Assistant Director of OSURO.

A total of over 19,000 discrete radio sources were measured, more than half of which had not been previously measured by any observatory. This Ohio Sky Survey was unique in that it covered a larger portion of the entire sky (about 70%) in more detail and at a shorter wavelength (21 cm) than any previous large-area survey. Tables containing the coordinates and signal strengths (flux densities) plus contour maps of signal strengths were published in nine articles (seven installments plus two supplements) totaling 660 pages in the *Astronomical Journal*. I was a coauthor for the fourth, fifth and sixth installments.

John Kraus and Bob Dixon compiled lists of radio sources called Ohio Specials from their radio spectra. These spectra (signal strength (flux density) versus frequency) were obtained from: OSURO observations at wavelengths of 21 cm, 49 cm, and 11 cm; published measurements from many other radio observatories; and measurements at other observatories by OSURO personnel. The goals were to find radio sources that had unusual radio spectra, then determine the best position for each radio source, and finally to identify the object optically on photographic plates (optical identification), if possible.

One of the Ohio Specials was OH471.[1] This was the first object known to have a redshift[2] greater than 3 (it was measured to be 3.4), and was described by *Time* magazine as “the blaze marking the edge of the universe.”

Three months after the redshift of OH471 was determined, the redshift

---

1 “O” stands for Ohio or OSURO; “H” refers to a 1-hour block of right ascension beginning with 6 hours; “4” means that the declination was at least 40 degrees and up to but not including 50 degrees; and “71” indicates that the remainder of the right ascension (here in the 6-hour block) was about 71/100 of an hour. The measured right ascension was 6h 42m 54s and the declination was 44° 52’.

2 Redshift is the change in wavelength divided by the original wavelength (for an object moving away from us, such as a distant galaxy). For OH471, the redshift of 3.4 means that the observed wavelength of a spectral line was 3.4+1.0=4.4 times that of the wavelength of that line measured in a laboratory on Earth. From this redshift, OH471 was receding at 90% of the speed of light, and, thus, was located at (approximately) 90% of the distance to the “edge of the universe”.

of OQ172 was found to be 3.53 (at a recession speed of 91% of the speed of light). Hence, as a result of the Ohio Sky Survey and the Ohio Specials Project, the two most distant sources (at that time) were found.

Bob Dixon directed a project to generate “A Master List of Radio Sources”, which was published in the July 1970 issue of the *Astrophysical Journal Supplement*. This list was the first publication to contain information about all discrete radio source measurements at all frequencies. This became very helpful in determining Ohio Specials. He also directed a project to compile “A Master List of Non-Stellar Objects” which was published in 1980 by the Ohio State University Press.

In 1981 Bob Dixon completed a project that produced acetate overlays for the prints of the National Geographic Society–Palomar Observatory Sky Survey (which had used the 48-inch Schmidt telescope). Each overlay contained the names of radio sources, so that when placed over a Palomar Sky Survey photographic print, it was often easy to identify an optical object corresponding to a radio source. Sometimes, however, the optical identification was not obvious. Hence, better locations were often needed.

Bob joined a group discussing the design of an extremely large radio telescope called *Cyclops*. Bob described *Cyclops* in Chapter 3, “*Cyclops*: The Greatest Radio Telescope Never Built”.

## 4.2 Transition from Wideband to Narrowband Observations

To operate OSURO and its “Big Ear” radio telescope, John Kraus applied for and received annual grants from the National Science Foundation (NSF) for many years. Unfortunately, in 1972, he received word that no further grants would be forthcoming due to a change in policy, namely, that more financial support would be granted to the national observatories by significantly reducing support to several university-supported observatories in the US. This meant that our team members would no longer be supported by the grant; with the exception of John Kraus (whose salary came from the Ohio State University’s Department of Electrical Engineering), all personnel, including Bob Dixon and myself, lost that compensation. Bob found employment at the OSU Computer Center and I went to another university in Columbus, Ohio. However, Bob and I, as well as several others, continued on the team as volunteers (unpaid!) in our spare time.

Bob Dixon and John Kraus had discussions about how to continue using the “Big Ear”. It was obvious that continuing the Ohio Sky Survey and related projects was very labor intensive, and really couldn’t be done effectively with a few volunteers. Bob suggested to John that the receiver could be converted from wideband operation (8 MHz at 21cm) to narrowband operation. [Note that almost all celestial radio sources (like galaxies, stars, quasars, etc.) are wideband sources (often called *continuum sources*), generating photons in the radio, optical (including visible, infrared and ultraviolet) and even into the X-ray and gamma ray bands. Narrowband sources are almost always purposely generated by intelligent beings (examples include AM, FM, TV and “ham” radio broadcasts, satellite transmissions, and radar). Wideband operation was important when measuring radio sources that generate photons over the entire radio band and beyond in order to increase the sensitivity of detection.] OSURO had received two narrowband filter banks from NRAO (National Radio Astronomy Observatory) that they were no longer using. Both were 50-channel units, one with 100 kHz (100 kilohertz = 100,000 Hz) channels and the other with 10 kHz (10,000 Hz) channels. Bob decided to use the one with 10 kHz channels. For 50 channels each 10,000 Hz wide, a total of 500,000 Hz could be observed at once.

Bob decided to dynamically adjust the frequency band of observation to compensate for: the rotation of the Earth, the revolution of the Earth about the Sun, and the orbit of our solar system about the center of our galaxy. This had the effect of removing the Doppler shift of our observatory relative to the Galactic Center of Rest (GCR). The idea was that if an extraterrestrial intelligence (ETI) within our galaxy did the same when they transmitted a signal in our direction, we would be observing in the correct frequency band, assuming that the ETI had chosen the 21cm band to transmit. The 21cm band was a logical band to transmit and receive because any civilization that has the knowledge of radio waves at least comparable to ours would know about the line of neutral hydrogen at 21 cm (at a frequency of 1420.4056 MHz after Doppler shifts have been removed), and would also know that hydrogen is the most abundant element in the universe.

Bob wrote most of the computer program for acquiring and analyzing the data. I wrote additional software to extend the analysis. Some details of this analysis are in the next section.

John Kraus became increasingly interested in the detection and measurement of narrowband radio signals and, especially, in the search for those that could come from extraterrestial intelligences (called SETI, the Search for ExtraTerrestial Intelligence). He used his own funds to start a magazine called *Cosmic Search* to deal with the wide variety of topics related to SETI. Bob Dixon assisted with that effort and many well-respected authors contributed to the magazine. Although it was well received by the scientific community and some of the general public, the number of subscriptions did not reach the level necessary for sustainability. Hence, John discontinued the publication after 13 issues.

Several years after the demise of *Cosmic Search* magazine, I took on the task of putting all 13 printed issues online. They can be read with a web browser by going to the following web address: www.bigear.org/CSMO/HTML/CSIntro.htm.[3]

## 4.3 The "Wow!" Signal

The narrowband observing program with the "Big Ear" radio telescope was set up so that no person needed to be at the Radio Observatory, except to make certain adjustments every three days or so. An IBM 1130 computer was programmed to acquire and analyze data from the receiver, which operated in a frequency band on and near the 21 cm line of neutral hydrogen at 1420.4056 MHz. This computer contained 65,536 bytes (64 kB) of RAM (Random Access Memory). However, since it used 2 bytes minimum (a "word") to store a data point in memory, it could hold only 32,768 such data points (including the software program). The computer program, called N50CH, was written mostly by Bob Dixon (with some additions by me) using both Fortran and assembler languages. Because of the very small size of RAM (i.e., the same size as that of an Apple II personal computer), it was necessary to create four main modules that successively (sequentially) overlaid each other. Key parameters were kept in a COMMON area. Each module called several subroutines. All four modules and most of the subroutines were written in Fortran although it was necessary to have a few subroutines written in assembler language.

In a manner similar to that used for the Ohio Sky Survey, after data started to come in regularly, we began a systematic survey of the 100 degrees of declination visible to the radio telescope (from +64 degrees down to -36 degrees). The telescope was kept at the same declination setting for three or four days allowing the Earth's rotation to give us the three or four passes over each value of right ascension. Then our mechanical technician Gene Mikesell moved the flat reflector to change the declination by 20 arcminutes (½ of the half-power beamwidth in declination). Just before resetting the declination, Gene would stop the computer and save the computer printout. Once the new declination was set, he would restart the computer for the new run. Our IBM 1130 computer had a hard drive that held a "whopping" 1.0 MB (1 megabyte; i.e., less than that of a 3.5 inch floppy disk). Bob and I were very careful about deciding just what to store on the computer. This allowed the hard drive to hold about four days' worth of data.

Shortly after the narrowband program began, Bob asked me to look at the computer printouts from the N50CH program. At that time, although I

---

[3] Note that the web address: www.bigear.org/ will take you to the home page of the OSURO website that deals with the "Big Ear" radio telescope.

was a volunteer with OSURO, I was employed at a university in downtown Columbus. We arranged for Gene Mikesell to deliver the computer printouts to my home every few days. I would look at those printouts and record anything that looked interesting.

While looking at the computer printout of August 15, 1977, I saw the data from the strongest narrowband signal I had ever seen. I immediately recognized the pattern of data points as that of a signal in one channel that varied in a manner for a celestial source moving through the beam of the antenna (due to the Earth’s rotation). I was so astonished at this strong signal, I wrote, in red ink, the notation “Wow!” in the printout’s margin. After finishing the review of the rest of the printout, I called John Kraus and then Bob Dixon to let them know about this signal. A day or so later the three of us met on campus to look at this data. We were all astonished. This began an extensive discussion and a search of the literature for possible celestial objects that could be the the source of this signal. Later, John started calling this signal the “Wow! signal” and the unknown source the “Wow! source”.

Figure 4.4 shows a portion of the computer printout with my handwritten notation of “Wow!”. (This image is from a scan of a color copy of the original computer printout taken several years after the 1977 arrival of the “Wow!” signal, and after the printout had faded noticeably.) The IBM 1130 computer running the N50CH program interacted with the receiver to acquire digital intensity values from each of 50 10 kHz (10,000 Hz) channels once each

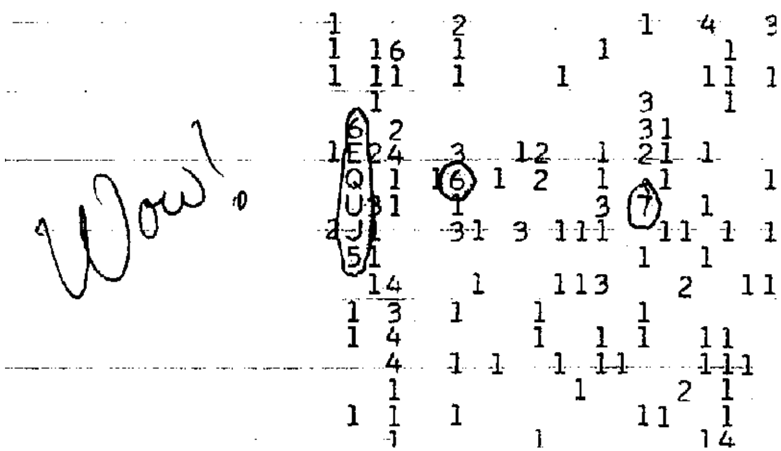



**Figure 4.4** “Wow!”

second. Ten of these values were combined to generate one number for each channel and the number for each channel was converted to a single digit or letter and printed out (two seconds were needed for the analysis and printout of each line of information). The sequence of six intensities thus lasted 6 times 12 = 72 seconds. This entire operation could be handled by the computer with no person present (except for the twice weekly stopping, resetting, and restarting the computer).

The “Wow!” source radio emission entered the receiver of the radio telescope at about 11:16 PM Eastern Daylight Savings Time on August 15, 1977. No one was at the observatory at the time. The set of characters “6EQUJ5” represents the sequence of signal strengths. The first value “6” means that the signal strength was between 6.0 and 6.99... times the background noise (standard deviation). Because of limited space on the computer printout to represent the signal strength for 50 channels, it was necessary to use only one print position. Hence, only one character could be used for the signal strength. The fractional portion was dropped (truncated). For signal strengths of 10 or more, letters of the alphabet were used. Thus “E”, being the fifth letter of the alphabet, represents a signal strength of 14.0 to 14.99... (or 14 after truncation (9+5 = 14)). Hence, the sequence of “6EQUJ5” is the set of truncated values: 6, 14, 26, 30, 19, and 5.

Figure 4.5 shows a plot made by John Kraus of the sequence of signal strengths. He noted that the pattern matched that expected from a source of

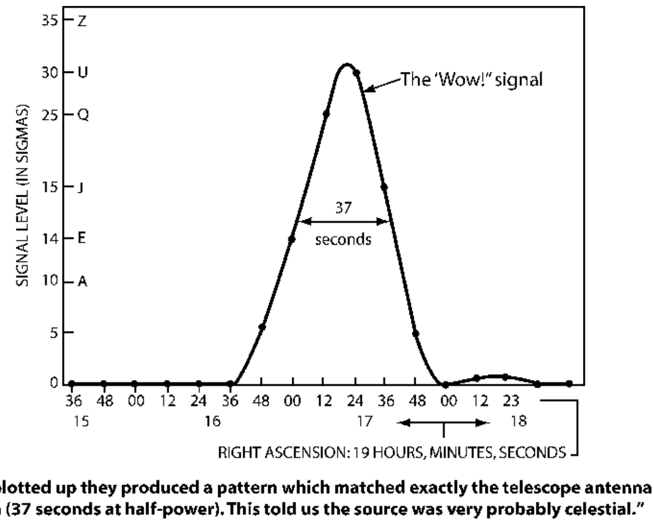



**Figure 4.5** The sequence of signal strengths of the “Wow!” radio emission.

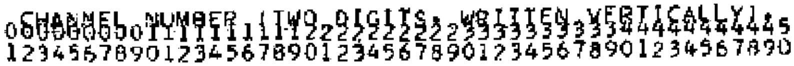



**Figure 4.6a** Left side of a 3-line header.


**Figure 4.6b** Right side of a 3-line header.

angular diameter much smaller than that of the beam of the telescope. Later, I did an analysis that showed the correlation of the “Wow!” data points with the antenna pattern was 99.14% (very close to “perfect”).

I noted previously that we didn’t have much space on the computer printout. We used the line printer attached to the IBM 1130 computer which was designed to print out one line of characters at a time. Each line was 120 characters at 10 characters per inch (fixed font). Figures 4.6a and 4.6b show the left and right sides, respectively, of a 3-line header.[4]

The information for each row of the printout is in sequence: signal strength for each of the 50 channels (truncated value in standard deviations of the background noise); a column to represent the signal strength for the continuum (wideband ) component (not used in 1977); right ascension and declination (for epoch 1950); the second local oscillator frequency (which was used to determine the center frequency for each channel); the galactic latitude and longitude; the Eastern Standard time; and the remaining columns to list nearby objects (which never got implemented).

Figure 4.7 shows about three minutes’ worth of the computer printout centered on the “Wow!” signal. (My notation of “Wow!” was written with a red pen.)

---

4 Note that the paper didn’t always move smoothly through the printer.

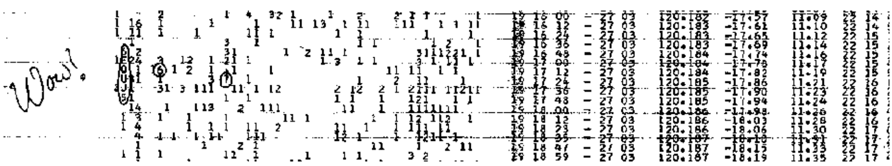


**Figure 4.7** Three minutes’ worth of the computer printout centered on the “Wow!” signal.

As I’ll now explain, several of the values shown on the printout needed adjustments, so they should not be used directly from this printout.

**Sequence of signal strengths: “6EQUJ5”:** Because of truncation (i.e., displaying only the integer portion of the signal strength or an alphabetic letter to handle signal strengths of 10 or more) a value of 0.5 was added to the printed signal strength when doing various analyses to account for this truncation. Thus, an uncertainty of ± 0.5 must be assigned to each signal strength value.

**Right ascension and declination:** Two feed horns were used. Rapid switching between the two horns was done electronically. The difference signal between the two horns was thus obtained. Because the two horns were physically separated from each other, a radio source would enter first one horn and then the other. The time between the two “peaks” (a negative response followed by a positive response) was about 150 seconds at the Equator (declination=0) but progressively larger at greater distances from the Equator. For some reason, the signal was seen in only one horn. At the time, we did not indicate on the computer printout the sign of the difference signal (we soon corrected this oversight by overprinting a minus sign for each negative difference). Thus, we had no way of knowing in which horn the “Wow!” signal arrived, so we had to calculate two separate values of right ascension.

We also converted these coordinates to epoch 2000 to account for the *precession of the equinoxes.* The Earth is like a spinning top in which the north polar axis successively points to a different location in the sky (now located near the star Polaris). It takes about 26,000 years to make one complete rotation.

**Galactic latitude and longitude:** The adjustment of right ascension and declination to epoch 2000 required adjustment of the corresponding galactic coordinates.

### 4.3.1 Summary of the “Wow!” Signal Parameters

Many adjustments were made to the printed values to obtain corrected values. Some of these were mentioned above. However, a complete detailed description is made in a web-based article I wrote entitled “The Big Ear Wow! Signal (30th Anniversary Report)”. The web address for it is: “www.bigear.org/Wow30th/wow30th.htm”. This report was written in 2007, 30 years after the August 15, 1977 occurrence of the signal.[5]

Based on all of the information presented in my reports, here is a summary of the final parameters for the “Wow!” signal:

**Epoch 2000 right ascension and declination:**

R.A. (positive horn): 19h25m31s +/ - 10s
R.A. (negative horn): 19h28m22s +/ - 10s
Declination: - 26d57m +/ - 20m

**Galactic latitude and longitude:**

**Latitude:**

Positive Horn: -18d53.4m +/ - 2.1m
Negative Horn: -19d28.8m +/ - 2.1m

**Longitude:**

Positive Horn: 11d39.0m +/ - 0.9m
Negative Horn: 11d54.0m +/ - 0.9m

**Eastern Standard Time (EST) and Eastern Daylight Time (EDT) for source peak:**

EST: 22h16m01s (or 10:16:01 PM)
EDT: 11:16:01 PM

**Frequency of observation (center of channel 2):**

1420.4556 +/ - 0.005 MHz

**Flux density:**

Flux density (inside a 10 kHz band): Results from two independent analyses: 54 Jy or 212 Jy (where Jy means the unit of jansky).

**Constellation:**

Both possible positions of the signal were located in the constellation of Sagittarius.

---

5 10 years earlier, in 1997 I wrote a report entitled: *“The Big Ear Wow! Signal - What We Know and Don’t Know About It After 20 Years”*. The 30th anniversary report is more detailed than the 20th anniversary report.

### 4.3.2 Summary of “Wow!” signal characteristics

Before discussing various speculations about the possible origins of the “Wow!” signal, let’s review the major characteristics that must be accounted for when hypothesizing such possible origins.

Here is a list of those major characteristics:

- narrowband signal: less than 10,000 Hz wide (in one channel only);
- signal observed in only one (of two) horns, but without the capability of determining which horn;
- signal observed only once (not seen on subsequent observations at the "Big Ear" radio telescope nor by observations made later at various observatories);
- within each 10-second observing interval, the average signal strength remained constant;
- for the entire observing interval of 6 data points lasting 72 seconds, the average signal strength remained constant (because the 6 data values follow the antenna pattern to better than a 99% accuracy);
- modulation (signal strength variation) on a time scale less than 10 seconds or more than 72 seconds could not be measured;
- the “Wow!” source didn't move (or moved very little) with respect to the celestial sphere (again because the 6 data values closely follow the antenna pattern).

### 4.3.3 An unsupportable conclusion

It has come to my attention that several persons in the scientific and engineering community have decided, more or less definitively in their minds, what the origin of the “Wow!” signal is. Their typical “conclusion” is that the signal is RFI (i.e., Radio Frequency Interference). Unfortunately, these conclusions are made by ignoring one or more of the characteristics listed above. Since these persons are well-respected and technically trained members of the scientific and engineering community, it is surprising, even shocking, to me that they would ignore some of the data (characteristics) and form such conclusions.

Here may be some reasons why otherwise intelligent persons would violate a basic tenet of the Scientific Method by selectively ignoring some data:

- They want to state a definitive conclusion, and are not able to deal with the uncertainty of open (unresolved) questions; or
- They have some bias against the Ohio State University Radio Observatory

(OSURO) and/or its Director, Dr John D. Kraus, and/or some other person associated with OSURO; or

- They want some organization other than OSURO to find and claim the first possible detection of a signal from an ETI.

## 4.4 Speculations, Hypotheses, and Investigations

After I showed the computer printout of the "Wow!" source to John Kraus and Bob Dixon, we immediately talked about it, speculating and making hypotheses. Quickly, John and Bob began to investigate the various possibilities (I wasn't heavily involved early on in this aspect since I was continuing to examine the incoming data from the telescope; in later years, I did pursue various analyses). I'll now discuss some of the possibilities. Some were ruled out and I will state why they were ruled out. Note that, in scientific parlance, the words "ruled out" mean "to assign a very low probability to".

**Planets and their moons:** The positions of all of the planets in our Solar System were looked up in an ephemeris. None of the planets were close to the "Wow!" source position. Of course, one would not expect a Solar System planet or its moons to be generating a narrowband radio emission (and we had not previously detected characteristics of ETIs on any of our Solar System planets and moons). Normally, when a planet/moon is observed in the radio band, we detect the radio emission over the entire radio band (assuming the telescope is sensitive enough). That radio emission is usually "thermal emission" due to the temperature of the object, although Jupiter generates non-thermal emission as well. Not only did the "Wow!" source emission not fit the pattern of this Jupiter-style emission or the thermal-type emission, but, in addition, none of the planets were in the proper position in the sky.

**Asteroids:** Asteroids are typically small rocky objects. They have negligible magnetic fields and hence negligible non-thermal radiation. Since their masses and surface areas are so much smaller than our planet's, they generate much less thermal radiation (but again thermal emission is not narrowband). However, the ephemeris was consulted for the locations of some of the larger asteroids, but none were in the vicinity.

**Satellites:** If a satellite from the US or Russia (formerly the Soviet Union) or other country were broadcasting around 1420 MHz, the "Big Ear" would have easily been able to detect it when it was in the beam. The frequency band around 1420 MHz (a few MHz on either side) was declared off limits for satellite transmission or Earth-based broadcasting over the entire world. Thus, no satellite should have been sending out any transmission in this protected band. If a country sent up a satellite that was violating this agreement, it is quite possible for the signal to be narrowband. For example, the AM (amplitude modulated) radio stations in the frequency range of around 0.54 -1.6 MHz (540 -1600 kHz) transmit over a bandwidth of approximately10 kHz, the same bandwidth as each of the 50 channels in our receiver. (Note that the bandwidths of FM radio and television are much wider than 10 kHz.) An investigation of the orbits of all known satellites revealed that none were in our beam at the time of the "Wow!" source. Also, satellites move with respect to the celestial sphere, while the pattern from the "Wow!" signal did not show such movement.

**Aircraft:** There are two major ways to rule out airplanes and other aircraft: (1) no aircraft transmitters operate in the protected radio band around 1420 MHz; and (2) aircraft move very rapidly with respect to the celestial background. The "Wow!" source intensity pattern received matched almost perfectly the pattern expected from a small-angular-diameter (point) radio source on the "celestial sphere" (i.e., at such a large distance that there is no perceptible motion relative to the background stars). An aircraft, which would show a significant motion with respect to the stars, would also cause the received pattern of intensities to depart noticeably from that expected for a celestial "point source".

**Spacecraft:** A check was made for known spacecraft and none were near the direction of "Wow!". In addition, a spacecraft is not supposed to be transmitting in the protected band. Again, spacecraft move with respect to the celestial sphere.

**Ground-based transmitters:** No transmitter on Earth or in space should have been transmitting in the protected band around 1420 MHz. I have already stated how a transmitter in space (an aircraft, a satellite, or other nearby spacecraft) would not be able to generate a point-source type response in our receiver. But how about a ground-based transmitter?

A ground-based transmitter is fixed to the ground. The "Big Ear" radio telescope is also fixed to the ground. Therefore, even if a signal from such a transmitter were getting directly into our receiver, there would be no relative motion and, hence, no way to have the signal intensity almost perfectly reproduce the antenna pattern. Thus, broadcast transmitters (AM, FM, TV, as well as ground-based radars) could not possibly generate the type of signal we saw.

On the other hand, if a ground-based transmitter were sending a signal out into space and it reflected off a piece of metallic space debris, couldn't that signal come back into the "Big Ear" receiver? The answer is yes! In fact, this hypothesis was one that I kept in the back of my mind as being slightly possible. However, now my belief is that it is much less likely than I earlier thought. For an Earth-based signal to be reflected from a piece of space debris and give us the response that we saw in the "Wow!" signal, several things would have to be true: (1) the ground-based transmitter would have to be transmitting in the protected band around 1420 MHz (and this is not supposed to be happening); and (2) the piece of space debris would have to be metallic (very possible), not tumbling (quite unlikely), and not moving significantly with respect to the celestial sphere (not likely for nearby debris and also not likely for almost any debris either in Earth orbit or debris that has escaped the Earth).

Even though a ground-based transmitter is not supposed to be transmitting in the 1420 MHz band, it is possible for a harmonic of a lower frequency transmission, or a spurious emission from a transmitter on an entirely different frequency, to occur in the 1420 MHz band. For example, if a transmitter were designed to send a narrowband signal at 710 MHz, it unavoidably would also send a much weaker version of that signal at twice that frequency (i.e., 1420 MHz). Similarly, if a transmitter were designed to send a narrowband signal at 473.33 MHz, it unavoidably would also send a much weaker version of that signal at triple that frequency (again, 1420 MHz). In other words, weak signals are always generated by a transmitter at integer multiples (harmonics) of the fundamental frequency. Filters are used to significantly reduce the intensity of these harmonics, but the intensities cannot be reduced to zero. Since the “Big Ear” contained a very sensitive receiver, it could have detected such harmonics. Note that most of these fundamental frequencies (e.g., 710 MHz, 473.33 MHz, etc.) occur in the bands used by television and radio; TV and radio signals are nearly always much broader in bandwidth than the 10 kHz width signal of “Wow!”

In order to generate an intensity response virtually identical to that of a celestial source of small angular diameter (point source), a piece of space debris could not be tumbling except at a very slow rate of one turn every hour or slower, and it couldn’t be moving with respect to the celestial sphere (background of stars) more than about one arcminute during the 72 seconds the “Wow!” signal was observed. These two constraints are uncharacteristic of most space debris. Thus, for the reasons stated above, I now place a low probability on this alternative as the explanation for the “Wow!” source.

**Gravitational lensing:** When an electromagnetic wave (such as light or radio waves) travels past a star or galaxy or other condensation of matter, that wave is deflected slightly. If a radio source (including a radio beacon from an intelligent civilization) were located in the same line of sight but further away than this condensation of matter, it is possible for the waves to be seen (or imaged) as a ring or multiple points of enhanced light or radio waves. This phenomenon is called “gravitational lensing”. Many instances of this phenomenon have been reported in recent years, both in optical and radio images. Could this be involved with the “Wow!” source? I think the short answer is “Not likely!”

Typically, the lensing phenomenon (rings or bright spots) remain in the images taken over a period of many days or months or even years, depending on the motion of the source and the condensed matter. On the other hand, the "Wow!" signal, which should have been seen twice (two beams) in about five minutes, was seen only once. The lensing effect probably would not have changed significantly in five minutes. Of course, if "Wow!" were a signal from an intelligent civilization, the beings responsible for transmitting the signal could have pointed it to another direction in their sky, or could have turned off their transmission within the five-minute period.

**Interstellar scintillation:** When we look at the stars in our sky, we see them "twinkling". That twinkling is due to each photon coming from the point source experiencing a slightly different travel path on the way to our eyes than other photons. The Earth's atmosphere accounts for nearly all of the differences imposed on these photons. We do not see the planets twinkle because a planet has an observable angular diameter and the effects applied to the photons from the various directions of the planet tend to average out.

When radio and optical waves travel through the interstellar medium (which is somewhat like our atmosphere except much more rarefied), those waves (photons) experience a kind of twinkling effect called "interstellar scintillation". It is possible for there to be an enhancement of the signal passing through this interstellar medium due to a partial coherence effect. If this effect did occur for the "Wow!" source, it still points to a signal originating many light-years away from us, thus tending to give more support for the hypothesis of a signal of an extraterrestrial origin.

**ETI (Extraterrestrial Intelligence):** Since all of the possibilities of a terrestrial origin have been either ruled out or seem improbable, and since the possibility of an extraterrestrial origin has not been able to be ruled out, I must conclude that an ETI might have sent the signal that we received as the "Wow!" source. The fact that we saw the signal in only one beam could be due to an ETI sending a signal in our direction and then sending it in another direction that we couldn't detect. Of course, being a scientist, I await the reception of additional signals like the "Wow!" source that are able to be received and analyzed by many observatories. Thus, I must state that **the origin of the "Wow!" signal is still an open question for me**. There is simply too little data to draw many conclusions. In conclusion, **I am not able to prove that either we did receive a signal from an ETI or that we did not.** Thus, more than three decades after its appearance, the "Wow!" signal remains a fascinating enigma.

# 5

# SETI: The NASA Years

John Billingham,
*Senior Scientist, SETI Institute*

This chapter, on the years of SETI in NASA, was initially prepared in 2000 for the celebration of Frank Drake's 70th birthday, but has never been previously published. All the material in these pages remains as valid today, in 2010, as it was 10 years ago. So it fits well into this volume on SETI Past, Present, and Future, with only minor revisions, and I am delighted that it is now seeing the light of day.

I am also delighted that Frank, whose name appears in my chapter (and, indeed, in this book) more than any other name, continues to be a key figure and active leader on the Board of Trustees of the SETI Institute

## 5.1 Introduction

This volume is dedicated to Frank Drake, to celebrate his 80th birthday, and the 50th anniversary of *Project Ozma*. I have been asked to contribute a chapter on the story of SETI in NASA. Since I was involved in the story from the very beginning to the very end, 1969 to 1994, I can relate here only the highlights of the story. What follows is therefore something of a personal story of SETI in NASA, told in sequential form and, of necessity, not including every detail. Should I have omitted important names and events, it is through lack of space. Watch how many times the name Frank Drake appears in the saga.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_5, 

For anyone who wishes to read the story in more detail, go to the beautifully written article by Steven Dick, historian of space science at the US Naval Observatory, in *Space Science Reviews* (Dick, 1993). For real detail, turn to the References at the end of Dick's article. The events of the final year, 1993-1994, when the NASA SETI was cancelled by the Congress, are well chronicled by Stephen Garber of the NASA History Office (Garber, 1999).

Frank Drake provided some of the original stimulus for SETI in NASA. At every stage throughout the next quarter-century he participated in making the story a reality. As the "Father of SETI", he played an active role - especially in the scientific community - in bringing the NASA project to fruition. In the beginning, *Ozma* was a bold and imaginative new departure in the exploration of the cosmos, but considered by many to be on the fringes of the scientific norm. By 1994, SETI was basically accepted by the scientific community as an exciting new intellectual and technical challenge, and Frank was firmly established as the Chairman of the Board of Directors of the SETI Institute in Mountain View, California.

## 5.2 1959-1969: Ten Years of Prologue

Cocconi and Morrison published their seminal paper on "Searching for Interstellar Communications" in 1959, establishing the radio region of the electromagnetic spectrum as a logical place for carrying out searches for signals of extraterrestrial origin (Cocconi and Morrison, 1959). In the very next year, Frank Drake independently conducted *Project Ozma*, the first search for such signals, at the National Radio Astronomy Observatory at Green Bank in West Virginia (Drake, 1960). In 1961, the National Academy of Sciences Space Science Board sponsored a small meeting at Green Bank to "examine the prospects for the existence of other societies in the Galaxy with whom communications might be possible; to attempt an estimate of their number; to consider some of the technical problems involved in the establishment of communication; and to examine ways in which our understanding of the problem might be improved" (Pearman, 1963). The meeting was notable for many things, but especially the genesis of the "Drake Equation", the presence of Bernard (Barney) Oliver at the meeting, and the conclusion that "the number of communicative civilizations in the Galaxy might range from less than 1000 to one billion".

In 1963, Kardashev conducted the first Soviet Union search for signals from advanced civilizations (Kardashev, 1963). The following year saw the conference on Extraterrestrial Civilizations at Byurakan in Armenia, under Ambartsumian and Kardashev, and attended entirely by radioastronomers (Tovmasyn, 1965). May of 1965 saw the first use of the term CETI –

Communication with Extraterrestrial Intelligence – by Rudoph Pesek of the Czech Academy of Sciences, in his proposal to the Board of Trustees of the International Academy of Astronautics to establish an international symposium on the subject. In 1966 Carl Sagan collaborated with Iosif Shklovskii on an English language version of an earlier book in Russian by Shklovskii. The new book was called *Intelligent Life in the Universe* (Shklovskii and Sagan, 1966). At this time I was Chief of the Biotechnology Division at NASA's Ames Research Center in the San Francisco Bay Area, and becoming aware of scientists in a sister Division at Ames called Exobiology, which had been formed a few years earlier by Harold (Chuck) Klein. They introduced me to the Shklovskii-Sagan book late in 1968, and it changed my whole life.

## 5.3 1969: The Embryogenesis of SETI in NASA

Through 1969, mulling over "Intelligent Life in the Universe", I began to realize that NASA Ames might be an ideal home for a definitive program to actively pursue interstellar communication, as it was then known, by designing and using a large-scale radiotelescope system to search for signals of extraterrestrial intelligent origin. NASA had been given specific responsibility in the Space Act of 1958 to conduct the exploration of space. The Exobiology Program had been established at Ames under Chuck Klein and Dick Young. *Project Viking* was being defined, which was to include biology experiments to search for evidence of microbial life on Mars. Klein was Project Scientist for these. Ames already had a strong program in space science. Perhaps it might be possible to build SETI telescopes in space, or on the moon. NASA had the capabilities to carry out all the necessary large-scale science and engineering, and one of Ames's roles was to be at the cutting edge of space exploration. Not least, I thought, NASA and Ames would have the vision and courage to explore the opportunities, and perhaps to turn them into an active new venture. I was right.

In September, Hans Mark became Director of the Ames Research Center. He believed strongly in personal contact, so came to visit people in their offices and labs and engineering shops. When he came to find out about my Division, I put to him the notion of beginning a study effort on interstellar communication. He thought it was a good idea, but advised proceeding slowly and thoroughly, since it would be such a new topic in NASA. With the agreement of Chuck Klein, then Director of Life Sciences at Ames, we carried out a small initial in-house feasibility study in the summer of 1970, and concluded that there were no show-stoppers. Concurrently, we ran a large Ames summer lecture series on Interstellar Communication, with Drake, Sagan, Oliver, Cameron, Bracewell and others as speakers (Ponnamperuma and Cameron, 1974). In the Fall, I met again with Hans Mark, and we decided to carry out a larger-scale conceptual study in the summer of 1971 under the aegis of the Summer Faculty Fellowship Program in Engineering Systems Design, run jointly every year by Ames and Stanford University, and funded by NASA through the American Society of Engineering Education. I was Co-Director of these Programs, together with Jim Adams, Professor of Mechanical Engineering at Stanford. Neither of us had the right technical background for the topic, so we decided to co-opt a third person who knew radio science and engineering. The two principal candidates were Barney Oliver and Frank Drake. Barney won out because of his vast knowledge of radio-engineering. He was Vice-President for Research and Development at Hewlett-Packard. I approached him in October and asked if he would take the job. He agreed, with enthusiasm.

## 5.4 1971: Project Cyclops

For 10 weeks during the summer of 1971, 20 physical scientists and engineers, professors in various appropriate disciplines in colleges around the country, gathered at Ames to conduct "A Design Study of a System for Detecting Extraterrestrial Intelligent Life". Under the inspiring leadership of Barney Oliver, and with consulting advice from visiting experts in radio science and engineering (including Philip Morrison), the team put together a landmark report, which Barney dubbed "Project Cyclops". Among the 15 conclusions were: that signaling was vastly more efficient than interstellar travel (the ratio is actually tens of orders of magnitude); that the microwave region of the spectrum was the best; that the quiet region between the spectral lines of hydrogen (1420 MHz) and the hydroxyl radical (1665 MHz) was a natural "water hole" for communication between species; and that it was technologically feasible to build a ground-based phased array for interstellar communication over galactic distances.

The team completed their conceptual design for *Cyclops*. It comprised an expandable phased array of 100-meter, fully steerable radiotelescopes and a signal processing system using an optical spectral analyzer to examine the 200 MHz of the water hole with a resolution not exceeding one Hertz. Should it be necessary to build a complete system to achieve the sensitivity necessary to detect faint narrowband signals from star systems within a sphere of radius 1000 light-years, namely 1000 of the 100-meter antennas, then the cost would be between 6 and 10 billion dollars, spread over some 10-15 years. The team also recommended that NASA initiate further scientific and engineering studies which would lead to a more detailed engineering systems design over a 3-5 year period.

Interestingly enough, the National Academy of Sciences of the US and the Academy of Sciences of the USSR sponsored a joint conference on CETI in Byurakan, Armenia, that same September. Some of the key US delegates were Drake, Sagan and Oliver (Sagan, 1993).

Oliver worked for more than a year to edit and refine the *Cyclops Report* before it was published (Oliver and Billingham, 1973). Over the succeeding years it came to be recognized as a visionary and technological *tour de force*, and was much in demand. Ten thousand copies were printed. (It was recently reprinted by the SETI League and the SETI Institute). At my instigation, the document included an artist's concept of the 1000-antenna phased array, occupying a circle 16 kilometers in diameter. This awesome picture led to a misunderstanding, which evolved into a myth, that the full array was necessary to detect extraterrestrial intelligence. Many people looked at the picture, and looked at the price tag for the full array, and without reading the fine print jumped to the black and white conclusion that $6 to 10 billion dollars was going to be needed to detect an extraterrestrial civilization. They were wrong on two counts. First, the array was to be built in stages, with searches carried out after each stage was completed. So it was possible that a signal would be found with only one dish, at a cost of a few millions of dollars instead of billions. Second, even the full-up array might not have detected a signal. In any case, the myth is still around today But I believe it is on the wane, as *Cyclops* is gradually superseded by the SETI Institute's brand new 1-hectare Allen Telescope Array and by the proposed international Square Kilometer Array.

## 5.5 1972-1974: Early Steps at Ames

Next, I had to find out if NASA would support further studies. With the blessing of Mark and Klein, I put together a Committee on Interstellar Communication at Ames. We were nine, drawn from different divisions and branches. Dave Black was our expert on planetary systems. My Deputy was John Wolfe, a space physicist of note. On accepting my invitation to serve, he told me that he had read *Cyclops* from cover to cover all night, having been unable to put it down. At this stage we received a boost. The National Research Council published its 1972 decennial report on *Astronomy and Astrophysics for the 1970s* (see under National Research Council in the references). Prepared under the Chairmanship of Jesse Greenstein, it included for the first time encouraging words on the future significance of interstellar communication, and on studies that might be undertaken in the area. Frank Drake played a major role in preparing this material. By 1974, the Ames committee had produced, and sent to NASA Headquarters, a comprehensive

"Proposal for an Interstellar Communication Feasibility Study". We briefed John Naugle, the NASA Chief Scientist, and his advisors from the scientific community. Barney and I also briefed the NASA Administrator, James Fletcher, and the Associate Administrator for Space Science, Homer Newell. In August of 1974, we received our first funding, to the tune of $140,000, from the NASA Office of Aeronautics and Space Technology.

At this stage, it was clear to us that interstellar communication was still generally considered to be a novel idea, outside the norms of respectability in most of the scientific community. We therefore decided to conduct a series of Science Workshops through 1975 and 1976, specifically to examine in greater detail all aspects of a program to detect extraterrestrial intelligence.

## 5.6 1975 and 1976: The Science Workshops on SETI

In 1974, after nine years of directing aviation and space biomedical and bioengineering research, I now decided to take a year off and devote my time to the nascent SETI program at Ames. Chuck Klein approved, and gave me a position to hire a secretary. Vera Buescher came on board, as the planet's first full-time interstellar secretary. (She has remained in SETI ever since, as the glue which held us all together). She and I worked together to plan the meetings of the Science Workshops. Philip Morrison agreed to take the chair, and together we worked out our goals and objectives, and decided whom to invite onto the team. The final membership was Ronald Bracewell, Harrison Brown, A.G.W. Cameron, Frank Drake of course, Jesse Greenstein, Fred Haddock, George Herbig, Arthur Kantrowitz, Kenneth Kellermann, Joshua Lederberg, John Lewis, Bruce Murray, Barney Oliver, Carl Sagan, and Charles Townes. I was Executive Secretary. Bruce was not on the original list, but called me from CalTech to offer his services, which we were glad to accept. It turned out he had heard a lecture on interstellar communication by Barney at CalTech, and was very intrigued with it. It turned out also that he was soon to become the Director of the Jet Propulsion Laboratory in Pasadena.

During 1975 and 1976, we had six three-day meetings, and accomplished much. It became apparent that we needed two additional splinter workshops on Extrasolar Planetary Detection, a neglected field at the time. Jesse Greenstein was chair, and David Black the Executive Secretary. We also had one splinter workshop at Stanford on the Evolution of Intelligent Species and Technological Civilizations, another topic neglected in the new domain of exobiology. It was chaired by Joshua Lederberg.

At the fourth SETI Science Workshop, in early December of 1975 in Puerto Rico, we discussed names for the new endeavor, and accepted the proposal by John Wolfe to use "Search for Extraterrestrial Intelligence" instead of "Communication with Extraterrestrial Intelligence". Communication could mean two-way or many-way exchanges, which were not our immediate goal. Our priority was the search. The acronym SETI stuck, and is now in common parlance the world over.

The Report of the SETI Science Workshops (Morrison, Billingham and Wolfe, 1977) confirmed the microwave window as a promising place to begin the search, and noted that progress in large-scale integrated circuit technology had been so rapid that million-channel fast Fourier analyzers could be used instead of the optical signal processing of Project Cyclops. Four conclusions emerged:

1 It is both timely and feasible to begin a serious search for extraterrestrial intelligence

2 A significant SETI program with substantial secondary benefits can be undertaken with only modest resources

3 Large systems of great capability can be built if needed

4 SETI is intrinsically an international endeavor in which the United States can take a lead

5 The Workshop members made the point that the search fell under NASA's mandate. Philip Morrison wrote a stimulating section on "The Impact of SETI", and concluded his Preface with the words, "We recommend the initiation of a SETI program now"

In the middle of the workshops, Chuck Klein asked me if I would take over the recently-vacated position of Chief of the Exobiology Division at Ames. I was delighted, and changed careers forthwith. Dick Young, Chief of Exobiology at NASA Headquarters, privately protested that I was "only an M.D." But I think Klein saw a potential expansion of Exobiology to incorporate SETI. In any case, Dick and I had been, and remained, close friends. With the encouraging words of the Morrison report in hand, I now established in the Division a formally constituted SETI Program Office, with John Wolfe, astronomers Mark Stull and Charles Seeger, sociologist Mary Connors, who was to study the societal aspects of SETI, and Vera Buescher. Barney Oliver and Frank Drake were always participating, and Hans Mark continued his support from on high, as did Chuck Klein. Without them there may have been no SETI in NASA.

## 5.7 1977: JPL Joins In

Early in the SETI Science Workshops, everyone assumed that the search would focus the radiotelescope beam for some minutes at a time on selected target stars, thus achieving high sensitivity, as in *Cyclops*. However, Murray argued forcefully for an additional approach - namely, to sweep the sky with the beam so that total sky coverage could be realized (at the cost, though, of a reduction of sensitivity of about one thousandfold). At the fifth meeting in 1976, Oliver gave in - "All right Bruce, have it your own way" - and the stage was set for the bi-modal search strategy which dominated SETI in NASA from then on. Murray was now director of JPL, and suggested that the laboratory join with Ames to conduct SETI.

Discussions between the Centers began in 1976. Bob Edelson took charge of the JPL program, and worked with me for several years. It became apparent that Ames had a strong preference for targeted searches, and JPL for sky surveys. Since the approaches were complementary, it made sense to divide responsibility between the Centers. Over the next two to three years, the outline of the signal detection system, based on a multi-channel signal analyzer (MCSA), was developed by the engineers who were beginning to come on board. The original plan was to use the same detection system for both searches, though later this proved too difficult, and each Center developed its own. For antennas, JPL was to use the telescopes of their Deep Space Network at Goldstone in the Mojave Desert, while Ames was to use existing large telescopes around the world.

Edelson and I made constant trips to NASA Headquarters for all the programmatic and funding discussions. By 1978, the Office of Space Science had taken over the funding of SETI. At Ames, astronomer Jill Tarter came from Berkeley on a National Academy of Sciences Post-Doctoral Fellowship for just a year, and then stayed for 15 more. (She is currently Director of SETI at the Institute.) During this time, she gradually took over the science of SETI. At JPL, the same function was happily in the hands of Sam Gulkis, a distinguished radioastronomer. In 1979, I organized a two-day conference on "Life in the Universe (Billingham, 1981) at Ames, which attracted an overflow crowd. At this meeting Ames and JPL were now able to present a joint paper on "SETI: Plans and Rationale" (Wolfe et al., 1981). The proposed NASA search system would achieve a ten million-fold increase in capabilities over the sum of all previous searches. The MCSA and its algorithms, at the heart of the system, would now allow a reasonable search of Jill Tarter's "cosmic haystack" for its "needle", a signal of indisputably extraterrestrial intelligent origin.

## 5.8 1980–1981: The SETI Science Working Group

Ames, JPL, and NASA Headquarters decided that the emerging SETI Program should be carried out with the continuing input at a working level from leading radio scientists and engineers in the academic community. Accordingly we formed the SETI Science Working Group under the chairmanship of John Wolfe and Sam Gulkis. It met on six separate occasions, and produced a report with 17 conclusions and recommendations. It confirmed the microwave region as preferable, endorsed the bimodal strategy, and envisaged a five-year R&D effort to design, develop and test prototype instrumentation. Its first conclusion was "the discovery of other civilizations would be among the most important achievements of humanity". Its last was "It is recommended that SETI be supported and continued as a long-term NASA research program". The members of the Group were Peter Boyce, Bernie Burke, Eric Chaisson, Thomas Clark, Michael Davis, Frank Drake, Woody Sullivan, George Swenson, Jack Welch, and Ben Zuckerman. Significant contributions came also from Michael Klein, who took over from Edelson as Manager of the JPL SETI Program in 1981, Kent Cullers, leader of the Ames MCSA signal detection/algorithm development team, Paul Horowitz from Harvard, (who had spent a year on sabbatical at Ames and developed "Suitcase SETI"), Allen Petersen from Electrical Engineering at Stanford, George Morris and Ed Olsen from JPL, and two other post-docs who had spent a year at Ames, Ivan Linscott and Peter Backus (both of whom were to join the Ames team), and of course, Barney Oliver and Jill Tarter. The report of the Group was put together by Frank Drake (Drake, 1983).

## 5.9 Dissidents Emerge

By now SETI was becoming better known and more respected in the scientific community. However, there were a few skeptics (Hart and Zuckerman, 1982; Tipler, 1980) who argued on a number of grounds that the number of co-existing civilizations in the Galaxy was vanishingly small. In 1978, the program received a "Golden Fleece" award from Senator William Proxmire, and our funding suffered accordingly. Our position was always that we do not know the number of other civilizations, and that the only way to answer the question is to carry out a search. Drake (1980) and Oliver (1992) argued that interstellar travel and colonization were too expensive, and that radio communication was vastly more efficient over interstellar distances. Morrison spoke out for the empiricism of Western science: "It is fine to argue about the number of extraterrestrial civilizations. After the argument, though, I think there remains one rock-hard truth: whatever the theories, there is no easy substitute for a real search out there, among the ray directions and the wavebands, down into the noise. We owe the issue more than mere theorizing".

Nevertheless, in the Fall of 1981, Proxmire introduced an amendment into the NASA budget, removing all 1982 funding for SETI. At this stage, I had to prepare a Termination Plan, which was somewhat disheartening. But Hans Mark, then Deputy Administrator of NASA, called a key meeting in Washington with all the senior people from the Agency and leaders from the scientific community. The decision was made to put SETI back into NASA's 1983 budget request to the Congress. So I prepared a Reinstatement Plan. As the budgetary process continued through 1982, Carl Sagan and others talked to Proxmire and convinced him of the validity of the endeavor, so he did not oppose it again.

SETI was always, and still is, an easy target to snipe at. Although we can hold that life is widespread through the galaxy, we cannot give figures for the probability of success of SETI. What we can say is that an unequivocal discovery of the existence of extraterrestrial intelligence would be of the most profound significance for humankind. In spite of this, we continued over the years to face opposition from a few skeptics in the Congress. Much of it was of a political nature, and happened because SETI was such a small element of the NASA budget, ultimately 0.1%, that it lacked the nationwide political support of the large NASA projects (Garber, 1999) It also was of such high interest to the public at large that it often figured prominently in the media, who would sometimes make fun of our search for mythical "Little Green Men". What we are actually searching for, of course, is unassailable evidence of the existence of an extraterrestrial technological civilization, born of cognitive intelligence. The anatomical and physiological structure of the extraterrestrials is a question of major theoretical interest, but what matters for the search is that they have figured out, almost certainly a long time ago, how to build powerful radio transmitters.

## 5.10 1982: Good News

In 1982, Carl Sagan published in *Science* a petition signed by 70 scientists, including seven Nobel prize winners, from around the world calling for international cooperation in and support of a systematic SETI program. They said: "No *a priori* arguments on this subject can be compelling or used as a substitute for an observing program. We urge the organization of a coordinated, worldwide, and systematic search for extraterrestrial intelligence" (Sagan, 1982).

In 1983, the Decennial Report of the 1982 Astronomy Survey Committee (The Field Report) strongly supported SETI as one of seven Moderate New Programs for the 1980s (see under National Research Council in the references). Their specific recommendation was for: "An astronomical Search for Extraterrestrial Intelligence, supported at a modest level, undertaken as a long-term effort rather than a short-term project, and open to the general participation of the scientific community". The Committee had a special Subcommittee on SETI, which interacted at some length with our academic leadership, Drake, Oliver, Tarter, and many others. At this time the new Director of Life Sciences in the Office of Space Science and Applications at NASA Headquarters was Jerry Soffen, who had been the Project Scientist for the Viking mission to Mars. Encouraged by the increasing support from the scientific community, he accepted our proposal for the first of the five years of R&D funding that had been recommended by the SETI Science Working Group, so our budget for 1983 came in at $1.65 million. Don Devincenzi, a key figure in exobiology science management at Ames, went to join Soffen in the Life Sciences at NASA Headquarters, and became Chief of Exobiology there, and a most capable SETI Program Manager. Also at this time, and in spite of some competition between the Centers, Ames and JPL and Headquarters got together and agreed that Ames would be the Lead Center for SETI in NASA, and so it was until the program was cancelled in 1993.

A major event occurred in 1983. Barney Oliver retired from Hewlett-Packard and accepted my invitation to join Ames as Deputy Chief of the SETI Program Office. I found a special Civil Service position which fitted him perfectly - it was called "Expert". I was delighted with his decision, especially since he had no great love for the federal bureaucracy. He used to say that he was not really suited for the job because he was "neither civil nor servile". He had always been close to us, as our principal technical colleague. Now it became a formal arrangement, and everyone benefited. He was the only person in NASA to hold memberships in the National Academies of Sciences and Engineering. Our standing rose in the world. Barney wanted to be a volunteer, but the rules would not allow that, so he was forced to accept a salary!

In 1984, another major event was the formation of the SETI Institute. This was a brainchild of Tom Pierson, then the administrator of graduate research studies at San Francisco State University. He consulted with Barney, me and Jill Tarter, and went ahead to establish the Institute as a California research and education Non-Profit Corporation. Tom next wanted the best person to be President and Chairman of the Board. The best person turned out to be Frank Drake. After serving for many years as Director of the Arecibo Observatory, followed by many more years as Professor of Astronomy at Cornell, Frank was now Dean of Science and Professor of Astronomy at UC

Santa Cruz. Frank accepted the position, part-time of course, and everyone was delighted. Jack Welch, Professor of Astronomy at UC Berkeley, and Head of the Radio Astronomy Lab there, became Deputy Chair of the Institute. Tom Pierson became Executive Director and ran the Institute with his astonishing flair for leadership, just as he does today. Jill Tarter joined the Institute to spearhead the science, and Vera Buescher followed to become the Research Assistant to the Institute Management.

## 5.11 1983–1987: Five years of R&D

Unhappily for us, Chuck Klein retired from NASA Ames in 1984. By then he was widely recognized as "The Father of Exobiology". But I have kept in touch and sought his advice ever since, as have we all.

With funding at about $1.5 million a year, Ames and JPL embarked on an intensive program to define all aspects of SETI in NASA. It was now formally titled the Microwave Observing Program (MOP). I worked with Mike Klein on the programmatic aspects, Barney oversaw the technology, and Jill Tarter and Sam Gulkis were the chief scientists. Elyse Murray joined the Ames team in 1983, and it was not long before we realized she was a super secretary. New spectrometers with millions of channels were needed. Some of the original thinking about ways of solving this difficult problem came from Bob Machol, Professor of Operations Research at NorthWestern University, who had joined us over the years on a sequence of sabbaticals. He talked with Alan Despain of UC Berkeley. Then Despain and Allen Petersen and Ivan Linscott at Stanford developed the digital technology for the first Ames MCSA. At Ames, Kent Cullers led the signal detection team in the design of very sophisticated algorithms to search for both continuous wave and pulsed signals and to reject radiofrequency interference, one of SETI's major and continuing problems (Cullers, 1988).

The prototype narrowband (1Hz) signal detection system had 74,000 channels, and was tested on a 26-meter telescope at Goldstone from 1985 to 1987. It succeeded in detecting the one Watt transmitter on the Pioneer 10 spacecraft at a distance of 4.5 billion miles. At JPL, Mike Klein, ably assisted by engineer Bruce Crow, supervised the corresponding development of their Wide Band Spectrum Analyzer,which was tailored to the needs of the Sky Survey. From 1985 on, he succeeded in obtaining support from the NASA Office of Telecommunications and Data Acquisition to use part of the Deep Space Network and for some of their engineering development work. This support was to continue for the remainder of the program.

During this period there was a reorganization at Ames, and I became head of an expanded Life Science Division which included exobiology and

SETI, ecosystem science and technology, and space biology, physiology and medicine. In SETI, Ames and JPL wrote a formal Program Plan, signed off by Barney Oliver for Ames and Mike Klein for JPL, which we submitted to Headquarters, and which was adopted in March of 1987. Jill Tarter played a major role in putting it together, and it was a major milestone. The plan proposed a 10-year, $73.5 million search for narrow band signals, composed of two complementary components. One was the Targeted Search, to be carried out by Ames, and the other was the Sky Survey, to be carried out by JPL. In addition to the technical, managerial and administrative details, we made sure that the Program Plan included sections on the following additional material: the intimate link between SETI and exobiology; evaluations from the scientific community; use of the sophisticated instrumentation for radioastronomy and other possible areas; a summary of the manifestations of considerable interest by the public and the media, and of the appearance of SETI in college courses around the country; and an annotated bibliography by Charles Seeger, which included references to the extensive bibliography of SETI which had appeared in the Journal of the British Interplanetary Society (Paprotny, 1985), and then continued to appear there for several more years. I insisted that we include in our NASA budget a Program Plan line item for R&D for future SETI telescopes, searches and systems at one tenth of the budget. Although approved at the time, this was unfortunately later to disappear in a funding crunch.

## 5.12 SETI at Large

I shall now depart from the ongoing story to discuss general SETI matters which emerged over the years. Although the NASA program was by far the largest, SETI had appeared over the years in many other places. Drake had carried out his own searches, and sponsored others at Arecibo. Beginning in 1973, Kraus and Dixon had the longest continuously running US observational project at the Ohio State Radiotelescope. In the early 1990s, Dixon started the imaginative Project Argus, a wide sky, broad frequency, low sensitivity search with small telescopes. Paul Horowitz developed extremely narrow channel (.05 Hz) instruments for the Harvard radiotelescope, beginning in 1980 with *Project Sentinel*, then progressing to META – the Megachannel Extraterrestrial Array (Horowitz, 1985), and finally to the current BETA, with a billion channels. Bowyer and Wertheimer at UC Berkeley have been running *Project SERENDIP* as a piggyback operation on radioastronomy projects at Arecibo since 1980.

Outside the US, SETI projects were carried out in France, Argentina, Italy, Germany and Japan. These programs and others came to a total of

61 searches worldwide (Tarter and Klein, 1991). It should be noted that the sum total of all searches had examined only a minute fraction of all of astronomical multi-dimensional search space. In 1991, SETI was still in its infancy. On the other hand, *a real signal might have been detected at any time by any SETI observing project anywhere on Earth.*

It had always been our policy to provide, where we could, some small level of financial support for some of these other SETI activities, and we did just this over the years. Another policy was to aim for the highest professional standards in the science and engineering of SETI. To this end we always engaged with the scientific and engineering communities, and made sure that we had a continuing presence at national and international professional society conferences, delivering papers and then submitting them to appropriate peer-reviewed journals. In the International Academy of Astronautics, review meetings on SETI have been held every year at the Annual Congress of the International Astronautical Federation since 1972. I was Chairman of the IAA SETI Committee from 1977 to 1994. Every four or five years, we would collect the best papers read at the Congresses, have them peer reviewed, and publish them as a Special Edition of *Acta Astronautica on SETI* (see under *Acta Astronautica* in the references) In the International Astronomical Union, a new Commission (51) on Bioastronomy was established in 1982, and since then has held scientific meetings triennially. Frank Drake was President from 1986 to 1989, and Jill Tarter from 1987 to 1990.

It had been apparent to us from the beginning that the unequivocal discovery of a signal of extraterrestrial intelligent origin would have profound consequences for humankind. Since this is obviously an international question, we brought it up over the years in the SETI Committee of the International Academy of Astronautics, and also with colleagues in the International Institute of Space Law. We devised a set of "Principles for Activities Following the Detection of Extraterrestrial Intelligence", and called it, somewhat loosely, the SETI Post-Detection Protocol (Tarter and Michaud, 1990). It was a list of recommendations to SETI investigators. It was endorsed by six major international space societies, and later by nearly all SETI investigators around the world. In the following years, the Committee worked on a second "Protocol", which examined questions dealing with the transmission of messages from Earth to extraterrestrial civilizations, and recommended that these questions be forwarded to the Committee on the Peaceful Uses of Outer Space (COPUOS) of the United Nations for their consideration. The basic issues were whether to transmit, either *de novo* or after the detection of a signal, what the message should say if transmissions were sent, and how these decisions were to be made. The Position paper included, for discussion purposes, a draft of a Declaration of Principles. Our document became a formal Position Paper of the Academy, and was endorsed by the International Institute of Space Law. It has now been formally received by COPUOS. It is available from the Academy, and is on the SETI Institute website at www.seti.org.

At the 1987 IAF Congress in Brighton, England, Dr James Fletcher, Administrator of NASA, presented a paper on what he imagined his successor would say about space achievements 30 years into the future. In discussing SETI, he pronounced that the discovery would eclipse all other discoveries in history.

It had been obvious to us for 20 years that there were many questions dealing with the implications of SETI for society that had not been addressed. So I asked the distinguished social psychologist Roger Heyns, Director of the Hewlett Foundation and former Chancellor of UC Berkekey, if he would work with me to chair a series of Workshops on the Cultural Aspects of SETI. We gathered together a team of specialists in history, theology, anthropology, psychology, sociology, international law, relations and policy, political science, the media, and education. We met three times in 1991 and 1992, and generated a report on the "Social Implications of the Detection of an Extraterrestrial Civilization" (Billingham et al., 1999). The report concluded that the issues were important, and should be addressed in extensive further studies. The Executive Summary, Principal Findings, and Recommendations can also be found at www.seti.org.

## 5.13 1988: The Buildup Begins

1988 saw the signing of the Project Initiation Agreement by NASA, another major step in the bureaucratic approval process. Lynn Griffith had replaced Don Devincenzi as Program Manager at NASA Headquarters. John Rummel became the HQ Project Scientist. Funding was now running at just under $3 million a year. At Ames, there was another reorganization, and in 1989 I became the full-time Chief of the SETI Office, with Barney Oliver at my side as Deputy. My first action was to appoint Jill Tarter as our Project Scientist. The SETI Institute, under Drake and Pierson, was playing an increasingly important role (Pierson, 1993). We were completing the R&D phase. Program reviews intensified at the Centers and in Washington. In 1990, SETI took on the status of an approved NASA Project, and we began Final Development and Operations. The budget for 1990 was $6 million. The final Project Plan was for a 10-year search at a total cost of $108 million. We had 140 people working on SETI at Ames and JPL. The search was to begin on October 12, 1992, the 500th anniversary of the discovery of America by Columbus. And so it did. For a textbook description of SETI at this time, including science rationale, observational plans, and signal detection system designs, see our chapter in the joint US-USSR publication on Fundamentals of Space Biology and Medicine (Billingham and Tarter, 1993).

In 1991, the National Research Council published its Astronomy Survey Committee Report for the 1990s, and again recommended SETI (see under The Decade of Discovery in Astronomy and Astrophysics in the references).

Speaking of Columbus reminds me that attempts of one sort or another were always being made to reduce our budget. We had constantly to be on guard. We continued to have sniping from individual members of Congress, though also much support. Some in the astronomical community saw SETI as having the potential for competing with them for funds. A standard question was, "Why don't you delay this project until the cost of digital signal processing has come down to a fraction of what it is today?", to which Oliver replied, "Columbus didn't wait for jets". We did actually have another strong argument for not delaying, and used it effectively. If we did not get on the air soon, we would have more and more difficulty in detecting faint signals from other civilizations because of the increasing saturation of the radiofrequency spectrum by interference, which in turn would cost progressively more millions of dollars to overcome.

In 1991 we began building and testing the actual search systems. Tarter and Gulkis finalized the observational plans, advised by an Investigators Working Group of scientists. The 1991 budget rose to $16.8 million. The Targeted Search was to be conducted at the Arecibo Observatory in Puerto Rico (having been approved by an NSF peer review process), and the Sky Survey on one of the Deep Space Network telescopes at Goldstone in the Mojave desert. I tried at this time to have Michel Klein formally named as Deputy of the NASA SETI Program, but Headquarters said it could not be done. We needed a full time overall Project Manager, and brought on David Brocker from the Space Science Division at Ames. Reporting to him were Larry Webster, Targeted Search Manager at Ames, and Mike Klein, Sky Survey Manager at JPL. The able Gary Coulter became Program Manager at NASA HQ, replacing the able Lynn Griffith.

In 1992, the name "Microwave Observing Program" was changed to "High Resolution Microwave Survey" by order of the US Congress. It was moved from the NASA HQ Life Sciences Division to the Solar System Exploration Division, along with Coulter and Rummel. The 1992 budget rose again, to $17.5 million. The signal detection systems were shipped to the telescopes for final testing. The Ames system was built into a Mobile Research Facility – a trailer. It was trucked to Travis Air Force Base, loaded onto a C-141 transport, flown to Puerto Rico, then trucked again to the Arecibo Observatory, and hooked up to the telescope. The basic idea behind the Mobile Research Facility was to be able to take the Targeted Search to any large telescope anywhere in the world. At the same time, scientists and engineers at JPL assembled and tested their Sky Survey instrumentation at Goldstone. Preparations were made for the inauguration of the search. A series of talks were to be given by distinguished people. Invitations went out to them and to the media and the activity level rose to a crescendo. The brunt of the organization fell on Vera Buescher, who did a wonderful job. We were *very busy*.

## 5.14 1992: NASA SETI Comes of Age

It was noon on Columbus Day, 1992, at the Arecibo Observatory in Puerto Rico. After a morning of inauguration speeches, including a rousing one from Frank Drake, Jill Tarter formally initiated the NASA High Resolution Microwave Survey, and pulled the switch to turn on the Targeted Search system. In a two-way hook-up with the JPL team at Goldstone, where there was a corresponding inauguration, David Brocker did the same for the Sky Survey. As I said in my briefing to the audience, these new systems were so powerful that they would eclipse the sum of all previous searches in the first few minutes of operation. And so it was.

Both teams worked for a year exploring the sky for signals of extraterrestrial intelligent origin, and learned how to deal with the vast flows of data that were analyzed in near real-time. Procedures were worked out for dealing with the ever-present radiofrequency interference. Teams of observers and engineers rotated back and forth between the NASA Centers and the observatories. The Targeted Search completed 200 hours of observations of selected nearby F, G, and K stars. The Sky Survey conducted observations at X-band, and completed a sequence of maps of the galactic plane, primarily at L-band. In August 1993, Jill Tarter and Mike Klein presented a summary of their results at a Bioastronomy Symposium in Santa Cruz, California (Tarter and Klein, 1995). They said, "At both sites the equipment has worked well, with minor, mostly low-tech glitches. These initial observations have verified the transport logistics for the Targeted Search, and provided the first platform for remote observations to the Sky Survey. As a result of the data which has been collected, modifications have been made or planned to the hardware, software, and observing protocols. Both observing programs have encountered signals that required additional observations because they initially conformed to the detection pattern expected for an extraterrestrial signal, but no signals persist as potential candidates at this time. This paper will discuss the lessons we have learned, the changes we are making, and our schedule for continued observation".

Alas, there was to be no continued observation.

## 5.15 The Dissolution of SETI in NASA

Shortly after the Santa Cruz meeting, Senator Bryan (Democrat, Nevada), introduced an amendment to the FY 1994 budget eliminating the HRMS program. His argument was based on deficit reduction, and he explained that 150 new houses could be built in Nevada for the same cost. In spite of a vigorous defence of HRMS by Senator Mikulski (Democrat, Maryland) and others, the motion was carried. The political complexities of all the issues are covered in detail in *Searching for Good Science: The Cancellation of NASA's SETI Program* (Garber, 1999).

I now had the unhappy task, for the second time, of putting together a Termination Plan. Slowly and surely, all the grants and contracts had to be wound down, and our team dissolved. It took six months. The total budget for SETI, over all the years, was close to $78 million. In March of 1994 the doors were closed on SETI in NASA.

## 5.16 Epilogue

We had successfully executed the first comprehensive search for extraterrestrial intelligence on this planet. We suspect there have been, still are, and will be in the future, searches by other intelligent species in the universe. Perhaps some of these searches have been successful, and perhaps communication now exists between these extraterrestrial societies. One day we may join in.

The Targeted Search was taken over and continued in 1994 by the SETI Institute with funding from private sources. The following year (also with private funding) Project Argus, a new All Sky Survey, was initiated by the nonprofit SETI League, on whose advisory board Frank Drake serves.

So Frank Drake, who began it all, now held the torch in his hands. In the year 2010, he still does.

## References

*Acta Astronautica*: Special Issues on SETI: 1979, **6, No. 1–2;** 1986, **13, No. 1;** 1989, **19, No. 11;** 1990, **21, No. 2;** 1992, **26, No. 3–4;** 1998, **42, No. 10–12:** Pergamon Press.

Astronomy and Astrophysics for the 1970s: 1972, **1:** *Report of the Astronomy Survey Committee.* National Research Council, National Academy Press, Washington, DC.

Astronomy and Astrophysics for the 1980s: 1982, **1:** *Report of the Astronomy Survey Committee.* National Research Council, National Academy Press, Washington, DC.

*The Decade of Discovery in Astronomy and Astrophysics: 1991*. National Research Council, National Academy Press, Washington, DC.

Billingham, J. (Ed.) 1981. *Life in the Universe.* NASA CP 2156; also published by MIT Press, Cambridge, Mass.

Billingham, J. and Tarter, J.C. 1993. Fundamentals of Space Biology and Medicine. Chapter 8 in *SETI: The Search for Extraterrestrial Intelligence*. American Institute of Aeronautics and Astronautics. Washington D.C., and Nauka Press, Moscow.

Billingham, J., Heyns, R., Milne, D. et al. (Eds) 1999. *Social Implications of the Detection of an Exterrestrial Civilization.* SETI Press.

Cameron, A.G.W. (Ed.) 1963. *Interstellar Communication.* W.A. Benjamin, New York.

Cocconi, G. and Morrison, P. 1959. Searching for Interstellar Communications, *Nature,* **184**: 844–46.

Cullers, D.K. 1988. Three Pulse/Multiple Stage Continuous Wave Detection Algorithm. In G. Marx (Ed.) *Bioastronomy: The Next Steps.* Reidel, Dordrecht, Holland, 371–376.

Dick, Steven J. 1993. The Search for Extraterrestrial Intelligence and the NASA High Resolution Microwave Survey (HRMS): Historical perspectives. *Space Science Reviews* **64**: 93–139.

Drake, F. 1960. How Can We Detect Radio Transmission? *Sky and Telescope,* **19**: 26–28, 87–89 and 140–43.

Drake, F. 1980. N is Neither Very Small nor Very Large. In M.D. Papagiannis (Ed.) *Strategies for the Search for Life in the Universe*. Reidel, Dordrecht, 27–34.

Drake, F., Wolfe, J. and Seeger, C. 1983. *Report of the SETI Science Working Group.* NASA Technical Paper 2244.

Garber, Stephen J. 1999. Searching for Good Science: The Cancellation of NASA's SETI Program. *Journal of the British Interplanetary Society,* **52**: 3–12.

Hart, M.H. and Zuckerman, B. (Eds) 1982. *Extraterrestrials – Where Are They?* Pergamon Press, New York.

Heidmann, J. and Klein, M.(Eds) 1991. *Bioastronomy: The Search for Extraterrestrial Life – The Exploration Broadens.* Springer-Verlag, Berlin.

Kardashev, N.S. 1964. Transmission of Information by Extraterrestrial Civilizations. *Aston. Zhurnal,* **41:282**, trans. *Soviet Astronomy AJ,* **8**, 217, reprinted in Goldsmith (1980), *The Quest for Extraterrestrial Life: A Book of Readings*, 39–47.

Morrison, P.M., Billingham, J. and Wolfe, J. 1977. *The Search for Extraterrestrial Intelligence.* NASA SP 419.

National Research Council, 1973, 1983, 1993. *Astronomy and Astrophysics for the Seventies, Eighties, and Nineties.* National Research Council, Washington, D.C.

Oliver, B.M. and Billingham, J. 1972. *A Design Study of a System for Detecting Extraterrestrial Intelligent Life.* NASA CR-114445.

Oliver, B.M. 1992. *Galactic Colonization and Other Flights of Fancy.* Internal manuscript, SETI Institute.

Papagiannis, M.D. (Ed.) 1980. *Strategies for the Search for Life in the Universe.* Reidel, Dordrecht.

Paprotny, Z. and Lehmann, J. 1985. Interstellar Travel and Communication Bibliography, 1982 update, *J.B.I.S.*, **36**, 311–329.

Pearman, J.P.T. 1963. Extraterrestrial Intelligent Life and Interstellar Communication: An Informal Discussion. Chapter 28 in A.G.W. Cameron (Ed.) *Interstellar Communication.* W.A. Benjamin Inc, New York, 287–293.

Pierson, T. 1995. SETI Institute: Summary of Projects in Support of SETI Research. In: *Progress in the Search for Extraterrestrial Life,* 1993 Bioastronomy Symposium, Santa Cruz, California, August 16–20. ASP Astronomical Society of the Pacific, San Francisco.

Ponnamperuma, C. and Cameron, A.G.W. (Eds) 1974. *Interstellar Communication: Scientific Perspectives.* Houghton-Mifflin, Boston.

Sagan, C. (Ed.) 1973. *Communication with Extraterrestrial Intelligence.* MIT Press.

Sagan, C. 1982. Extraterrestrial Intelligence: An International Petition. *Science*, **218,** No. 4571**.**

Shklovskii, I.S. and Sagan, C. 1966. *Intelligent Life in the Universe.* Holden-Day, San Francisco.

Tarter, J. and Michaud, M.A. 1990. SETI Post-Detection Protocol*., Acta Astronautica,* **21**, 2.

Tarter, J. and Klein, M. 1991. SETI: On the Telescope and on the Drawing Board. In J. Heidmann and M.J. Klein (Eds), *Bioastronomy: The Search for Extraterrestrial Life – The Exploration Broadens*. Springer-Verlag, New York, 229–235.

Tarter, J. and Klein, M. 1995. HRMS: Where We've Been, and Where We're Going. In *Progress in the Search for Extraterrestrial Life.* ASP Conference Series **74**, 457–469. G.S. Shostak (Ed.). Astronomical Society of the Pacific, San Francisco, 1995.

Tipler, F.J. 1980. Extraterrestrial Intelligent Beings Do Not Exist. *Quart. Journ. Royal Astron. Soc.,* **21**, 267–281.

Tovmasyan, G.M. 1965. *Vnzemnye tsivilzatsii.* Erevan, English trans., *Extraterrestrial Civilizations,* Jerusalem, 1962.

Wolfe, J., Edelson, R.E., Billingham, J. et al. 1981. In J. Billingham (Ed.) *Life in the Universe*, The MIT Press, 391–417.

# 6

# From HRMS to *Phoenix*: Up from the Ashes

Peter Backus,
*Observing Programs Manager, SETI Institute*

The story of NASA and SETI begins with John Billingham, whose chapter (most appropriately) precedes mine. John was an RAF physician who worked on the Apollo missions to the moon, and came to the Life Sciences division at NASA Ames research center in Mountain View, CA. There, he became interested in SETI. In 1970, he invited Barney Oliver, then Vice President of Research and Development at Hewlett Packard, to head up a summer engineering study called Project Cyclops. The goal of Project Cyclops was to design a system to detect Earth-level technology at a distance of 1000 light years.

## 6.1 The *Cyclops* Study

The *Cyclops* workshop produced a report (Figure 6.1), and some very impressive artwork (Figure 6.2) of a huge array that many people to this day think was actually built. I still get many questions asking about that huge array in the southwest. Sadly, there is no such huge array, despite the fact that it has been depicted by Hollywood. In fact, Project Cyclops proposed to build a single antenna and do a complete search of the sky. If that proved unsuccessful, the plan was to expand to 10 antennas and repeat the survey. If unsuccessful at that level, expand to 100 antennas, and eventually to 1000 antennas. So, it was to have been a staged survey. But many people just remember the artwork and the 1000 antennas.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_6, © Springer-Verlag Berlin Heidelberg 2011

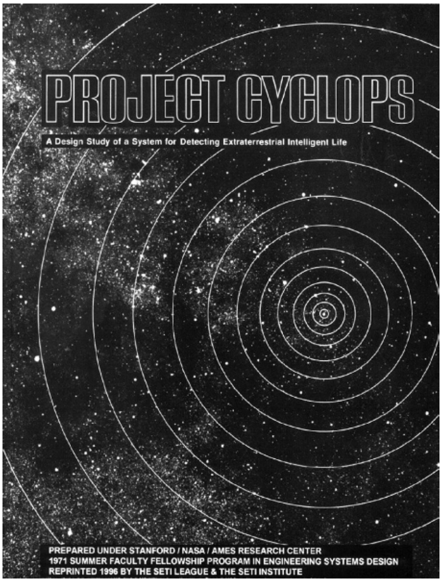



**Figure 6.1** The report on Project Cyclops.

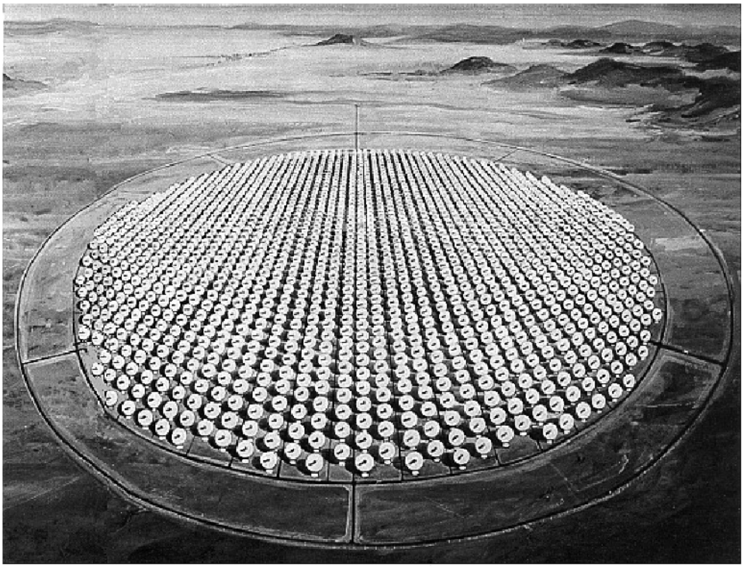


**Figure 6.2** Artwork depicting a huge array of 1000 antenna which was never built.

One of the important outcomes of Project Cyclops was that it proposed to focus in on a part of the microwave spectrum that is called the terrestrial microwave window. Assuming the extraterrestrials live on a planet like the Earth, they would have a atmosphere similar to the Earth's, and the upper end of their microwave window would be partially clogged by atmospheric effects. Still, there would be a broad range in the microwave spectrum, from 1 to 10 GHz, that would remain a very quiet zone. In particular, one part of the spectrum, the 300 MHz marked off by the spectral emission lines of hydrogen and hydroxyl, would be a quiet zone. This 300 MHz segment was dubbed the "waterhole", since H and OH are the disassociation products of water. Water being key to life on the Earth, it was assumed it would be key for life elsewhere in the galaxy. Searches in the waterhole spectrum, first proposed by Project Cyclops, still characterize many SETI studies today.

## 6.2 The Morrison Workshops

In 1974 John Billingham was able to establish the Office of Interstellar Communications at NASA Ames Research Center. With just a small staff, he then proceeded to organize a series of workshops in the mid 1970s, to answer the question of whether and how to conduct a search, and whether

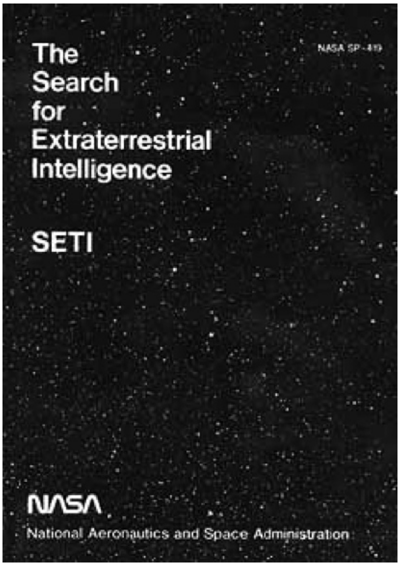



**Figure 6.3** The NASA Report into The Search of Extraterrestial Intelligence.

NASA was going to move forward with SETI. These workshops were headed up by Philip Morrison, and produced a landmark report (Figure 6.3). Several other interesting scientists participated, including Frank Drake and Carl Sagan. The outcome of the workshops was that, yes, indeed it was reasonable for NASA to pursue a search, and the participants identified two modes or strategies to pursue.

The first of these two proposed strategies was a sky survey. The idea is that you assume that extraterrestrial civilizations will have transmitters whose power might follow a power law, just like the luminosity of stars. We see most of the stars in the sky not because they are close, but because they are intrinsically bright. Perhaps, it was reasoned, there are a few bright, very powerful transmitters out there that can be viewed from anywhere in the galaxy. If that is the case, you want to cover the entire sky, since signals from such powerful transmitters could be coming from any position on the sky, at any distance. Because of that, you can cover the sky most efficiently if you use a smaller antenna that has a broader beam, a broader coverage of the sky. Hopefully, the reduced sensitivity of small antenna would be balanced by the intrinsically powerful transmitters that we hope are out there. And, since you are covering the entire sky, you might as well cover the terrestrial microwave window.

This, then, was the goal of the NASA SETI sky survey. But, the workshops also recognized that, when trying to detect transmitters across many light years, you are naturally going to detect transmitters that are only miles away. So terrestrial interference was going to be a problem. The plan was to use the 34-meter diameter antennas that are part of NASA's deep space network and scattered in three locations around the world.This way, full sky coverage could be obtained. These antennas would be able to search the entire terrestrial microwave window, 1 to 10 GHz, but special electronics would need to be developed in order to handle tens of millions of channels simultaneously. To cover the entire 1 to 10 GHz band, it would have to be processed at about 300 MHz at a time. Thus, the sky survey was going to develop a spectrometer for that, with electronics that would analyze the data and keep track of statistics as to which channels had unusually high levels of power, more than would be expected from noise.

The sky was to be divided into large rectangular areas that could be scanned very quickly in a sliding racetrack fashion, so that each spot within the rectangle would be covered twice. If a signal appeared twice at the same frequency and the same position on two separate observations, then that was considered a candidate ETI signal.

But there was also another strategy recommended by the Morrison workshops. Since we know that life *as we know it* originated at least once on a planet near a star like the Sun (our own planet, near our own sun), perhaps we should observe sunlike stars. But, to cover our bets, let us also observe the nearest 100 stars, regardless of the type of the star, in order to cover the possibility of life as we *don't* know it.

In order to get increased sensitivity, the workshop participants proposed to trade off on the amount of frequency covered, focusing on the lower end of the microwave window where the sensitivity is best. Once again, you are going to have to be very sensitive to signals coming from light years away, so you have to be able to handle interference from the nearby, very strong signals produced by our own technology.

## 6.3 The High Resolution Microwave Survey (HRMS)

At NASA AMES, Billingham started searching for ways to implement the recommendations in the Morrison report. In order to get great sensitivity you have to use large antennas. A plan was formulated to use the largest available antennas in the world: the 1000-foot diameter Arecibo observatory in Puerto Rico, the Parkes observatory's 210-foot telescope in Australia, and the National Radio Astronomy Observatory 140-foot telescope in West Virginia. The frequency range would be constrained but even at just the low end of the microwave window. However, that still required covering 2 GHz of spectrum. Custom electronics were needed to produce the tens of millions of narrow channels required to cover at least 20 MHz at a time. Special detectors were developed to be sensitive to narrow band signals that could either be continuously present, a CW tone, or something that might flash on and off like a rotating lighthouse.

Because you are observing for an extended period of time, several minutes at every frequency, you have to allow for the relative acceleration between the transmitter and the receiver. The Earth is a rotating planet, so the observatory is changing its velocity along the line of sight to a particular star. And, the transmitter might be on a rotating planet as well, or on a particular spacecraft somewhere in their solar system. So, we have to allow for a possible drift in frequency in these signals, up to 1 part in $10^9$.

NASA's goal was to observe about 800 stars from a list of approximately 2000 candidates. That number was chosen to basically try to accommodate for the amount of telescope time that was expected to be available on the largest telescopes in the world. If the signal was detected that appeared to be coming from a star, and it followed that star as it moved across the sky, then it was considered to be a candidate ETI signal.

From the late 1970s through the 1980s, NASA developed the technology for this search, at the AMES Research Center and at the Jet Propulsion Laboratory (JPL). The name High Resolution Microwave Survey (HRMS) was introduced because, to some politicians, SETI was a four-letter word.

Both the sky survey and the targeted search were launched on the carefully chosen start date of Columbus Day 1992, the 500th anniversary of Columbus' discovery of the new world. Sadly, after just one year, the US Congress canceled the funding for the NASA high resolution microwave survey, when they figured out that HRMS was *also* a four-letter word.

## 6.4 The Birth of Project Phoenix

The decision to terminate HRMS came on a Friday afternoon. That was a rather dark weekend for the members of the project, but on Monday morning one of my colleagues, Dr John Dreher came in with a different attitude He said, "You know, if this project was a good idea on Friday, then it is still a good idea today, and we ought to find out a way to do it." With that inspiration the SETI institute created Project Phoenix from the ashes of the NASA search.

A general fundraising search was started, trying to get donors at any level. When Barney Oliver, who had chaired the Project Cyclops study, retired from Hewlett Packard, he had joined NASA to be part of the NASA search. After cancellation of the NASA HRMS, Oliver became a part of the SETI Institute. There, he started a personal fundraising campaign, and called up a few of the friends he had made in the technology industry. Those he contacted included Bill Hewlett and Dave Packard of HP, Gordon Moore of Intel and Paul Allen of Microsoft. In a matter of weeks, we had pledges of funding sufficient to continue the search.

The Institute also arranged for a long-term loan of the equipment that had been developed for the NASA targeted search. This was appropriate, since most of the equipment had been developed either by SETI Institute scientists and engineers, or by contractors. Unfortunately, the sky survey equipment had been developed as a co-project between JPL and the NASA Deep Space Network (DSN). Most of the sky survey equipment was destined for use in the DSN, so it was unavailable for further use in SETI.

We had to cut back on a few of the plans from the NASA program. We had only a single 20 MHz bandwidth system, composed of two 10 MHz units - the original plan had been to have a total of six 10 MHz units. Project Phoenix ended up scaling back on the frequency coverage slightly, partly because the receiver that we had available lost some sensitivity between 1000 and 1200 MHz. Also, based on our earlier NASA experience at Arecibo, we realized that the interference environment was more complex and dynamic than we had anticipated.

We had planned to do some simple on-off pointing of the antenna to determine whether a signal was associated with a star or could be seen anywhere on the sky. But, as you might be able to derive from Murphy's Law, the interference tends to vary on a time scale consistent with the on-off pattern. So, whatever timescale you choose, the interference will be there. It became obvious that we had to come up with a better scheme for doing interference mitigation.

## 6.5 RFI Mitigation

The system that we had developed under NASA processed the data in three stages. To do spectrum analysis, it broke down 20 MHz of spectrum into 28.8 million channels unique to two polarizations. The channels were separated by .7 Hz, but each channel was 1 Hz wide, so there was a little bit of overlap between the channels. Each spectrum that was generated every 1.5 seconds was overlapped with the previous spectrum by 50% in time, so that there would be no cracks in our data sampling. An actual ETI signal, we reasoned, would not slip through the cracks, and we would be sensitive to signals whether they were tuned to our particular frequency channels, or had pulses arrayed differently from the way we were taking the data.

Project Phoenix had special electronics to do signal detection of both continuous wave and pulsed signals, following them if they drifted, and clustering them because a strong signal would result in more detections. We had a 2-stage RFI mitigation scheme. We kept a database of the signals we saw in the previous week. Any signals that matched what we found in the database were assumed to be the same signal and ignored, just updating the database. We had also developed a new system called two site pseudo-interferometry. which I'll explain shortly. This process was a real-time pipeline so that, while analysis was going on for one observation, the previous observation was being analyzed for signals, and the observation before that one was being checked for interference using the two site strategy.

Now, this two site strategy took advantage of the rotation of the Earth. Imagine that you have two antennas – say, one in Puerto Rico at Arecibo, and one in the UK at Jodrell Bank. These two observatories are located in quite different positions of latitude and longitude on the Earth. So, as the Earth rotates, the two observatories will experience two quite different velocities, as seen in Figure 6.4. The Earth's rotation causes a Doppler shift in frequency, which is proportional to relative velocity. The Doppler velocity component at Arecibo, in this example 950 mph, will be at quite a different frequency from the signal at Jodrell Bank, where the relative velocity is 570 mph. Also, they are in different longitudes on the Earth: England is four hours ahead of Puerto Rico. Thus, the phase of the Doppler shift will have a different slope at each location. The rate of change of frequency will be

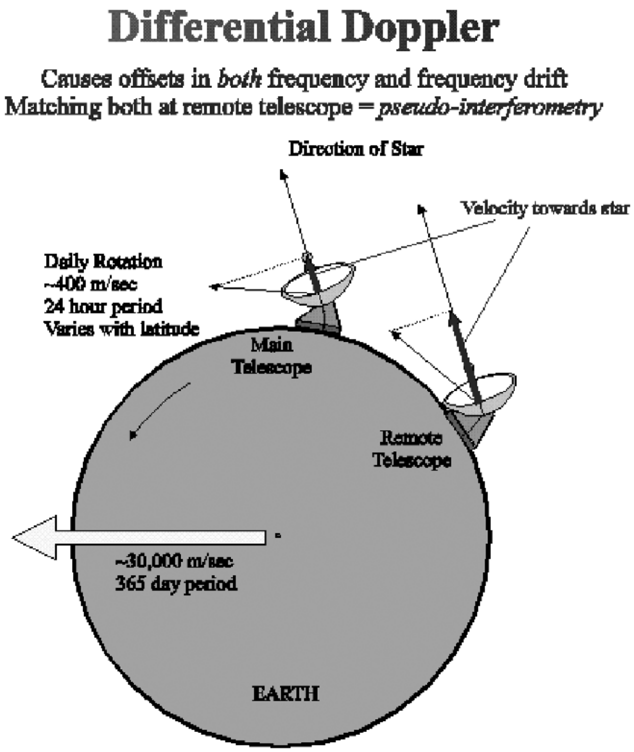



**Figure 6.4** Differential Doppler shifts.

different. Based on what we see for a signal detected at Arecibo, we can predict exactly on what frequency, and at what rate of change of frequency, we should expect to see it at Jodrell Bank.

Figure 6.5 shows graphically the rate of velocity for the two observatories. You can see not only that the velocity is different, but also that the rate of change is quite different. This gives us a very effective tool to differentiate a valid signal from RFI. But we have to make sure it is not *too* effective, else we experience excessive false negatives. We have to be sure that the equipment at both the observatories is locked to the same time and frequency standards. Fortunately, NASA provided us with an excellent test signal from the Pioneer 10 spacecraft at over 100 AU from the Earth. We did a daily test of Pioneer 10 or one of its sister spacecraft, to ensure that both observatories were working in synchrony.

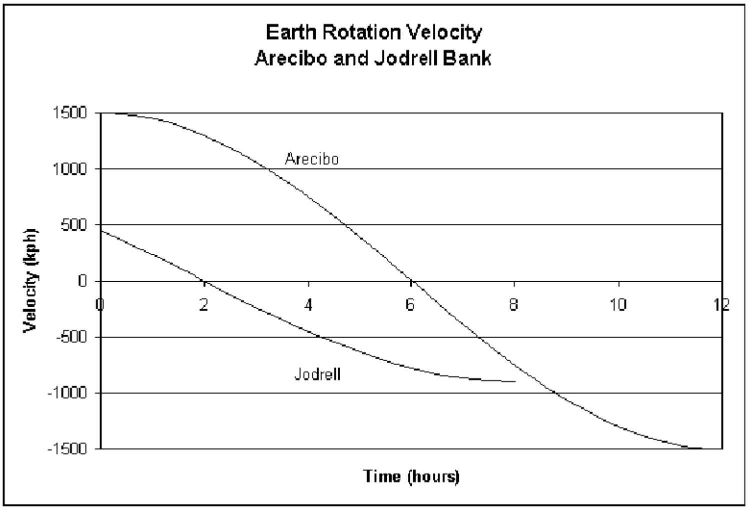



**Figure 6.5** Earth rotation velocity, Arecibo and Jodrell Bank.

## 6.6 Project Phoenix Observing Runs

Project Phoenix observations started in February 1995 at the Parkes Observatory in Australia, just about a month later than the original NASA plan. The 64-meter diameter Parkes antenna was paired with a smaller 22-meter antenna at Mopra, for radio frequency interference (RFI) mitigation in real time. The two antennas together formed a pseudo-interferometer. We had 16 weeks of observing time at Parkes, followed by two more weeks of collaborative observing with Australian astronomers.

After the Australian observing run, we then moved the equipment back to the US, upgraded the monitor and control software and, in September 1996, moved it all to the National Radio Astronomy Observatory (NRAO) in Green Bank, West Virginia (the locale of Frank Drake's pioneering *Project Ozma* observations). We operated at the Green Bank 42-meter telescope for roughly 50% of the time over the next 18 months. Our second observatory in this case was a 30-meter telescope operated by Georgia Tech in the Woodbury GA communications facility.

Next, in 1998, we moved to Arecibo's 305-meter spherical reflector, where we observed in parallel with the Jodrell Bank 76-meter radio telescope in the UK. We had two observing sessions per year, running for about three or four weeks each in the spring and the fall. Ultimately, this used up the

2400 hours of telescope time that had been awarded to the NASA program and then transferred to Project Phoenix, since the scientists on the original proposal were still the same.

## 6.7 The Mobile Research Facility

The way we moved our equipment around between these observatories was by employing another piece of NASA equipment, called the Mobile Research Facility. This is a 40-foot facility container with special RF shielding, a solid steel liner that attenuated radio signals caused by the equipment inside the trailer by a factor of 100 million. The Mobile Research Facility also had its own air conditioning and humidity controls, and racks of equipment. The flatbed trailer that was developed for this container became the first spin off from the NASA SETI program. It is now in use by the Air Force for other purposes.

Figure 6.6 is a view inside the trailer. You can see the equipment racks and some of the equipment we used. Of course, since this project spanned more than 10 years, there was quite a change in technology during its run. Figure 6.7 shows the type of technology that we used at the beginning of Project

**Figure 6.6** Inside the Mobile Research Facility.

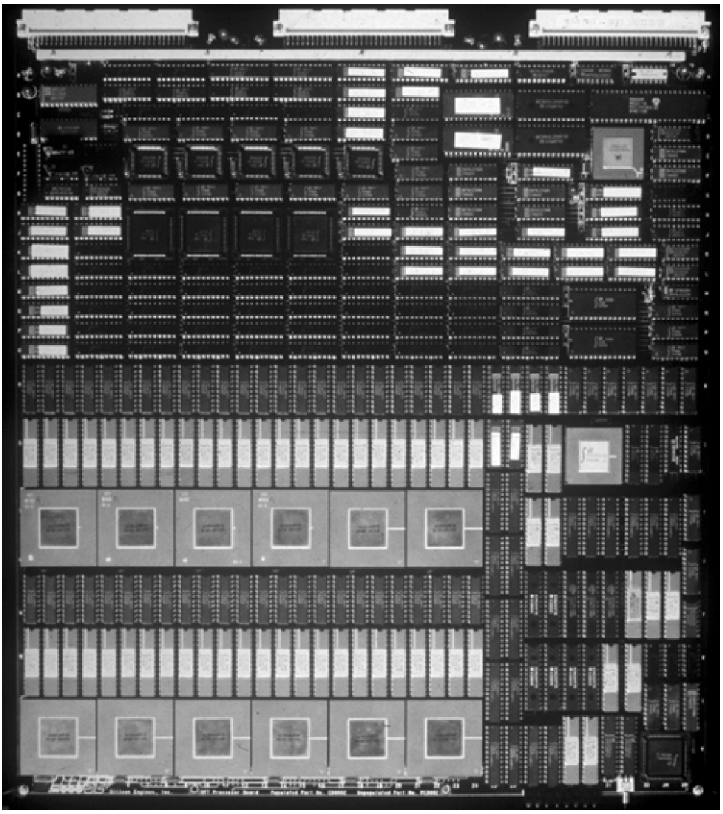


**Figure** 6.7 Technology in use at the start of Project Phoenix.

Phoenix. This equipment was originally developed by NASA for HRMS, a 9U VME circuit board, roughly 15 by 16 inches, filled with chips. The 12 chips in the lower left are custom Very Large-Scale Integration (VLSI) chips the network developed for Fourier transforms. This system performed 1.2 giga floating point operations per second (1.2 GFlops), dividing 10 MHz of spectrum into 14.37 million channels.

Ten years later, we were able to develop custom technology using commercial PCs, Pentium and Xeon processors, with a couple of commercial Digital Signal Processing (DSP) boards, and one custom FPGA board, as seen in Figure 6.8. We are now using this equipment at the Allen Telescope Array (about which you will read in a subsequent chapter), and are already in the process of replacing that. We will soon be going to an all software system based on commercial servers, and we hope to have that operating before the end of 2010.

**Figure 6.8** Technology 8 years into Project Phoenix.

## 6.8 The Legacy of Project Phoenix

Project Phoenix succeeded in completing, under private funding, the targeted search component of the former NASA high resolution microwave survey. After 11,000 hours of observation, about a year and a quarter of total time, Project Phoenix observed 800 stars out to a distance of about 250 light years, over the 1.2 to 3 GHz frequency range. We detected more than a million signals, but none have been proven to be from ETI. From that, we can set an upper limit on the number of narrow band transmitters that might be emanating from those stars.

The equipment and techniques developed for the NASA HRMS, refined for use in Project Phoenix, have laid the groundwork for the development of future SETI experiments and facilities, including the new Allen Telescope Array, about which you will read in Chapter 9.

### Acknowledgments

This chapter was adapted from a presentation made to the American Academy for the Advancement of Science (AAAS), at its February 2010 annual meetings in San Diego, CA.

# 7

# SERENDIP: The Berkeley SETI Program

Stuart Bowyer,
*Professor of Astronomy (Emeritus), and Professor in the Graduate School, University of California, Berkeley*

UC Berkeley's SERENDIP program is an ongoing effort in the search for extraterrestrial intelligence. It searches for stable narrowband spectral features in the radio frequency spectrum which could conceivably be a signaling beacon sent by an alien civilization. SERENDIP is a backronym [1] for the Search for Extraterrestrial Radio Emissions from Nearby Developed Intelligent Populations.

The SERENDIP search strategy is based on an unobtrusive piggyback observing model in which SETI observations are conducted alongside an observatory's regularly scheduled astronomical observation programs. Data acquired in piggyback mode are analyzed off-line at the UC Berkeley Space Sciences Laboratory. A commensal SETI program such as this is not free to choose observing frequencies and sky coordinates. However, in view of the plethora of postulated frequency regimes for interstellar communication and the large number of potential sites for civilizations which have been suggested, this is not necessarily a disadvantage.

[1] Editor's note: unlike an acronym, in which the initial letters of the words in a phrase are strung together to make a new word, a backronym works the other way around: an existing word is broken down into its letters, to which a phrase is then fitted. The term SERENDIP (referring to the old Persian fairy tale "The Three Princes of Serendip") denotes an unintentional discovery made by accident and sagacity. The name seemed most appropriate for the Berkeley commensal SETI program, It was Jill Tarter who made up the name "Search for Extraterrestrial Radio Emissions from Nearby Developed Intelligent Populations" to fit those initials.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_7,  

The SERENDIP program began with a simple set of hardware with limited capabilities. But it was begun at a time when very few searches were being attempted, and those that were being carried out were only done intermittently. We were able to obtain substantial amounts of observing time and collected large quantities of data. Even more important was our continuing work to develop more powerful instrumentation, and to find ways of placing this instrumentation on ever larger telescopes. I describe these efforts in the following sections.

## 7.1 SERENDIP I

The SERENDIP I hardware and processing software were developed at the Berkeley Space Sciences Laboratory with volunteer help and a small faculty research grant from the University. It was first installed at the University of California's Hat Creek Radio Observatory in 1980, and operated at the Deep Space Network at Goldstone in 1981 and 1982. The data acquisition system employed a 100-channel spectrum analyzer with a resolution of 100 Hz per channel and an integration time of 30 seconds. A 20 MHz band of the intermediate frequency spectrum was scanned over a period of 100 minutes. The power spectrum was calculated using an analog autocorrelator and microprocessor, and the results were searched for spectral peaks with an amplitude exceeding a preset threshold. If such a peak were found, the power spectrum and time was recorded. Subsequent application of cluster analysis techniques to the space, time, peak frequency, power, right ascension, declination, hour angle, azimuth, and elevation provided insight into the nature of the sources detected.

Although the system was crude, it obtained a substantial amount of data at a time when there were few SETI searches under way. It also had some unexpected benefits. A very bright Berkeley Astronomy PhD student, Jill Tarter, had just completed her thesis on Galaxy Clusters, and was looking for work. Although she was unacquainted with SETI, she thought it sounded like great fun and she volunteered to develop the software for the system. Thus began her new career as a SETI researcher!

## 7.2 SERENDIP II

SERENDIP II employed the same search strategy as SERENDIP I but achieved a major improvement in sensitivity by upgrading from 100 channels each 100 Hz wide, to a system with 65,536 channels each 0.98 Hz wide. This resolution was chosen so that Doppler drifts due to the Earth's motion would not smear a narrow band signal in a 10-second integration time. An array processor performed a 65,536 point complex Fast Fourier Transform on the data. The resulting power spectra were scanned for peaks above a previously chosen threshold. Upon detection of a peak, SERENDIP II recorded civil time, telescope coordinates, bin number, power, intermediate frequency, and synthesizer frequency on a disk for off-line analysis. SERENDIP II covered a total bandwidth of 3.5 MHz, 65 KHz at a time, by taking 50 KHz steps along the band before processing the next power spectrum.

The SERENDIP II system was installed at NRAO's 300-foot telescope in Green Bank, West Virginia in 1986, where it operated until the unfortunate collapse of the telescope. Its performance, and comparisons with other searches being carried out at that time, are shown in Table 7.1.

| | **Ohio State** | **NASA (test MCSA)** | **Sentinel** | **Meta** | **Serendip II** |
|---|---|---|---|---|---|
| Telescope diameter (feet) | 175 | 84 | 84 | 84 | 300 |
| System temperature (degrees K) | 100 | 25 | 65 | 65 | 25 |
| Total bandwidth (KHz) | 500 | 74 | 2 | 420 | 3500 |
| Single channel Bandwidth (Hz) | 1000 | .5 | 0.03 | 0.05 | 0.98 |
| Channels per beam | 500 | 74,000 | 65,536 | 8,388,608 | 3,571,428 |
| Integration time (sec) | 10 | 1000 | 30 | 20 | 1 |
| Beams/hr | 8 | 3.6 | 20 | 20 | 10 |
| Fraction of time in operation (percent) | 80 | 40 | 80 | 80 | 40 |
| Relative probability of detection per unit time | 0.0002 | 0.2 | 0.007 | 0.7 | 0.7 |

**Table 7.1** Summary of SETI programs, circa 1986.

## 7.3 SERENDIP III

SERENDIP III was installed on the world's largest radio telescope at the Arecibo Observatory in Puerto Rico in 1991. SERENDIP III utilized a $4\times10^6$ point Fast Fourier Transform spectrum analyzer with a 1.7 second integration time. In this configuration SERENDIP provided 0.6 Hz spectral resolution over 2.5 MHz of instantaneous band coverage. In order to cover the entire 12 MHz intermediate frequency signal from the telescope, SERENDIP analyzed successive 2.5 MHz sub-bands by mixing with a signal generated by a frequency synthesizer phase-locked to the observatory's maser. The signal from the frequency synthesizer was stepped by 2.5 MHz after each integration time until the entire 12 MHz intermediate frequency signal had been scanned. This step-look process was repeated indefinitely while observing. It took about eight seconds to complete one 12 MHz intermediate frequency sweep.

The Fast Fourier Transform unit converted the time-sampled vector into the frequency domain by computing squared Fast Fourier Transforms with the application of twiddle factor coefficients between the column and row operations. In the frequency domain, spectral bins exhibiting power above 16 times the local mean baseline power ($16\sigma$) were logged for further off-line analysis.

SERENDIP III used a receiver located in carriage house one, while the primary science observations were carried out with a receiver in carriage house two. While carriage house two tracked a target, carriage house one slewed across the sky at twice the sidereal rate. In this observing mode, a source remained within SERENDIP III's half-power beam for 20 seconds. Off-line, SERENDIP III data were filtered for radio frequency interference and then analyzed for the presence of extraterrestrial intelligence signals. The data analysis system identified interference by observing signal-persistence over sky angle several tens of arc minutes in extent (four half-power beam widths). The receiver slewed over the sky several beams faster than the duration of most radio frequency interference signals. Consequently, the SERENDIP III data analysis system could easily detect and remove radio frequency interference signals.

## 7.4 SERENDIP IV

SERENDIP IV was installed at the Arecibo Observatory in Puerto Rico in 1997. The SERENDIP IV instrument consisted of 40 spectrum analysis/post-processing boards working in parallel. Each of the 40 boards utilized dedicated hardware to perform a 4 million point Fast Fourier Transform on individual sections of the intermediate frequency band. The power spectra were passed to the post-processing portion of the board where baseline normalization, coarse resolution spectra computation and event thresholding were carried out. Baseline normalization was achieved with a sliding 8000-channel local mean boxcar. Thresholding was performed based on mean spectral power, Signals exceeding the threshold were reported by the post-processors to a control computer, which logged these signals to a UNIX workstation via a dedicated Ethernet. The workstation functioned as a short-term archive and data transfer machine. As data arrived, they were simultaneously sent across the Internet to UC Berkeley. Off-line data reduction consisted of a set of cluster analysis algorithms that removed radio frequency interference. Instrument health was also monitored at this stage by the detection of an artificial signal that was periodically injected into the instrumentation. A suite of pattern detection algorithms was used to identify a collection of statistically interesting events. These algorithms looked at signal persistence, telescope beam pattern matching, pulsed signals, and very high power events.

## 7.5 SERENDIP V

SERENDIP V is the most powerful instrument yet built in the SERENDIP project. It covers 2 GHz of instantaneous bandwidth with 1.5 Hz resolution. It was mounted at the Gregorian focus of the Arecibo telescope in 2009 where it continues to obtain data to the present day. It utilizes Arecibo's seven-beam L-band receiver array and processes data from all seven beams simultaneously. The instrument employs 21 spectrometer boards. Each board carries out a 64 million point Fast Fourier Transform on both polarizations of a 100 MHz sub-band of the input spectrum. After processing the 100 MHz sub-band is stepped to the next adjacent sub-band. Each board utilizes four 200 MHz 8 bit analog-to-digital converters for analog input and 176 digital I/O lines for transfer and readout. Signal processing is carried out by a unique gate array chip. Another gate array chip is used as a backend processor that passes data off to an independent computer for signal analysis. In total, the SERENDIP V system results in a 2.7 billion-channel instrument. This is a 42-fold increase in capability over the SERENDIP IV system.

## 7.6 Summary

A summary of the evolution of the SERENDIP program's capabilities is provided in Table 7.2. The signal detection and analysis software employed in these programs underwent a similar increase in sophistication, but this development is hard to quantify in a simple tabular form.

| Program | Bandwidth (MHz) | Resolution (Hz) | Number of channels | Date/Location |
|---|---|---|---|---|
| SERENDIP I | 0.1 | 1000 | 100 | 1979–1982 Hat Creek and Goldstone |
| SERENDIP II | 0.065 | 1 | 64K | 1986–1990 Green Bank |
| SERENDIP III | 12 | 0.6 | 4M | 1991–1997 Arecibo |
| SERENDIP IV | 100 | 0.6 | 168M | 1997–2006 Arecibo |
| SERENDIP V | 300 | 1.5 | 2 G | 2009–present Arecibo |

**Table 7.2** Summary of the various SERENDIP programs, 1979 to the present.

## 7.7 A Personal Appraisal of the Development of SETI and the Future of this Endeavor

The idea that a credible search for extraterrestrial intelligence could be carried out evolved soon after capabilities in radio astronomy developed rapidly beginning in the middle of the last century. In parallel to these developments in radio astronomy, technology to carry out systematic SETI searches also developed rapidly. As this work became more and more sophisticated and no signal was detected, searches were developed and carried out in other bands of the electromagnetic spectrum. Although these searches are intriguing, it is my opinion that they are not as likely to yield results as compared to searches in the radio band.

Extrapolating on the fact that no SETI signals were found in intermittent searches of a very small portion of our Galaxy, some people speculated that we must be alone in the Universe or some other such silliness. Undeterred by these arguments, SETI researchers continued their efforts. By the end of the century radio searches were still improving but by then had reached a point where giant steps in detection capabilities were no longer the norm.

Although incremental improvements are continuing to be made in SETI search technologies, other fields of research are emerging with intriguing implications for SETI. A major discovery affecting SETI was the detection of a surprising number of planetary systems. Although there was virtually universal agreement that these systems must exist around a substantial number of stars, this was, in fact, conjectural. A dramatic change occurred with the actual detection of planets. This area of research is currently undergoing very rapid expansion, and it is clear that huge leaps in this field will continue. The results obtained to date provide a complex picture in regards to the likelihood of extraterrestrial intelligence. On the positive side is a demonstration that a huge variety of stars have planetary systems. This increases the potential locations for extraterrestrial intelligence. However, the detailed characteristics of the systems that have been discovered show that the formation of planetary systems result in complex sets of outcomes many of which are unstable. At this point there is no sound theoretical basis to provide an estimate for the fraction of stars that will produce systems stable enough to support life. In a worst-case scenario for the field of SETI, there will be far more stars with planetary systems than expected but fewer planets with orbits that are sufficiently stable to allow for the development of intelligent life.

There are suggestions that other areas of research may rise in prominence and dominate the future SETI landscape. Recent laboratory work on the origin of life has produced results that are quite promising, but future work may not be able to progress beyond this point. On the other hand, discoveries in this field may show that life forms are easily produced. Work in the fields of brain research, anthropology, and psychology is also advancing rapidly and may provide key insights regarding the possibility of extraterrestrial intelligence.

Hold your breath. These are exciting times for SETI!

## Acknowledgements

The SERENDIP project has been supported by funding provided by the University of California Faculty Research Grant Program, the Planetary Society, Mrs Elizabeth Bowyer, Sun Microsystems, Toshiba, Intel Corp. Xilinx, the Friends of SERENDIP headed by Arthur C. Clarke, NASA, and NSF.

# 8

# Millions and Billions of Channels

## A History of the Harvard SETI Group's Searches

Darren Leigh,
*Research Scientist and Consultant*

and Paul Horowitz, *Professor of Physics and Electrical Engineering, Harvard University (who is in no way responsible for the Introduction)*

The history of the Harvard SETI group is inextricably linked with the history of Paul Horowitz. Horowitz became enamored with SETI as a student at Harvard, reading Ed Purcell's paper "Radio Astronomy and Communication Through Space" (Purcell, 1960), discussing with his roommates a class that Carl Sagan was teaching there using a draft of Shklovskii and Sagan's "Intelligent Life in the Universe" (Shklovskii and Sagan, 1966) as a text, and finally attending a Loeb Lecture series at Harvard by Frank Drake (Drake, 1969). The series was officially about pulsars but Drake did manage to slip in one inspiring talk about SETI. Horowitz says that "It was this lecture that launched me into this field; it was a revelation that you could go beyond idle speculation – you could actually calculate stuff."

As faculty in the Harvard physics department, Horowitz took his first sabbatical at The University of Colorado studying the flagellar motor of the *E. coli* bacterium; and he mentioned to Frank Drake that he wanted to "look for life in Puerto Rico". This led to Horowitz's first SETI project in 1978: looking for narrow-band microwave "carriers" at the Arecibo Observatory in Puerto Rico.

Although Horowitz's first Arecibo search scanned only a few hundred stars, and these only over a very small part of the microwave spectrum, it

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_8, © Springer-Verlag Berlin Heidelberg 2011

was the most sensitive search ever performed at the time, and this record lasted for two decades after the project ended. The search was implemented using the observatory's existing hardware and computers, which were designed for flexible astronomical use but were not optimized for SETI. Horowitz would follow his original 1978 search with four more microwave SETI programs and two optical ones. All of these used custom hardware and software, developed to provide the specific features needed to search for extraterrestrial signals and differentiate them from terrestrial ones.

This history will discuss our thoughts on the design of SETI projects, as well as provide details about each of the Harvard group's searches

## 8.1 The Design of SETI Projects

Looking for signals from an alien civilization is a lot like trying to answer a telephone call from a complete stranger when you have to design, build and connect the telephone yourself. There are so many different ways of communicating that we need to narrow down the possibilities substantially, mostly through educated guesses. A good method of doing this is to try to think like the aliens who might be doing the sending. Why is the signal being sent? What is being sent? How is it being sent? Where are good places to look for a signal? The answers to these questions involve physics, parsimony, and the technological capabilities of our own civilization.

First, we assume that transmitting alien civilizations have a grasp of the physical sciences comparable to or greater than our own. Numerous experiments and astronomical observations have indicated that the laws of physics are the same everywhere we can observe. These laws appear to be universal, so a civilization living in another star system would learn and understand them the same way we do. Because we are arguably the youngest technological civilization in our galaxy (it was only a few decades ago that we became capable of communicating beyond our own ionosphere), any signals we might receive were sent by an older civilization with a knowledge of physics at least as great as our own. Our shared knowledge of physical law is therefore a good basis for educated guesses about how to receive alien signals.

Second, we assume that the other civilization wants to communicate with us, and is not only trying to make the process possible, but as simple as possible. There are many different ways to communicate information, but some have fewer parameters than others, allowing for a smaller search space and requiring less guesswork.

Third, we have to work with the technology that is available to us. While it's possible that alien civilizations are currently beaming coherent neutrinos, gravitational waves, or "subspace radio" at us right now, we have no capability to receive such forms of communication. Given our current knowledge of physics, electromagnetic (EM) radiation – such as radio, microwaves and light – seems like an excellent means of communication. EM radiation travels at the velocity of light, which is the fastest possible speed for transferring information. EM radiation is easy to generate and receive, and it propagates with little distortion over interstellar distances. All of our projects have searched for signals which might use EM radiation.

## 8.2 Beacons or Leakage?

There are two types of signals that could be received from an extraterrestrial civilization: intentionally transmitted beacons and unintentional leakage radiation.The design and search strategy of a particular program will depend on which type of signal we are looking for.

A beacon is a signal that has been designed and radiated for the express purpose of initiating interstellar communication. It is meant to attract attention and may carry no information other than its very existence. The signal will be designed to appear artificial, to be easy for a search to detect, and to be able to cross the interstellar medium with minimal corruption.

Leakage radiation includes signals that are designed and radiated for use by the transmitting civilization itself and are not specifically intended to be received by others. Some examples from Earth are television carriers (effective power of about $10^7$ watts) and the Arecibo S-band radar (effective power of about $10^{13}$ watts). One advantage of leakage is that the transmitting civilization need not do anything special for the receiver to take notice. A major disadvantage, however, is that leakage from an advanced civilization may be weak, intermittent, or non-existent. Even if a leakage signal were constant and strong, we might have difficulty distinguishing it from natural noise. As a civilization improves its technology, signals are transmitted more efficiently, redundancy is removed, and their communication appears increasingly noise-like. It is unlikely that an advanced civilization would transmit strong signals inefficiently. Even the signals normally transmitted by our own civilization have been optimized for efficiency, making them less easily received from far away. For example, radio and television signals are transmitted in flat "pancake" beams with most of the energy going toward the horizon and very little being sent toward the sky. An alien civilization might be able to receive these signals if its own solar system happened to be parallel with the pancake beam but, as the Earth turned, the beam would change angle and transmit far less energy toward the aliens. And, beginning in 2009, all television transmissions in the US have been encoded in a noise-like digital format, making *a priori* detection far more difficult. Because receiving leakage is quite difficult, all of our searches have been designed to receive deliberate beacons.

## 8.3 Targeted or All-sky?

Search strategies are usually divided into two categories: targeted and all-sky. A targeted search concentrates on individual targets (usually stars), tracking them over a period of time. This has several advantages: radio telescopes with high directivity can be used because the target is an unresolved point on the sky with a known position. Integration techniques (long dwell times) can be used to gain higher signal-to-noise ratios. The Doppler shifts – at the reception site, relative to an inertial frame – are well known and the search can completely compensate for them. On the down side, only a small portion of the sky can be searched so, if the civilization is not within the set of targets, it will be missed.

An all-sky search can observe the entire sky visible from the observatory so it does not share its designers' target prejudices. However, if one wants to search the sky within a reasonable amount of time, this kind of strategy cannot use long dwell times or telescopes with high directivity (because the latter "see" only a tiny patch of the sky at one time). Also, since the target position is unknown, it will not be possible to completely compensate for Doppler shifts and the search will have to use compromise techniques.

The Harvard group's searches have employed both targeted and all-sky strategies.

## 8.4 Polarization

Radio waves, like light and all other EM radiation, are polarized. An EM wave's polarization can be in one of two opposite states – e.g., horizontal or vertical – which depend on antenna design and orientation. An antenna set up to transmit or receive one of these will be largely blind to the other. Even if we knew the exact type and orientation of the antenna being used by the transmitting civilization (so that we could build a matching receive antenna), linearly polarized radio waves are altered in the interstellar medium and the Earth's ionosphere by the Faraday Effect. This rotates the plane of polarization, making it appear that the transmitting antenna is in a different orientation. Fortunately, there is another form of polarization called circular, which is immune to rotation from the Faraday Effect. Using circular polarization requires a different type of antenna but, as long as that antenna is pointed in the right direction, its orientation is irrelevant. We believe it likely that extraterrestrial beacons will be transmitted using circular polarization.

Circular polarization also comes in two opposite states – left hand and right hand – and there is no way to know ahead of time which of those a transmitting civilization will choose. A good SETI receiver needs to monitor both polarization states. One way to do this is to double the amount of equipment used and monitor both polarizations simultaneously. Another way is to alternate between them, spending half of the observing time on each. This keeps the equipment cost down but decreases the effective observing time by a factor of two. A third option is to receive with a linearly polarized feed. This will pick up both circular polarizations without affecting equipment cost or time resolution, but the received signal strength will be a factor of two lower.

Some of the Harvard Group's searches have used linear polarization and some have used circular.

## 8.5 Waveforms and Modulation

Because we have no *a priori* way of knowing what an extraterrestrial signal would look like, we have to make educated guesses based on our knowledge of physics and signal processing. As mentioned before, we are assuming that the transmitting civilization would design its signal to be easily received, efficient and as simple as possible to detect.

What type of signal should we listen for? There are an infinite number of possibilities, and even Earthly technology employs a large variety, from amplitude modulation (AM) and frequency modulation (FM) to exotic digital modulation schemes such as "code-division multiple access with direct-sequence spreading" (CDMA-DSS), or "orthogonal frequency division multiplexing" (OFDM). The important thing to remember here is that we are searching for a beacon signal, whose sole purpose is to attract attention to its own existence; it need not send any other information. While a complicated waveform or modulation scheme still might be used, our assumption of simplicity suggests that we look for signals with only a small parameter space to search. Two signal types come readily to mind.

### 8.5.1 Pulses

The basis vectors (fundamental components of this signal type) are short pulses encoded via their timing. The process of detecting pulses is to compare the incoming signal level to a threshold, time-stamping over-threshold events for further analysis. A big advantage of this signal type is that the received signals are already in a format suitable for digitization and input to a computer, so very little computation needs to be done. A disadvantage, however, is that dispersion caused by ionized interstellar medium smears out radio frequency pulses, making them harder to distinguish from noise. Also, there are natural radio sources that emit pulses (pulsars) which could make distinguishing an artificial pulse source confusing. On the other hand, a pulsed beacon might be picked up serendipitously during a pulsar search. SETI programs have previously looked for pulses in both the radio and optical regimes.

### 8.5.2 Sinusoids

The basis vectors are sinusoids (waves) with different frequencies and phases. Because a constant phase carries no information, we can remove it by taking the square magnitude in the frequency domain, which leaves us with a single, linear parameter space to search. This meets our simplicity criterion.

The process of detecting a sinusoid is more complicated than detecting a pulse, but still straightforward. If the incoming signals are transformed into the frequency domain first, the rest of the detection is just as described above for pulses. Converting to the frequency domain can take a lot of computation, but the burden is eased by two things. First, the Fast Fourier Transform (FFT) algorithm can be employed. It is a computationally efficient method of converting signals from the time domain to the frequency domain, requiring only that slightly more than double the number of operations be performed every time the problem size is doubled. (For the technically inclined, the FFT is an $O(n \log n)$ algorithm. This is compared to previous Fourier Transform techniques which required four times the number of computations every time the problem size doubled because they were $O(n^2)$ algorithms). Searching a wider frequency band requires a larger problem size, so the FFT is a great boon to wider bandwidth searches.

Second, Moore's Law has allowed much more powerful computational hardware to become available at reasonable cost. This increase in the availability of computational power is nicely illustrated by the series of the Harvard SETI programs that searched for sinusoids. The earliest project used an off-the-shelf computer to do the FFT and could only search a small bandwidth (1 kHz). Later projects used large amounts of dedicated FFT hardware and could search much wider bandwidths (40 MHz).

Unlike pulses, sinusoids propagate well through the interstellar medium and are unlikely to be confused with natural radio sources. The natural radio sources with the narrowest bandwidths are microwave masers (bandwidths on the order of a kilohertz). Because an artificial radio transmitter can easily generate sinusoids of bandwidth much less than one Hertz, extreme narrowness in bandwidth is an excellent indicator that a signal is non-natural. The interstellar medium is kind to sinusoids, with dispersion having a negligible effect and scintillation (an effect similar to patterns of sunlight flickering at the bottom of a swimming pool), causing much less than 1 Hertz of broadening.

In addition to their good propagation characteristics, sinusoids seem to be a “natural” signal type. They are seen in spectral lines, orbital motion, pendulums and other elementary physics, because they are solutions to simple second-order linear differential equations. Pulses, chirps and pseudo-random spread spectrum are encountered less frequently in nature.

All of the Harvard Group’s radio searches have looked for sinusoids. By contrast, all of the optical searches have looked for pulses, for reasons explained later.

## 8.6 Why Narrow-band SETI?

All of the Harvard Group’s radio projects have searched for narrow-band sinusoids, basically very pure tones at radio frequencies. If such a signal were translated to audio frequencies and played on a speaker, it would sound like a flute. The narrower its bandwidth, the purer a sinusoid becomes, and this purity makes it easier to distinguish the sinusoid from other natural signals.

As we mentioned above, the natural signals with the narrowest bandwidth of which we are aware are microwave masers, but an artificial radio transmitter can produce sinusoids with far narrower bandwidths. While it is possible that we have just not discovered them yet, there is excellent reason to believe that very narrow-band natural radio sources do not exist. Any such source powerful enough to be received over interstellar distances would need to be either very large, or very hot.If it were large, different parts of the source would move at different speeds and directions and each of these parts would radiate a signal at a different frequency due to the Doppler Effect. The varying frequencies would broaden the signal’s bandwidth. If the source were hot, random thermal motion of the particles making up the source would also cause Doppler broadening.

The other type of natural signal that could interfere with our reception of an extraterrestrial beacon is thermal noise, the broad-band hiss or “static” that can be heard on a radio tuned between stations. Thermal noise is caused by heat; just as a microwave oven uses radio waves to make heat, any source of heat will also create radio waves in the form of thermal noise. Some of the thermal noise seen in a SETI receiver comes from outer space, some of it comes from the warm Earth, and some of it comes from the receiver’s own amplifiers.

Any deliberate signal that we receive will be mixed with this noise so, to be confident that we have received an actual signal and not just a random noise fluctuation, the received signal must be reasonably larger than the competing noise. This makes it very important to filter out as much of the noise as possible.

Thermal noise is proportional to the bandwidth over which it is being received, so a radio receiver with a channel width of 10 Hertz will receive only a tenth of the thermal noise of one with a channel width of 100 Hertz. To receive the least amount of noise possible, our radio receiver should have a channel width that is the same bandwidth as the signal that we intend to receive. If it is any wider, the receiver will pick up too much noise. If it is any narrower, part of the signal itself will be filtered out. The narrower the bandwidth of the signal that we intend to receive, the narrower the receiver channel can be and so less thermal noise will compete with the signal. This is why a very narrow bandwidth would be an important characteristic for an interstellar radio beacon.

## 8.7 Doppler Effects

Because the planet Earth is in constant motion, the frequency of any signals we receive from an extraterrestrial source will be altered by the Doppler Effect. There are three main components of the Earth's motion for which we need to account and, because these components are circular, we have to be able to handle both the Doppler shift due to the velocity along the circle and the change in Doppler shift due to the centripetal acceleration which is fundamental to circular motion.

### 8.7.1 The Earth rotating on its axis

The Earth rotates once per day,[1] from west to east. At the latitude of our observatory (42° N) this amounts to a velocity of about 340 meters per second, toward the east. If we are observing at the famous "hydrogen line" of 1420 MHz, for example, then the frequency of a signal coming from the east will be increased by 1.6 kHz and one coming from the west will be decreased by 1.6 kHz. Of course, the apparent position of a star changes as the Earth turns, so this Doppler shift varies with time, reaching a maximum of –0.16 Hz/sec (for a source in the equatorial plane).[2]

---

1 Per *sidereal* day, that is: 23h 56m, approximately. You're off by 4 minutes if you mistakenly use the sun as your reference direction.

2 You can think of this as deriving from the observatory's centripetal acceleration, if you like.

### 8.7.2 The Earth orbiting around the sun

The Earth revolves around the sun once per year, with an orbital speed of about 30 kilometers per second. If we are observing at 1420 MHz, a signal coming from ahead of us along the orbit will be increased by 140 kHz and one coming from behind will be decreased by 140 kHz. Because Earth's year is so much longer than its day, the change in Doppler shift due to the Earth's centripetal acceleration will be a maximum of 0.0028 Hz/sec – much smaller than the similar term for the Earth's rotation.

### 8.7.3 The Earth and moon orbiting around each other

The Earth and its moon orbit each other once a month, and their center of gravity (around which they orbit) is actually inside the Earth. At 1420 MHz, this motion will cause a maximum Doppler shift of 57 Hz and its acceleration will cause a maximum change in Doppler shift of 0.00016 Hz/sec – both of which are much smaller than the effects of the previous two motion components.

If we assume that the transmitting civilization corrects the frequency of its transmission so that, in our solar system, the signal appears to be at some guessable frequency with no changes due to acceleration of the transmitter, then we only have to handle the Doppler shifts associated with the motion of our own planet. We can do this by adjusting the frequency of our receiver, essentially "changing the station" to track any changes in the signal's apparent frequency.

The best way to compensate for Doppler Effects will depend on the parameters of the particular search program, including the total bandwidth over which the search operates and the amount of time spent observing each point on the sky. A search looking over a small total bandwidth (less than the 280 kHz spread caused by the Earth's orbital velocity) and with a long observing time for each target will require careful compensation for both the velocity and acceleration components of the Earth's motion. A search with a large total bandwidth and a short observing time for each point on the sky (e.g., an all-sky search using the Earth's rotation to sweep across the celestial sphere) may require compensation for only some of the acceleration components. More details about the Doppler consequences of each of our radio searches are included below.

## 8.8 Interference Rejection

A good candidate for a signal from an alien civilization needs to have two qualities: it must be artificial and it must not come from Earth. A signal with a bandwidth of less than 1 Hertz is undoubtedly artificial so a good interference rejection strategy is to perform a Fourier transform, as described above, and then examine the results for narrow features.

Proving that a signal did not come from Earth is trickier. The Earth has many transmitters in the frequency range used by SETI projects so there is a serious interference problem. A robust search needs to filter all of the received narrow-band features and pass only those which have some quality that an interstellar signal would uniquely possess.

One thing we know for certain about an interstellar signal is that it would come from a fixed location on the celestial sphere. All SETI projects take advantage of this characteristic in various ways to discriminate terrestrial interference from true signal candidates. Some of these ways include:

- *Using the known motion of the Earth to ensure that a candidate displays the proper characteristic Doppler shifts of an extraterrestrial signal*

The Doppler "chirp" (changing Doppler shift) caused by the centripetal acceleration of the Earth as it rotates around its axis is a very robust method of filtering out interference. Terrestrial signals are transmitted in the same frame of reference as the observatory and therefore do not show such a chirp. Our META search, for example, swept its local oscillator at precisely the right rate to cancel any Doppler shift for non-terrestrial signals while adding that shift to terrestrial ones. META had a channel width of 1/20 Hz and an integration time of 20 seconds, so the energy from terrestrial signals was spread across about 50 adjacent channels, while that from non-terrestrial ones would remain in one or two adjacent channels.

- *Using the time history of a signal to ensure that its power profile matches that of the antenna aimed at the right spot in the sky*

Any radiotelescope we use to receive signals from space has a characteristic gain pattern: signals received near the center of the beam (where the telescope is pointed) will be stronger than signals received near the beam's edges. As the beam sweeps past the point on the sky from which the signal is being transmitted, the received signal will start out weak, get stronger until the center of the beam is reached, and then get weaker again. Having a time history matching the telescope's gain pattern and rate at which the beam is sweeping across the sky is good evidence that the signal is coming from a point on the sky and not from Earth.

- *Deploying multiple antennas or antennas with multiple beams to ensure that the received signal is consistent for the entire system*

Some combination of the above two, the search setup can comprise multiple antennas or antennas which look at the sky with multiple beams. A successful candidate will have a Doppler shift and time history consistent with all of them. A terrestrial veto antenna might also be used to discriminate against local interference. The terrestrial antenna would look at its local horizon using a low-gain “pancake” beam; any signals received by the radiotelescope that are also received by the terrestrial antenna are unlikely to come from another star system.

Some extraterrestrial signals do not come from alien civilizations, namely those being radiated by satellites and spacecraft launched from Earth. Such spacecraft travel at high speeds, and so their signals tend to present time and Doppler profiles very different from those expected by an interstellar beacon. Our BETA search frequently picked up satellite signals, some so strong that they were also received by the system’s terrestrial veto antenna.

Interference rejection is a difficult task, especially when working outside of protected radio astronomy bands, and the situation has become progressively worse in our modern era of wireless communication. Television and radio stations, mobile telephones, satellites, etc., have plagued our various projects with unwanted interference. In all of our radio searches we devoted a large fraction of the resources and effort to filtering out terrestrial interference.

## 8.9 Specific Searches

The Harvard SETI group has operated seven search projects: five looking for narrow-band signals at radio frequencies near the neutral hydrogen line (1420 MHz) and two looking for nanosecond-scale pulses in the optical regime. Each succeeding search has built upon what we learned in previous projects to make the new system more robust and easier to operate.

### 8.9.1 Arecibo (1978)

The group’s first SETI project began in 1978 with a search, at the Arecibo 1000-foot dish, of 200 interesting candidate objects. The project used existing equipment at the observatory and performed its 64K-point Fourier transforms in software on a general purpose computer (Figure 8.1). Because the transform was done in software and took longer than the observation, the search could not be done in real time.

The Arecibo search observed 1 kHz instantaneous bandwidth segments centered on the frequency of the neutral hydrogen line, and had a resolution bandwidth of 0.015 Hz. That was the highest resolution and sensitivity ever achieved in SETI at the time, and the sensitivity record lasted for decades afterward. The project used real-time compensation for Doppler shifts relative to the sun, effectively producing a chirped receiver (tracking the change in Doppler due to the Earth’s acceleration) which afforded excellent immunity to non-chirped terrestrial radio interference. The tiny bandwidth of the search (1 kHz at 1420 MHz: a part in a million) required a transmitting

**Figure 8.1** Paul Horowitz in the control room of the Arecibo Observatory, plugging together his first SETI receiver from the observatory's existing equipment.

civilization to precompensate their beacon frequency for their own motion relative to our sun. This is a rather restrictive scenario, although it is a task an advanced civilization could accomplish, if they so desired. This search was described in Horowitz (1978).

### 8.9.2 "Suitcase SETI" (1981–1982)

With support from NASA and The Planetary Society, Paul Horowitz spent a year as a National Research Council (NRC) postdoctoral fellow at NASA Ames Research Center (1981–1982), where he and colleagues from Stanford University and NASA built a high-resolution hardware spectrometer that could handle, in real time, the kind of signal processing that was used in the earlier Arecibo search. Specifically, this "Suitcase SETI" hardware implemented the FFT in firmware on a Motorola 68000 microprocessor, achieving 64K-channel spectrum analysis (0.03 Hz resolution bandwidth, 2 kHz instantaneous bandwidth) simultaneously in each of two polarizations, along with the capability to search for narrow-band features and archive them. The hardware included a phase-continuous programmable local oscillator that could "chirp" to compensate for changing Doppler in real-time. Suitcase SETI travelled to the Arecibo Observatory in March 1982, where it searched 250 candidates (stellar and other), mostly at the second harmonic of the neutral hydrogen line (2840 MHz). Once again, radio frequency interference rejection was impressive; and once again, there were no confirmed signals. However, as a test of the hardware, we looked at the maser source W49(OH), and produced a spectrum of such detail that, if plotted at 200 dpi, would stretch across the 1000-foot dish. This search was described (with others) in Horowitz et. al. (1986).

### 8.9.3 Sentinel (1983–1985)

With sponsorship by The Planetary Society, and with the permission of NASA, we reconfigured Suitcase SETI as a dedicated search at the Harvard/Smithsonian 84-foot steerable radiotelescope at the Oak Ridge Observatory in Harvard, Massachusetts. This search, known as “Sentinel”, was the first dedicated high-resolution SETI, covering the northern sky in a “transit mode” (in which the Earth’s rotation scans the beam from west to east) at the neutral hydrogen line. Unlike the earlier targeted searches at Arecibo, we chose an all-sky transit search because the larger beam size (30 arc minutes, compared with 3 arc minutes at Arecibo) corresponds to a full search of the visible sky in about 200 days. Once again, the system’s receiver was “chirped” so that it was only sensitive to transmissions whose Doppler profile was from beyond the Earth. And because the instantaneous bandwidth was only 2 kHz, the system could only receive beacons whose frequency had been precompensated to the frame of reference of the sun. As in the previous two searches, the system had good interference rejection but found no confirmed signal sources. Sentinel is described in the Icarus article referenced above as well as in Horowitz and Forster (1985).

### 8.9.4 META (1985–1994)

Sentinel and its predecessors achieved high resolution bandwidths at the expense of total frequency coverage: Suitcase SETI and Sentinel covered only 2 kHz of bandwidth, and the earlier off-line search at Arecibo covered only 1 kHz. They required a transmitting civilization to target our star specifically in order to permit Doppler precompensation to the sun’s frame of reference. What was needed was a spectrometer of much greater bandwidth, in order to cover contiguous bands centered on “magic” frequencies as seen in “magic” inertial rest frames. Good choices for the rest frames are:

1. The galactic barycenter, i.e., the center of mass of our galaxy
2. The local standard of rest, which follows the velocities of the group of stars near our sun
3. The heliocenter, or our sun’s frame of reference
4. The rest frame of the cosmic microwave background (CMB).

The uncertainties in our knowledge of these frames were of order 30 km/s, corresponding to +/− 150 kHz of Doppler uncertainty at 1420 MHz. META's spectrometer would therefore need an instantaneous bandwidth of at least 300 kHz.

Furthermore, the required long integration times (30 seconds) of the previous searches prevented immediate re-observations of interesting candidates, because points on the sky drifted through the radio telescope's beam in only about two minutes.

We thus embarked on a project to build an 8-million channel spectrometer, to achieve 400 kHz of instantaneous bandwidth at 0.05 Hz resolution bandwidth.This was the Megachannel ExtraTerrestrial Assay, or META, funded by The Planetary Society through a gift from film director Steven Spielberg. META was a dedicated all-northern-sky transit search, with successive spectra alternating among the rest frames listed above. As with its predecessors, META used an agile local oscillator to compensate for Doppler chirp caused by site acceleration, which provides a characteristic changing Doppler signature for narrowband signals of extraterrestrial origin. During each 20-second integration (the time needed to collect the amount of data necessary to achieve the 0.05 Hz frequency resolution), the Doppler chirp amounts to some 50 frequency channels, thus nicely discriminating against terrestrial radio interference. This dedicated search covered most of the northern sky (–30 degrees to +60 degrees declination) with the Harvard/Smithsonian 26-meter radiotelescope operating in "meridian transit" mode. Each potential source passed through the antenna beam pattern in approximately two minutes, during which the three reference frames were covered once in each antenna polarization. META's hardware, designed in 1983, consisted of GaAsFET low-noise amplifier frontends for both observed polarizations, image-reject downconverters with programmable phase-continuous second local oscillator (to provide the needed Doppler chirp), 7-bit quadrature digitizers, a 144-channel digital filter bank (built by Ivan Linscott at Stanford) feeding an array of 144 Motorola 68000-based 64K-point FFT processors (based on the dedicated FFT hardware of Suitcase SETI), and a central workstation of modest performance. META was the first megachannel SETI, and ran for a decade before being replaced by BETA in 1995. In an analysis of five years of data, during which 60 trillion channels were searched, we found 37 candidate events exceeding the average detection threshold of $1.7 \times 10^{-23}$ W/m$^2$, none of which has been detected upon repeated reobservations. In spite of lack of a confirmed signal, META permits us to set some interesting limits on the prevalence of advanced civilizations that transmit in ways that the search would have detected. For a technical summary, see Horowitz and Sagan (1993); a non-technical version appears in Horowitz (1993).

### 8.9.5 BETA (1995–1999)

Given the results of META and its predecessors, and the fact that SETI elsewhere has similarly found occasional candidates that have the right characteristics but do not repeat in observations made much later (a characteristic that led to a meeting on "Intermittency in SETI" at the SETI Institute in January 1994), we felt that the next search system should incorporate means for:

1 Rapid and automatic reobservation of candidate events with two antenna beams pointed at the sky, one aimed slightly to the east and the other slightly to the west. A successful candidate would first appear in the east beam and, after transiting that, do the same in the west beam. The time over which it did this should be consistent with the rotation speed of the Earth.

2 Better discrimination of interference, through a simultaneous 3-beam configuration, including the two "sky beams" above, and a third terrestrial "veto" antenna, aimed at the local horizon.

3 Coverage of the full 1400-1700 MHz "water hole" band of frequencies.

Thus was born the Billion-channel ExtraTerrestrial Assay, or BETA, which was switched on in October 1995 (Figure 8.2). BETA took four years to design and build; the project was funded by The Planetary Society, NASA, the Bosack/Kruger Charitable Foundation, and the Shulsky Foundation. It used the Oak Ridge Observatory 26-meter dish with dual (east-west) feedhorns and a third low-gain terrestrial "veto" antenna to feed a 240-million channel FFT spectrometer (80 million channels of 0.5 Hz resolution and 40 MHz instantaneous bandwidth for each of those feeds). The spectrometer outputs fed an array of programmable "feature recognizers" that sifted through 250 megabytes per second of spectral data, seeking distinctive spectral features that transited from the east to the west horn without appearing in the low-gain terrestrial antenna. BETA's hardware consisted of HEMT low-noise frontend amplifiers, an array of 63 quadrature mixer/digitizers with GPS phase-locked local oscillators, and an array of 63 4-million channel complex FFT boards feeding a flexible state-machine based feature recognizer array resident in a set of Pentium motherboards. The latter communicated over Ethernet with a UNIX workstation that performed final processing and archiving functions. BETA searched the 1400-1720 MHz spectrum in eight hops of 40 MHz bandwidth, with each hop taking 2 seconds yielding a 16 second time for a full cycle through the water hole. Thus each potential source was visited eight times at each frequency hop, in each sky beam. A good candidate (seen first in the east beam, then in the west beam, but never in the terrestrial beam) triggered the antenna to leapfrog a few beam widths to the west, inviting the source to perform an encore. If that ever happened,

**Figure 8.2** Paul Horowitz seated at the control console of Project BETA at Oak Ridge Observatory. Much of BETA's equipment was custom designed and constructed for the search.

the antenna would break off its survey and go into sidereal tracking mode, repeatedly moving on and off the candidate source, archiving all integrated spectra.

Because BETA observed outside of protected radio astronomy bands, it was subject to more terrestrial interference than our previous searches. Even so, BETA found no confirmed extraterrestrial signal sources. A complete description of the search can be found in Leigh and Horowitz (1996; 1999) and Darren Leigh's doctoral dissertation (Leigh, 1998).

## Oak Ridge dish damaged by wind

On March 23, 1999 the 26-meter radiotelescope at Oak Ridge Observatory was blown over by strong winds.[3] While the dish did not collapse, the wind banged it into the ground several times, severely damaging the dish surface

[3] Google Earth shows the radio telescopes location at 42°30'22" N, 71°33'12" W, and you can use its historical imagery feature to see the dish site as it used to be and as it looks now.

and secondary support structure. We investigated the possibility of repair, but it was eventually decided that the cost was too great for such an old facility. This ended the run of BETA, which became the group's final radio SETI project.

## 8.10 Optical SETI Projects

There's more to the electromagnetic spectrum than radio waves, though the latter have been the perennial favorite for SETI. That choice has been partly historical (in 1959, at the time of Drake's pioneering *Project Ozma*, there were no lasers, but there were already megawatt radio transmitters), and partly scientific (radio waves are efficient, the galaxy is transparent to them, and the radio sky is quiet). But 1960 saw the invention of the laser and, a year later, the suggestion (Schwarz and Townes, 1961) that it could be used for interstellar communication.

In the following years the capabilities of lasers grew exponentially. Townes published a remarkable paper (Townes, 1983). Making reasonable assumptions about antenna apertures and accuracies, detection methods, transmitter power and so on, he showed that optical methods are comparable to, or perhaps slightly preferred, in the single figure of merit of delivered signal-to-noise ratio for a given transmitter power. And Dan Wertheimer at Berkeley demonstrated the benefits of an optical twin-detector arrangement to suppress background starlight.

Consider this: It is a remarkable fact that our most powerful pulsed lasers, if pressed into service as interstellar calling cards, would generate a flash of light that, seen by a distant observer in its slender beam, would appear 10,000 times brighter than our sun during its brief flash. And this does not assume that the recipient is looking for a particular wavelength – the flash would outshine the sun by four orders of magnitude in "broadband visible light."

### 8.10.1 The targeted optical search

Inspired by these ideas, the Harvard group designed an optical SETI apparatus, piggybacked it onto an existing survey program running on the 61" telescope at the university's Oak Ridge Observatory (the largest optical telescope east of the Mississippi), and began observations in 1998. This search ran for more than five years, making some 16,000 observations of 5000 stars (including several years of simultaneous observations with Dave Wilkinson's group at Princeton, observing with an identical system). In this search we looked for nanosecond-scale pulses, compact in time rather than in frequency (as in the radio searches), because the situation is fundamentally different: at visible wavelengths a short pulse travels undistorted through the galaxy, it stands out from the steady light of a star, and it's easy to detect with simple apparatus; by contrast at radiofrequencies a pulse is smeared out as it travels (dispersion), and it must compete with backgrounds such as lightning and "cultural" impulsive artifacts.

This "targeted" search was reported in Howard et. al. (1999; 2004) and, at a less technical level, in Horowitz et al. (2001).

### 8.10.2 The all-sky optical search

Our first optical SETI worked well, in the sense that it was refreshingly free of false events, as it looked at 5000 sun-like stars in our neighborhood. Of course, the bad news is that it was likewise free of any detections of the elusive flash – a flash that, we are sure, would be heard around the world. But it did set some interesting limits on the prevalence of civilizations sending laser flashes our way. And it broke ground for our biggest project to date: a search for optical pulses from a source anywhere in the northern sky.

Five thousand stars may sound like a lot – but it's not. There are roughly a million stars similar to our sun within a thousand light-years of Earth, and our optical search covered only about 1/100,000 of the heavens. It looked at the sky through a soda straw. Hence the motivation for the next step: why not cover the whole sky visible from our site?

Thus was born the "all-sky OSETI," supported (as with previous searches) by The Planetary Society. This turned out to be a gargantuan project: we had to build our own dedicated wide-field telescope (a 72" f/2.5 spherical primary "light bucket," with a flat 36" secondary; the new largest 'scope east of the Mississippi) and observatory building, and we had to replicate the twin-detector scheme 500-fold. The telescope was cast for us by Ray Desmarais, and housed in a custom roll-off roof observatory. The "camera" consists of a giant beamsplitter and array of pixilated photomultiplier tubes, followed by custom electronics wrapped about an array of full-custom ICs and downstream processing electronics. The system is sensitive to nanosecond light flashes in its observing stripe of 0.2 × 1.6 degrees, which is swept across the sky by Earth's diurnal motion. A tremendous amount of processing takes place in real time: the system makes a trillion measurements per second (about 3.5 terabits/sec), about the same amount each second as the contents of all books in print. It takes about 200 clear nights to cover the northern sky. The search commenced, full-time, in 2006. As of this book's publication we have covered the sky three times, so far without a confirmed detection. This system is described in Betts (2006) and Horowitz et. al. (2001; 2008).

## 8.11 Conclusions

Although some of our searches picked up tantalizing prospects, we have never received a signal of clear extraterrestrial origin. While this is personally disappointing, we understand that all of the world's SETI programs have barely scratched the surface of the search space, given their limitations in bandwidth, sky coverage, sensitivity, duty cycle, etc. We are not discouraged because we understand the true size of the problem. Space is vast and the time scales for communication across interstellar distances boggle the human mind.

SETI is a waiting game that can only be won by the extraordinarily patient. It is our hope that the human race will demonstrate the necessary patience so that we can join other patient and long-lived civilizations.

**Acknowledgements**

Over more than three decades many people have been part of the Harvard SETI Group, either directly or indirectly:

Ian Avruch, Derrick Bass, Joe Caruso, Chris Clearfield, Ken Clubok, Chip Coldwell, Paul Eremenko, John Forster, Jason Gallicchio, Greg Galperin, Will Galvin, Neil Hendin, James Higbie, Andrew Howard, Dave Latham, Chris Laumann, Dylan Manna, George Marcus, Brian Matthews, Curtis Mead, Costas Papaliolios, Nick Schectman, Suhail Shah, Al and David Sliski, Pratheev Sreetharan, Robert Stefanik, Anne Sung, Jonathan Weintroub, Eric Wey, Jonathan Wolff, and Joe Zajac.

In addition, the following contributed substantially to the Harvard searches:

Peter Backus, Kok Chen, Mike Davis, Martin Gimersky, Ed Groth, Jon Hagen, Norm Jarosik, John Kraus, Ivan Linscott, Al Peterson, Robert Pound, Ed Purcell, Carl Sagan, Cal Teague, Dan Werthimer, Dave Wilkinson, Alexander Willman, and Bill Yerazunis.

Finally, we received financial support and donations of equipment from:

The Planetary Society and its members, the Bosack/Kruger Charitable Foundation, the Shulsky Foundation, NASA, Steven Spielberg, Micron Technology, Hewlett-Packard, Fluke, AMD, Intel, MOSIS, and Xilinx.

## References

Betts, B. 2006. We Make It Happen! The Planetary Society Optical SETI Telescope. *The Planetary Report*, **XXVI: 3**, May/June, 4-7.

Drake, F. 1969. Concepts in the Problem of Detecting Distant Civilizations. Colloquium as part of the Loeb Lecture Series, Department of Physics, Harvard University, 1969.

Horowitz, P. 1978. A Search for Ultra-narrowband Signals of Extraterrestrial Origin. Science, **201**, 733-735.

Horowitz, P. 1993. Project META: What Have We Found? *The Planetary Report*, **XIII: 5**, September/October, 4-9.

Horowitz, P. 2000. Flash! Optical SETI Joins the Search. *The Planetary Report*, **XX: 2**, March/April, 8-15.

Horowitz, P. and Forster, J. 1985. Project Sentinel: Ultra-narrowband SETI at Harvard/Smithsonian. In M.D. Papagiannis (Ed.) *The Search for Extraterrestrial Life: Recent Developments*. Proceedings of IAU Symposium No. 112, Dordrecht, D. Reidel Publishing Co., 291-303.

Horowitz, P. and Sagan, C. 1993. Five Years of Project META: An All-Sky Narrow-Band Radio Search for Extraterrestrial Signals. *Astrophysical Journal*, **415**, 218.

Horowitz, P. et al. 1986. Ultranarrowband Searches for Extraterrestrial Intelligence with Dedicated Signal-Processing Hardware. *Icarus* **67**, 525-539.

Horowitz, P et al. 2001. Targeted and All-sky Search for Nanosecond Optical Pulses at Harvard-Smithsonian. *Proc. SPIE* **4273**, 119.

Horowitz, P. et al. 2008. The Planetary Society's All-Sky Optical SETI: Where Are We Now? *The Planetary Report* **XXVIII: 6**, November/December, 4-7.

Howard, A. et al. 1999. Optical SETI at Harvard-Smithsonian. In *Proceedings of the 6th International Conference on Bioastronomy*, Kohala Coast Hawaii, August 2-6.

Howard, A. et al. 2004. Search for Nanosecond Optical Pulses from Nearby Solar-Type Stars. *Astrophysical Journal* **613**, 1270.

Leigh, D. 1998. An Interference-Resistant Search for Extraterrestrial Microwave Beacons. Doctoral dissertation, Division of Engineering and Applied Sciences, Harvard University.

Leigh, D. and Horowitz, P. 1996. Millions and Billions: The META and BETA Searches at Harvard. In *Proceedings of the 5th International Conference on Bioastronomy* (IAU Colloquium 161), Capri, Italy, 1-5 July.

Leigh, D. and Horowitz, P. 1999. Strategies, Implementation and Results of BETA. In *Proceedings of the 6th International Conference on Bioastronomy*, Kohala Coast Hawaii, August 2-6.

Purcell, E. 1960. Radio Astronomy and Communication Through Space. Brookhaven Lecture Series, BNL 658 (T-214).

Schwartz, R.N. and Townes, C.H. 1961. Interstellar and Interplanetary Communication by Optical Masers. *Nature* **190**, 4772, 205-208.

Shklovskii, I.S. and Sagan, C. *Intelligent Life in the Universe*. Holden-Day.

Townes, C.H. 1983. At What Wavelengths Should We Search for Signals from Extraterrestrial Intelligence? *Proceedings of the National Academy of Sciences of the United States of America* **80: 4**, 1147-1151.

# Part II: The Spirit of SETI Present

# 9

# ATA: A Cyclops for the 21st Century

Jill Tarter,
*Director, Center for SETI Research, SETI Institute*

## 9.1 Introduction

The Search for ExtraTerrestrial Intelligence (SETI) finally has its own full-time telescope. The Allen Telescope Array (ATA) in Northern California was dedicated on October 11, 2007. This array, which will eventually be comprised of 350 small radio antennas, each 6.1 meters in diameter, is being built as a partnership between the SETI Institute and the University of California Radio Astronomy Laboratory. At the dedication, Paul G. Allen (who provided the funds for the technology development and the first phase of array construction) pushed a silver button and all 42 antennas of the current ATA-42 slewed to point in the direction of the distant galaxy M81. Specialized electronic backend detectors attached to the ATA began making a radio map of that galaxy and simultaneously began SETI observations of HIP48573, a G5 V star near M81 on the sky and a distance of 264 light years from Earth. The Allen Telescope Array will greatly improve the speed of conducting SETI searches over the next few decades, and it will allow a suite of different search strategies to be undertaken. This paper summarizes some of the earliest SETI observations from the array, and describes the search strategies currently being planned.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_9, © Springer-Verlag Berlin Heidelberg 2011

## 9.2 Commensal Observations

The ability to conduct multiple concurrent observing programs on the ATA is enabled by the large field of view of an array constructed from small 6.1-meter antennas, and its flexible electronics. The **Full-Width Half-Maximum** (FWHM) of the primary beam is 3.5°/f (in GHz). A spectral imaging correlator designed and built at the UC Berkeley Radio Astronomy Lab is used to make radio images (or maps) of this large field of view. The maps contain about 20,000 independent pixels, and for each pixel there are 1024 spectral channels. A series of beam formers add the voltage outputs from each of the antennas, adjusting the phases and time delays so that the information from just one pixel is selected to be processed by electronic backends, such as SETI signal processors.

Figure 9.1a illustrates the formation of 32 beams within a 2.5° field of view at a frequency of 1.4 GHz, with the large galaxy M31 and the moon shown for scale. Also shown is the field of view of the large Arecibo telescope in Puerto Rico. For a single dish, this is also identically equal to the resolution beam of the telescope. For the 42-dish implementation of the Allen Telescope Array (hereafter called the ATA-42), or any interferometer, the resolution beam size is determined by the largest baseline in the array (300 m for the ATA-42, 900 m for the fully built out ATA). Today, the first two dual-polarization beam formers, constructed from Field Programmable Gate Arrays (FPGAs),

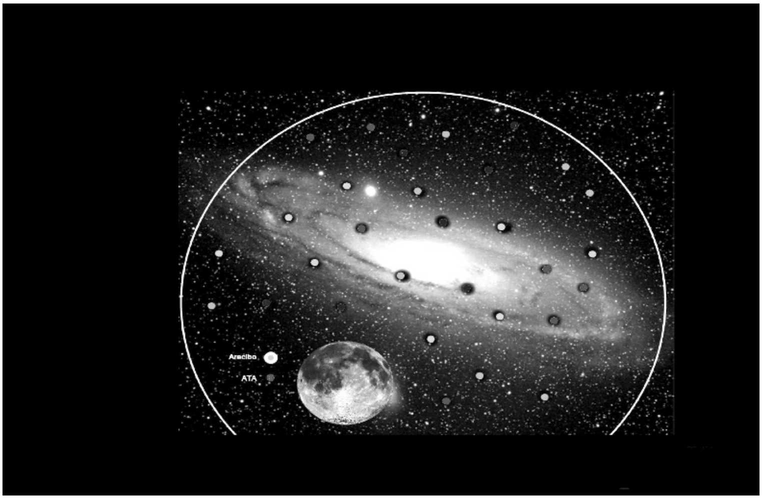



**Figure 9.1a** The field of view of the ATA at 1.4 GHz, with 32 individual resolution beams. The moon is 0.5° in diameter, and the Andromeda galaxy is 2.5° across.

operating on two separate data channels on the ATA-42, are allowing us to begin SETI observations. Over time, if it becomes possible and more cost effective to construct software beam formers within commodity processors, that number will grow to as many as four dual-polarization beam formers on each of the four independently tunable frequency channels available at the ATA (32 beams in all).

Figure 9.1b illustrates conceptually how the field of view would be mapped by a spectral imaging correlator. In fact, each pixel in this grid should be the same size as the resolution beam of Figure 9.1a. Today on the ATA-42, there are two dual-polarization correlators, each capable of correlating 32 antennas, on each of two independently tunable frequency channels.

The ability to observe multiple objects within the large field of view, or to observe the same object at multiple frequencies makes commensal observations feasible.

## 9.3 Targeted SETI Searches

For SETI observations that select target stars which are likely to host habitable planets, multiple beam formers and a large catalog of targets are the keys to efficient observing. As Figure 9.2 illustrates, for a given size catalog, the average number of target stars within the field of view of the ATA increases with the size of the catalog and decreases with frequency.

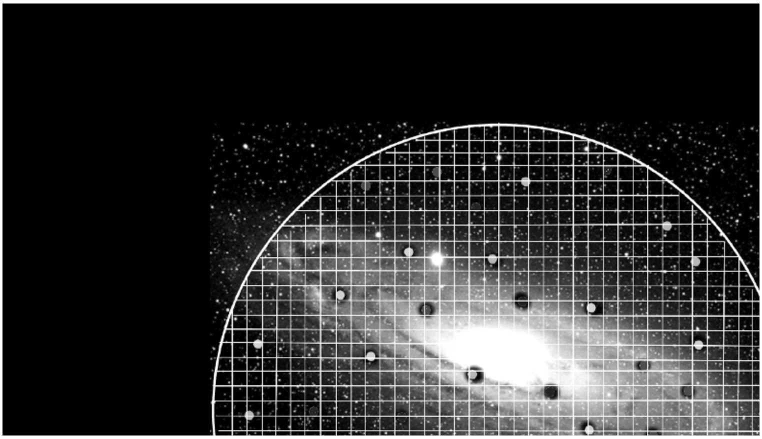


**Figure 9.1b** A spectral-imaging correlator can make a map of the entire Andromeda galaxy in one pointing, with ~20,000 pixels the size of the resolution beam, and each pixel having 1024 channels of spectral data.

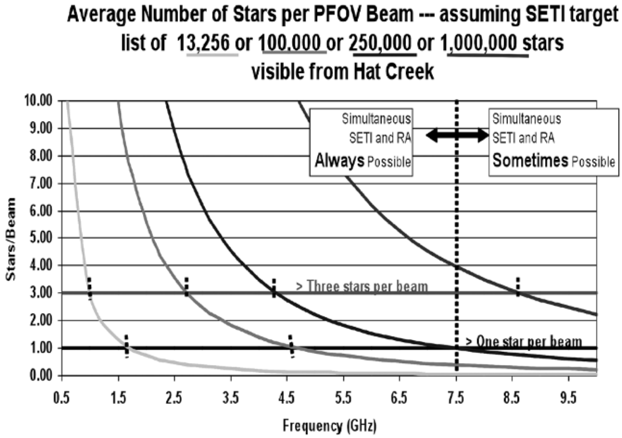



**Figure 9.2** Average number of stars per primary field of view of the ATA, assuming different stellar catalogs.

It is desirable to observe multiple stars within each field of view simultaneously to increase the speed of the SETI search, and also enable strategies to discriminate against interfering sources in near-real-time. Any signals detected simultaneously in the directions of more than one star are actually interfering signals entering the sidelobes of the array and can be disregarded.

For the ATA, Turnbull and Tarter (2003a; 2003b) have used data from the Hipparcos and Tycho-2 catalogs of stars to select the most likely stars to be hosts of habitable planets on which technologies may have evolved. Selected stars must be older than 3 Gyr, must not have companion stars that preclude stable planetary orbits, must have sufficient heavy element abundance to permit the formation of rocky planets, and must not flare in luminosity by more than a few percent.

The total numbers of stars for the curves in Figure 9.2 represent the stars selected from the Hipparcos catalog (13,256), the stars selected from the Tycho-2 catalog, whose distances are less accurately determined and therefore some of the selected spectral type F, G, and K stars may in fact be distant giants rather than nearby main sequence stars (250,000), and the anticipated results from the Nomad catalog now being sorted, or the data from the astrometric Gaia spacecraft in the next decade (~1,000,000).

At frequencies below 4.5 GHz, there will usually be at least three Habstars derived from the Tycho-2 catalog within the primary field of view of the ATA. The Tycho-2 catalog yields enough target stars so that at frequencies below 7.5 GHz, there will (on average) be at least one Habstar within the primary field of view of the array, ensuring that commensal observations will be possible, although perhaps not efficient.

Figure 9.3 is a visual demonstration of sharing the sky for targeted SETI observations and cosmography, the precise astronomical mapping of local hydrogen in the nearby universe. The underlying image is a map of the Andromeda galaxy made in the neutral hydrogen line at 1.42 GHz using 42 ATA antennas. Superimposed on this map are three stars from the Habcat I list, derived from Hipparcos data, and 66 stars drawn from Habcat II based on the Tycho-2 data. All of these target stars are within the Milky Way Galaxy, less than 1000 light years from Earth, and are projected on the sky within 2.5° of the center of Andromeda. In principle, they could have been observed at frequencies at or below 1.42 GHz while this HI map is being made. In fact, this particular map was made in the fall of 2007, when the correlators first became usable but before the beam formers became fully operational, and so commensal observations in the M31 field have not yet taken place.

The first opportunity to perform substantial commensal targeted search observations occurred earlier in the summer of 2010, picking up on the

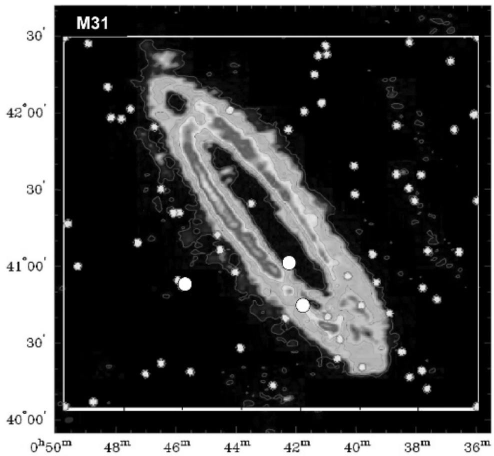



**Figure 9.3** Target SETI stars within the primary field of view of the ATA at 1.4 GHz.

observations demonstrated during the ATA-42 dedication, when the field centered on the galaxy M81 was again mapped, along with its companion M82. Figure 9.4 (top) shows the optical image of the field along with a, outer circle indicating the field of view at 5 GHz, and an inner circle marking the field of view at 1.42 GHz. In the optical image the galaxy M81 (center) is

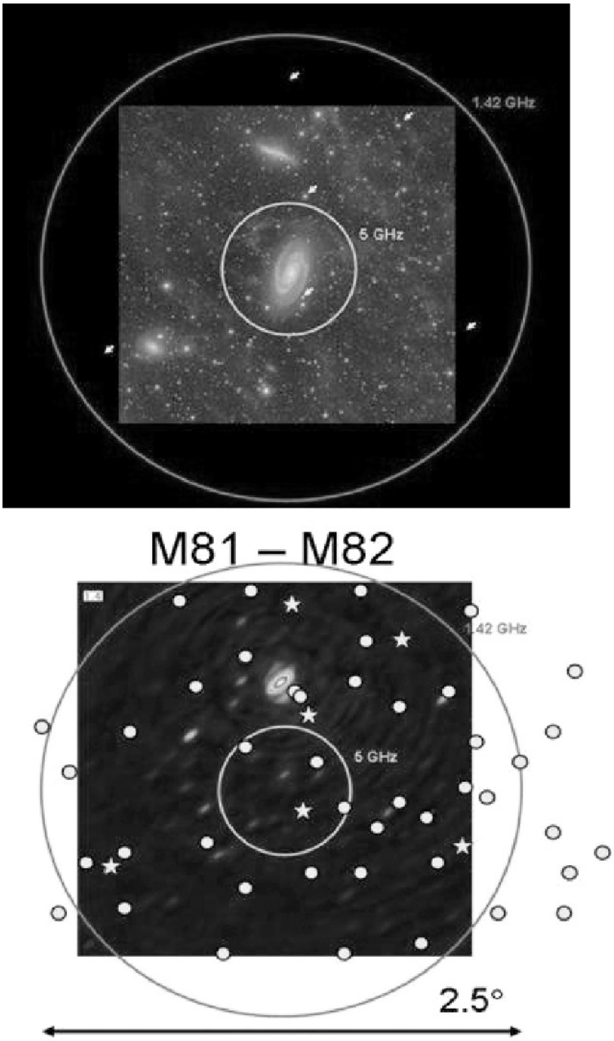



**Figure 9.4** Optical (top) and radio continuum maps (bottom) of M81 and M82, with SETI target stars from Habcat 1 and Habcat II superimposed.

much brighter than its companion M82. Figure 9.4 (bottom) shows the radio continuum map of these two galaxies in which only M82 is visible. Six stars from Habcat I and 44 stars from Habcat II are shown. While the continuum map was being made, these stellar targets were observed pair-wise over the frequency range from 1400–1426 MHz for 50 seconds each, using two dual-polarization phased beam-formers and a 176,160,768 channel Prelude spectrometer (see below) with 1 Hz resolution. Narrowband pulses and CW signals drifting up to ± 1 Hz/second, could have been detected if a transmitter with a power equivalent to 1-20 times the EIRP of the Arecibo Radar ($2\times10^{13}$ W) had been operating near these stars. The pair-wise observation of target stars is intended to help mitigate against terrestrial RFI signals.

Since the ATA-42 is still in its commissioning phase, these observations were hand-tended - that is the signal detection reports were observed and interpreted by humans rather than by the SETI System Executive (SSE) software that is normally responsible for detected-signal classification, selection of candidate signals and their automated follow up. This SSE functionality has been developed over the past decade of observing with Project Phoenix (Backus, 2004), and is being integrated into the ATA SETI detection system during the on-going commissioning tasks.

## 9.4 Prelude and SonATA

The current generation of SETI signal detectors is called Prelude and is installed in the signal processing room of the ATA along with correlators, beam formers, and pulsar processors. Figure 9.5 shows the 28-rack mounted PCs in that system. Each PC contains two custom accelerator cards that facilitate the near-real-time calculation of the high resolution spectra. In all 176,160,768 spectral channels with noise-equivalent bandwidths of 1 Hz are produced every second. Each spectrum is divided by a recent measure of the instrumental baseline, and the unit-normal data are statistically analyzed to detect narrowband carriers or pulses that may be changing frequency as fast as (d-/dt)/- $< \pm 10^{-9}$ Hz/s /Hz (or 1 Hz/s at an observing frequency of 1 GHz) due to Earth's diurnal rotation as well as any acceleration at the transmitting source. Powers are summed along all possible paths through the frequency time plane with very efficient algorithms for carriers and pulses (Cullers and Stauduhar, 1997). Detected signals are compared against a dynamic database of known RFI, and automated reobservation is conducted on interesting candidate signals. There currently exists a C++ software instantiation of the signal detection functions of the Prelude system running on commodity servers from Sun Microsystems. A side by side, on the sky demonstration of this software detector and the Prelude system is planned in the near future.

**Figure 9. 5** Prelude system hardware based on rack-mounted PCs, each with two custom accelerator cards.

Over the next few years, Prelude will be replaced by the software SonATA (SETI on the ATA) system, and that in turn will morph into a Software Defined Radio Telescope (SDRT) where other SETI detection algorithms and software beam formers can be implemented, with the involvement of the open source community.

## 9.5 Nulls and Beams

Beams are now formed on the ATA using FPGA-based reconfigurable computing components produced by the Berkeley Wireless Research Center (BEE2 boards) and analog/digital converters and signal distribution boards (ADC iBOBs and DAC iBOBs) that are the results of open source projects carried out at the Center for Astronomy Signal Processing and Electronics Research (CASPER) on the Berkeley campus. Figure 9.6 shows the signal flow through the tree summation of the two dual-polarization beam formers.

Before the individual voltage streams from the antennas can be summed together, they must first be corrected in phase and delay for fixed geometrical corrections and time-varying instrumental corrections that will change with pointing direction and frequency. This alignment of the phasors from each antenna requires a small correlator within the beam former, which calibrates each antenna against a reference antenna while observing an astronomical (or other) point source. This calibration results in a series of complex weights to be applied rapidly to the input data streams from the antennas to form a single pixel beam.

By adjusting the coefficients that weight the calibrated data from each antenna within the beam former, it is possible to form not only a single pixel beam in a particular direction, but a null beam at an arbitrary spatial offset from the primary beam. Figure 9.7 shows an example of this capability. The solid curve is a one-dimensional cut through the beam profile, in the absence of any nulling. The dashed curve is the resulting profile when a 1.2 arc minute wide, -25 dB null is created at an azimuthal offset of 12 arc minutes from the primary beam direction.

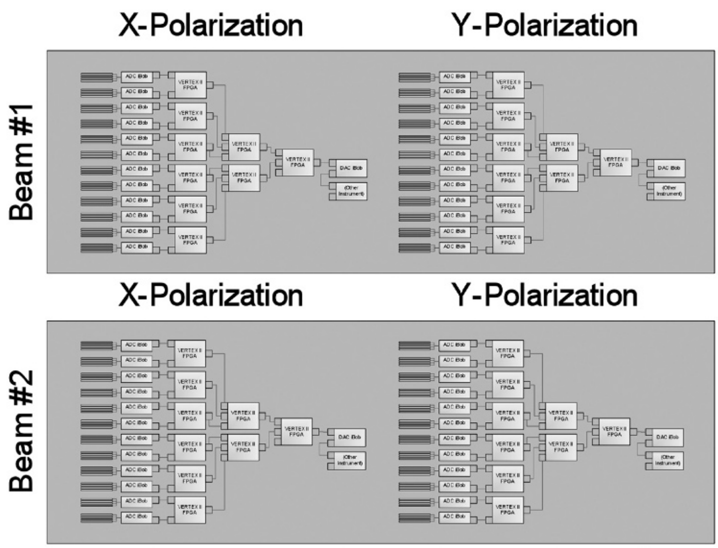



**Figure 9.6** FPGA-based summation of antenna data in two polarizations, to yield four independently tunable beams.

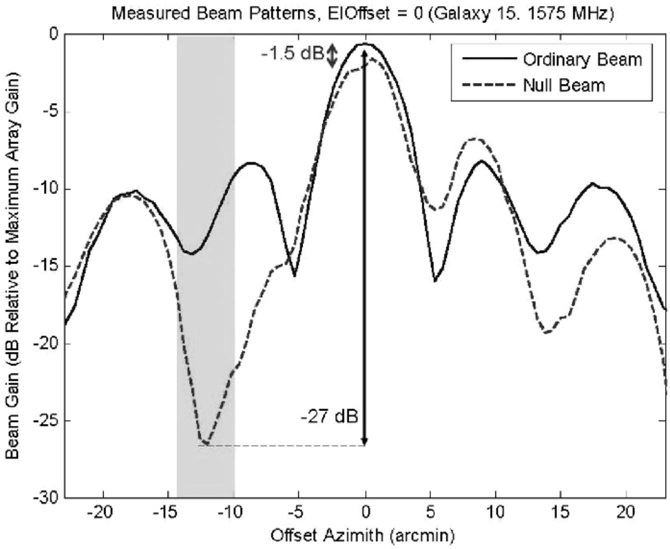



**Figure 9.7** Solid curve is the gain as a function of azimuth for ordinary beam profile, dashed curve shows the resulting gain profile when a null is placed 12 arcmin away from boresight.

Note that the forward gain of the primary beam is reduced by about 1 dB and the peak of the opposite first sidelobe is increased by the same amount as a result of forming this null. This seems like a small price to pay for a potential increase in observing efficiency in an RFI-filled environment. Although the Hat Creek Radio Observatory sits in a radio-quiet valley with a low population density, and is completely rimmed by mountains, it is subject to interference from orbital and over-flying transmitters.

To discriminate against interference, the Phoenix project historically used a pseudo-interferometer, consisting of two widely spaced (>200 km) antennas, to measure the differential Doppler signature of any candidate signal and compared that signature with what was expected for a signal actually arriving from the distance of the target star. The wrong Doppler signature excluded local stationery, orbital, and over-flying sources of RFI very effectively. With time, it is possible to generate this same sort of differential Doppler signature across the much shorter (<1 km) baselines of the ATA. However, nulling should provide a much quicker first line of RFI defense.

Observing multiple stars, each star serves as an ‘off-source’ observation for the others. If each beam former places a primary beam on one star and a spatial null at the position of another, then a good candidate ETI signal should be present in the stellar beam, but not in the null formed by another beam former on that same star. Any signal seen in the null is actually coming in through a sidelobe or is coming from the star and is much stronger than the null is deep. We have yet to model or empirically determine the depth of the nulls and the effective detection thresholds for RFI rejection. As we begin to 'educate' our beam formers and SETI signal detectors to the interference environment, we will select thresholds that allow for efficient observations, but are consistent with a probability of <50% of missing a real signal.

## 9.6 Fiducials on the Sky

Because there are no known ETI signals against which to calibrate the performance of new instrumentation, it is important to use sources produced by terrestrial technologies for that purpose. The underlying assumption of the SETI signal processing now undertaken at the SETI Institute is that an ETI signal will appear to be emanating from a point source moving at sidereal rate on the sky (note this excludes transmissions from alien spacecraft roaming our solar system). Distant spacecraft serve as proxies for such signals. Project Phoenix only observed its target stars from 1-3 GHz, and it routinely observed the S-band transmitter onboard the Pioneer 10 spacecraft to assure system performance. At the ATA, we plan to observe from 1 to 10 GHz, and calibrating the beam former at higher frequencies is more challenging. Therefore we have decided to use X-band transmitter onboard the Voyager 1 spacecraft as our current fiducial to guarantee functionality. Figure 9.8 (top) shows the first successful detection of Voyager 1. Though faint to the eye, this signal detection is statistically very significant when the power is summed along the correct path, as shown in the plot in the bottom of Figure 9.8. This detection of a distant spacecraft carrier demonstrates that the beam formers and signal processors are working correctly.

## 9.7 Exoplanet Survey

The first targeted search mini-survey to be made with the ATA-42 concentrated on the 155 known extrasolar planetary systems (comprising 193 planets) visible from Hat Creek. With current signal processing capability a search starting with the Cosmic Waterhole, requiring approximately 150 hours of ATA time, will be repeated several times during the next five years as sensitivity improves and speed increases due to enlargements of

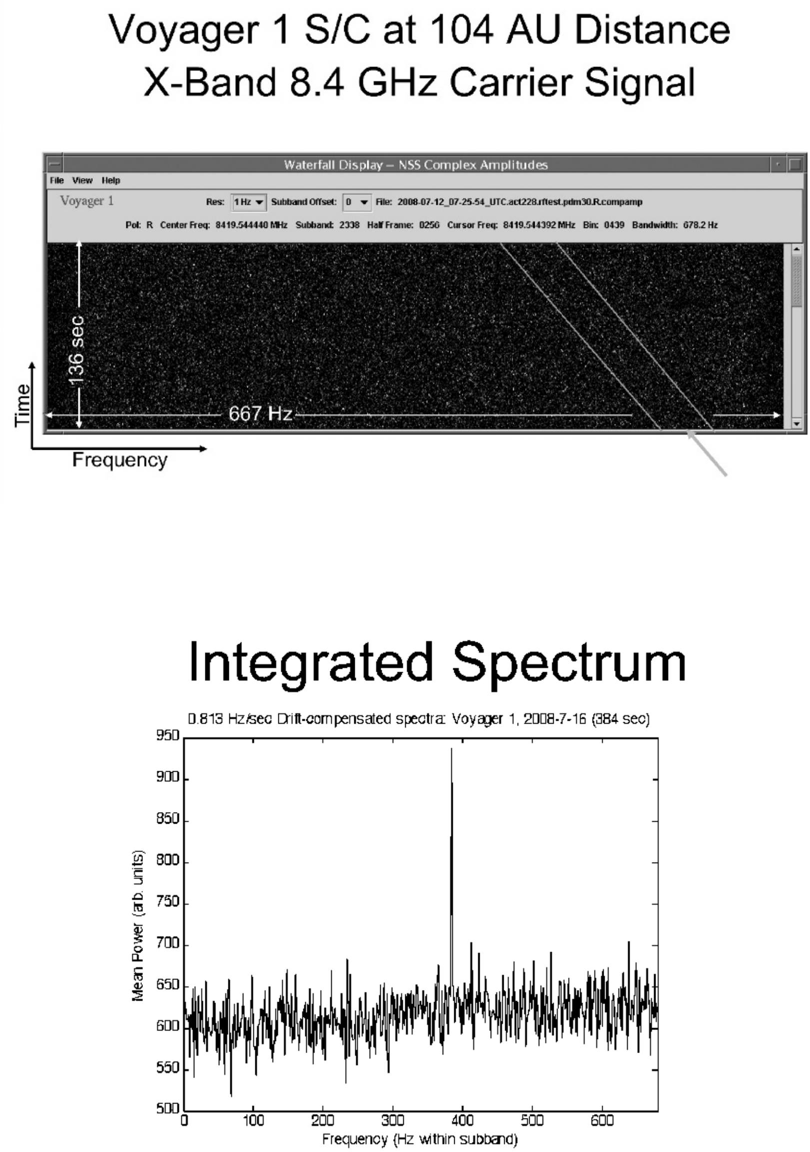



**Figure 9.8** (Top) waterfall plot of Voyager 1 spacecraft X-band carrier, with a drift rate of -0.813 Hz/s, lines and arrow are to guide the eye to find the weak signal. (Bottom) the integrated power along the correct drift path over time.

the processing bandwidth. While the migration of hot Jupiters in some of these systems may have ejected habitable worlds, Mandell, Raymond, and Sigurdsson (2007) have shown that it is possible for terrestrial planets to survive or re-form in the habitable zones of others. Habitable moons might accompany super-Jupiter-mass planets in yet other exoplanet systems, so it is worth keeping all discovered planetary systems on our target list until and unless additional information excludes them. Routine, commensal targeted searching will also begin during the first year of this mini-survey. Highest priority will be given to the nearest 100 stars whenever they appear within the primary field of view (PFOV), as these are the targets for which the weakest transmitters can be detected with ATA-42 sensitivity. For the more distant targets, observing times will be lengthened.

## 9.8 Galactic Plane Survey

As this chapter is being written, we are conducting a survey of the inner 20 square degrees of the galactic plane, the region with the highest column density of stars. The survey will require steering phased-array beams along the galactic plane for five hours per night (during the seven months per year when the galactic center region is visible to the ATA) eventually observing 3519 target positions. This program permits commensal radio astronomy observations of interesting sources in the galactic plane, and can be scheduled at times complementary to the Five GHz Sky Survey (FiGSS) observations for transient sources (the 90 second cadence of the FiGSS. program is not conducive to commensal SETI observing). The galactic plane survey will examine $\sim 4 \times 10^{10}$ stars over the frequency range from 1420 to 1720 MHz (the so-called Cosmic Waterhole). It will have sufficient sensitivity to detect a transmitter at the distance of the galactic center with an effective isotropic radiated power (EIRP) of $5 \times 10^{17}$ W (the equivalent of 25,000 Arecibo planetary radars). This survey samples more stars than a targeted search without bias as to what constitutes a “good” target, and is an optimal strategy in the circumstance that a small number of very powerful galactic transmitters exist. Figure 9.9 illustrates the region of the galaxy to be sampled.

Figure 9.10 illustrates how each one square degree is gridded into 186 target centers for the ATA-42 beam. Multiple beams (two or three) are automatically selected for observation using criteria that help to discriminate against RFI. With this region of the galactic plane visible for about five hours per day from March through October, it should be possible to complete the survey within one observing season.

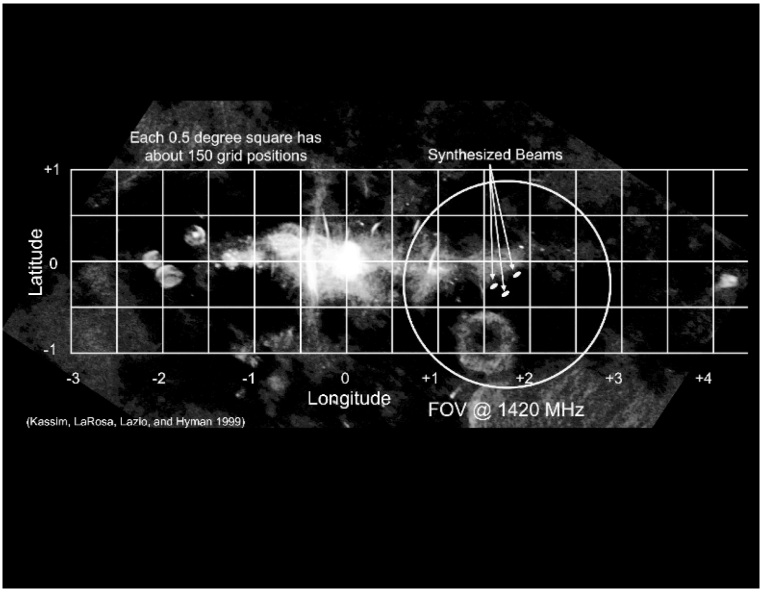



**Figure 9.9** 20 square degrees along the galactic plane from -4°<l<+6° and -1°<b<+1°. Lines of sight to this area intersect about 40 billion stars, mostly distant.

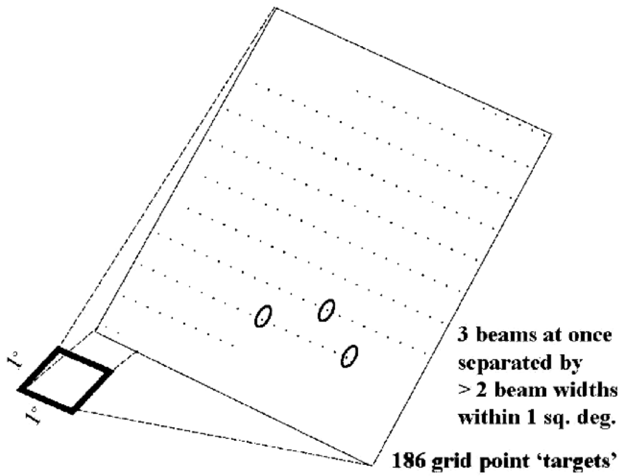



**Figure 9.10** Each square degree contains 186 target centers for synthesized beams.

**Acknowledgments**

My SETI Institute colleagues Robert Ackermann, William Barott, Peter Backus, Michael Davis, John Dreher, Gerald Harp, Jane Jordan, Tom Kilsdonk, Seth Shostak, and Ken Smolek contributed significantly to the preparation of this chapter, and are hereby acknowledged as co-authors. This work has been supported in part by NSF grant AST-0540599, and by generous donations from the Paul G. Allen Family Foundation, Nathan Myhrvold, Greg Papadopoulos, Xilinx Corporation, and many other individual and corporate sponsors.

This chapter is based on a presentation at the International Academy of Astronautics 37th Symposium on the Search for Extraterrestrial Intelligence (paper A4.1.03) during the International Astronautical Congress in Glasgow, Scotland, October 1, 2008.

## References

Backus, P.R. 2004 Project Phoenix: A Summary of SETI Observations and Results, 1995–2004. *Bulletin of the American Astronomical Society*, **36**, 805.

Cullers, D.K. and Stauduhar, Richard P. 1997. Follow-up Detection in Project Phoenix. Conference Paper, Astronomical and Biochemical Origins and the Search for Life in the Universe, IAU Colloquium 161, 645.

Mandell, Avi M., Raymond, Sean N. and Sigurdsson, Steinn. 2007. Formation of Earth-like Planets During and After Giant Planet Migration. *The Astrophysical Journal*, **660: 1**, 823-844.

Turnbull, Margaret C. and Tarter, Jill C. 2003a. Target Selection for SETI. I. A Catalog of Nearby Habitable Stellar Systems. *The Astrophysical Journal Supplement Series*, 1**45: 1**, 181-198.

Turnbull, Margaret C. and Tarter, Jill C. 2003b. Target Selection for SETI. II. Tycho-2 Dwarfs, Old Open Clusters, and the Nearest 100 Stars. *The Astrophysical Journal Supplement Series*, **149: 2**, 423-436.

# 10

# Optical SETI: Moving Toward the Light

Monte Ross and Stuart Kingsley

## 10.1 Introduction

In 2009, the SETI community celebrated a half-century since the classic paper by Giuseppe Cocconi and Philip Morrison in Nature, that described how we might look for radio transmissions from extraterrestrial civilizations.[1] It is propitious that the publication of this book in 2010 marks both the 50th anniversary of Frank Drake's Project Ozma, and the 50th anniversary of the demonstration of the first (ruby) laser by Theodore Maiman.[2] The invention of the laser was based on the maser work by Arthur Schawlow and Charles Townes[3] and the simultaneous work of Gordon Gould.[4] During this first half-century of SETI, most observing has been carried out in the radio spectrum, during which time there have been enormous developments in laser technology. Only during the past two decades has the optical approach to SETI, otherwise known as optical SETI, been given the attention it deserves. In 1961, a year following the invention of the laser, Robert Schwartz and

1 Cocconi, G., and Morrison, P. (1959). Searching for Interstellar Communications, Nature, Vol. 184, No. 4690, pp. 844–846, http://www.coseti.org/morris_0.htm

2 Maiman, T. (1960)., Stimulated Optical Radiation in Ruby, *Nature*, Vol. 187, No. 4736, pp. 493–494, http://www.coseti.org/maiman.htm

3 Schawlow, A. L. and Townes, C.H. (1958). Infrared and Optical Masers, Physical Review, Vol. 112, pp. 1940–1949, http://www.coseti.org/schawlow.htm

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

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_10, 

Charles Townes published a paper in Nature which explained the potential for continuous-wave CO2 lasers for extraterrestrial communications in the infra-red.5,6 It is of note that the same Charles Townes of Nobel Laureate maser fame has been actively involved in optical SETI for many years.7 This chapter discusses some of the history of optical SETI, what is presently being done, and what we might do in the future, both on earth and on space-based observatories.

SETI has been of interest to one of the authors, Monte Ross (MR) since 1965, soon after the birth of the laser in 1960.[8] However, little was done while attempts to detect signals concentrated on radio frequencies. Interest in optical SETI was kept alive by a few people, notably Ross, Shvartsman,[9] Beskin,[10] Betz,[11] Kingsley, Lemarchand and Bhathal, until the lack of success at radio frequencies forced reconsideration in 1998.[12] Monte Ross pointed out in 1965 that laser signals would be best sent by short pulses, since this would allow a modest transmitter to readily overcome the brightness of the host star.[13] The other author, Stuart Kingsley (SK), through The Columbus Optical SETI website[14] and three international optical SETI conferences arranged by SPIE in 1993, 1996 and 2001 respectively, did much to re-ignite

---

5 Schwartz, R.N. and Townes, C.H.(1961). Interstellar and Interplanetary Communication by Optical Maser, Nature, 190: 205, http://www.coseti.org/townes_0.htm

6 Cameron, A.G.W. (ed.) (1963). Interstellar Communications, W. A. Benjamin, 1963.

7 Townes, C. H. (2001). Reflections on Forty Years of Optical SETI – Looking Forward and Looking Backward, in Kingsley, S. and Bhathal, R. (eds) The Search for Extraterrestrial Intelligence in the Optical Spectrum III, Proceedings SPIE – The International Society for Optical Engineering, Vol. 4273, pp. xix–xxii.

8 Ross, M. (2006). New Search for Extraterrestrial Intelligence, IEEE Spectrum, November.

9 Shvartsman, V.F. (1977). Communications of the Special Astrophysical Observatory, No. 19, pp. 5–39.

10 Beskin, G.M. (1991). Results of Searches for Optical Signals of Extraterrestrial Intelligence, USA-USSR Joint Conference On The Search for Extraterrestrial Intelligent Life, University of California, Santa Cruz, August 5–9.

11 Betz, A. (1991). A Search for IR Laser Signals, USA-USSR Joint Conference On The Search For Extraterrestrial Intelligent Life, University of California, Santa Cruz, August 5–9.

12 Naeye R. (1992). SETI at the Crossroads, Sky & Telescope, November 1992.

13 Ross, M. (1965). Search Laser Receivers for Interstellar Communications, Proceedings IEEE, Vol. 53, No. 11, http://www.coseti.org/ross_02.htm

14 COSETI website, www.coseti.org

interest in the optical approach to SETI.[115,16,17] Today, there is a serious effort at Harvard devoted to pulsed laser signal detection, but all the efforts to date at all wavelengths are still quite limited in relation to what can be done.

## 10.2 The Benefits of Pulsed Laser Optical SETI

An argument can be made that we are more likely to find alien signals in the optical spectrum than at either radio frequencies or microwaves. For one thing, it is easier to deal with noise at optical wavelengths. Attempting SETI at radio frequencies means contending not only with interference from terrestrial sources such as radars and radio stations but also with cosmic sources, including the background left over from the Big Bang. And, of course, there is noise intrinsic to the receiver itself. Although the most sensitive detectors are cooled to almost absolute zero to minimize internal noise, this cannot be eliminated entirely. The only significant terrestrial source of interference for optical SETI is lightning, which is at worst a sporadic problem with a very low probability. In the early days, many investigators dismissed optical SETI, believing that the sender's star would be an overwhelming source of noise. But they did not appreciate that if a short-pulse laser is used instead of a continuous one then it is possible to outshine a star during the time the pulse transmitter is 'on'. With a short-pulse laser, both spectral and temporal discrimination in the receiver can be readily attained since a laser shines at a single wavelength whereas a star shines in a broad spectrum, which enables the laser receiver to reject much of the spectrum but still have a much wider spectral acceptance than microwave receivers. The laser can deliver very large peak powers for brief intervals. And because direct detection for SETI at optical wavelengths is not obliged to preserve the coherence of the signal at the detector, large collectors can be made more cheaply than those of the same size which must preserve the phase of the signal across the face of the collector in order to produce an image. In other words, optical SETI receivers do not require expensive, diffraction-limited optics and lower cost 'photon-buckets' are preferable.

---

[15] Kingsley, S.A. (ed.) (1993). The Search for Extraterrestrial Intelligence (SETI), in the Optical Spectrum, Proceedings SPIE _ The International Society for Optical Engineering, Vol. 1867, www.coseti.org/spiepro1.htm

[16] Kingsley, S.A., and Lemarchand, G.A. (eds) (1996). The Search for Extraterrestrial Intelligence (SETI), in the Optical Spectrum II, Proceedings SPIE – The International Society for Optical Engineering, Vol. 2704, www.coseti.org/spiepro2.htm

[17] Kingsley, S.A. and Bhathal, R. (eds) (2001). The Search for Extraterrestrial Intelligence in the Optical Spectrum III. Proceedings SPIE _ The International Society for Optical Engineering, Vol. 4273, www.coseti.org/spiepro3.htm

It has sometimes been presumed that optical SETI will employ the same method of detection as for radio frequencies, namely heterodyne or coherent detection, whereas it is actually better to use direct detection in which the phase of the carrier frequency is discarded and only the amplitude is detected. Coherent detection is essential for an application such as high-resolution spectroscopy, however, because a laser is used as a local oscillator and mixed with the signal to create a microwave frequency which is analyzed in the same manner as a radio- telescope signal. Owing to the differences in noise at radio frequencies and at optical wavelengths, direct detection can achieve a sensitivity approaching that of heterodyne detection without the complexities and limitations of coherent detection. The type of noise which limits an optical system is the uncertainty inherent in detecting the signal, which is denoted as ‘quantum noise’ or ‘photon noise’. The major external source of noise is the light of the target star, but this is minimized by seeking nanosecond-like pulses which provide a high peak power at the source without imposing an excessive average power rating on the laser. Mankind has produced picosecond-type pulses with a peak power as high as a petawatt, which is $10^{15}$ watts. We have also made lasers with several megawatts of average power. Hence, it is reasonable to expect that an alien civilization will be able to produce laser signals consisting of high-peak-power brief pulses at a reasonable pulse rate. The number of photons received in a short pulse can far outshine the natural light from the target star during the time of a pulse by many orders of magnitude. In its simplest form,

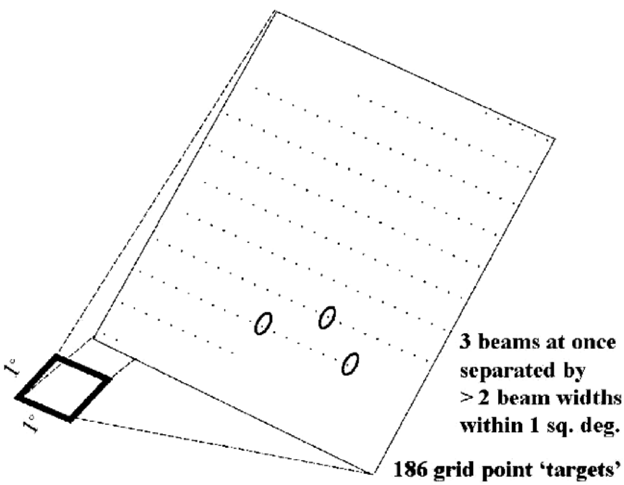



**Figure 10.1** Coincident pulses using two detectors to reject false signals.

direct detection consists of a telescope with a photodetector. If a second photodetector is added for coincident detection and the measurement time is short then it will prevent internal noise from triggering false detections. The difficulties of coherent detection include knowing the signal frequency and handling Doppler shifts, but for direct detection, a photodetector such as the photocathode of a photomultiplier has a spectral response broader than any Doppler shift likely to be encountered in SETI. The 1-GHz shift resulting from a radial velocity of 1 km per second would have no impact on a photodetector, but a heterodyne detector would have to perform the difficult task of tracking this shift and eliminating it. Also, whereas the interstellar medium causes radio frequencies to suffer dispersion, this effect is negligible at optical wavelengths and eliminates the requirement to reconstruct the signal. Although starlight suffers interference in the Earth's atmosphere, this is minor on the 1-ns time scale. It can be argued that only optical frequencies can overcome the severe data-rate limitations produced by interstellar dispersion present at microwave frequencies, so transmitting the 'encyclopedia galactica' in a short period of time is only possible at optical frequencies.

It is important to recognize that because SETI is not an imaging task the optics are not required to be state-of-the-art. The laser receiver requires only that any signal in the field of view reaches the detector, that the 'blur circle' at the focus of the optics lies wholly in the detector aperture, and that the focal ratio of the optics give a field-of-view sufficiently narrow that when the target star is centered there are no other stars to significantly illuminate the detector. And, of course, the collection area should be as large as possible within the budget. It turns out that the optical performance suited to this application is much less costly than for an astronomical telescope. A 0.6-meter-diameter mirror made to SETI specifications will cost many times less than one of the same size required to produce a pin-sharp image.

The light is reflected by the primary mirror to a smaller secondary mirror that sends it to the focal point and the detector assembly, which for SETI is really two parallel channels of photo-detection, amplification and threshold detection. The outputs of the threshold detectors are sent to a coincident detector that responds only if both channels 'see' a pulse at the same time. This scheme is illustrated in Figure 10.1. This strategy prevents internal noise events in each channel from generating false detections. An optical filter in front of the beam-splitter may be used to help reduce the background light. Although one such system is unlikely to yield results, 10,000 systems distributed around the globe – all aimed at the same time at a particular target star in order to achieve a large effective collector area – should be able to detect with great sensitivity any signal from the target star.

Hence, a 1-meter-diameter telescope will get 200,000 photons per second from a solar-type star at a distance of 1000 light-years, and for a quantum efficiency of 0.2 over the visible spectrum this equates to $4 \times 10^{-5}$ background photoelectrons in each nanosecond. As low-probability random events, the arrival of the photons will be in accordance with Poisson statistics. With a threshold set at five or so photoelectrons, the likelihood of receiving five background photoelectrons in the same nanosecond with such a low background rate is infinitesimal. The most likely source of noise is internal to the detector, which is why the receiving beam is split into two paths and a second photodetector added for coincident detection. We can set the threshold at two photoelectrons in each detector. The likelihood of two noise pulses occurring in the same nanosecond is given by:

$$R = (r_1^2 t)\ (r_2^2 t)\ t$$

where $r_1$ and $r_2$ are the arrival rates and t is the coincident window. If the rates are $10^5$ photoelectrons per second and t is 1 nanosecond, then the rate of false detection from the background is $10^{-7}$ per second, or a little less than once per year. If desired, this can be reduced much further by adding a third detector. The point is that with so few background photoelectrons arriving in each nanosecond, there is an exceedingly small likelihood of there being two events in each of two detectors during the same nanosecond. To express this another way, consider that since there are many fewer photoelectrons per second than there are nanoseconds in a second, the likelihood of a random event occurring in a particular nanosecond is quite low, and the odds of a second random event occurring during the same nanosecond are extremely low. Of course, the larger the collection area of any single telescope the greater will be the background rate, and at some point a third detector may be needed to eliminate false signals. If the same area were achieved by arraying smaller telescopes, then the need for a third detector is reduced. A signal level of 5-10 photoelectrons per pulse would avoid false detections and yet guarantee that a signal detection would not be missed. This raises the issue of the receiver's size. All we can do is set a limit. The example was for a solar-type star at a range of 1000 light-years, which is the maximum reasonable distance for SETI. Although stars which are closer will contribute more light to the background, for a given laser power the reduced distance will allow more signal to be detected. This would not be so for more luminous stars such as spectral classes O, B and A, but these appear to be inconsistent with the requirements for the development of advanced life. We can therefore be reasonably certain that light from the parent star of a civilization will not impede our detection of a short-pulse laser signal.

The development of laser communication for submarines, aircraft and military satellites proved that short bursts of laser light are far more efficient than continuous waves at carrying information. Although each pulse has a high peak power, the laser is inactive for most of the time and therefore has a low total power consumption. It is reasonable to believe that an alien civilization must have figured this out as well. With transmissions in brief bursts, each pulse could readily outshine any star in the field of view of the collector. What is more, the shorter the pulse, the less background light there is per pulse to compete with the signal. Reducing the pulse to nanosecond intervals makes any signal detected even more obviously of artificial origin, as such short flashes are unlikely to occur naturally. (SETI faced this dilemma in 1967 when a radio source was discovered to be 'ticking' with the regularity of an atomic clock. Until it was realized to be a rapidly rotating neutron star, and natural, the signal was labeled LGM-1, with the acronym standing for Little Green Men.) There is little to be gained from reducing the duration of the laser pulse below 1 ns, because (1) electronics and detectors function well down to nanosecond levels, but they have difficulty at much shorter times; (2) the optical background is already quite low at nanosecond intervals, with typically less than one background photo-electron per interval; and (3) multiple paths through the atmosphere will spread pulses out and make it more difficult for the system to function with proper accuracy.

Another reason to prefer optical methods over radio SETI is that it is much easier to produce a narrow beam. In crossing interstellar space, a signal will travel many trillions of kilometers. If the sender were to broadcast in all directions simultaneously, i.e., omnidirectionally, then the power required would be prohibitive at any wavelength. George Swenson of the University of Illinois at Urbana-Champaign has calculated that if a radio transmitter were 100 light years away and radiated its energy in this manner, it would require 5800 trillion watts to provide us with a detectable signal; an amount

| | OPTICAL | | INFRARED | | MICROWAVE | |
|---|---|---|---|---|---|---|
| PARAMETER | A | B | A | B | A | B |
| Wavelength | 1.06 μm | 1.06 μm | 10.6 μm | 10.6 μm | 3 cm | 3 cm |
| TRANSMITTER | | | | | | |
| Antenna Diameter | *22.5 cm* | *22.5 cm* | 2.25 m | 2.25 m | 100 m | 3 km |
| No. of Elements | 1 | 1 | 1 | 1 | 1 | 900 |
| Element Diameter | 22.5 cm | 22.5 cm | 2.25 m | 2.25 m | 100 m | 100 m |
| Antenna Gain | $4.4 \times 10^{11}$ | $4.4 \times 10^{11}$ | $4.4 \times 10^{11}$ | $4.4 \times 10^{11}$ | $1.1 \times 10^{8}$ | $9.8 \times 10^{10}$ |

**Figure 10.2** Table taken from page 50, July 1973 revised edition (CR 114445) of the Project Cyclops design study of a system for detecting extraterrestrial life. This study was prepared under the Stanford/NASA/Ames Research Center 1971 summer facility fellowship program in engineering systems design.

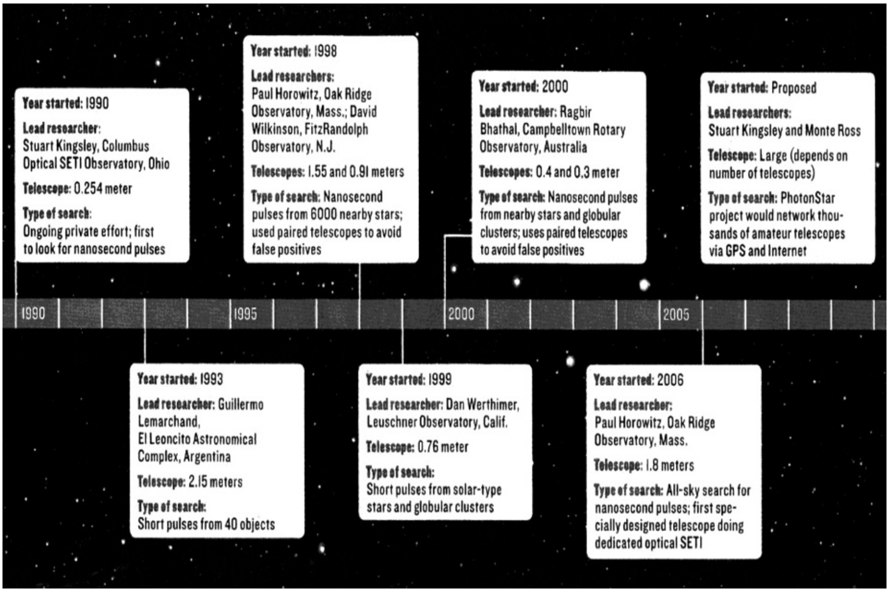



**Figure 10.3** Optical SETI through the years.

which, Swensen points out, is more than 7000 times the total electricity-generating capacity of the USA. And there is little chance of there being a communicating civilization as close as 100 light years. Thus, directionality of the beam is essential. In general, the narrower the beam the better, but in targeting a particular star system it should be confined in order to deliver all of the energy inside the radius of the star's habitable zone. A beam that narrow can only be achieved at short wavelengths below, at or near the visible regime. The attainable beamwidth is approximately the wavelength that is being transmitted divided by the diameter of the transmitter's antenna. The wavelength of light is six orders of magnitude shorter than it is for microwaves. So in a targeted approach, the physics of beamwidth supports the use of lasers rather than radio frequencies.

## 10.3 A Brief History of Optical SETI

Although Charles Townes and Robert Schwartz first suggested the idea of searching for optical signals from extraterrestrials in a paper in the journal *Nature* in 1961, the fact that laser technology was not nearly as mature as radio-frequency technology meant it was several decades before optical

SETI caught on.[18] Now microwave detectors have almost reached their fundamental limits in terms of noise, but there is significant scope for improving laser systems. Another reason for the delay in pursuing optical SETI was the insistence on comparing radio frequencies with continuous-wave lasers, a preference for small laser and receiver apertures, and a study which showed radio frequencies to be more appropriate. This Project Cyclops report was published a little over a decade after the invention of the laser. Copies are still available from the SETI League at www.setileague.org.. Figure 10.2 shows a table from the report which well illustrates the bias. The maximum collector diameter for a laser signal was only 22.5 centimeters, whereas the antennas for radio frequencies were up to 3 km in diameter. In addition, *Cyclops* did not consider short-pulse lasers, even though there was literature available - including a paper by one of the authors (MR) published in a leading journal in 1965. The inherent assumption that aliens could not make use of the very narrow beams produced by large optical apertures is the reason why the optical transmitter antenna diameters were so small. This very poor assumption helped 'cripple' the optical link efficiency compared to its microwave counterpart. Further information about this may be found at www.coseti.org/cyclops.htm. The history of terrestrial electromagnetic-wave SETI these past 50 years would likely have been very different but for the *Cyclops* report's poor comparisons between the efficacy of the two approaches.

Although Charles Townes and Robert Schwartz first suggested the idea of searching for optical signals from extraterrestrials in a paper in the journal *Nature* in 1961, the fact that laser technology was not nearly as mature as radio-frequency technology meant it was several decades before optical SETI caught on.[19] Now microwave detectors have almost reached their fundamental limits in terms of noise, but there is significant scope for improving laser systems. Another reason for the delay in pursuing optical SETI was the insistence on comparing radio frequencies with continuous-wave lasers, a preference for small laser and receiver apertures, and a study which showed radio frequencies to be more appropriate. This Project Cyclops report was published a little over a decade after the invention of the laser. Copies are still available from the SETI League at www.setileague.org.

In 1998, SETI investigators with resources began to look for nanosecond pulses. Paul Horowitz of Harvard built a nanosecond detection system that

[18] Oliver, B.M. (1962). Some Potentialities of Optical Masers, Proceedings Institute Radio Engineers, 50: 135.

[19] Oliver, B.M. (1962). Some Potentialities of Optical Masers, Proceedings Institute Radio Engineers, 50: 135.

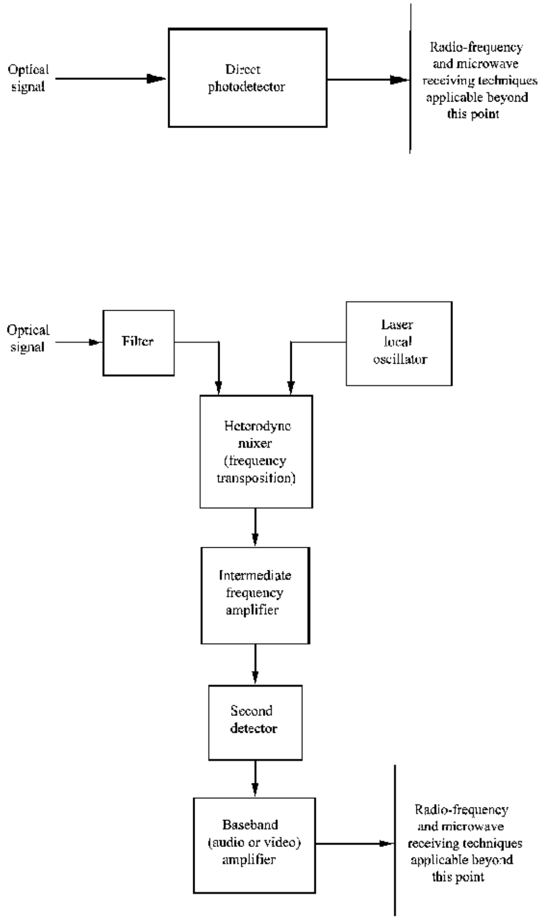



**Figure 10.4** Optical SETI by direct detection (top) and heterodyne (photomixing) detection (bottom). A direct detection receiver is obviously simpler.

used a telescope at the nearby Oak Ridge Observatory in Massachusetts.20,21 He and his team operated this system intermittently for several years. Later, David Wilkinson at the Princeton Fitz-Randolph Observatory studied the

[20] Horowitz, P., Coldwell, C., et al. (2001). Targeted and All-Sky Search for Nanosecond Optical Pulses at Harvard-Smithsonian, in Proceedings of the Search for Extra-terrestrial Intelligence in the Optical Spectrum, Kingsley, S. (ed.), SPIE - The International Society for Optical Engineering, Vol. 4273, pp. 119–127.

[21] Howard, A., et al. (2003). All-Sky Optical SETI, in Proceedings 54th International Astronautical Congress, Bremen, Germany.

same star at the same time in order to aid in distinguishing a real signal from a false alarm. The Harvard team used a 1.5-meter reflector, and about one-third of the light was deflected into the SETI receiver. The Princeton team's telescope was only 0.9 meters in diameter but they used almost all of the collected light. The two installations therefore had a similar amount of light available. Harvard passed the incoming light through a beamsplitter on to two Hybrid Avalanche Photodiodes (APDs) whose outputs fed high-level discriminators which had levels corresponding to roughly 3, 6, 12 and 24 photoelectrons. Approximate waveforms could be recorded to a precision of 0.6 nanoseconds by time-stamping level crossings. Coincident pulses triggered the microcontroller to record the arrival time and waveform profiles. A 'hot event veto' eliminated a class of high-amplitude bipolarity signals apparently due to breakdown effects in the APDs. Pulse counters and miscellaneous electronics allowed test apparatus to confirm proper operation. In particular, fiber LEDs were used to test the coincidence electronics prior to a search. The receivers were set to a sensitivity of 3 photoelectrons in each 5-nanosecond window. Taking into account the optical system losses and quantum efficiency, the receiver sensitivity signal level for the Harvard system was estimated at 100 photons per square meter and for the Princeton system at about 80 photons per square meter. These numbers apply to the visible spectrum between 450 and 650 nanometers, for which the quantum efficiency is about 20 per cent; outside this range it is closer to 1 per cent. During 2378 hours, between October 1998 to November 2003, a total of 15,897 observations were made of 6176 stars, typically for periods of between

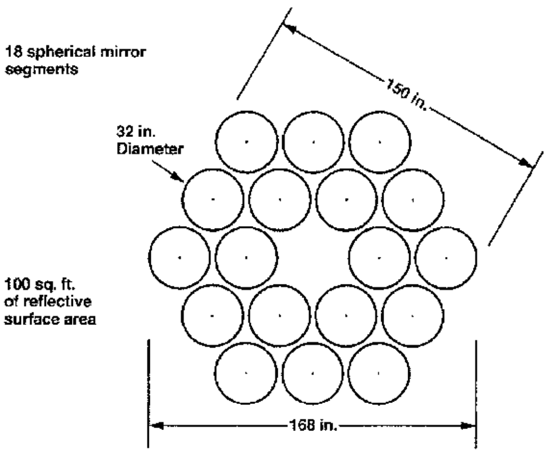



**Figure 10.5** A segmented primary mirror for an optical telescope.

2 and 40 minutes per session. The targets included all the main sequence dwarf stars between spectral classes A and early M within a radius of 100 parsecs and, owing to physical limitations, at celestial declinations between -20 and +60 degrees. A number of false detections occurred, including with the domes closed. In some cases these were equipment problems, but most appeared to have natural explanations, including the possibility of strikes by cosmic-ray muons. A lot was learned about the practicalities of doing optical SETI. In particular, a second observatory coupled with precise event timing completely eliminated background events and provided for the strongest possible confirmation of a signal. The lessons learned were applied in designing the dedicated all-sky system discussed later in this chapter.

A large number of small telescopes scattered geographically, as in the PhotonStar concept described later, will have no background problem. The Harvard and Princeton experiments showed that making the instrument robust against many failure modes with reliable software, good telemetry, and optical fibers for lightning protection is key for nearly automatic operation with low maintenance. The need for diagnostic checks prior to each observation is self-evident. In an experiment seeking a rare result, it is essential that the apparatus operate at its best during an observation. End-to-end testing with a source in the far-field is highly desired.

In 1999, Dan Wertheimer of the University of California at Berkeley began to look for optical signals using the 0.76-meter telescope of the Leuschner Observatory in California. In 2000, Ragbir Bhathal in Australia resumed the investigation of the southern hemisphere. He designed a system to detect short pulses and used a pair of telescopes to avoid false alarms. His telescopes were 0.3 and 0.4 meters in diameter, and were located at the Campbelltown Rotary Observatory.[17]

One can see from this summary that optical SETI has attracted far less resources than were made available to radio-frequency research. The dedicated all-sky system developed by Harvard represents a leap forward in experimental investigations, but its collector area is far less than could be achieved relatively cheaply by the ‘photon bucket’ approach. Figure 10.4 is a general block diagram of an optical receiver. The two basic methods are either direct detection or heterodyning (photomixing). As is evident from Figure 10.4, heterodyning is far more complex. As noted earlier, direct detection is the technique of choice for optical SETI not only because it is simpler and avoids many problems, but also because it can come close to the performance of an optimal heterodyne system. To summarize: direct detection eliminates the need to know the exact wavelength, does not require all parts of the received signal to be in phase, and is insensitive to Doppler shifts. The low background due to examining time slots of nanosecond duration allows the use of large collectors. The efficiency is improved by eliminating narrowband optical filters. All that is required of the optics is that the field of view be narrow enough to exclude other stars as bright as the target star, and that the light be focused on a spot no larger than the input of the photodetector. The fact that the phase information is not needed makes it possible to use less accurately figured optics that are considerably cheaper than optics of comparable size intended for imaging. In particular, large-segment low-cost mirrors can be used without the need for adaptive real-time alignment correction to eliminate the fluctuations caused by the light's passage through the Earth's atmosphere.

By way of example, Figure 10.5 shows a 4.2 meter-diameter optical system capable of being steered at speeds sufficient to take only a few seconds to slew from one star to the next. The main mirror is made up of 18 hexagonal segments, each approximately 0.8 meters across. The design is based on three assumptions. The first assumption is that the sender will aim its signals only at star systems that are likely candidates for hosting intelligence (with us included on their list). The second assumption is that a signal will be sent in such a way as to prevent accidental discovery by others. Only when one star system at a time is encompassed by the beam

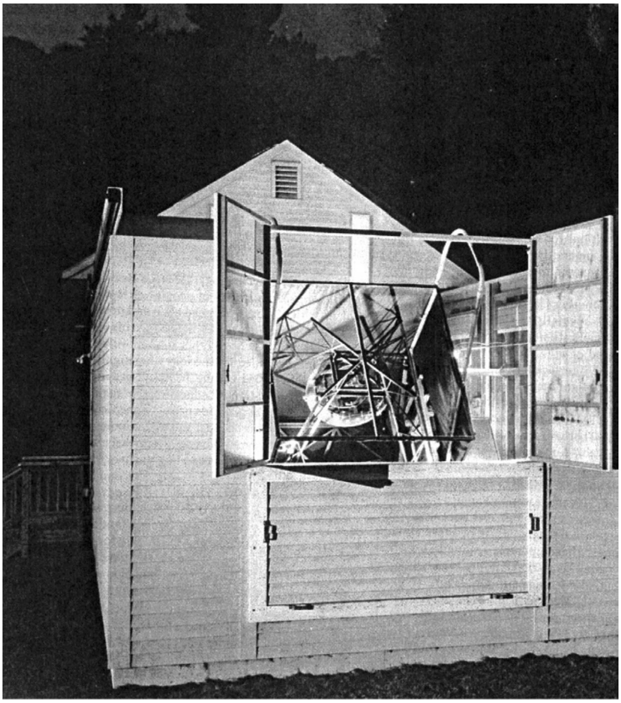


**Figure 10.6** Harvard's all-sky search telescope for optical SETI.

can accidental discovery be minimized. The broad beams typical of radio-frequency signals would inevitably cover many star systems simultaneously. Although sub-microradian radio-frequency beams could be achieved using enormous antennas, why would the sender proceed that way when a small-diameter optical system could readily achieve the same purpose? The strength of the laser can be tailored so that when the signal reaches the target it will be at the level consistent with the expected capability of a society possessing the technology to make a detection. The third assumption is that there is no need to transmit on a continuous basis to a single star system, so the signal will be offered to the receiver only for part of the time. The sender will slew the laser sequentially from one target star to the next, covering all of the likely candidates in a tiny fraction of the time it takes for the signal to travel through space. This enables a single time-multiplexed laser to cover many hundreds of candidates. (There is no need to build a separate laser for each one.) By applying these assumptions, we arrive at an optical receiver design that satisfies the requirements within our technological capability, but has not yet been implemented. It has four major features:

- The receiving apparatus must be of at least a certain size, since otherwise the signal will go undetected
- The receiving apparatus can rapidly slew from one star to another to inspect all likely sources in a time period during which a sender, when aiming a laser at us would (if existing) send a signal
- A short-pulse detection system enables the signal to outshine the host star.
- It must be dedicated to this task so signals (if present) are not missed.

No existing laser receiver satisfies these requirements. The closest match, the new Harvard system, meets two requirements but falls short in terms of receiver size and the ability to slew rapidly.

Non-imaging telescopes that have large collectors include the old Mount Hopkins 10-meter-diameter instrument in Arizona originally utilized for Cerenkov radiation studies, and the McDonald Observatory in Texas which has an 11-meter-diameter collector that consists of 91 segmented mirrors. The segmented primary mirror in Figure 10.5 provides a collection area of 9.2 square meters. As the targeted stars will be only a few degrees apart, rapid slewing is possible by electro-mechanical means. Our technology is already sufficient to build such a collector. A 10.4-meter-diameter mirror consisting of 36 hexagonal segments, each of which has its own computer-controlled actuators for maintaining the mirror's shape as it turns to track stars, was inaugurated in 2007 on the island of La Palma in the Canaries. To facilitate imaging, this aptly named Great Canary Telescope provides much higher resolution than is necessary for SETI. It should retain the record for

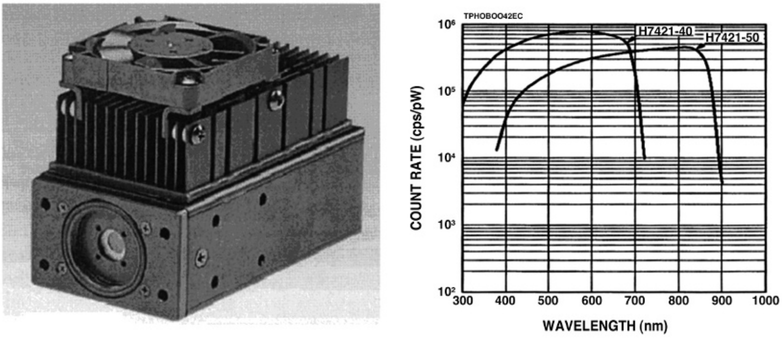



**Figure 10.7** Response versus wavelength for a Hamamatsu H7421 photomultiplier photon counter.

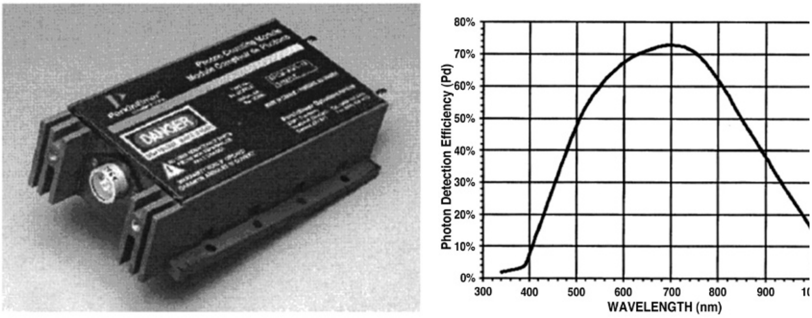



**Figure 10.8**. Response versus wavelength for a Perkin-Elmer solid-state avalanche photodiode single-photon counter employing Geiger mode detection.

size until the European Extremely Large Telescope enters service in 2017 with a mirror 42 meters in diameter made up of 984 hexagonal segments. Perhaps we will require optics of such size to detect an alien laser, but in the meantime, we will have to make the best of the new Harvard system.

### 10.3.1 All-sky optical SETI at Harvard

Optical SETI entered a more mature era with the introduction in 2006 of a dedicated facility with which to detect nanosecond pulses.[22] This telescope and signal detection and processing capability were designed by a team at Harvard led by Paul Horowitz and built at the Oak Ridge Observatory in Massachusetts. This dedicated observatory is supported by The Planetary Society. Significantly, this system is designed to search by scanning the entire sky rather than targeting individual stars. It exploits advancements in optical and laser technologies over the past two decades; for example, it uses improved quantum efficiency photodetectors with nanosecond response times. A key feature of the system is the use of multi-pixel photomultiplier tubes to examine more of the sky simultaneously. The primary mirror is 1.8 meters in diameter and the secondary mirror is 0.9 meters in diameter (Figure 10.6). The non-imaging optics (that is, a 'photon bucket') were cheaper than a comparable imaging system. The telescope is designed to cover the northern sky in 150 days. However, this assumes that a laser is pointing in our direction continuously in a pulse mode. If, as is more likely, the laser points at us only intermittently as it works through a list of possibly thousands of stars as potential hosts of civilizations, then the probability that it will be sending to the Sun at the same time as we look at the host is miniscule. Nevertheless, the increase in capability from sporadically examining a small number of stars to conducting an all-sky survey is a significant step towards a serious search for alien signals.

The Harvard system divides the sky into patches of 0.2 × 1.6 degrees. It observes each for 48 seconds before moving to the next. The telescope is able to move only in declination, and as the Earth's rotation sweeps the field of view in right ascension the system scans a strip of declination on a continuous basis. Photomultipliers convert photons into electrons with very little added noise. A series of stages amplify the number electrons in a cascade until they exceed the electrical noise and are output. A multi-pixel photomultiplier divides the collection area into tiny squares, with each channel acting like a separate detector. In this case there are 64 squares per tube. This facilitates looking at more than one star system at a time. Because the electronics are looking independently at each nanosecond, an enormous amount of processing takes place during those 48 seconds. All the signals from each pixel are fed into 32 microprocessors custom-designed for the project by Horowitz's graduate student Andrew Howard. These chips crank through the data at a rate of 3.5 trillion bits per second in search of a large spike in the photon count that could be a laser pulse from space.

22 Howard, A., et al, 'Initial Results from Harvard All-sky Optical SETI,' Proceedings 57th International Astronautical Congress, Valencia, Spain, 2006.

The receiver's detectors are divided into two arrays so that if one array detects an interesting signal it can be checked against the other one. With this implementation, internal system noise is unlikely to cause a problem. If a pulse is detected in one channel, it will have to be detected in the same nanosecond in the second channel. Stray noise from outside the system could still cause a problem, but adding a second system some distance away which simultaneously examined the same piece of sky would enable random false detection events of this type to be discarded. Whilst two photodetector channels reduce false detections to perhaps one occurrence per night, adding a third channel reduces the rate to only one per year or so.

With the recent introduction of Avalanche Photodiodes (APDs) as alternatives to photomultipliers, APD arrays may substitute for photomultipliers, especially in the near-infrared in which the quantum efficiency of photomultipliers drops off rapidly whereas APDs still offer high quantum efficiency and gain with little added noise. Figures 10.7 and 10.8 show the responses of sensitive photomultipliers and APDs with wavelength. Photomultipliers have been made with 40 per cent quantum efficiency at green, but typically are less than 1 per cent at wavelengths of 1 μm in the near-infrared. The low photomultiplier quantum efficiency makes APDs attractive even if they add a small amount of noise by having a quantum efficiency of up to 80 per cent. Future improvements in solid-state detector arrays are expected to facilitate single-photon detection using an APD in the self-quenching Geiger mode, where it acts as a trigger. In this mode, single-photon detection events can improve sensitivity. Although it will also render the detector essentially inoperative for several nanoseconds, this should not pose a serious handicap to a system with a low duty cycle. An example would be a system of 1000 slots, with each slot of 1 ns contributing to a total duration of 1 μs. Reference pulses would be provided every 3 μs. The first microsecond after a reference pulse would not be used, nor the microsecond before the reference pulse. Since the data pulse can only happen in the middle microsecond, this prevents any problem arising from the Geiger mode operation. The Geiger mode detector recovers in much less than a microsecond when it is triggered by a nanosecond pulse.

### 10.3.2 Beamwidth and habitable zones

One of the benefits of a laser is that it can form a very narrow beam. This facilitates sending a beam just wide enough to cover the habitable zone of the target star, making the most efficient use of the energy. When considering using narrow beams to signal a star system, there are several issues which affect the pointing accuracy: pointing misalignment, pointing jitter, and point-ahead error. The first two represent minor imperfections in hardware and limit how narrow the beamwidth can usefully be. It is desirable that the combined effect of misalignment and jitter produces a pointing error no

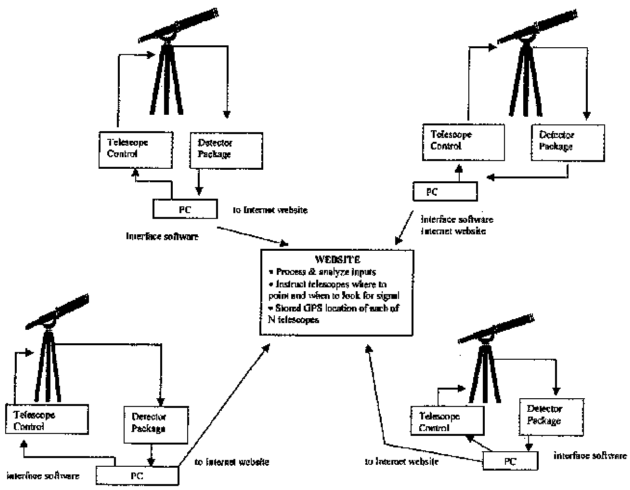



**Figure 10.9** Conceptual diagram of a multiple-telescope PhotonStar system.

greater than a small fraction of the beamwidth, say no more than 10 per cent. As the beamwidth gets narrower, this becomes harder to accomplish. The third issue is an error arising from imprecise knowledge, and could well be the largest uncertainty. It results from the fact that the targeted star is not only so distant that a signal traveling at the speed of light may require centuries to reach it, the star is also moving independently through space. If the laser were to be aimed precisely at the star, the star will have moved by the time the beam arrives. Both the speed and direction of a star must be known very accurately to calculate how much to offset the laser. Obviously, this depends on the relative positions of the two stars in the galaxy. The worst case would be where the point-ahead error was greater than the beamwidth, and the signal missed the target completely. Since we are discussing such small angles and rapidly moving targets at such great distances, serious errors can result. If each of the three factors were limited to 10 per cent uncertainty of the beamwidth, and all errors tended in the same direction then a 30 per cent increase in the beamwidth would be needed. This would translate (by squaring 1.3 to get 1.69) to a 69 per cent increase in laser power in order to maintain the number of photons per square meter. A plausible calculation of the point-ahead accuracy indicates that it must be 1 part in 100,000 to avoid an impact on the minimum useful beamwidth. We have considered links in which the minimum is as small as $10^{-7}$ radians. At 10 per cent, we need to know the point-ahead to less than $10^{-8}$ radians. At a distance of 100 light years, this is a displacement perpendicular to the line of sight of $4.5 \times 10^{-4}$ light years, which is large relative to the diameter of a habitable zone of a solar-type star. An accuracy of 1 per cent is preferable. Whilst this may be difficult to achieve, it should be feasible for an advanced civilization that will know much more about its stellar neighborhood than we do at the moment. The point-ahead angle will be different for each targeted star. However, this is a transmitter-only issue. In seeking a laser signal we look with a field of view of typically 1 milliradian, which renders insignificant a shift measured in microradians, and in any case, we see the light as it arrives - we do not need to make any allowance for how its point of origin has moved since the light began its journey; at least not until we attempt to reply, at which time we must make the corresponding calculation.

### 10.3.3 Weather and other issues

Although there would appear to be no celestial events which could be mistaken for a 1-nanosecond laser pulse, terrestrial weather poses a problem. Obviously the best site to place an optical receiver would be a location which has clear skies, but the situation improves if a number of telescopes are distributed geographically and work in a coordinated way.

At optical wavelengths there is a certain amount of loss arising from light being scattered and absorbed. Optical scattering is caused by interstellar dust grains. This is directionally dependent and can be significant over great distances, but is minor within the search radius considered for SETI. The effect on a laser pulse is to reduce the pulse height, and simultaneously yield delayed tails on two timescales: a close-in tail from forward scattering by large dust grains, and a much longer tail from diffuse scattering. The major leading edge retains its pulse shape, but having lost photons to scattering its amplitude is reduced. For signals sources 1000 light years away, the maximum expected reduction is 40 per cent. At slightly longer wavelengths than visible, such as near-infrared, this effect is reduced. Optical SETI by short-pulsed laser should therefore be perfectly viable within the radius we have selected for our search.

## 10.4 The Future

### 10.4.1 The PhotonStar project

The PhotonStar SETI project to detect alien laser signals involves a large number of small telescopes acting together in a geographically

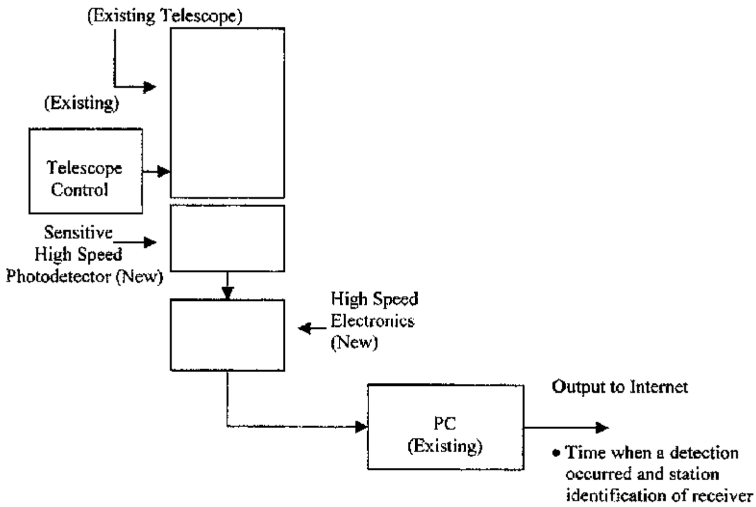



**Figure 10.10** Receiver station hardware configuration of the PhotonStar system.

dispersed array.[23] Figure 10.9 shows a conceptual view of the system. In acting together, looking at the same star system at the same time, the large overall collection area increases the chance of detecting a signal, should one be present. The position of each telescope can be located by GPS so that the differential distance from a given target star to each telescope can be determined to calculate timings, and the Internet used not only to coordinate the switching of one target to the next but also to send the data to a central station. This became feasible with the advent of relatively low-cost single-photon detector technology for use by amateurs.

The timing of the receivers can yield better than 10 nanoseconds accuracy – i.e., once the physical locations of the telescopes had been allowed for, if a pulse were to occur then every receiver that detected it would measure the time of detection as being within 10 nanoseconds of any other telescope in the array, which is sufficient. There are thousands of amateurs with telescopes of 20 cm or more in aperture. Arraying offers a total collection area greater than the largest optical telescope. The signal photon flux from the target is obviously the same irrespective of whether there is a single collector

[23] PhotonStar website: www.photonstar.org

with an area of 10,000 square meters or an array of 10,000 collectors, each providing an area of 1 square meter. If signal photons are received by the single large collector, scaling dictates that the same number of photons will be collected in total by the array. If the flux is low, then a few detectors of the array will still receive photons even if the majority receives nothing. Poisson statistics will apply, and some detectors must receive a photon such that the average number of photons being detected will be the same whichever architecture we use. We have no control over the incoming photon density. In this way, each detector sends its data to the central station via the Internet for analysis in real-time. Specific directions prior to observing will enable each telescope to look at the proper star and know the times at which to do so. Each receiver has its own set of coincidence detectors to reject internal noise results. Although it is unlikely that a pulse detected simultaneously by widely separated receivers is not extraterrestrial, it will be thoroughly examined to determine whether it could have been otherwise.

Using plausible numbers, if the sender is able to confine the beam to the habitable zone there will be roughly 0.01 photons per square meter. This photon flux density will ensure that at least one photon per pulse reaches the receiver. It is based on the sender expecting that the receiver collector will have an area of at least 100 square meters. But optical system losses and the quantum efficiency of the detector will require at least five photons per pulse to reach the receiver. Hence laser energies of the order of 1000 joules per pulse are needed (without considering optical path losses that could contribute a factor of two). For a rate of 10 pulses per second, the sender will need a laser with a modest average power of 20 kilowatts. This is based on the beam being confined to precisely match the habitable zone. The beam should not be made narrower than this, or it might miss the receiver. If it is wider, then either the power must be increased to compensate for the dilution of the energy or the recipient will make the receiver larger than the minimum of 100 square meters in recognition of the difficulty a sender may face in precisely pointing a beam which is tailored to the habitable zone of a solar-type star. What we can certainly conclude, is that we cannot reasonably expect a few meters of collection area to be sufficient. A telescope with a diameter of 30 cm has an area of about 0.07 square meters, so an array of 1000 such telescopes would provide an area of 70 square meters and 10,000 telescopes would provide 700 square meters. Given the impracticality of an enormous single collector, arraying would seem to be the way to go. The requirement for large collection areas raises two points. First, the energy-per-pulse requirements for a useful pulse rate are readily attainable and do not impose a serious burden on the transmitting society. They could have a number of systems, each of which is signaling to hundreds of stars in sequence. Second, a potential recipient must make a serious effort to detect such a signal, but this, too, is readily attainable.

| Large Telescopes | Area (meters$^2$) |
|---|---|
| 3 meter diameter | ~7.0 |
| 10 meter diameter | ~77 |
| 30 meter diameter | ~700 |
| **Small Telescopes** | |
| 12 inch diameter | ~0.07 |

**Figure 10.11** Photon collection area.

| N | Area (meters$^2$) |
|---|---|
| 100 | ~7 |
| 1,000 | ~77 |
| 10,000 | ~700 |

**Figure 10.12** Collection area of N small telescopes.

### 10.4.2 Minimum useful beamwidth

The minimum useful beamwidth is that required to precisely match the habitable zone of the target star. It will be different for stars of various spectral classes, and at differing distances from the sender. In the case of the Sun, the habitable zone has a diameter equivalent to approximately 30 minutes of light travel. If, as we expect, the sender requires us to make a serious effort, then the photon density will be such that the number of photons per pulse will be less than the number of square meters in the zone, and the photons collected will certainly be less than the number of receivers in an array. Whereas a single large collector may receive ten or so photons, in the array of small collectors with the same total area typically ten of the receivers will report single-photon detections to yield the same photon count. However, in the case of the array, each transmitted pulse will be detected by a different set of ten receivers. But the single large collector sees more background light than any of the collectors of an array by a factor which is the ratio of its collection area to that of any given smaller one. The background will therefore be less of an issue for smaller collection areas. A receiver which does not presume any knowledge of the laser's wavelength becomes more feasible.

### 10.4.3 Participation of many

PhotonStar enables the participation of both amateur and professional astronomers at the cost of a laser receiver per telescope. If the receivers were to be standardized and thousands made, then the cost would be relatively low. The requisite software can be downloaded from the Internet. The data output has to be packaged and sent via the Internet to a central station that looks at the data from the full array in real- time. The system offers the attractions of: (1) avoiding the necessity of building a large-optics multi-million dollar system to seek extraterrestrial pulsed lasers; (2) it enlists anyone who wishes to participate; (3) due to its geographical diversity it is less susceptible to weather constraints; and (4) it can grow as more receiving stations are added. But it was impractical prior to the advent of GPS, the Internet and single-photon detector technology. The laser receivers should be designed to be readily used in conjunction with existing computerized telescopes, such as those by Meade and Celestron. If the software is user-friendly, then no special knowledge will be needed to participate. It ought to be possible to enlist thousands of these individual systems and operate them

**Nanosecond pulse**

**Area of beam at 200 lys ~7 x $10^{23}$ meters squared**

**10 meter diameter collector for ~70 meters squared for receiver**

**Quantum efficiency at 0.5 micrometers (green) ~0.2**

**Assume 5 photoelectrons needed for detection.**

**$E_r$ = hfn ~4 x $10^{-19}$ joules per photon x 25 photons received per pulse = $10^{-17}$ joules per pulse**

**$E_t$ = transmitter energy per pulse = $E_r$ x $A_t$ / $A_r$ where $A_t$ / $A_r$ is $10^{21}$**

**Resulting in 10,000 joules per pulse** (=100,000,000,000,000 watts for 1 nanosecond)

**Figure 10.13** Laser energy per pulse requirements for a link at 200 light-years.

in a coordinated manner. The central station should be designed to enable the array to expand without disturbing the existing users. By knowing each small telescope's location precisely in terms of latitude, longitude and altitude, the central station will direct each telescope to enable them all to aim at the target star at the same time and work through its list of candidates.

If a pulse detection occurs, the time of detection is sent to the central station, which notes how many receivers detected it at the same time. For the expected weak signal flux of optical SETI, a small number of receivers in the array will detect pulses and most will not. However, the average total number of photons detected should be the same as for a single large collector with the same area as the array. Each time a weak pulse is detected, a different set of receivers will detect the pulse. Because long-term monitoring of a fixed site by GPS can determine its position to within a few centimeters, the difference in distance for each telescope to a location in space can be calculated. The range will vary over time, as the Earth's rotation causes the angle of viewing to change, so all of this must be tracked in real-time. Figure 4.10 shows the hardware configuration of a site, involving the telescope, the PC, and the single-photon detector. Each receiver has its own identification number to enable the central processor to determine which telescopes are reporting detections. The detectors are fast enough to distinguish 1-nanosecond pulses. At light-speed, 1 nanosecond is equivalent to a distance of 0.3 meters. If one receiver is 1500 meters further from the source than another, it will detect a pulse 5000 nanoseconds later. Once the geometries of the sites are normalized, the signal reception timings can be correlated to determine whether they were simultaneous. False pulse detection will occur with an array of small collectors, but the central processor should be able to sort out false signals through time of detection and lack of simultaneity.

### 10.4.4 Small telescope arrays versus one large collector

Figure 10.11 shows the collection area of a large-diameter telescope. Figure 10.12 shows a number of much smaller telescopes attaining an equivalent area. Arraying offers a number of benefits over a single large telescope. Obtaining time on large telescopes for SETI is always difficult. Whilst a dedicated system can be built, as Harvard has done, it involves significant initial cost and is subject to outages due to weather and essential maintenance. A diverse array of small telescopes makes use of an existing infrastructure of telescopes and PCs, is able to expand as stations are added without the need for reconfiguration, and is less susceptible to weather. With a large number of collectors tied to a worldwide network, new strategies can be used. In particular, instead of having the entire array work through the target list in concert, it would be possible to assign subgroups to particular stars and upon a detection being made the entire array swings onto that star to maximize the effort. The receiving stations can operate automatically, with the central station controlling where the telescope points and being alerted by any detections, so once set up they do not require the constant attention of the owner.

As the technology improves, the best choice of detector today may not be the best tomorrow, so an array might incorporate a variety of detectors. Today's commercial single-photon detectors are made by Hamamatsu and Perkin-Elmer and sold as Single Photon Detector packages. Candidates include the HPMT (Hybrid Photoelectron Multiplier Tube) and the solid-state Hybrid-APD. High-speed circuitry is needed in order to properly process the pulse output to precisely determine the time of signal detection. If the pulse is broadened owing to insufficient detector/electronics bandwidth, the accuracy of the time of detection will be impaired. This, in turn, will make it harder to avoid false alarms. The likelihood of false detections increases as the pulse width broadens. For example, if a receiver seeks 1-nanosecond pulses and the receiving capability can only resolve 10 nanoseconds, it is uncertain which of the nanosecond periods it actually occurred in. Any background photons that were in the 10 nanosecond period could falsely be counted as signal photons, as on the basis of time received the receiver was incapable of telling the difference. A large collection area enables a relatively low energy per pulse laser to be detected at a range of many light-years.

Figure 10.13 shows a laser of only 10,000 joules per pulse being detected at 200 light-years. For 10 pulses per second, the laser would need an average

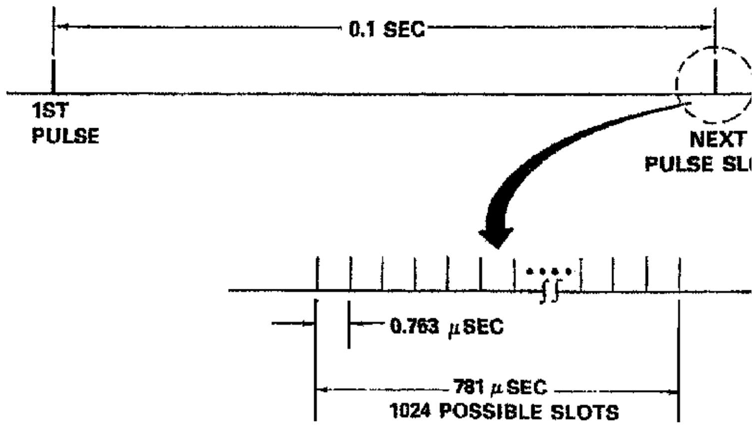
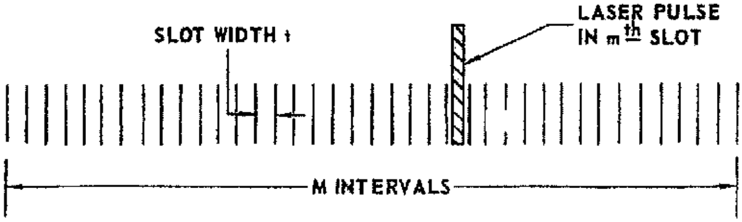



**Figure 10.14** Pulse interval modulation with a 'window' showing how a small part of the interpulse can be used to convey information.




| M INTERVALS | DUTY CYCLE 1/M | BITS/PULSE |
|---|---|---|
| $10^4$ | $10^{-4}$ | 13.3 |
| $10^5$ | $10^{-5}$ | 16.6 |
| $10^6$ | $10^{-6}$ | 20.0 |
| $10^7$ | $10^{-7}$ | 23.3 |
| $10^8$ | $10^{-8}$ | 26.6 |

$$\frac{\text{BITS}}{\text{PULSE}} = \text{LOG}_2 M$$

**Figure 10.15** A pulse interval modulation waveform as an example of a low-duty-cycle highly efficient signal design for laser communication. The table gives bits/pulse and duty cycle in terms of M possible pulse periods or intervals.

power of 100,000 watts. Signal detection is determined by a sufficient number of the stations detecting a photon at the same time. Let us say it is calculated from the spectrum of a star that in each 0.1 second interval this background should prompt two of 100 stations to detect a false pulse at the same time. Suddenly ten stations report a pulse, making this an event worthy of examination. If it is a true signal, it can be expected that a different set of about ten stations will report when the next pulse arrives. Other factors being equal, the stations that report ought to be random. If a given station is reporting detections in excess of the average then it is probably sending false alarms owing to excessive internal noise or is malfunctioning. The signal processing at the central station must examine every nanosecond period and determine whether, based on the data supplied by many receivers, there were real pulses. Fortunately, only the central station requires such computing power, the individual stations do very little processing and this makes the system feasible.

Although there is currently no effort under way to implement the PhotonStar array, it is likely that this approach will eventually be attempted owing to the fact that it is a viable way of achieving a large collection area for short-pulse signals. And, as SETI@home proved, the public is willing to participate directly in the search for extraterrestrial signals.

## 10.5 Other Parts of the Spectrum, Decoding the Data and Forming Pictures

We initiated our search for extraterrestrial intelligence as soon as we developed the means to do so, and it expands in pace with technological development. Although it was dispiriting not to have had an immediate result, it is likely that we will continue our efforts as long as our civilization exists. However, civilization is a fragile thing. It is possible, if unlikely, that we will be wiped out like the dinosaurs. A more likely scenario is that warfare, pandemics, pestilence or climate change causes civilization to fall into a new 'dark age' which lasts for hundreds of years. But if we are able to avoid such dangers, the search will continue, perhaps in fits and starts, until success is achieved. Other issues that may play a role in a successful SETI program include other parts of the electromagnetic spectrum, receiving stations off Earth, recovering the information in the pulse train, forming and interpreting pictures from such data, and methods that will resolve the transmitter/receiver directionality issue.

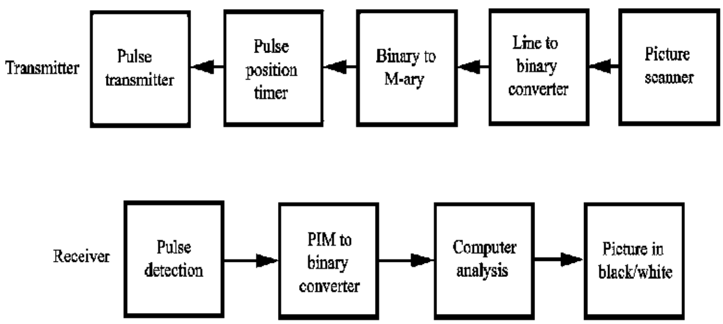



**Figure 10.16** Sending pictures using low-duty pulses. Computer analysis in the receiver determines the number of pixels per line and lines per frame.

### 10.5.1 Many bits per pulse

As stated earlier, it is desirable to send short laser pulses which can readily outshine a host star, for easier detection. As the pulse rate need not be high, the average power transmitted is reasonable. A low pulse rate facilitates sending data at many bits per pulse. In a standard continuous-wave communication system, each pulse or absence of a pulse represents a single bit, with an equal chance of it being a '1' or a '0'. But if an M-ary system is used instead of a binary system then it is possible to send more than one bit per pulse. The fact that physics at laser and infrared wavelengths favors short pulses and a low duty cycle lends itself naturally to an M-ary system. Consider a system in which M intervals or slots are present. Each slot can represent a unique number. If we restrict ourselves to precisely one pulse in the M intervals of time T, then each pulse represents log2 M bits. This type of system does not have to take up the full period between pulses, but can function when only a small percentage of the maximum time between pulses is utilized, as shown in Figure 10.14. If we send F pulses per second, then the product of F and log2 M is the bit rate expressed in bits per second. Figure 10.15 relates duty cycle, intervals, and bits per pulse. Stating the bits per pulse in terms of M intervals and base 2 logarithm simply specifies that a certain number of '1' and '0' bits will convey the same information.[24,25]

[24] Ross, M., Laser Receivers, Wiley, 1966.

[25] Ross, M. (ed), Laser Applications, Vol. 1, Academic Press, 1971. http://www.coseti.org/ross_03.htm

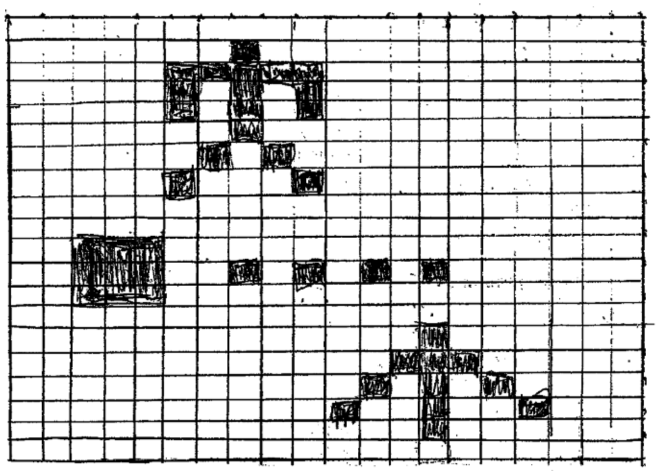


**Figure 10.17** A simple picture example using 20 × 20 pixels 20 M-ary pulses at 20 bits per pulse.

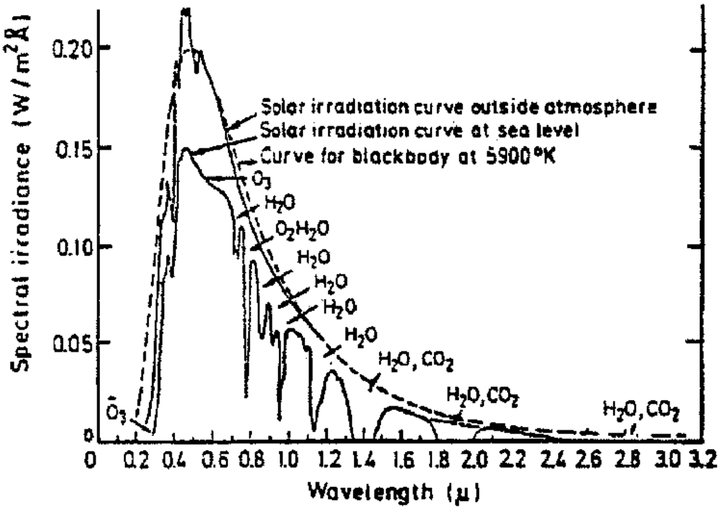



**Figure 10.18** The blackbody curve for the Sun ranging between the ultraviolet and the infrared, the actual solar flux impinging on the atmosphere, and a series of atmospheric absorption bands in the infrared.

If there are M choices for the placement of a single pulse, then as M increases it is possible to carry more information in that pulse. In a system which we designed, we would know M and the pulse rate. Here we do not know M, but the position and rate of detected pulses can be subjected to powerful computer analysis to determine whether it is an M-ary system and then estimate the value of M and the likely pulse time slot. Here we have assumed 1 nanosecond, but computer analysis can choose different time slots and then determine the best fit of the data based on the time of detection. And as the sender will desire the signal to be interpreted, there will undoubtedly be a reference pulse to enable the recipient to establish the timing and determine M. A pulse train which repeats at some measured rate constitutes a reference pulse for another pulse whose position with respect to the reference pulse changes each time. The pulse after the reference pulse can contain a number of bits owing to the many potential time slots in which the information pulse might occur. We may not know whether the transmitter is sending using nanosecond time slots, or longer time slots, but for a short pulse of nanoseconds or less it is more likely the former, as otherwise the communication efficiency is needlessly low. It is theoretically possible to operate without a reference pulse, but this makes analysis of the data much more difficult. As an example, if there is a reference pulse once every millisecond then there will be 1000 pulses per second, and an M of

1000 with nanosecond time slots for the information pulse allows 10 bits per information pulse. With 1000 pulses per second, we receive 10,000 bits per second. In such a system, the duty cycle of the transmitter remains low at $10^{-6}$. This could be exploited either to limit the average power required by the transmitter, or, more probably, to enable a single transmitter to send low-duty-cycle pulses to a large number of star systems.

### 10.5.2 Sending and receiving pictures with pulses

A low-duty-cycle pulse system can send pictures with many bits per pulse. This can be converted into a binary number to represent a line of pixels. A succession of lines constitute a black and white picture. 10. 16 shows this process. At the transmitter, the picture is scanned line by line, and each line represented by a series of '1's and '0's that are converted into an M-ary number which causes the pulse to be sent at the appropriate time. For 1-nanosecond time slots, the uncertainty of 1 million choices allows 20 bits per pulse. By sending 20 reference pulses and 20 information pulses at appropriate times it is possible to send a 20 × 20 picture. Interpretation of such a simple picture might be difficult. For example, Figure 10.17 could be interpreted as showing two life forms (perhaps male and female) from the fourth planet of the star system. Greater resolution on the part of the sender would be helpful. A case of greater resolution would be if we reconstruct smaller sub-areas for a 1000 × 1000-pixel view. Software can examine the detected pulse train and work out if it contains a picture, determine the dimensions and can reconstruct it. Of course, the meaning of the image may elude us! One million 1-nanosecond time slots equals 1 millisecond. This is the maximum time between reference pulses, and based on 1000 information pulses at 20 bits per pulse this puts an upper limit on the data rate of 20,000 bits per second. This could send a 1-megabit picture of 1000 lines with 1000 pixels per line in 50 seconds. To put this into context, 1 megabit per frame is about equivalent to a high-definition TV. This can be accomplished by a laser which operates with a duty cycle of only $2 \times 10^{-6}$ by

| Wavelength (micrometers) | UV improvement |
|---|---|
| 2.0 | 64 |
| 1.0 | 16 |
| 0.5 | 4 |

**Figure 10.19** Ultraviolet lasers improve antenna gain over other wavelengths. The wavelength of UV is around 0.25 micrometers.

firing 2000 pulses per second. Although black and white pictures are easy to 'see', color pictures might be very difficult to interpret because even if the data provided three color channels, we would not know the wavelengths of these colors.

### 10.5.3 Other parts of the electromagnetic spectrum

Although the microwave and optical parts of the electromagnetic spectrum appear to be the most probable choices for extraterrestrial signals, we should also consider the millimeter band, infrared and ultraviolet. Absorption by a planetary atmosphere is an issue in the choice of wavelength. In the near-infrared the predominant absorption is due to vibrational modes of the water molecule. Carbon dioxide also contributes to near-infrared absorption in bands centered at 2.7 micrometers and weak bands at 1.6 and 1.4 micrometers. Figure 10.18 shows the absorption bands in the infrared. In fact, a laser could operate in this region, avoiding the absorption bands. Solid-state lasers at 2.1 and 3.9 micrometers have been approaching a level of practicality that shows us that these wavelengths would be available to an advanced society for laser signaling.

In summary, in using ground-based receivers the visible and near-infrared are better options than the ultraviolet or mid-to-far-infrared bands that are heavily absorbed by the Earth's atmosphere.

The millimeter-wave band suffers both atmospheric absorption and noise, and does not offer the antenna gain to compete with shorter-wavelength lasers. We could probably overcome the technical issues with enough research and development, but just as microwave technology was not developed until a requirement was perceived for it during the Second World War, we will probably not make a major investment in millimeter-wave technology until there is a need for it. Useful systems currently go up to around 100 gigahertz. In addition to having few atmospheric windows, the infrared faces the problem that its photons are not powerful enough for direct photon detection to provide sensitive noise-free devices capable of overcoming the thermal noise without the detector being cooled to around 77K. Even when cooled, infrared detectors used to suffer from significant noise owing to the lack of noiseless gain in the devices, but infrared APDs such as HgCdTe devices are now becoming available which offer high sensitivity because of their internal gain. It is still necessary to cool the detector significantly for optimum performance, however. Progress in low-noise infrared detectors in the bands below a wavelength of 4 micrometers offers another possible regime. Noiseless gain is also possible by going to a heterodyne receiver, but this requires an infrared laser local oscillator inside the receiver and this in turn obliges one either to make an assumption about the specific wavelength to detect, or to search the spectrum by tuning the local oscillator – and tuning means the receiver may miss a signal unless it is a continuous wave, because it could well be tuned to a different part of the infrared spectrum when a pulse shows up. However, with the availability of low-noise detectors with internal gain, there is little reason to take on the complexity of heterodyning. Owing to the limited receiver sensitivity for direct detection at 10.6 micrometers, a heterodyne system has been developed for infrared laser communication, but tuning it is much easier when the transmitting frequency is known! Again, when we use the term infrared here we mean from 2 micrometers to 1000 micrometers, where it becomes the millimeter band. At shorter wavelengths, the near-infrared is more akin to the visible range because the energy of a photon is sufficiently high for devices to offer almost noise-free gain, for sensitivity without resorting to complexities of heterodyning in the receiver. Heterodyning requires the incoming signal and the laser local oscillator in the receiver to be aligned with great precision so that an uncorrupted input wave can be mixed with the phase front of the local oscillator to produce a useful intermediate frequency. But if at the wavelength in question the atmosphere can seriously affect the alignment of the incoming wave then an active correction must be incorporated to eliminate phase front changes. In addition, Doppler shifts must be tracked and eliminated.

All of the above implies that infrared wavelengths longer than a few micrometers would not be a good choice for SETI searching. Ultraviolet at a wavelength of 0.35 micrometers offers possibilities. Although a ground-based receiver is impractical due to absorption by the atmosphere, a receiver on a space station or on the Moon is attractive. The shorter wavelength of ultraviolet increases the antenna gain of the transmitter, which means the beam spreads out less in crossing interstellar space. As a result, when the

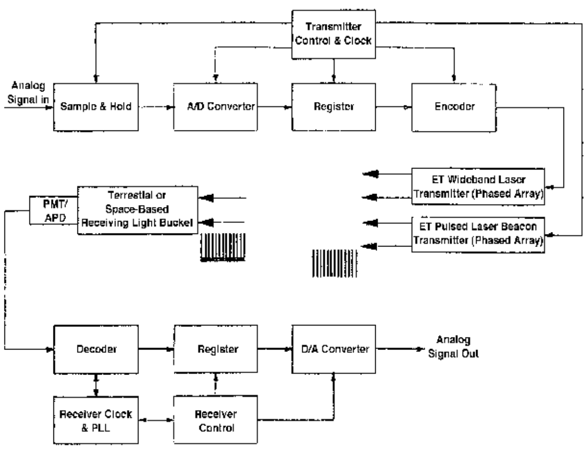



**Figure 10.20** Block diagram for beacon and high-rate information channels.

beam reaches us all of its power is confined into a smaller area, which means the same signal density can be obtained for a reduced transmitter power than in the case of a longer wavelength. Figure 10.19 shows the increase in antenna gain with shorter wavelengths when compared to other wavelengths. Arthur C. Clarke, who gave us the concept of the geostationary satellite and a great deal of award-winning science fiction, considered ultraviolet to be a much better option than either infrared or visible (personal correspondence with author, 1996). Since there is no air on the Moon there is no wind, so a large receiver will be able to be constructed using very lightweight materials. NASA's Goddard Space Flight Center recently reported that it had built a 3-meter-diameter dish using a concrete-like substance which consisted of crushed rock and an epoxy of carbon nanotubes. The dish was then spun in vacuum and coated with aluminum. This process could probably be scaled up to manufacture a telescope on the Moon with a diameter of 20 to 50 meters. Even if the figure of the mirror was not perfect, it would still make a practical ultraviolet 'photon bucket' for SETI. Owing to their high energy, ultraviolet photons offer high detector sensitivity. They are detected by photo-multipliers of high quantum efficiency which offer noise-free gain to overcome internal noise issues. If optical SETI from the Earth's surface proves fruitless, a reasonable next step would be to operate an ultraviolet receiver in space.

### 10.5.4 Space-based optical SETI

A team of internationally renowned astronomers and opticians has proposed to make very large telescopes on the Moon using liquid mirrors. A parabolic mirror could be created using a slowly rotating ionic solution (molten salts) with an ultra-thin coat of silver no thicker than 100 nanometers. It is estimated that all the materials for a 20-meter-diameter lunar telescope would weigh only a few tons. A single Ares V rocket could boost this to the Moon in the 2020s. Future telescopes might have mirrors as large as 100 meters in diameter. They could peer back in time to when the first stars and galaxies were forming. Although such telescopes would be designed as imagers, they would certainly address the question of 'bigness' for optical collectors for SETI. That said, however, based on past history SETI would not be the observing priority for a telescope of this size on the Moon.

NASA has been considering for a while now how to unfurl antennas in space for a variety of applications at microwaves and millimeter waves. In addition, large solar arrays have been deployed to draw power from the Sun. The solar arrays that NASA installed on the International Space Station span 73 meters from tip to tip. Although the design of a solar array has some similarity to that of a large photon bucket, there are unique design challenges involved in deploying and using lightweight optics in space. In particular, the shape of the collector surface must be precisely maintained.

Solar radiation and thermal distortions will require sophisticated solutions. The low residual pressure of the Earth's atmosphere at satellite altitudes induces sublimation of materials, evaporation of lubricating fluids, and vacuum welding. A stabilization system will have to overcome instability effects of torques created by solar pressure acting on large surfaces.

After spending half a century seeking SETI signals at radio frequencies, we are expanding the search to the optical spectrum. After a few decades we might come to the conclusion that it is preferable to build large optical collectors either in space or on the Moon. Perhaps an advanced civilization will signal using ultraviolet lasers precisely because requiring the recipient to be capable of operating in space is seen as a mark of its worthiness.

### 10.5.5 Beacon and high date-rate channels

Generally, SETI is about detection of an extraterrestrial attention-getting beacon, not a high data-rate channel. There is the question of whether we will first receive a beacon that will lead us to a channel where a reasonable data-rate will provide detailed information. It is possible to receive a beacon on the same channel as an information channel; one just needs to build a much larger collector, because the signal per pulse is much less owing to the greater number of pulses sent per second to convey the information. Figure 10.20 shows that the same block diagram can apply. At the transmitter, a separate laser source at the same wavelength is utilized with a reduced peak power and a much greater pulse rate. The larger collector permits detection of a weaker signal pulse as the pulse rate of the channel increases, allowing greater information flow. A major advantage of combining the beacon and the information channel is that the beacon needs only to make its existence evident; it does not need to send much, if any, data. Perhaps, the only data the beacon needs to convey is how big the receiver should be to receive the information. A very low repetitive pulse rate can be used where pulses are either there or not, conveying '1's and '0's.

## 10.6 In Conclusion

The Allen Telescope Array to detect microwave signals and the Harvard all-sky system for laser signals are the most ambitious SETI enterprises of their types to date. One of them might detect a signal. What if they fail? We will face either the Fermi Paradox that we are alone in the galaxy, or the argument that our collectors are still too small. If we have to make yet another jump in capability, we will face the decision of where best to invest our resources. A key issue will be the choice of wavelength. Let us end by reviewing this contentious issue.

Let us say that you are an alien on a planet around another star, and you desire to contact intelligent beings in some other system. There is near-zero likelihood of two civilizations being at the same level of technology. Mankind has had radio for about 100 years and lasers for 50 years. Let us say that your technology is several hundred years ahead of humans. You can only make yourself known to a civilization whose technology would enable them to detect your signal:

1 Are you going to build a 1000-meter-diameter radio-frequency antenna, or a 1-meter laser transmitter?

2 Are you going to make one that slews slowly from star to star, or one that can do so very rapidly?

3 Are you going to choose an approach where the recipient needs to know the precise frequency, or one that can be detected using a broadband receiver that looks at a significant portion of the electromagnetic spectrum?

4 Are you going to choose a system that can search many hundreds and possibly thousands of light-years away?

5 Are you going to avoid broadcasting the fact that you exist?

6 How will you protect yourself while actively searching for an alien civilization?

7 Will you start off with the most appropriate technology for point-to-point communication once you have found someone to talk to?

Considered from this perspective, it is probably an Earth-centric delusion for us to seek a radio signal that is deliberately aimed at us and maintained continuously. But even if you chose to use a laser, perhaps you would not use one in the portions of the electromagnetic spectrum we have been considering. Perhaps you would prefer an X-ray laser. If not a laser, then it might be something we have yet to discover. But we, as the potential recipient, must choose where to make our main effort. Half a century of listening at radio frequencies has been fruitless.

We have only just begun to search for laser signals. In a sense the Harvard optical system is only the opening gambit. We may well have to undertake a larger project before we have a real chance of detecting a signal, if one is there. We should apply our best intelligence to the problem. To search in a comprehensive way requires very little of our society's resources. We cannot conclude that we are alone until we have made a satisfactory search, and even in that situation we should ask ourselves what else we could and should have done. To conclude that we are alone would lead nowhere. To find other intelligent beings around other stars would be as momentous a discovery as is it possible to make.

Finally, for a more detailed account of microwave and optical SETI see Monte Ross's recent book on the subject.[26]

[26] Ross, M. (2009). *The Search for Extraterrestrials – Intercepting Alien Signals*, Praxis Publishing, Chichester, UK.

# 11

# Distributed Processing of SETI Data

Eric Korpela,
*Research astronomer and Project Scientist, University of California, Berkeley*

As you have read in prior chapters, researchers have been performing progressively more sensitive SETI searches since 1960. Each search has been limited by the technologies available at the time. As radio frequency technologies have become more efficient and computers have become faster, the searches have increased in capacity and become more sensitive. Often the limits of the hardware that performs the calculations required to process the telescope data in order to expose any embedded signals is what limits the sensitivity of the search. Shortly before the start of the 21st century, projects began to appear that exploited the processing capabilities of computers connected to the Internet in order to solve problems that required a large amount of computing power. The *SETI@home* project, managed by myself and a group of researchers at the Space Sciences Laboratory of the University of California, Berkeley, was the first attempt to use large-scale distributed computing to solve the problems of performing a sensitive search for narrow band radio signals from extraterrestrial civilizations. (Korpela et al., 2001) A follow-on project, *Astropulse*, searches for extraterrestrial signals with wider bandwidths and shorter time durations. Both projects are ongoing at the present time (mid-2010).

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_11, 

## 11.1 Computation in Radio SETI

Why would an enormous supercomputer be necessary to detect radio signals from an alien civilization? It might seem to be a fairly simple signal processing task. One reason is that the parameters of any alien signal are unknown. Some of these parameters are intrinsic to the signal: frequency, frequency changes, bandwidth, encoding, and duration. Others are properties of how the signal is sent and received: Is the transmitter on a planet, in orbit, in interstellar space? Is it directional or omnidirectional? Still others are unavoidable properties of how the signal has propagated through space to arrive at Earth. To perform a thorough search, we need to investigate a wide variety of these parameters.

One typical assumption made in SETI is that an alien civilization wishing to make contact with other races would broadcast a signal that is easily detectable and easily distinguishable from natural sources of radio emission. One way of achieving these goals is to send a narrow band signal. By concentrating the signal power in a very narrow frequency band, the signal can be made to stands out among the natural sources of noise which are broad band. A second way is to send a signal of short time duration which, in principle, would be detectable above the background noise for the duration that the signal is on. For reasons to be described later, much more processing power must be employed in order to detect this second type of signal.

In part because of this, radio SETI efforts have concentrated on detecting narrow band signals. When searching for narrow band signals it is best to use a narrow search window (or channel) around a given frequency. The wider the channel, the more broad band noise is included in addition to any signal. This broadband noise limits the sensitivity of the system. Early systems used analog technology to create narrow bandpass filters that could observe at a single frequency channel. More recent systems use massive filter-banks of dedicated Discrete Fourier Transform[1] (DFT) processors to separate incoming signals into up to a two billion spectral channels, each of width ~1 Hz.

There are, however, limitations to this technique. One limitation is that extraterrestrial signals are unlikely to be stable in frequency due to accelerations of the transmitter and receiver. For example, a receiver listening for signals at 1.4 GHz located on the surface of the earth undergoes acceleration of up to 3.4 cm/s$^2$ due to the Earth's rotation. That may not seem like much, but it corresponds to a Doppler drift rate of 0.16 Hz/s. If uncorrected for this drift, an alien transmission would move outside of a 1 Hz channel in about 6 seconds, effectively limiting the maximum integration time to 6

[1] Other chapters in this book may use the term Fast Fourier Transform (FFT) which is a specific type of DFT. For these purposes, the terms are interchangeable.

seconds. Because of the inverse relationship between maximum frequency resolution and integration time ($\Delta\nu=1/\Delta t$) there is an effective limit to the frequency resolution that can be obtained without correcting the received signal for this effect. ($\Delta\nu \sim 0.4$ Hz) Most of the recent generation of SETI spectrometers have channel widths between 0.5 and 1.7 Hz.

In principle a correction can be made for most of the drift due to motions of the Earth, but how does one correct for motions of an unknown planet? An alien civilization beaming signals directly at the Earth could correct the outgoing signal for the motions of the transmitter, but a civilization transmitting an omnidirectional beacon could not make such an adjustment.[2] Therefore, to search for this type of signal at very narrow bandwidth ($\ll 1$ Hz) and with the highest possible sensitivity, the correction for Doppler drift must be made at the receiving end and a search for signals performed at multiple Doppler drift rates. Repeating an analysis at multiple Doppler drift rates becomes compute intensive.

Other parameters of the signal are also unknown, for example: At what frequency will it be transmitted? What is the bandwidth of the signal? Will the signal be pulsed, if so at what period? Fully investigating a wide range of these parameters requires proportionally larger computing power.

In addition to detecting a signal, we must be able to determine whether a signal is truly of celestial origin. The vast majority of the narrow band signals received by a radio telescope will be radio frequency interference (RFI) generated locally. Fortunately RFI often has properties that allow it to be distinguished from extraterrestrial emission. RFI elimination requires some level of computing resources.

Performing all of these these calculations for even a small portion of the radio spectrum requires as much computational power as is available in the largest existing supercomputer. However, such computers are not typically made available to SETI researchers.

---

[2] Actually, with enough expense, they could. Rather than building a single omnidirectional beacon they could build enough directional transmitters to cover the sky. Sixteen million five hundred thousand Arecibo class telescopes would do the job nicely, but that's probably overkill. Because small telescopes have a larger field of view, but require more power to send the same effective isotropic radiated power, there is a tradeoff between number of telescopes, the amount of uncorrected Doppler drift and the total power that could be transmitted without melting the transmitters. Calculating the optimum number is left for the reader, a colleague or another time.

## 11.2 SETI@home

Fortunately, searching for signals in a data stream from a radio telescope is a task that is easily distributed. Data from an observation can be broken up into frequency bands that are essentially independent of one another. In addition, an observation of one portion of the sky is essentially independent of an observation of another part of the sky. This allows a large data set to be divided into small chunks that can be analyzed by a personal computer in a comparatively short time, making possible the distribution of the work to people willing to donate their spare CPU cycles.

SETI@home conducts its observations at the National Astronomy and Ionospheric Center's 305-meter radio telescope in Arecibo, Puerto Rico (Figure 11.1). The project uses ALFA, an array of seven receivers arranged in a hexagonal pattern with one in the middle, which is mounted in the enclosed dome-like structure seen suspended above the Arecibo telescope. SETI@home makes its observations in conjunction with other uses of the ALFA array. Currently this array is used to search for pulsars near the plane of the Galaxy, to map the distribution of hydrogen in all parts of the

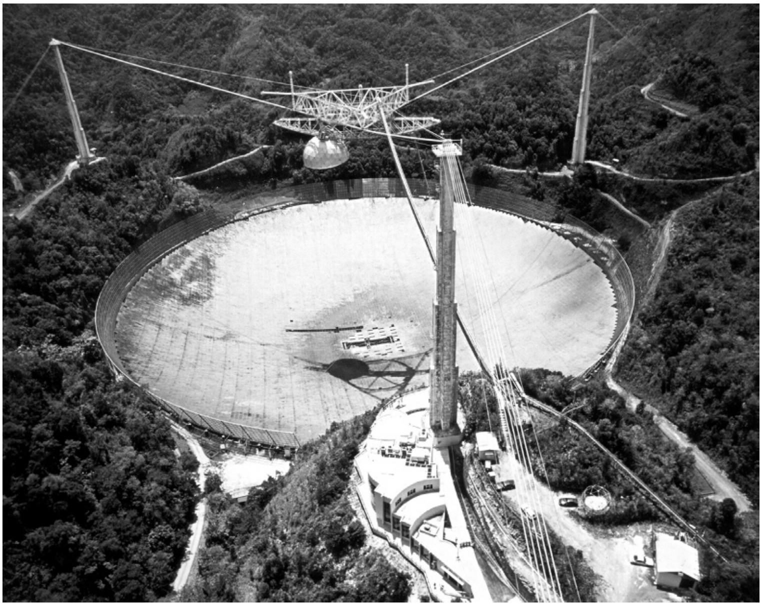


**Figure 11.1** SETI@home and Astropulse use the National Astronomy and Ionospheric Center's 305 meter telescope at Arecibo, Puerto Rico. Photo courtesy of the NAIC-Arecibo Observatory, a facility of the NSF.

Galaxy visible from Arecibo, and to search for extragalactic hydrogen gas in isolated clouds or in nearby galaxies. This results in three main modes of observation. The pulsar surveys tend to track positions in the sky while accumulating data for 30 seconds to tens of minutes. The other surveys either utilize a drift scan mode where the receivers are held in position while objects in the sky drift by during the earth's rotation or a "basket-weave" mode in which the receiver tracks north and south while the sky drifts by, resulting in a zigzag path.

If the primary feed is stationary, objects in the sky pass through the fields of view of the ALFA receivers (0.05°) at the rate of the rotation of the Earth (also known as the sidereal rate). An object would require about 13 seconds to transit the field. When used in basket-weave mode, less time is required for transit. When tracking, objects can remain in the field of view for large durations.

During the course of the these projects, SETI@home will view most portions of the sky visible with the Arecibo telescope three or more times. This includes stars with declinations (the celestial equivalent of latitude) between –2° and 38° thoroughly covering about 25% of the sky.

The SETI@home system records a 2.5 MHz wide band from each of the two polarizations of the seven receivers (14 data streams in all) centered at the 1420 MHz Hydrogen line. Because the Hydrogen line would be of interest to astronomers of any species who were studying the Galaxy, this frequency is considered one of the most likely locations for deliberate extraterrestrial transmissions. These 2.5 MHz bands are recorded continuously onto hot-swappable serial ATA disk drives using 2 bit complex samples. A 2TB drive holds the data for about 57 hours of observing. We are accumulating data at a rate of about 50 TB per year. This data is archived at the National Energy Research Scientific Computing Center at the Lawrence Berkeley Laboratory.

The full drives are shipped to Berkeley where they are subdivided into small 'work units' using software appropriately known as a 'splitter'. The 2.5 MHz bandwidth data is divided into 256 sub-bands by means of a 2048 point DFT followed by 256 eight point inverse transforms. The 9766 Hz wide sub-bands are divided into lengths of 220 samples. Each work unit corresponds to about 10 kHz of bandwidth and 107 seconds of duration. When the project began in 1999, these sizes were chosen such that a common desktop computer could perform our analysis procedure in less than a week. Thanks to Moore's Law[3] (Moore, 1965), which successfully predicted that processing power would double every 18 months during that

[3] The original formulation of Moore's Law concerns the number of components that can be fitted on a silicon chip. Here we use a common extension of that law to computer processing capabilities.

decade, a current (2010) 4-core processor can typically process four of these work units in two hours.

Subsequent work units overlap by 20 to 30 seconds to allow full analysis of signals that may be within a beam transit time of the end of a work unit. Each of the work units data files are transferred to temporary storage (which typically holds one to three million work unit files) for distribution to users. The work unit files are stored there until the results for that work unit have been received.

## 11.3 BOINC

The structure of the SETI@home server hardware has evolved over time from a single underpowered workstation to what is now several six-foot tall racks of computers and disk drives. The software has evolved even more.

The original SETI@home server was a relatively small program that spoke a limited subset of hypertext transfer protocol (HTTP). Although it communicated over the standard HTTP port, it was only capable of processing requests from the SETI@home application, storing results, and returning a single work unit file. Despite this simplicity, it was easily overwhelmed when request rates became high. It had no means of monitoring behavior of users or validating that the result returned belonged to the work unit file that had been sent. It was easy for malicious people to attempt to both boost their credit standings or attempt to damage the integrity of our science database by returning invalid data for a large number of results. This server was also very specific to SETI@home. If we wanted to develop other volunteer computing applications, we would have needed to develop a new server for each.

To alleviate some of these issues we have developed the Berkeley Open Infrastructure for Network Computing or BOINC (Anderson, 2004). Rather than using a special purpose HTTP server, BOINC utilizes standard web servers that support the Common Gateway Interface (CGI) or FastCGI which can be used to call external programs for web page generation. Handling and monitoring connections is done by the web server, which is typically well optimized for the task. The BOINC software is divided into (1) 'work generators', which in our case are our splitters described above; (2) a CGI 'scheduler' which handles request from volunteers' computers and decides what work to distribute to each - downloads of the work units are performed using standard HTTP from any web server; (3) a CGI 'file upload handler' that collects the results that are returned; (4) a 'validator' which determines whether the returned results are likely to be correct, in our case by comparing results returned from two or more machines; and (5) an 'assimilator' which stores the valid results. In our case our assimilated results are stored in our science database.

BOINC allows easy distribution of these tasks across multiple machines. In addition, it maintains statistics on each computer including processing speed estimates which are used to determine how much work to send. It maintains an estimate of the error rate for a machine, so a trustworthy machine might be trusted to generate a correct result without sending the same work to another computer, while an untrustworthy machine would always have its work checked by another machine.

Once a result has been returned to our server and validated, the assimilator process stores the time, sky coordinates, frequencies, etc. for each of the potential signals that was returned. The largest portion of the science database capacity is used for storing these parameters of potential signals. This database is currently (May 2010) about 2TB in size, and holds about 4 billion potential signals. Later, we'll discuss how we sift through that many signals to try to find the extraterrestrial ones.

## 11.4 The SETI@home Application Program

SETI@home volunteers download the BOINC client software through a link provided on the SETI@home website (http://setiathome.berkeley.edu). Standard versions are available for Microsoft Windows, Apple Macintosh, and Linux systems. Versions ported to many other systems are available through the BOINC website (http://boinc.berkeley.edu). After installation, the BOINC client will provide a list of projects that the volunteer can join. After joining SETI@home little user interaction is required. The BOINC client software will automatically contact the SETI@home server to request work. The server will reply containing the URLs at which the BOINC client will download the SETI@home application and each of the work unit files to be processed.

If the user wishes to have more control over how work is processed they can set preferences as to whether work will be processed while the computer is in use or to set hours of the day when processing can be performed. For Microsoft Windows and Apple Macintosh the user has the option of using a BOINC screen saver (Figure 11.2). This screen saver allows the processing application to generate graphics that will be displayed when the screen saver is active. If the running application does not generate graphics, the screen saver displays statistics about the running application.

After receiving a work unit file, the application performs a baseline smoothing on the data to remove any wide band ($\Delta\nu > 2$ kHz) features. This prevents the application from confusing fluctuations in broad band noise

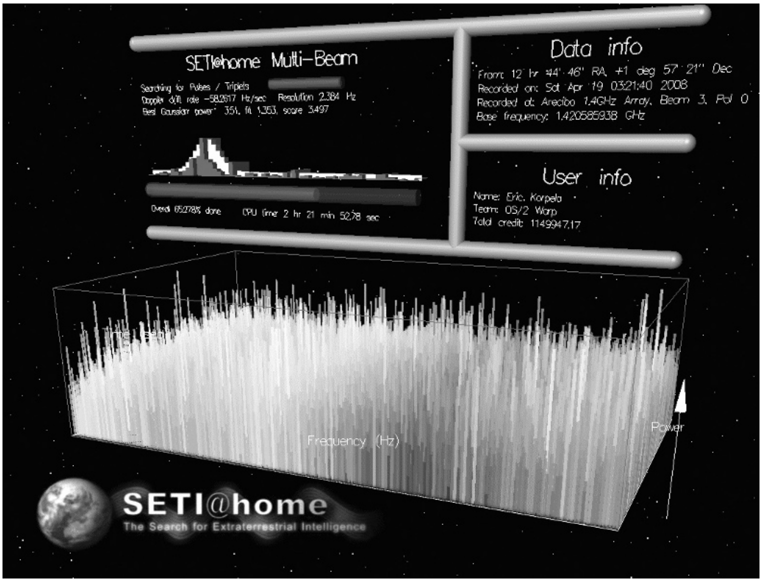



**Figure 11.2** A screenshot of the SETI@home application graphical display. The bottom half of the screen presents the power spectrum currently being analyzed. The upper left section shows analysis state and the results of the current analysis. The upper right section shows information about the data being processed and the user's statistics.

```
for Doppler drift rates from -100 Hz/s to +100 Hz {
   for bandwidths from 0.075 to 1220 Hz in 2X steps {
      Generate time ordered power spectra.
      Search for short duration signals above a constant threshold (spikes)
      for each frequency {
         Search for faint signals matching beam parameters (Gaussians)
         Search for groups of three evenly spaced signals (triplets)
         Search for faint repeating pulses (pulses)
      }
   }
}
```

**Figure 11.3** A pseudo-code representation of the SETI@home processing method.

(due in part to variations in the Hydrogen line emission as the field of view transits the sky) with intelligent signals. The application then begins the main data analysis loop, which is shown schematically in Figure 11.3.

At the start of each passage through the loop, the data is transformed into an accelerated frame of a given Doppler drift rate. The drift rates at which the application searches the data for signals vary from –30 Hz/sec to +30 Hz/sec (accelerations expected on a rapidly rotating planet) in steps as small as 0.0009 Hz/sec. The application also examines the data at Doppler drift rates out to ±100 Hz/sec (accelerations of the magnitude that would arise from a satellite in low orbit about a super-earth), but at a more coarse step of 0.015 Hz/sec. A signal from an alien world would be most likely to have a negative drift rate (as the accelerations involved would be away from the observer). Despite this, we examine both positive and negative drift rates for the purpose of statistical comparison and to leave open the possibility of detecting a deliberately chirped extraterrestrial signal.

At each drift rate the application searches for signals at one or more bandwidths between 0.075 and 1221 Hz. This is accomplished by using DFTs of length $2^n$ (n=3,4,..,17) to transform the data into a number of time ordered power spectra. In order to avoid repeating work, not all bandwidths are examined at every Doppler drift rate. Only when the change in drift rate becomes significant compared to $1/\Delta\nu^2$ is another DFT of that length computed. Therefore 32k-point transforms are performed one quarter as often at 64k-point transforms.

The transformed data is examined for signals that exceed 24 times the mean noise power. This threshold corresponds to 2.0×10-25 W/m2 at our finest frequency resolutions, or the equivalent of detecting a cheap cell phone on one of the moons of Saturn. The SETI@home application reports any such ‘spike’ signals when it transmits the results of the data processing.

If there is sufficient time resolution in the transformed data (n<15) and the SETI receiver is not tracking an object in the sky, the application examines it for signals which match the parameters of the telescope beam. As a radio source drifts through the field of view, the measured power will vary depending upon the beam profile of the telescope. This profile is approximately Gaussian. The SETI@home application performs a $\chi^2$ curve fit on any signals which exceed 3.2 times the mean noise power and reports those for which the goodness of fit is better than a certain level. This power level typically corresponds to $2.1\times10^{-25}$ W/m$^2$.

The application then divides transformed data at each frequency into chunks with duration equal to the time required for an object to transit the telescope’s field of view. These chunks are examined for pulsed signals using two algorithms. The first algorithm, the triplet finder, searches each chunk for three evenly-spaced signals that each exceed 9.1 times the mean noise power (as little as $2.5\times10^{-25}$ W/m$^2$), and reports any detected signals.

The second algorithm is a modified fast folding algorithm (FFA). A folding algorithm divides the data into chunks of duration equal to the period being searched and co-adds them in order to improve signal-to-noise ratio. An FFA performs this function on a large number of periods without duplicating additions. The SETI@home folding algorithm searches roughly N log N pulse periods, where N is the length of the input array, between 2 samples and N/3 samples. During a typical run of the application this typically means half a million periods between 2 ms and 10 s. The threshold for detection of a pulsed signal is computed dynamically to match the number of co-added samples, and can be as low as 0.04 times the mean noise power for pulses with periods less than 10 ms. This corresponds to pulse energies of about $4.4\times10^{-27}$ J/m$^2$.

This processing loop requires over 5 trillion floating point operations (teraFLOP). For an average work unit the SETI@home application would report one spike signal, one Gaussian signal, one pulsed signal, and one triplet signal.

## 11.5 Astropulse

One advantage of the BOINC infrastructure is that adding an additional application that processes a different data format is relatively straightforward. All that is necessary is to build the application, a work generator, a validator, and an assimilator. I mentioned earlier that it is possible that an extraterrestrial wishing to attract attention might send a short (microsecond) duration broad band pulse rather than a narrow band signal of long duration.[4] Detecting such a pulse presents some challenges because of how a broad band signal interacts with the tenuous gas that fills interstellar space. In most of the volume of space the gas through which a signal would traverse is at least partially ionized into a plasma of positive ions and free electrons. As the radio wave passes an electron, the electric and magnetic fields in the wave try to shake the electron at the frequency of the wave. The longer the wavelength (which also means the lower the frequency) the more the electron is able to interact. This interaction tends to slow the speed at which the radio wave propagates. This process spreads out a wide band signal by delaying the low frequencies more than it delays the high frequencies (see Figure 11.4). This process is called dispersion, and it can be reversed with mathematical manipulations similar to those used by SETI@home to correct for Doppler

[4] There are potentially natural sources of such pulses as well. Evaporating mini-black holes left over from the big bang are theorized to emit them. Some pulsars emit short timescale giant pulses. And there may be new types of objects we have not yet discovered.

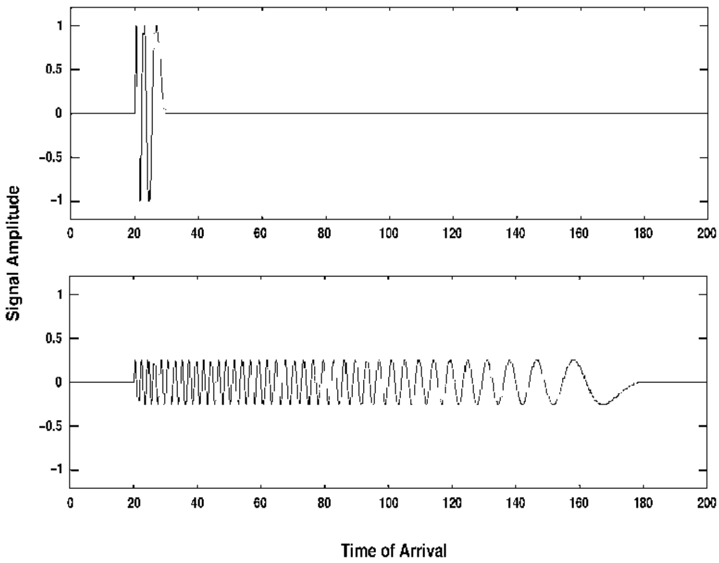



**Figure 11.4** An illustration of dispersion of a broadband pulse. The upper figure shows the waveform of a pulse that has undergone a small amount of dispersion. The lower figure shows a pulse after more dispersion has slowed the low frequency components. Note the high frequencies are located at the left side of the figure which indicates they arrive first.

drift. Fortunately, this process only depends upon how many free electrons lie on the line of site from the transmitter to the observer, rather than on the details of that distribution. This quantity is, in detail, the integral of the electron density along the line of sight. Astronomers call it the Dispersion Measure (DM) and usually report its value in $cm^{-3}$ parsecs. For example, if the density of electrons in space was 1 per cubic centimeter ($cm^{-3}$), an object at a distance of 27 parsecs (pc) would have a dispersion measure of 27 $cm^{-3}$ pc.

Unfortunately, we don't know where the transmitter is, so we don't know how many electrons are between it and us. So we correct for reasonable values of galactic dispersion where a signal might be seen, from 49.5 to 830 $cm^{-3}$ pc. Because an extraterrestrial might transmit a signal that is negatively dispersed either as an indication the signal is artificial, or as precompensation for dispersion toward the target of the signal, and because seeing negatively chirped interference helps us to characterize the interference in our data, we also look at the same range in negative dispersion as well.

In one way, Astropulse uses a simpler method than SETI@home; because we are looking for a broad band pulse we don't want to divide the recoded data by frequency. The work generating splitter for Astropulse merely needs to divide the data in chunks of about 13 seconds duration (a typical beam transit time). Both these and the Astropulse application are sent out to our volunteers. No additional action is required on the part of SETI@home volunteers to receive Astropulse work, although they may opt out if they wish. Because Astropulse work units take ten times longer to process than SETI@home work units, the BOINC server must check to be sure the volunteer's computer is capable of processing the data in a reasonable amount of time before assigning Astropulse work to it.

The algorithm is fairly simple. We disperse the data at a specific dispersion measure, which generates a time series representing the signal power in the dedispersed frame with a time resolution of 0.4 μs. If any events are above threshold, they are reported. We then co-add adjacent bins, to improve sensitivity to longer timescale pulses, again looking for events above threshold. We repeat this co-add 8 more times, examining the data for signals at timescales from 0.4 to 204.8 μs. Then we move on to the next dispersion measure (usually stepping 0.05 $cm^{-3}$ pc). At some dispersion measures we perform a folding algorithm similar to that used by SETI@home to detect repeating pulses. More details of the Astropulse algorithms have been presented by Von Korff (2010). We have set the thresholds such that ~1 pulse will be detected in a work unit filled with random noise. As we have discovered, there are many dispersed terrestrial signals that result in many signals being detected in an average workunit.

## 11.6 Post-processing

When the applications have done their work, the job isn't done. Typically the application programs return a few potential signals per work unit. Of course, not all of these signals are evidence of extraterrestrial intelligence. Some of the signals are due to errors made in the processing computers. Numeric processors, memory and disk systems are fairly reliable. However SETI@home and Astropulse represent thousands of years of CPU time per day, magnifying even low error rates. Event if undetected errors occur only on average every $10^{18}$ machine instructions, SETI@home would see ten per day. To combat these effects our validator examines each signal to see if the parameters match their permitted values. We also send each work unit to multiple volunteers, and cross check the returned values to verify accuracy.

A large number of the signals in the database are evidence of terrestrial intelligence. Sources of narrow band radio emission are ubiquitous where human technology is present. The sources of dispersed broad band emission (primarily radars) are even stronger. Even at the Arecibo observatory,

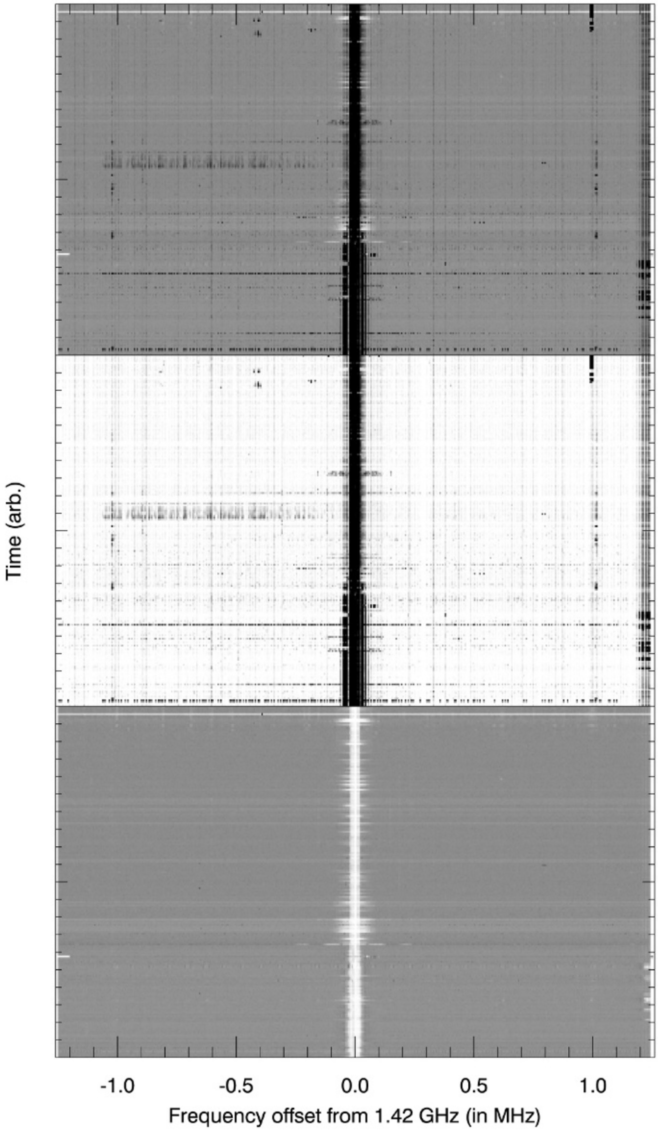



**Figure 11.5** These plots show the frequency distribution of pulses detected by SETI@ home. The upper panel shows all pulses. The middle panel shows pulses determined to be due to persistent interference sources. The lower panel shows the pulse frequency distribution after the interference has been removed.

where care is taken to minimize interference, this noise is present, due to airport and air defense radars, local equipment, aircraft, satellites, and other transmitters. Most of the time these terrestrial emissions are fairly easy to distinguish from an extraterrestrial signal.

It's possible to mitigate the effect of radars both at the telescope, and in our data processing pipeline. At the Arecibo telescope, the observatory maintains an antenna which monitors the most powerful radar and a device known as the 'radar blanker' that predicts when the radar pulses will arrive. An observer can use the prediction to replace the telescope data with a noise-like signal during the times when a radar pulse might arrive. We have developed a second system that works in a similar fashion after the fact by examining our recoded data for the radar signals. We can then fit the known radar patterns to what is seen and remove one or more of the contaminating radar patterns. This has greatly reduced the number of radar signals that are being stored in the SETI@home and Astropulse databases.

A large fraction of RFI consists of continuous narrow band signal generated at or near the observatory. We use this property to detect it signals in the zones containing it. The RFI frequency zones are typically quite narrow. We have identified 35,000 frequencies, covering less than 1% of our total bandwidth, which are subject to frequent interference. These zones contain between 5% and 20% of the detected signals depending upon signal type. For example, the top panel of Figure 11.5 shows the frequency distribution of 378,362,077 potential pulse signals detected by SETI@home between July 5, 2006 and September 16, 2009. The vertical bands that are present indicate frequencies that are over-represented and are probable RFI frequencies. We use a statistical analysis to determine which frequencies appear too frequently on differing sky positions to be due to noise processes. Those frequencies define the exclusion zones. Pulses determined to be within these zones (6.6% of the total) are shown in the middle figure. The lower figure show the distribution of pulses that remain after those within zones have been removed.

Other RFI sources are of short duration and repeat on time scales of hours to days. So any signal that repeats after a short time when the telescope is viewing a different portion of the sky should also be rejected. After RFI is removed, the bulk of the remaining signals are due to random fluctuations in the noise background mimicking an extraterrestrial signal. One means of sorting out the true extraterrestrials is by looking for persistent signals. We expect that an extraterrestrial signal will be present at a similar frequency the next time the same celestial location is examined. We have developed a program called a Near Time Persistency Checker (NTPCkr) that sorts through the database looking for persistent signals. When it finds one, it sends it off to the RFI removal program, to make sure that it is not due to RFI. Nearly every time, the RFI removal program finds that the signal was due to an RFI event. For those haven't been flagged as RFI are added to a candidate list. We then propose for telescope time to reobserve them. Thus far no reobservation has confirmed the detection of a candidate.

## 11.7 Distributed Thinking

Each candidate on the candidate list must be verified by a human being before being confirmed as a target for reobservation, primarily because automated means of RFI detection are insufficient. Temporary RFI sources often appear and disappear, or shift frequencies in ways that cannot be easily detected by automated software. We are hoping to develop a means for our volunteers to help identify RFI in time vs. frequency 'waterfall plots' by first training them on manufactured data. They will then be able to examine our candidates and give an opinion on whether each signal is clean or the result of interference. We're hoping this will help to speed the process of identifying the candidates for reobservation from among the thousands of possibilities.

## 11.8 Distributed Development

There are other ways to distribute the SETI workload. One is distributed software development. Shortly before transitioning to the BOINC infrastructure, we released the source code to SETI@home under the General Public License (GPL). This enabled several developments. First, many bugs within the source code were brought to our attention. Most were minor, but some would have limited our ability to correctly identify candidates if we hadn't corrected for their effects. Second was optimization of the code. There has always been an element of competition to SETI@home. People compete to see who can do the most work in the least time. Many of these volunteers developed optimized versions of SETI@home. Some found new algorithms to perform the same functions; others added support for single instruction multiple data (SIMD) instruction sets such as AltiVec and SSE. Many of these contributions have been returned to us and included in the application we distribute. The current SETI@home application runs in about one-twelfth the time that the original version would take, despite doing many times as much work.

SETIQUEST, a project run by the SETI Institute (http://www.setiquest.org), has even more ambitious distributed development goals. In addition to providing the source code for the existing SETI Institute data processing routines, they invite participants to download data and develop their own algorithms for detecting signals within the data. They are hoping to develop a group of citizen scientists which will help to improve current and future SETI searches.

## 11.9 The Future of SETI@home

SETI@home was originally slated to process two years' worth of data from the Arecibo telescope. The strong public response and new improvements to the application software have kept us going for 11 years. Recently we've started deploying versions of SETI@home that run on graphics processing units (GPUs) that are capable of highly parallel operations. SETI@home can compute on the GPU up to 30 times faster than the CPU on systems that contain a compatible GPU.

Despite this, SETI research lives in a perpetual state of being starved for computation resources. In the past 12 months we have discussed three new algorithms with other SETI researchers. One we will probably implement soon. It will make very little change to the time required to process a work unit, but will perform a search for a different type of signal. The second would increase our processing time by factors of 10 to 100, but with the possibility of proportionally higher science return. We are considering it for the future. The third, if implemented fully, would easily require all of the compute cycles executed by all of the computers that have ever existed on Earth in order to examine a small fraction of our data. If Moore's law continues to apply, perhaps this will be possible before we realize.

SETI@home currently samples only a small portion of the radio spectrum, and a small portion of the sky. The two most obvious means of expanding its capabilities are to expand the sky coverage and widen the frequency bandwidth. The primary impediment to larger bandwidth is the SETI@home data recorder (which can record at 80 Mbps for a total recorded bandwidth of 40 MHz), the available storage for maintaining the data, and the available computing power. Very large baseline interferometry (VLBI) data recorders in use at many observatories can record at 4 Gbps (total recorded bandwidth up to 2 GHz). That's enough to fill a 2TB disk drive in one hour. At current prices it would cost $1.2 million to buy disks to hold one year's worth of data when recorded at that rate. Needless to say, SETI@home doesn't have the financial resources to do that. But incremental improvements can be achieved for less cost. The required computing power is roughly proportional to the recorded bandwidth.

The best means of expanding the sky coverage would be to add a SETI@home recorder system to a southern hemisphere radio telescope. This would allow us to increase our sky coverage from about 25% to 75%. We have considered this quite often, but thus far the resources needed for us to do so have not presented themselves.

As in any voluntary organization, it's important that SETI@home be responsive to the desires of its volunteers. The success of SETI@home is entirely dependent on the volunteers who provide the computing resources. We will continue working to keep our volunteers informed of our progress and to share with them the science behind SETI.

**Acknowledgments**

I'd like to thank the rest of the SETI@home/Astropulse and BOINC staff at the Space Sciences Laboratory: Josh Von Korff, Andrew Siemion, Jeff Cobb, Matt Lebofsky, Robert Bankay, Dan Werthimer, and David Anderson. SETI@home and Astropulse are supported by grants from NASA and the NSF and donations from the friends of SETI@home. We'd like the thank the volunteers who helped optimize SETI@home, many of whom can be found at *http://lunatics.kwsn.net/*. We would also like to thank SETI@home volunteers everywhere for their valuable contribution of the processing power that makes SETI@home work. The observations for these projects were made at the NAIC Arecibo Observatory a facility of the NSF operated by Cornell University. This research uses resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## References

Anderson, D. 2004. BOINC: A System for Public-Resource Computing and Storage. 5th IEEE/ACM International Workshop on Grid Computing, November 8, 2004, Pittsburgh, USA, 1-7.

Korpela, E.J., Werthimer, D., Anderson, D., Cobb, J. and Lebofsky, M. 2001. SETI@home - Massively Distributed Computing for SETI. *Computing in Science and Engineering*, **3:1**, 78-83.

Moore, Gordon E. 1965. Cramming More Components onto Integrated Circuits. *Electronics*, **38:8**, 4-7.

Von Korff, J. 2010. Astropulse: A Search for Microsecond Transient Radio Signals Using Distributed Computing. Ph.D. Thesis, University of California, Berkeley.

# 12

# Project Argus: Pursuing Amateur All-Sky SETI

H. Paul Shuch,
*Executive Director Emeritus, The SETI League, Inc.*

Project Argus, a global effort of the non-profit SETI League, Inc., seeks to achieve continuous microwave monitoring of all four pi steradians of space, in real time. Initially, Project Argus was planned as the most ambitious SETI project ever undertaken without benefit of government support, ultimately to involve 5000 small radio telescopes worldwide, built, maintained and operated by private individuals (primarily radio amateurs and microwave experimenters), coordinated so as to miss no likely candidate signals, and providing independent verification of any interesting signals detected. Though prototype stations went into operation in 1996, and second-generation stations by 2000, full sky coverage is yet to be achieved.

In this chapter, sensitivity and range of the typical Project Argus station are assessed by comparison of hardware and software capabilities to those in place at the Ohio State Radio Observatory in 1977, when the so-called "Wow!" signal was detected. The "Wow!" signal serves as a convenient benchmark, even though its exact nature remains unknown. Should a similar candidate signal appear during the fully deployed phase of Project Argus, it will not evade detection.

Though utilizing just a small satellite TV dish as its antenna, each station achieves range and sensitivity on a par with the Ohio State Big Ear radio telescope at the time of the "Wow!" detection. This chapter outlines the technological breakthroughs which made that level of performance possible, and explores the reasons that this global network fell far short of initial expectations.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_12, 

## 12.1 Probabilistic Considerations

Fifty years after implementation of the first modern SETI study, it has begun to appear that electromagnetic signals from extraterrestrial civilizations (if, indeed, they exist at all) are likely to be highly intermittent in nature, of just a few seconds to a few minutes in duration, and never repeating. If such signals are the norm, then even the most advanced temporally displaced signal verification scheme (such as the Follow Up Detection Device scheme used by Project Phoenix, detailed elsewhere in this book) will be prone to a high incidence of false negatives. Further, as large radio telescopes scan only a tiny fraction of the cosmic sphere at any given time, we can expect that the overwhelming majority of interesting candidate signals will evade initial detection.

The probability of detection of any given extra-terrestrial radio emission is a multivariate parameter space with at least six degrees of freedom, three spatial and three temporal. Spatial parameters include sky coverage (expressed in ranges of azimuth and elevation, or alternatively right ascension and declination), and capture area. Temporal factors include frequency coverage, resolution bandwidth and observing time. The latter two factors are highly correlated through integration time constant. There are also thermodynamic factors, specifically involving sky noise and receiver noise, but these can be negated. Judicious choice of operating frequency (such as within the transparent portions of the microwave window) can minimize noise sources associated with the interstellar medium, and the state of the art in receiver design makes equipment noise contributions almost negligible.[1]

Much attention has been given in past SETI efforts to maximizing frequency coverage, through the use of elaborate multi-channel spectrum analyzers (MCSAs), and capture area, through the use of very large parabolic reflector antennas. This emphasis has traditionally been at the expense of sky coverage and observing time. As MCSAs are inordinately expensive and in short supply, they have thus far been utilized at only our largest radio telescopes. Such large antennas are appropriate for targeted searches of specific stars, but perhaps not for all-sky surveys. When a drift-scan sky survey is performed, capture area and observing time (or its complement, sky coverage) are mutually exclusive.

It is hypothesized that the most likely microwave signals of intelligent extraterrestrial origin will be highly intermittent in nature. The best-known candidate signal to date, the Ohio State "Wow!" signal, is a case in point. It was explored in great detail in an earlier chapter of this book. The duration

---

[1] An additional consideration, much discussed elsewhere in this book, is the overall sensitivity of the receiving system. However, sensitivity is established as a function of noise temperature, capture area, bandwidth and integration time, all factors covered above.

of this signal was insufficient to be captured by both of the feedhorns employed on that antenna for terrestrial interference elimination. Over 100 follow-up observations of the same region of sky, at the same frequency, failed to turn up a repetition of this tantalizing candidate. If the "Wow!" represented incidental radiation leakage rather than an interstellar beacon, we should not reasonably expect it to repeat over time scales consistent with the human life span.

Consider that the late Big Ear radio telescope at the Ohio State Radio Observatory was narrow in beamwidth, viewing about one part in $10^6$ of the sky at a given time. Let us imagine that the "Wow!" emanated from a similar antenna some tens to hundreds of light years distant. As both antennas can be assumed to be situated on rotating planets, the likelihood that each will be pointing at the other is found simply as the square of the sky coverage of either antenna; that is, one part in $10^{12}$.

Granted, there could be other equally interesting candidate signals emanating from other directions, the reception of which would be just as significant as a reprise of the "Wow!" That, after all, is the justification for the sky survey approach to SETI. But wherever a signal might originate, the narrow beamwidth of Big Ear suggests that if we happen to be listening on exactly the correct frequency, at exactly the instant an interesting signal arrives at Earth, there's still a 99.9999% chance our antenna will be pointed the wrong way!

One answer is to scatter a million Big Ears across the surface of the Earth. We would surely then be able to look in all directions at once, but at tens of millions of dollars per antenna, the costs could quickly exceed the Gross Planetary Product. And, in fact, history has (sadly) shown that this planet can no longer afford a single Big Ear, much less a million of them. So, consideration must be given to smaller, less costly radio telescopes, if true all-sky coverage is to be achieved.

## 12.2 Quantifying the "Wow!" Signal

Any radio telescope which we might propose for any SETI survey must, of course, be capable of detecting signals of likely power levels. Let us assume for this analysis that the "Wow!" signal is a valid SETI candidate, of just such a likely level. We know the signal-to-noise ratio (SNR) of this candidate signal as received in 1977, and can easily compute the sensitivity of the Ohio State Radio Observatory at that point in time. Thus, we can readily determine the flux density of the Wow!, which establishes for us a practical minimum sensitivity requirement for future SETI instruments.

**The SETI League, Inc.** **Radio Telescope Sensitivity Analysis**

(Input variables shown in **Bold** )

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | = | **1420** | MHz; | | | | λ = | 21.1 | cm |
| Eff. antenna diam. | = | **52.5** | meters | = | 172.2 | ft | | | |
| Illum. Efficiency | = | **50** | % | | | | | | |
| Computed Antenna Gain | | | | = | 3.0E+05 | | = | 54.8 | dBi |
| Antenna Half Power Beamwidth | | | | = | 4.1E-03 | radian | = | 2.4E-01 | degrees |
| Drift Scan time (zero declination) | | | | = | 0.9 | min | = | 57 | sec |
| Effective Capture Area | | | | = | 1082.38 | m^2 | | | |
| System Noise Temperature | | | | = | **100** | K | : | -4.6 | dB/To |
| Detector Noise Bandwidth | | | | = | **10000** | Hz | : | 40.0 | dB/Hz |
| Receiver Noise Threshold = kTB | | | | = | 1.4E-17 | W | = | -138.6 | dBm |
| Incident Isotropic Power Threshold | | | | = | 4.5E-23 | W | = | -193.4 | dBm |
| Sensitivity | | | | = | 1.3E-20 | W/m^2 | | | |
| Integration Time Constant | | | | = | **10** | sec | : | 100000 | samples |
| Integration Power Gain | | | | = | 316 | | = | 25.0 | dB |
| Incident Isotropic Power Threshold | | | | = | 1.4E-25 | W | = | -218.4 | dBm |
| Sensitivity w/ integration | | | | = | 4.0E-23 | W/m^2 | | | |

**Table 12.1** Radio telescope sensitivity analysis - "Big Ear" circa 1977.

It is reported that when the "Wow!" signal was intercepted, the gain of the Big Ear radio telescope was roughly equivalent to that of a circular parabolic reflector 52.5 meters in diameter (Dixon, 1995). Wow! discoverer Jerry Ehman has indicated that the equivalent capture area of the antenna was roughly 1000 square meters (Ehman, 1995). At the 21 cm operating wavelength, the two figures correlate well if we assume a dish illuminated at roughly 50% efficiency, which is consistent with a feedhorn system designed to minimize sidelobes and antenna noise temperature (see Table 12.1). The bin bandwidth, noise temperature, and integration time used during reception of the "Wow!" signal are widely reported in the literature, and are also reflected in Table 12.1. It can be seen that the resulting sensitivity of the Ohio State Radio Observatory on 15 August 1977 was on the order of $4 \times 10^{-23}$ W/m$^2$.

The amplitude of the "Wow!" signal is reported as 30 sigma above receiver background noise, for a SNR of +14.9 dB. The peak of the signal was concentrated in a single channel 10 kHz wide. This suggests that the signal's flux density in a 10 kHz bandwidth was 30 times ($4 \times 10^{-23}$ W/m$^2$), or $1.2 \times 10^{-21}$ W/m$^2$. Thus any SETI instrument with a sensitivity exceeding $1.2 \times 10^{-21}$ W/m$^2$ will, in theory, be capable of detecting a repeat of the "Wow!", or any similar signal which it should happen to intercept.

## 12.3 The Project Argu*s* Concept

Recall that a chief limitation of the Big Ear radio telescope (aside from its premature demise) is that it could 'see' only perhaps a millionth of the 4 $\pi$ steradians of space at any given time. Consider that at the 21 cm neutral hydrogen line, a 5-meter diameter parabolic antenna (such as was commonly used in the 1980s and 1990s for C-band satellite TV reception) will have a power gain perhaps 200 times less than that of a 'real' radio telescope such as Big Ear. The reduced capture area would also imply that such an antenna would enjoy 200 times the sky coverage, so a mere 5000 such antennas could, if properly situated, 'see' the whole sky at once. And such a global array of small telescopes could be constructed at a cost on a par with but a single Big Ear.

Unfortunately, this increase in angular coverage afforded by smaller antennas was accomplished by a reduction in their capture area. Hence, they deliver correspondingly less gain. Thus, as compared to our Big Ear example, these smaller antennas will experience a reduction in their effective communications range by that same factor of 200, all else being equal. A signal which could be detected by Big Ear at a range of, say, 20,000 LY, would be detectable to our smaller antennas at a distance of only 100 LY. For uniform distribution of candidate stars, the number of targets varies roughly with the cube of distance, so this sacrifice in sensitivity significantly reduces (perhaps by a factor of several million) the number of suitable stars which might be within range of our sky survey.

We can, however, buy back some of that lost range. It is axiomatic in astronomy that 'there is no substitute for capture area'. It turns out that, in fact, there is: integration time. Most all-sky surveys are performed with the antennas in meridian transit, or drift-scan, mode; that is, fixed in position, the rotation of the Earth bringing candidate stars within range. The narrow beamwidth of a large antenna limits the time which a given candidate signal will spend within its pattern, hence the length of time over which the signal can be integrated. The actual time of accessibility varies with declination, but in the case of the "Wow!" signal equaled 37 seconds at the half-power points (actual integration time used at Big Ear was, at the time, set for 10 seconds.) The proposed 5-meter dish, on the other hand, would, for the same signal, have enjoyed at least 10 minutes of signal duration within its half-power beamwidth. Since sensitivity varies with the square root of integration time, these wider beamwidth antennas can, through signal integration, compensate somewhat for their reduced gain. In this example, integration increases our sensitivity (hence our effective range) by a factor of 8.

Our small dish still falls short of Big Ear's range by a factor of 25. Is there anything else we can do to improve performance? It turns out there is. The state of the art in 1977 was such that the Ohio State Radio Observatory employed a 10 kHz bin width in each channel of its 50 channel filter-bank receiver. Today, digital signal processing (DSP) has advanced to the point that thousands of frequency bins, each a small fraction of a Hz wide, can be readily accomplished at vanishingly low cost. Employing a relatively modest personal computer, 10 Hz bin width was easily achieved with 1990s technology. This level of DSP reduced background noise in the earliest Project Argus stations by a factor of 1000.

Since maximum range varies with the square root of noise power, DSP gave early Project Argus stations greater than a factor of 30 in range improvement, approaching that achieved by Big Ear, circa 1977. In other words, those small amateur SETI stations had a range and sensitivity on a par with that achieved at Big Ear when the "Wow!" was received.

## 12.4 The Project Argus Prototype System

The first prototype instrument in the Project Argus all-sky survey went on the air on Earth Day, 21 April 1996, from SETI League headquarters in New Jersey. Simultaneously, four similar instruments went on the air in North America, Europe and the Pacific, launching one-tenth of one percent of the total proposed *Argus* system. The block diagram of these first instruments is seen in Figure 12.1. They each consisted of a small parabolic reflector of the type then used for satellite TV reception, a cylindrical waveguide feedhorn, a GaAs MMIC (gallium arsenide Monolithic Microwave Integrated Circuit) low noise amplifier, a commercial scanning microwave receiver operated

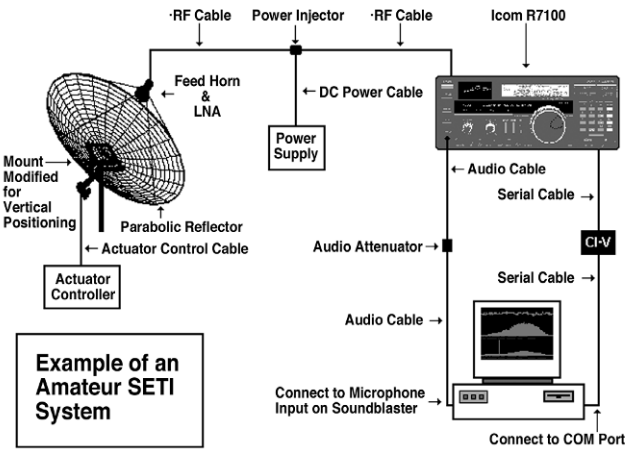



**Figure 12.1** Example of an amateur SETI system.

under control of a personal computer, a computer sound card used as an analog-to-digital converter, and Fast Fourier Transform software for Digital Signal Processing. The cost of each of these systems was a few thousand US dollars.

Although the design standard for Project Argus participants included a 5-meter diameter antenna, the demonstration station built at SETI League headquarters utilized an available parabolic reflector of only 3.7 meters. The first generation LNA utilized a GaAs MMIC with 23 dB of gain, exhibiting a modest noise temperature on the order of 150 K. We conservatively estimate the overall system noise temperature for this station at 200 K, and calculate its sensitivity accordingly. Using 10 Hz bin widths and 10 seconds of integration, the sensitivity of the prototype system is about $5 \times 10^{-22}$ W/m$^2$. This result, reflected in Table 12.2, is marginally adequate for reception of the "Wow!" signal, though it is about an order of magnitude short of the sensitivity achieved by Big Ear, circa 1977. Perhaps there really is no substitute for capture area.

Let's now examine the individual components of the typical amateur SETI receiving station

### 12.4.1 Parabolic reflector

Though many other antenna types have been used successfully, by far the favored antenna for amateur SETI use is the parabolic reflector ('dish'). The chief advantage of the parabolic reflector is that it operates over an extremely wide range of frequencies, limited at the low end by its diameter (which must be a respectable multiple of the longest wavelength being received, to

**The SETI League, Inc.** — Radio Telescope Sensitivity Analysis
(Input variables shown in **Bold** )

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Frequency = | **1420** MHz; | | | λ | = | 21.1 | cm |
| Eff. antenna diam. = | **3.7** meters | = | 12.1 | ft | | | |
| Illum. Efficiency = | **50** % | | | | | | |
| Computed Antenna Gain | | = | 1.5E+03 | | = | 31.8 | dBi |
| Antenna Half Power Beamwidth | | = | 5.9E-02 | radian | = | 3.4E+00 | degrees |
| Drift Scan time (zero declination) | | = | 13.5 | min | = | 808 | sec |
| Effective Capture Area | | = | 5.38 | m^2 | | | |
| System Noise Temperature | | = | **200** | K | : | -1.6 | dB/To |
| Detector Noise Bandwidth | | = | **10** | Hz | : | 10.0 | dB/Hz |
| Receiver Noise Threshold = kTB | | = | 2.8E-20 | W | = | -165.6 | dBm |
| Incident Isotropic Power Threshold | | = | 1.8E-23 | W | = | -197.4 | dBm |
| Sensitivity | | = | 5.1E-21 | W/m^2 | | | |
| Integration Time Constant | | = | **10** | sec | : | 100 | samples |
| Integration Power Gain | | = | 10 | | = | 10.0 | dB |
| Incident Isotropic Power Threshold | | = | 1.8E-24 | W | = | -207.4 | dBm |
| Sensitivity w/ integration | | = | 5.1E-22 | W/m^2 | | | |

**Table 12.2** Radio telescope sensitivity analysis - Project Argus, circa 1995.

provide reasonable gain), and at the high end by its surface accuracy (which must not deviate from the parabolic shape by more than a small fraction of the shortest wavelength being received, to maintain reasonable efficiency). Typical satellite TV dishes generally provide reasonable performance over the 1 to 10 GHz portion of the microwave window.

For reception in the 1.4-1.7 GHz region which is highly favored for much amateur SETI activity, the optimum dish size is on the order of three to five meters in diameter. In countries such as the US and Canada, where C-band satellite television distribution has been widely used for decades, suitable dishes are abundantly available at low to no cost. In other parts of the world they are harder to come by, and enterprising SETI League members have acquired surplus commercial telecommunications dishes, or even built their own from scratch.

The size of the dish and the operating wavelength together determine antenna gain. As a first order approximation, the *voltage* gain (as a ratio) is equal to the circumference of the reflector, measured in wavelengths. Consider, for example, a 3-meter dish, which has a circumference of (3×pi) = about 9.4 meters. At the 21-cm resonant wavelength of neutral hydrogen atoms (corresponding to the popular SETI frequency of 1420 MHz), the voltage gain of this antenna would approach (940/21) ~ 45. Since power ratio equals voltage ratio squared, the power gain of such an antenna would be about 2000, which equates to +33 dBi of gain. (In fact, since the efficiency of amateur SETI antennas is generally in the order of 50%, the actual gain realized is more like +30 dBi.)

Dish size also determines beamwidth, which dictates the degree of aiming precision required when targeting specific stars. As an approximation, half-pwer beamwidth in radians equals wavelength divided by antenna diameter. Thus, for our prior example of a 3-meter dish operated at 21 cm, the beamwidth is in the order of (21/300) ~ 0.07 radians, or 70 milli-radians, which is about four degrees.

If you choose to obtain a surplus antenna, dish condition becomes an important factor. The main consideration here is surface accuracy. In order to perform up to expectations, a dish's surface cannot deviate from the parabolic by more than a tenth of a wavelenght. At 1420 MHz, that's about 2 cm of allowable surface error. If the surface of the dish is dimpled, dented, or distorted beyond 2 cm, avoid that dish! Look for something which approximates a smooth parabolic curve. If panels are missing or bent, performance is going to suffer.

Next, look at the mounting hardware. If it's rusted, you're going to have trouble getting the dish apart, and more trouble reassembling it. Weight is sometimes a consideration, as is wind loading. If these are concerns to you, a mesh dish may prove more practical than a solid one.

Many of the accessories which come with a satellite TV dish will be of limited use for SETI, and therefore you should not pay extra for them. C-band or Ku-band feedhorns and preamps are only useful if you're going to search in C-band or Ku-band (some of our members do; most prefer to scan the Water-Hole, in L-band.) TVRO receivers are great sources of microwave components, but unless other civilizations utilize exactly the same TV transmission standards we do, they're not particularly useful as SETI receivers. And a motorized mount which tracks the Clarke (Geosynchronous) orbital belt is not particularly useful for drift-scan, meridian transit mount radio telescopes, except if modified per the instructions in the Antenna Mounts section below.

In the final analysis, your budget will likely be your chief limitation, so go with what you can afford. Any dish at all will receive better than no dish at all!

### 12.4.2 Antenna mounts

The beauty of mounting a parabolic antenna for SETI use is that you just can't go wrong. Since we are interested in monitoring the sky for artificial signals from beyond, the antenna merely need be pointed up - there are stars (with potentially habitable planets) to be found in all directions. So mounting an antenna for SETI use is considerably simpler than, for example, using the same antenna for satellite TV, where it must be precisely aimed at the satellite's location in the sky.

Because there are no *wrong* directions for SETI, many SETI antennas are simply set on the ground, "bird-bath" style, looking straight up. But a disciplined sky survey, such as The SETI League's Project Argus effort, requires coordinated sky coverage, and that in turn necessitates a limited steering ability for at least some of the antennas in the network.

Where steering of the antennas is desired, we need to consider two degrees of freedom: azimuth (the compass heading to which the antenna points), and elevation (the angle which the antenna's beam makes with respect to the horizon). In terms of celestial coordinates, azimuth of a radio telescope (along with a station's latitude and longitude, and the date and time) determines the Right Ascension (RA) of its target, while elevation (again, along with lat/lon, time and date) determines Declination (Dec).

Since we live on a rotating planet, the Earth itself makes a most cost-effective RA rotor, as long as we are willing to be patient and let the proper portion of the sky eventually rotate into view. But since (thankfully!) the Earth doesn't rotate north-to-south, the only way to acheive Dec control is to physically rotate the antenna along a north-south line. This can be accomplished by aligning a satellite TV antenna's position rotor as a vertical (elevation) rotor.

When antenna rotors are desired, the dish positioners commonly used for satellite TV will require some source of power. The operating voltage for these positioners can often be supplied by a surplus C-band satellite TV receiver, which may also have a digital readout of dish position. Alternatively, a separate DC power supply can be used. The required voltage is typically 24 to 36 VDC, and these rotors draw anywhere from one to four amperes. Polarity determines the direction of rotation, so switching should be provided to move the dish both up and down (or alternatively, both right and left).

### 12.4.3 Antenna feeds

When radio waves strike a dish antenna, the parabolic shape of its reflector directs all the energy to a single point out in front of the dish, called its focus. The purpose of the feedhorn, which is mounted at the focus, facing the reflector, is to scoop up all this energy, and apply it to the LNA and receiver for processing.

The most common feedhorn for amateur SETI use is a metal pipe, closed off at the end farthest from the dish, forming a shorted cylindrical waveguide. The horn contains a small metallic probe, connected to the center pin of a coaxial connector, to collect the energy and apply it to the input connector of the LNA. The horn may be surrounded by a metal ring, used to improve the efficiency of energy collected from the surface of the dish, or to block interference from entering the feed from beyond the periphery of the dish.

The chief drawback of the cylindrical waveguide feedhorn is that its large physical size actually blocks a part of the dish surface from view of its incoming signals, effectively reducing the size (and hence the gain) of the parabolic antenna. This blockage loss is most severe for small dishes, becoming almost negligible at the popular 1.4 to 1.7 GHz SETI frequencies when the dish diameter exceeds about 4 meters.

An alternative to the waveguide feedhorn is the helical feed, consisting of about three turns of heavy wire in a corkscrew shape, with a circumference of one wavelength at the operating frequency, and a spacing between turns of a quarter wavelength. A helix feed doesn't block the aperture of the dish to the extent that a waveguide horn does, but is more prone to interference from signals off to the side of the antenna. Both helix and waveguide feedhorn designs have been used successfully by SETI League members.

### 12.4.4 Low noise amplifiers

The Low Noise Amplifier, or LNA, is sometimes called a preamplifier, or preamp. Its function is to turn an impossibly weak signal into a merely ridiculously weak one. The critical parameters to consider in selecting an LNA are its frequency response, gain, and noise temperature.

Frequency response determines that portion of the electromagnetic spectrum over which a particular LNA will boost the received signal, with minimum distortion or added noise. You should select an LNA with a frequency range consistent with your particular SETI station requirements.

For example, C-band Satellite TV LNAs cover the frequency range of 3.7 to 4.2 GHz. Thus they are *not* suitable for use in SETI stations designed to monitor the 1.4 GHz hydrogen line. Some LNAs incorporate filtering, which reduces the overall range of frequencies amplified, but which can help to reduce out-of-band interference.

Gain, measured in deciBels (dB), is a measure of how much the LNA boosts the incoming signal. Although in many things 'if a little is good, a lot is better', this is *not* the case for preamplifier gain. In fact, excess LNA gain can actually reduce the sensitivity of your SETI receiver. The rule of thumb is that the gain of the LNA should equal the sum of the microwave receiver's noise figure (in dB) plus the RF cable insertion loss (also in dB), plus an additional 10 dB. For the average SETI station with a short coaxial cable between the LNA and the receiver, 20 dB of preamp gain is usually about right. If a very long or unusually lossy RF cable is used, a 30 dB gain LNA might be more appropriate.

Noise temperature is a measure of how much additional noise the LNA adds to your SETI system. Since any actual signal has to compete with a variety of natural and artificial noise sources, the lower the noise temperature, the better. The LNAs commonly used for amateur SETI typically have between 35 Kelvin and 100 Kelvin of internal noise. Noise is sometimes expressed not in Kelvins, but as Noise Figure (in dB) or Noise Factor (a unitless power ratio).

Many commercial LNAs are provided with a choice of coaxial input and output connectors. Most SETI League members prefer to standardize on the coaxial connector known as Type N, since this is the connector used on most feedhorns and microwave receivers. To minimize losses, the LNA should be mounted directly on the output connector of the antenna feedhorn, with the appropriate coaxial adapter (probably a Type N male-to-male barrel adapter).

An additional consideration is how to get the appropriate operating potential to the LNA. Most LNAs operate from a DC power supply, typically in the +12 VDC range. Some designs require that this operating voltage be applied via the center-conductor of the RF cable, and some LNA vendors give you a choice between internal and separate DC feed. DC feed via the transmission line requires that the microwave receiver be designed to provide this voltage, or that an accessory called a DC Inserter, or Bias Tee, be connected into the signal path ahead of the receiver, and tied in to an appropriate power supply. Although this is the scheme commonly used to power the antenna-mounted circuitry in commercial satellite TV receivers, many SETI experimenters prefer to run a separate DC cable (such as a telephone cable, speaker cable, or lamp cord) outside to the LNA, and to apply the required DC potential to it inside the SETI station. (Caution: double-check the polarity applied to this cable, as reversing the positive and negative power supply leads can damage the LNA. The center pin of the LNA's power feedthru capacitor is typically positive.)

Although most commercial (and many home-built) LNAs are metal-boxed to provide good shielding against Radio Frequency Interference (RFI), few are provided in weather-proof enclosures. To prevent damage from exposure to the elements, I like to put my LNAs in plastic Tupperware® sandwich boxes. It is necessary to drill or punch holes in the plastic for the input coax adapter, output cable, and power wiring. Be sure to seal these openings with room-temperature vulcanizing (RTV) silicon rubber, which you can obtain in a tube from most hardware stores.

### 12.4.5 RF cables

The most common SETI station configuration would place the microwave receiver, signal analysis computer and related accessories inside the house, with the antenna and LNA mounted outside, some distance away. To connect the two halves of a SETI station, we use an RF cable.

RF stands for radio frequency. The cables we use are usually coaxial (i.e., "coax" cable), and we prefer those with low loss at radio (specifically microwave) frequencies. The stuff used for cable TV is cheap (pennies per meter) but pretty lossy in the 1.4 to 1.7 GHz region of the spectrum typically used for amateur SETI. The kind you buy for, say, CB radio antennas is a little better, and a bit more costly. If you have a local Radio Shack store or similar, you can probably find what they call low-loss coax - it's larger (perhaps 1 cm diameter) than the CB or TV type, costs maybe a dollar or more per meter, and may go under such part numbers as Belden 9913, RG-8 Polyfoam, etc. It may take special connectors (the ones most of us use are called "Type N"), which require some experience to properly install.

For any type of coax, the longer the lossier. So we try to keep our antennas near the radio room. If this is not practical, we can do several things: use more gain in the preamp (to boost the weak signal before it suffers cable loss); follow the LNA with a satellite TV line amplifier; mount the whole receiver, or just the downconverter, outside on the dish (pumping a lower frequency through the cable is more efficient); or use specialized cables such as hardline or Andrew Heliax® (which can cost upwards of tens of dollars per meter).

### 12.4.6 Bias tee

Most low noise amplifiers used for amateur SETI and radio astronomy operate from a DC power supply, typically in the +12 VDC range. Some designs require that this operating voltage be applied via the center-conductor of the RF cable. DC feed via the transmission line requires that the microwave receiver be designed to provide this voltage, or that an accessory called a DC Inserter, or Bias Tee, be connected into the signal path ahead of the receiver, and tied in to an appropriate power supply.

The bias tee is typically a small box with two coaxial connectors, and a place to hook up the LNA's required DC operating voltage from a power supply. One of the connectors is marked "from LNA" or "RF + DC", and will have DC appearing on its center pin. The other connector is marked "to receiver" or simply "RF", and will not have DC on it. Caution: It is essential that you hook the bias tee into your system the right way around, or damage to your microwave receiver could result.

The internal circuitry of the typical bias tee is rather simple. There's usually a 50 ohm microstrip transmission line connecting the two coaxial connectors, with a chip capacitor mounted in the middle to serve as a DC block. A radio frequency coil (rf choke) carries DC from a feedthru capacitor to one side (the LNA side) of this microstrip. To the feedthru capacitor, the user connects the LNA's operating potential (typically +12 VDC) from a lab power supply, or a "wall wart" type power adapter such as may be found at Radio Shack ® and other electronics retailers. Caution: It is essential that you observe proper polarity when connecting a power supply to the bias tee, or damage to your LNA could result.

A commercial bias tee favored in many amateur SETI stations is manufactured by Down East Microwave as their Model BT, and sells for $35 US (plus appropriate postage). The unit uses type N female connectors. It is compatible with LNAs from various commercial vendors, and is relatively foolproof - as long as attention is paid to correct power supply polarity, and to the direction in which the unit is inserted into the transmission line.

### 12.4.7 LNA power supplies

The low noise amplifier mounted at your SETI antenna feed operates from a DC power source, typically in the +12 VDC range. Whether you apply operating potential to the LNA via a bias tee through the center conductor of the coaxial cable, or directly via a separate power cable, you will need to obtain an appropriate power supply.

Although regulated laboratory power supplies are indeed suitable (and in fact are used in The SETI League labs to test LNAs), they are rather expensive. An alternative is to use a 'wall wart' type power supply which plugs directly into the household AC mains. The ripple from such inexpensive power supplies is admittedly higher than we'd like, but quality LNAs typically have on-board 3-terminal voltage regulators (12 VDC in, down to 5 VDC for the FET) which brings power-line hum down to negligible levels.

### 12.4.8 Microwave receivers

The microwave receiver takes a small, selected portion of the radio spectrum, and converts it to audio for signal analysis. Selection of the appropriate receiver leaves more to the discretion of the experimenter than any other portion of the amateur SETI system. Four distinct options present themsleves. In descending order of cost, they are:

1. **High-end microwave scanning receivers**. These units (typified by the Icom models R-7000, R-7100, and R-8500, as well as the AOR 3000 and 5000) are multi-mode receivers which can receive AM, FM, CW, SSB, and sometimes video and digital modes. Various IF bandwidths are usually available, and these receivers are normally programmable to scan a selected range of frequencies. They typically tune from a few hundred kHz all the way up to about 2 GHz, which actually exceeds our SETI needs. Prices are likely to start around $2000 US, making these receivers as expensive as all other portions of an amateur SETI station combined.

2. **Modified radio-telescope receivers**. One of the very few vendors of commercial radio astronomy receivers for the amateur market is Radio Astronomy Supplies. Their microwave receivers, which are designed specifically for continuum radio astronomy (that is, searching for natural astrophysical phenomena), can sometimes be modified for SETI use. Such modifications generally require considerable electronics expertise, but offer the ultimate in performance.

3. **Computer-controlled receivers**. The first generation were built on ISA cards, and plugged directly into one of the vacant slots on the motherboard of a personal computer. These units were prone to radio frequency interference generated by the computer itself. Later units, like the Icom PRC1000 and WinRadio 1500e, are separate boxes which plug into a computer via a serial, parallel, or USB port. They have many of the features of the high-end microwave scanning receivers, but since they rely on a companion computer for digital control, typically cost half as much.

4. **Downconverter/receiver combinations**. Several converters are available to shift a selected portion of the microwave spectrum down in frequency, for reception in a shortwave or VHF ham radio receiver. Popular units are available from Down East Microwave in the US, and VHF Communications in Europe. Downconverters are appealing for those who already own a high-performance communications receiver, which unfortunately doesn't tune to the SETI frequency of interest. Downconverters cost about half as much as the computer-controlled receivers, but require the user to couple them to an existing receiver.

Whichever receiver scheme is selected, present practice suggests operating it in single sideband mode (either USB or LSB), and leaving it fixed-tuned, rather than scanning it across the spectrum. The reason for avoiding frequency scanning is that the Earth is turning the antenna continually, so that the spatial dimension of the observation is always changing. Only by holding frequency constant for at least one rotational period of the Earth (that is, one day) can we avoid the problem of "too many variables."

The bandwidth of the receiver's audio stages will typically be the limiting factor, as far as instantaneous frequency span is concerned. Many SSB receivers cover as little as 3 kHz of spectrum at a time, which is an inefficient way to search for ETI. Advanced SETI experimenters sometimes modify their receivers for up to 22 kHz of instantaneous IF and audio bandwidth, while custom-built receivers can cover several hundred kHz all the way up to a few MHz of spectrum at a time.

### 12.4.9 Audio connections

Except for those specialized computer-controlled receivers which couple all signals directly to a personal computer through its serial or parallel port, an amateur SETI station requires some kind of audio interface between the receiver and the computer's sound card (see www.setileague.org/hardware). Some experimenting is usually required, since a receiver may have one of three kinds of audio output port, and a sound card may have two different types of audio input port.

The audio output options on a microwave receiver are Line Out, Speaker, and Headphones. The Line Out jack provides a low-level, high-impedance audio signal, which (when available) is usually the best option for interfacing the receiver to a computer sound card. Speaker and Headphone are generally both low impedance interfaces, with the former usually providing a higher amplitude signal than the latter. Since sound cards are prone to overload and distortion from high-amplitude signals, the Headphone output is generally preferred.

Computer sound cards typically sport Microphone and Line audio input connections. The Line In port is almost always the right choice for interfacing a receiver, as the audio levels present will usually overload the sound card if the Microphone connector is used. If the receiver has a Line Out connection, that will probably best match the sound card's Line In port.

Especially if the Microphone input on the sound card must be used, the receiver's audio level may be sufficient to overdrive the sound card, causing distortion. This problem can be alleviated by placing an audio attenuator circuit in the audio line between receiver and computer. Although you can purchase attenuators at a stereo shop, or build them yourself out of resistors, probably the simplest solution is to obtain an attenuating audio cable from your local stereo shop. These cables, which have built-in resistive attenuators, usually employ shielded wire and high quality plugs, and are used to patch the high-level outputs of turntables, tape decks and CD players directly into the low-level inputs of stereo preamplifiers.

If you choose not to use an attenuating audio cable, be sure to connect your receiver to your computer via a well-shielded audio patch cord. Lamp cord, telephone wire, and speaker wire should be avoided, as these may make your system susceptible to electromagnetic interference and hum.

### 12.4.10 Digital signal processing

The audio output from the SETI receiver is mostly noise, both natural and artificially generated. If we are very lucky, there may be buried somewhere in that noise an intelligently generated signal of extra-terrestrial origin. But it's likely to be buried so deep in the noise that no human sense can detect it. To separate the cosmic wheat from the galactic chaff, we employ a technique known as Digital Signal Processing, or DSP.

The first step in the DSP process is to feed the receiver's audio output into the computer, in a form which the computer can recognize - that is, as binary data. We need an analog to digital converter (ADC) to accomplish this, and the ADC of choice for amateur SETI is the PC Sound Card. Just about any SoundBlaster® compatible audio card will work with The SETI League's signal analysis software. These cards sample an audio waveform at least 44,000 times per second. One of the rules of information theory is that to digitize a signal, it must be sampled no less than twice for every cycle at its highest frequency. With 44 KSPS (kilo-samples per second) sound cards, for example, this means we can digitize and analyze audio components out of our receiver up to 22 kHz in frequency. With higher sampling rates, of course, wider baseband segments can be analyzed in software.

### 12.4.11 Control cabling

In addition to analyzing signals, some SETI League computers also control the station. Remember the computer-controlled radios discussed in the Microwave Receivers section above? They can often be tuned by software, driven from the PC's serial, parallel or USB port. Antennas can similarly be computer-aimed, if they are equipped with software-driven azimuth and elevation rotors. Some SETI computers make lights ring and bells flash whenever they detect something interesting. And the most advanced of the computers used by SETI League members also dial into the internet when an interesting candidate signal is received, automatically alerting other participants that their assistance in signal verification is required.

Some SETI receivers are designed to operate solely under computer control. For several others, it is an option. A few commercial receivers (notably the Icom model R-7000 series) require an accessory interface box to convert the receiver's control lines to RS-232 levels for connection to a computer's serial port. And, some SETI signal analysis software packages are designed specifically to take advantage of this interface capability.

Control cabling is largely optional, because computer control of the complete SETI station is presently practiced only by our most advanced members, and even by them, it is still experimental. Except when using a computer-tuned receiver or Software Defined Radio (SDR), all the necessary control functions can be performed manually, and usually are. But automation techniques will become more important in the years ahead, as more receivers, antennas, and networked signal verification protocols emerge which call for computer control.

### 12.4.12 Signal analysis computer

Even the simplest of today's personal computers is thousands of times more powerful than the ones NASA used to put men on the moon. Of course, the objective of SETI is not to reach the moon, but rather to reach much farther out into space for intelligently generated signals. To do so, we employ a technique known as Digital Signal Processing, or DSP.

The audio signal applied to the computer's sound card (if that's what you choose to use for Analog to Digital conversion) is likely to be so narrow in bandwidth that even an antique, 486-class computer can analyze in real time, with excellent resolution. The typical shareware DSP program chops the received audio band up into 2048 or 4096 individual channels, each about 10 Hz wide. The computer permits analyzing and displaying all those channels simultaneously, in real time. Thus, the computer turns the SETI station into a 2048 or 4096-channel receiver. This is, admittedly, a far cry from the millions and billions of channels analyzed by high-end research grade Multi-Channel Spectrum Analyzers (MCSAs) described in other chapters of this book. Then again, it is a quantum leap beyond the single-channel receiver used by Frank Drake for *Project Ozma*!

Much of the signal analysis software developed by SETI League members runs under various versions of the Microsoft Windows operating systems. It is shareware, offered at low or no cost to all participating SETI league members via the Software pages of The SETI League website. Its job is to identify signals which exhibit the hallmarks of artificiality, characteristics which distinguish it from natural phenomena, and then to help determine whether those characteristics might have come from some terrestrial source.

Our civilization pollutes its own radio environment, so we need to sift through any detected signals rather thoroughly in order to rule out manmade interference from our own transmitters, aircraft, spacecraft and orbiting relay stations. For example, the first candidate signal received by a Project Argus station, in May of 1996, is depicted in Figure 12.2. Upon close examination,

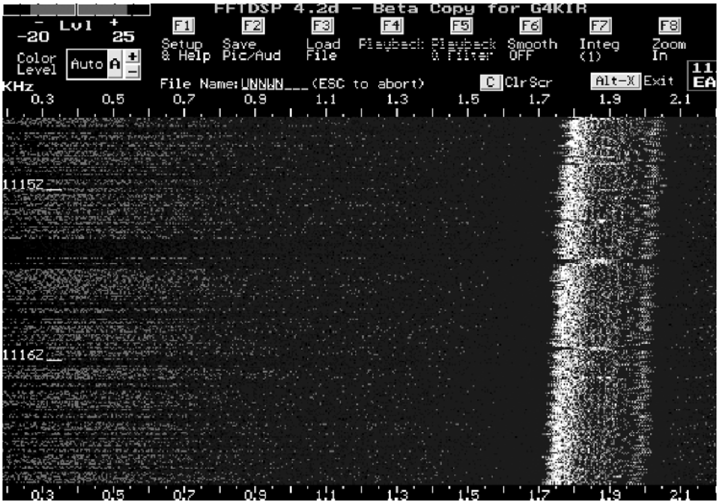



**Figure 12.2** The first candidate signal received by a Project Argus station, Mary 1996. It turned out to be interference from a classified military satellite.

it turned out to be interference from a classified military satellite. Determining this would have been challenging for the human observer, but was a trivial identification task for our computers.

## 12.5 Tinkering at the Margin

As I've already mentioned, there really is no substitute for capture area.

Or is there?

By making incremental improvements to the *Argus* station, we have, over the past decade, raised its sensitivity the dozen dB or so necessary to equal the performance of Big Ear when the Wow! was detected.[2] It is highly unlikely that these small stations will ever have the sensitivity of the world's great radio telescopes, given equivalent technology. However, by using the Wow! as a benchmark, we hope to show that even these small stations have sensitivity adequate for SETI success.

2 It should be noted that, in the intervening years, the sensitivity of the Ohio State Radio Observatory was also significantly improved. Such is the nature of technological progress. However, in 1997, that grand radio telescope was demolished, to make way for a commercial golf course. This, too, speaks of the nature of progress.

**The SETI League, Inc.** **Radio Telescope Sensitivity Analysis**
(Input variables shown in **Bold** )

| | | | | | | |
|---|---|---|---|---|---|---|
| Frequency | **1420** | MHz; | | λ | 21.1 | cm |
| Eff. antenna diam. | **5** | meters | 16.4 | ft | | |
| Illum. Efficiency | **50** | % | | | | |
| Computed Antenna Gain | | | 2.8E+03 | | 34.4 | dBi |
| Antenna Half Power Beamwidth | | | 4.3E-02 | radian | 2.5E+00 | degrees |
| Drift Scan time (zero declination) | | | 10.0 | min | 598 | sec |
| Effective Capture Area | | | 9.82 | m^2 | | |
| System Noise Temperature | | | **100** | K | -4.6 | dB/To |
| Detector Noise Bandwidth | | | **10** | Hz | 10.0 | dB/Hz |
| Receiver Noise Threshold = kTB | | | 1.4E-20 | W | -168.6 | dBm |
| Incident Isotropic Power Threshold | | | 5.0E-24 | W | -203.0 | dBm |
| Sensitivity | | | 1.4E-21 | W/m^2 | | |
| Integration Time Constant | | | **120** | sec | 1200 | samples |
| Integration Power Gain | | | 35 | | 15.4 | dB |
| Incident Isotropic Power Threshold | | | 1.4E-25 | W | -218.4 | dBm |
| Sensitivity w/ integration | | | 4.1E-23 | W/m^2 | | |

**Table 12.3** Radio telescope sensitivity analysis - Project Argus, circa 2000.

If we increase our antenna size to 5 meters in diameter (entirely consistent with our goal of achieving real-time full-sky coverage with 5,000 instruments), we improve system sensitivity by about 2 1/2 dB. A second generation preamplifier developed in 2000 utilizes a GaAs PHEMT (pseudomorphic heterojunction microwave transistor) in front of the existing MMIC stage, to lower preamp noise temperature to below 50 K. This noise reduction made an overall system noise temperature of 100 K or less entirely feasible, buying Project Argus participants another 3 dB of sensitivity. But the most dramatic improvement with these small terminals comes from integration gain, facilitated by their relatively wide beamwidths.

As shown in Table 12.3, the beamwidth of a 5-meter dish is such that, when operated in meridian transit mode, a signal will remain within its beam for no less than 10 minutes. By integrating for, say, 2 minutes, we can achieve five samples of any signal transiting our beam, which is indeed sufficient to trace out the pattern of the antenna over time.[3]

Of course, a 120-second integration time will only yield the promised improvement if the candidate signal is a pure continuous wave (or at least confined to a 10 Hz bandwidth), and present for a sufficiently long period.

[3] Since sensitivity varies with the square root of integration time, we can see that another roughly 5 dB improvement in sensitivity can be had relatively easily.

This is speculative at best. However, returning to the "Wow!" signal as a benchmark, even though its exact nature remains unknown, we do know that its bandwidth did not exceed 10 kHz , and that it had a sustained duration of at least 37 seconds. Let us hope that either nature, or other civilizations, will provide us with other interesting signals of like amplitude, and sufficiently narrow bandwidth.

There exists a potential incompatibility between desired integration time constant and bin width, in view of the anticipated Doppler shift on signals emanating from, and received on, rotating planets. It is common practice in SETI to chirp the receiver's local oscillator so as to correct frequency to the Galactic Standard of Rest, in hopes that a transmitting civilization will do the same. Software is currently under development to allow for computer tuning of the Project Argus receivers. Of course, this places a burden on the transmitting civilization, which we can only hope they will choose to shoulder.

When we're done tinkering at the margin, we see that we have the potential, utilizing today's readily available technology, to receive with an amateur SETI station a CW signal at a power density as low as $4.1 \times 10^{-23}$ W/m$^2$. This is on a par with the sensitivity of Big Ear, when the Wow! signal was detected. Whether this level of SETI sensitivity is adequate to bring us the existence proof we seek, only a fully implemented Project Argus network, and an indeterminate period of patient observation, can disclose.

## 12.6 Growth and Stagnation

The SETI League's initial goals for the activation of a global radio telescope network proved unrealistically optimistic. Four years after the launch of Project Argus, a mere 100 stations had come online. From there, project growth leveled off, and today total participation sits at just 144 stations in 27 countries, far short of the number necessary to achieve the stated goal of real-time full-sky coverage. We attribute this disappointing rate of growth to five factors:

**1. Global economic downturn.** Project Argus was initiated during a time of unsustainable economic growth, especially in the computer and technology sectors. The bursting of the dot com bubble at the end of the 20th century was followed by terrorist attacks at the beginning of the 21st, warfare on several fronts, market meltdowns, and a general reduction in discretionary capital for all but the wealthiest inhabitants of Planet Earth. Since, for privatized SETI, participation is at the hobby level for most, it was one of the first expenditures curtailed by many SETI League members in a declining economy. Similarly, the modest grant funds attracted by The SETI League from industry, governments, and nongovernment organizations dried up as other priorities were competing for limited resources.

**2. Lack of turn-key systems.** The earliest Project Argus stations were pieced together out of a combination of surplus and purpose-built electronics equipment, by a highly-trained and talented cadre of electronics experimenters (many with lifelong backgrounds in amateur radio). Globally, persons with these talents and abilities number in the low thousands, and The SETI League succeeded in attracting a significant number of them into our membership ranks early on. The explosive growth required for full-sky coverage would have had to draw from an entirely different population of enthusiasts, ones who would be depending upon off-the-shelf hardware, which we expected to become commercially available at modest cost. Unfortunately, even the hoped-for 5000 stations did not afford the economies of scale necessary for equipment manufacturers to commit to volume production. Early on, several small companies did indeed bring antenna feeds, low noise preamplifiers, microwave receivers, downconverters, and related products to market. Within just a few years, these products were all abandoned by their manufacturers when market volume proved disappointing. At one point, the author approached the largest electronics retailer in the US, offering for their private label production a proven turnkey system design on a royalty basis. Their (understandable) response was that, unless we anticipated a market volume of 1 million units per year, they were not interested.

**3. Negative economies of scale.** For their antennas the first Project Argus stations used decommissioned C-band home satellite television parabolic reflectors on the order of 3-5 meters in diameter. As the US, Canada, Europe, and Australia made the transition to Ku-band digital satellite television in the 1990s, these antennas briefly became abundantly available at low to no cost. However, with C-band analog TVRO services now largely phased out, the primary market for these antennas disappeared. Many of the established antenna providers abandoned this product line, and today, their availability even on the retail market is limited. Thus, over time and with the success of Ku-band direct broadcast services, the cost of the required antennas has failed to decrease, and in fact has risen sharply.

**4. No new "Contact" films.** The release in 1997 of the Warner Brothers film "Contact," based upon the science fiction novel by Carl Sagan, did much to publicize SETI, and resulted in explosive growth in public support and participation. In the months following the release of the film, SETI League membership and Project Argus participation both doubled. Unfortunately, within a year this public interest began to wane, and most of the new members acquired post-"Contact" failed to renew their membership. Another blockbuster movie depicting SETI science in a positive light would have been beneficial, but was not forthcoming.

**5. Computers are easier.** Shortly after the launch of Project Argus, Professor Woody Sullivan of the University of Washington presented at a Bioastronomy conference in Italy a concept for distributed processing of signals being gathered by the UC Berkeley SERENDIP project, long ongoing at the Arecibo Observatory. Implemented by Dan Werthimer and his colleagues at Berkeley, this proposal quickly grew into the well-known SETI@home project, detailed elsewhere in this book. SETI@home has attracted several million participants worldwide, growing into the world's most powerful supercomputer.

This experiment, though highly successful from a computer science standpoint, has proved a double-edged sword for observational SETI. On the one hand, it has raised public interest and awareness to unprecedented levels. On the other hand, it is far cheaper and easier to allow an idle home computer to crunch data than it is to build and operate an actual radio telescope. As a consequence, many potential Project Argus participants opted out, deciding instead that SETI@home participation was an acceptable contribution to SETI science. While we are delighted at the success of this UC Berkeley initiative, it is an unfortunate reality that several million participants are now analyzing data from a single radio telescope, operating in a single frequency band, which, though highly sensitive, can see only a minute fraction of the sky.

## 12.7 Where Do We Go From Here?

Since it appears unlikely that Project Argus will ever realize its full potential, or grow to meet its initial optimistic projections, one must ask what can realistically be expected for the future of this project in particular, and amateur observational SETI in general. I'd like to suggest five guidelines for this and future SETI League initiatives:

**1. Scale back expectations.** It is clear that the goal of real-time all-sky coverage is unrealistic. Instead of continuing to pursue it, Project Argus participants should consider what might be accomplished with the network of stations as it now exists. Although disappointingly short of our initial goal, 144 observatories still represent more operational radio telescopes than exist in the rest of the world combined (and is likely to remain so, until the next phase of the SETI Institute's Allen Telescope Array is funded, constructed, and activated).

Even with this fractional Project Argus network, given proper coordination of declination assignments, we can still create a drift scan array which is capable of monitoring the entire 4 pi steradians of space, for tens of minutes every day. Though not our initial goal, this scaled-back project still promises to make significant contributions to SETI science.

**2. Do more with less.** Even at its present modest level, we can expand the capabilities of our existing stations, so that they cover more search space. One promising area in which our reach has continued to expand is in the area of frequency coverage. Whereas Project Ozma, the world's first modern observational SETI experiment (circa 1960) monitored a single channel a mere 100 Hz wide, the first Project Argus stations were able to monitor about 1000 bins of 10 Hz each, for a total spectral coverage of several kilohertz.

In the years since the program's inception, SETI League members have increased both the bandwidth of their receivers and the resolution of their digital signal processing algorithms. Thus, the spectral coverage of Project Argus stations is now extended from the tens of kilohertz, into the tens of Megahertz. Though this is still far from the hundreds of Megahertz monitored in real time by the highend, purpose-built multi-channel spectrum analyzers developed by NASA and used at The SETI Institute, Moore's Law suggests we can expect to see amateur capabilities continuing to close in on the state of the art.

**3. Attract new constituencies.** Not surprisingly, given its origins as a ham radio club, The SETI League early on concentrated upon attracting those radio amateurs and microwave experimenters already skilled in the arts required of SETI observers. This constituency represents an extremely limited pool of admittedly high-level talent. SETI@home has already proven that the masses are seeking an appropriate level of SETI participation. If SETI populism is to succeed, we need to explore projects that can be pursued by enthusiasts of more modest technical background.

**4. Embrace new technologies.** The traditional SETI paradigm involves high-gain microwave antennas, sensitive receivers, and powerful digital signal processors. In half a century of observation (perhaps 1000 separate experiments conducted in dozens of countries), we have yet to detect a single clear, unambiguous signal of decidedly intelligent extraterrestrial origin. Perhaps it's time to seek a new approach to SETI.[4] This is an area in which The SETI League's members can provide leadership, given the highly interdisciplinary nature of our membership, and considering that enthusiastic amateurs are not personally or professionally committed to maintaining the status quo.

**5. Forego instant gratification.** We are beginning now to realize that SETI success is unlikely to come quickly or easily. We are, after all, in our technological infancy; a thorough SETI search may take generations.

---

4 What that new approach might be, I can't even begin to imagine.

SETI is not a science that offers much to those persons who demands instant gratification. To date, not only have we not yet scratched the surface; we haven't even felt the itch. If we do the search, and we do it right, sometime in the distant future we will have arrived at one of two possible conclusions: either we are not alone in the cosmos, or we are. Either possibility boggles the imagination.

## 12.8 Conclusions

During the last half-century, SETI has emerged out of the realm of science fiction and into the scientific mainstream. Every month we read about the discovery of yet another planetary system in space. We are beginning to learn about how life might have developed on other worlds. And we have completed the Copernican Revolution, finally realizing that we are not the center of all creation. Yet SETI programs continue to yield null results. Still, we must not allow ourselves to become discouraged. Humans have possessed suitable technology for less than an eyeblink on the cosmic timescale.

The non-profit, membership-supported SETI League launched its search on Earth Day, and flies the Flag of Earth, because SETI is an enterprise which belongs not just to one country, government or organization, but to all humankind. Like *Argus*, the guard-beast of Greek mythology who had a hundred eyes, we seek to see in all directions at once, that we might capture those photons from distant worlds which may well be falling on our heads even now.

Project Argus started with a mere five stations. This small step for humanity represents a humble beginning for what will ultimately be a global effort. At this time of writing, we have stabilized (some might say stagnated) at a mere 144 operational stations. Our long-term goal remains activating 5000 observing stations around the world, properly coordinated via the Internet to achieve full sky coverage, all the time, in all directions, 23:56/7. When we reach that level, there will be no part of the sky which evades our gaze. Then we can hope to find the answer to a fundamental question which has haunted humankind since first we realized that the points of light in the night sky are other suns: Are We Alone?

## References

Dixon, Robert S. 1995. A History of the Ohio SETI Program. *SETIQuest* 1(3): 3-6.

Ehman, Jerry. 1995. Personal conversation with the author, at Ohio State Radio Observatory, 19 August 1995.

Shuch, H. Paul. SETI League Technical Manual. Online at http://www.setileague.org/hardware/techmanu.htm

# 13

# Gravitational Lensing Extends SETI Range

Richard Factor,
*President, The SETI League, Inc.*

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

## 13.1 Background

Since the first tentative SETI experiment in the 1960s, increasingly larger radio telescopes and more powerful signal processing engines have been searching the sky for signals. Perforce these searches have been limited to looking largely for continuous or pulsed narrowband signals since these are the most likely to be detectable, and are most identifiable as being of unnatural origin. A number of ‘hits’ have been recorded, beginning with the famous “Wow!” signal and continuing to the present. After weeding

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_13, 

out cases of equipment problems and man-made interference, a number of candidate signals remained, any of which might have been of intelligent origin. None of them could be proved to be so, largely because they were not verifiable. Revisiting the signal's supposed point of origin failed to provide a repetition of the event, leaving the original signal as a tantalizing but scientifically useless phenomenon. In order for a detected signal to be accepted as of intelligent, extraterrestrial origin, it is generally agreed that it must meet two criteria:

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

Of course, it would be desirable for the signal to have, somehow, a modulation that would impart information to the observer. An on-off modulation in some obvious pattern such as sequential prime numbers would comfortably fulfill this desirable but not-strictly-necessary characteristic. With several decades of sometimes fitful, sometimes diligent searching, the results can be summed up in two sentences: We know that the sky isn't teeming with strong signals. And we have searched such a small percentage of the phase space that it would be foolish to conclude there's nothing to be found.

## 13.2 Searches

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

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

## 13.3 Targeted Search

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

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

## 13.4 All-sky Survey

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

Assume that one desires to cover the entire sky with as much sensitivity as possible. With appropriate location of the observatories, one could accomplish this with approximately 5000 “small” dishes on the order of 3-5 meters in diameter. This is the essence of the SETI League’s Project Argus, a ‘grass roots’ endeavor. There are literally millions of these dishes in the hands of TV watchers, at least in the United States and, due to the advent of DBS satellites, many of them are available for the price of carrying them away. Assuming the economic cost of recommissioning each dish is on the order of $1000, the antennas for the all-sky survey come in at only $5 million. This is a pittance compared to even the cost of a single research-grade radio telescope. However, the economics of scaling is very unfavorable. For instance, to only double the distance at which a given signal can be detected, one would need to double the diameter of the antenna, putting it in the 6-10-meter range. Because these dishes are no longer littering the landscape, they must bear their actual economic cost, on the order of $10,000 each. Just as bad, doubling the diameter halves the beamwidth in two dimensions, raising the required number of dishes to 20,000. Thus, doubling the sensitivity increases the cost from $5 million to $200 million. Doubling the sensitivity yet again requires 80,000 12-20-meter dishes, at perhaps $50,000 per copy.

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

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

## 13.5 Common Requirements

The two strategies were discussed without regard to the electronic instrumentation necessary. As divergent as the antenna requirements are, the receiver and signal detector requirements are very similar. For a research-grade radio telescope, the cost of the mechanical system is so high that any reasonable electronic detection ensemble has a cost - you should forgive the expression - in the noise level. This is emphatically not the case in the all-sky-survey scenario, in which the electronic requirements of the receiver and data reduction hardware can equal or exceed the cost of the antenna, and yet come nowhere near the capability of the larger instrument's electronics. Fortunately, there is great cause for optimism!

While the cost of constructing mechanical hardware increases slowly with time, the cost of constructing electronic hardware plummets with Moore's law. At the moment, professional electronic hardware exceeds amateur capability by perhaps a few dB in sensitivity (disregarding antenna size), two orders of magnitude in stability, and three to four orders of magnitude in frequency coverage. Advances in DSP in particular, as well as improvements in semiconductors and other technology, are likely to make today's professional capabilities within the reach of amateurs in only a few years.

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

## 13.6 A Third Strategy

The purpose of this extensive background discussion was to examine the implicit assumptions and requirements of both kinds of searches. Each has a distinctive vulnerability. If there are no nearby civilizations, the targeted search will fail. No matter how many civilizations there are, if nobody is transmitting a beacon the all-sky survey is unlikely to detect any of them. What if we happen to be in a deserted neighborhood? Too bad.

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

The phenomenon of gravitational lensing, a consequence of general relativity, is scientifically accepted and has proved a valid astronomical and astrometrical tool. A gravitational lens occurs when electromagnetic radiation passes a massive astronomical object such as a star or even a galaxy. Because of the large area of signal "collected" by the lens and the potentially small area of its focus, enormous signal gain is possible. Claudio Maccone has written a treatise on the subject, stating that our own star would have a gravitational focus at about 550 AU, allowing a spacecraft at this distance to take advantage of this lens to provide signal gain greater by far than that of the Arecibo dish. One of the purposes of this spacecraft would be to look for signals of intelligent origin. Sadly, most of us do not have our own space program and therefore cannot rely on the Sun to supplement our antenna. Is all lost?

No!

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

Since the focus is a line, it follows that this effect can be employed at any distance beyond 550 AU. While we have no immediate prospect of going

550 AU from the Sun, we are already more than 550 AU from every one of the billions of stars in our galaxy! Therefore, at any given time we could be in the line focus of some other star's gravitational lens, and could be receiving some other civilization's signals with relatively modest equipment. Perhaps we have already done so. One reasonable (but entirely conjectural) explanation of the SETI "hits" that we've received over the decades is that it was a transient gravitational lensing phenomenon.

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

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

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

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

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

The electronic package of a typical SETI system comprises, after the analog receiver components, a digitizing and analysis subsystem. An amateur system can be little more than a personal computer with a sound card. Such hardware can look for narrowband signals over a bandwidth of perhaps 40KHz. A professional system uses a number of dedicated processors to give several orders of magnitude more frequency range, on the order of tens or hundreds of MHz. In either case, however, the analysis system must make a decision: is there a narrowband signal present in the passband? If so, the immediate goal is to determine from whence the signal emanates. If it is coming from a point stellar source, it should show Doppler shift characteristic of the Earth's rotation, and should vanish if the radio telescope is pointed momentarily in some other direction. If these conditions are fulfilled, then another telescope at another location is advised of the signal and asked to verify its presence. Missing from all this excitement is any analysis of the signal itself! A narrowband signal is characterized by a single number: its frequency. This, plus or minus a few hundred Hz due to Doppler shift, is all you need to know. There's no point in recording the signal itself. To see what to expect from a gravitational-lens-enhanced signal event, consider what would happen if one were to aim an antenna at the Earth from space. As the Earth swam into the focus of the dish, a panoply of signals would reveal themselves. Among the strongest would be television transmitters and pulsed radars. Weaker signals used for point-to-point communications and radio navigation, for example, would be evident if the receiver had enough sensitivity. These signals would be all along the frequency axis. Depending upon time of day or night, frequencies below approximately 5 MHz to 50MHz would be filtered from the ensemble by ionospheric reflection and absorption.

Anything from 50 MHz to many Ghz would be fair game. For "internal" use by our civilization, there are no "magic" frequencies. In fact, the "water hole" is the least likely to have strong signals, since it is reserved for receiving weak signals! Whether or not another planet has a radio reflective ionosphere, such as ours does, isn't all that important, since for other reasons we will want to limit our search to a somewhat higher range of frequencies. Ideally, it would be desirable to search in the range of approximately 1 to 10 Ghz, or even lower and/or higher if antenna size and/or precision permits.

Would we detect the Earth with a receiver designed specifically for extremely narrow frequency bin detection? Maybe. Although there is little point in transmitting a totally modulation-free, extremely narrowband signal (except, perhaps, as a frequency standard or interstellar beacon), there is often enough energy transmitted at a "carrier" frequency used as a demodulation reference. It has been said that "a sufficiently advanced form of modulation is indistinguishable from noise" and we have been approaching that "ideal" almost since the beginning of electromagnetic communication. For example, television transmission in the United States has, over the recent years, shifted from a format with a strong carrier component to a digital format in which there will be no carrier at all. Another civilization's hope of detecting the next century's "I Love Lucy" will be greatly reduced. For the purpose of SETI it would be better to have a detector that could detect any artificial characteristic of a signal ensemble. Among the hallmarks of artificiality would be, in addition to frequency coherence, a broadened or otherwise interesting autocorrelation function, a non-Gaussian probability density function, a suddenly differing smoothed frequency spectrum, and an amplitude modulated, at whatever rate, intensity.

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

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

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

I have no desire to bury the Statue of Liberty in worthless data, which is what most of it would be. A better way to handle this is to be more judicious in our data recording habits.

First, we would only want to run the recorder when there is a candidate signal present. Based on the gravitational lens statistics, or, alternatively, the number of "hits" received in SETI searches in the past, this would be comfortably under one percent of the time. Of course, the time to initiate and terminate recording would be determined by a signal detector broadly described above. Next, recording the entire baseband, beyond the state of the art for a single recorder at present, isn't really necessary. Although it is conceivable that there would be a torrent of signals at all frequencies, it is more likely that they will appear in a more limited area. On Earth we allocate frequency bands for different purposes. Some have a few strong signals (broadcasting); some have many weak signals (portable telephony). Even with the enormous gain of a gravitational lens, it is unlikely that we can receive signals unless they have many kilowatts behind them. By setting up a number of recorders capable of an instantaneous bandwidth of, say, 50MHz, and a suitable number of signal presence detectors, we should be able to deal with whatever comes our way.

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

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

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

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

## 13.7 Summary

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

# 14

# Detection Algorithms: FFT vs. KLT

Claudio Maccone,
*Technical Director, Scientific Space Exploration, International Academy of Astronautics*

Given the vast distances between the stars, we can anticipate that any received SETI signal will be exceedingly weak. How can we hope to extract (or even recognize) such signals buried well beneath the natural background noise with which they must compete? This chapter analyzes, compares, and contrasts the two dominant signal detection algorithms used by SETI scientists to recognize extremely weak candidate signals.

We begin by introducing the use of the Karhunen-Loève Transform (KLT) to extract weak signals from noise of any kind. In general, the noise may be colored and over wide bandwidths, and not just white and over narrow bandwidths. We will show that the signal extraction can be achieved by the KLT more accurately than by the more commonly used fast Fourier transform (FFT), especially if the signals buried into the noise are very weak, in which case the FFT fails. This superior performance of the KLT happens because the KLT of any stochastic process (both stationary and nonstationary) is defined from the start over a *finite* time span ranging between 0 and a *final and finite* instant $T$ (contrary to the FFT, that 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 will seek to prove that this new result will lead to the immediate reconstruction of a signal buried in the thick noise. We have thus put on a strong mathematical foundation a set of important practical formulae that can

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_14, 

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.

## 14.1 A Bit of History

We argue that the Karhunen-Loève Transform (KLT) is the most advanced mathematical algorithm presently available 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 by 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 sometimes find it 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 for two mathematicians, the Finnish actuary, Kari Karhunen (1915-1992) (Karhunen, 1946) and the French-American mathematician, Michel Loève (1907-1979) (Loève, 1946, 1955), who proved, independently and about the same time (1946), that the series (2) hereafter is *convergent*.

Put this way, the KLT looks like a purely mathematical topic, but really this is hardly the case. As early as 1933 the American statistician and economist Harold Hotelling (1895-1973) had used the KLT (for discrete time, rather than for continuous time), so that the KLT is sometimes called the "Hotelling Transform". Long before these three authors, the Italian geometer Eugenio Beltrami (1835-1899) had discovered as early as 1873 the SVD (Singular Value Decomposition), 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 chapter, we’ll derive **heuristically and step-by-step** the many equations that make up the KLT. We think that this approach is much easier to understand for beginners than that which 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 chapter, where we describe the recently discovered (2007-2008) “Bordered Autocorrelation Method” (BAM) to easily compute the KLT.

## 14.2 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 $x(t)$ into a basis of orthonormal functions (sines and cosines, in this case), typified by the classical Fourier series (first put forward by the French mathematician Jean Baptiste Joseph Fourier (1768-1830) around 1807),

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[ a_n \cos(\omega_n t) + b_n \sin(\omega_n t) \right] \tag{1}$$

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 considering a random function $X(t)$ (notice that we denote random quantities by capitals, and that $X(t)$ is also called a “stochastic process of the time”). We now seek to expand this stochastic process into a set of orthonormal functions $\phi_n(t)$ according to the starting formula

$$X(t) = \sum_{n=1}^{\infty} Z_n \, \phi_n(t) \tag{2}$$

that is called **Karhunen-Loève (KL) expansion of X(t) over the finite time interval** $0 \le t \le T$ .

What then are the $Z_n$ and the $\phi_n(t)$ in (2)? To find out, let us start by recalling what “orthonormality” means for the Fourier series (1). Leonhard Euler (1707-1783) had already laid the first stone towards the Fourier series

(1) by proving that, if $T = t_2 - t_1$ is the assumed period of $x(t)$ and one sets $\omega_n = n \cdot \frac{2\pi}{T}$, then the coefficients $a_n$ and $b_n$ in (1) are obtained from the known function (or "signal") *x(t)* by virtue of the equations ("Euler formulae"):

$$a_n = \frac{2}{T}\int_{t_1}^{t_2} x(t)\cos(\omega_n t)\, dt \quad b_n = \frac{2}{T}\int_{t_1}^{t_2} x(t)\sin(\omega_n t)\, dt\ . \tag{3}$$

If the same result is going to be true for the Karhunen-Loève expansion, the functions of the time, $\phi_n(t)$ in (2) must be orthornormal, i.e., both orthogonal and normalized to one. That is,

$$\int_0^T \phi_m(t)\phi_n(t)\, dt = \delta_{mn} \tag{4}$$

where the $\delta_{mn}$ are the Kronecker symbols, defined by $\delta_{mn} = 0$ for $m \neq n$ and $\delta_{nn} = 1$

But what then are the $Z_n$ appearing in (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 the $Z_n$. In other words, the $Z_n$ must be random variables not changing in time, i.e., "just" random variables and not stochastic processes. By doing so we have actually made one basic, new step: 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!)

$$E\{X(t)\} = \sum_{n=1}^{\infty} E\{Z_n\}\,\phi_n(t)\ . \tag{5}$$

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 $0 \le t \le T$ :

$$E\{X(t)\} \equiv 0 . \tag{6}$$

In fact, if this wasn't, 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

$$E\{Z_n\} \equiv 0 . \tag{7}$$

This equation has a simple consequence: since the variance $\sigma^2_{Z_n}$ of the random variables $Z_n$ is given by

$$\sigma^2_{Z_n} = E\left\{Z_n^2\right\} - E^2\{Z_n\} \tag{8}$$

by inserting (7) into (8) we get

$$\sigma^2_{Z_n} = E\left\{Z_n^2\right\} . \tag{9}$$

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 $\lambda_n$ such that every $\lambda_n$ is the variance of the corresponding random variable $Z_n$, that is

$$\sigma^2_{Z_n} = \lambda_n = E\left\{Z_n^2\right\} > 0. \tag{10}$$

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 (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

$$E\{Z_m Z_n\} = \lambda_n \delta_{mn}. \tag{11}$$

In this equation, we were forced to introduce the positive $\lambda_n$ in the right-hand side in order to let (11) reduce to (10) in the special case $m = n$.

As for the KL equivalent of the Euler formulae (3) of the Fourier series, from the KL series (2) and the orthonormality (4) of the $\phi_n(t)$ one immediately finds that

$$Z_n = \int_0^T X(t)\,\phi_n(t)\,dt. \tag{12}$$

In other words: the random variables $Z_n$ are obtained from the given stochastic process $X(t)$ by "projecting" this $X(t)$ over the correspoding eigenvector $\phi_n(t)$. If one likes the language of mathematicians and of quantum physicists, 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 $0 \le t \le T$.

To sum up, we have actually achieved a remarkable generalization of the Fourier series by defining the Karhunen-Loève expansion (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 $t_2$:

$$R_{XX}(t_1,t_2) \equiv R_X(t_1,t_2) = E\{X(t_1)X(t_2)\}. \tag{13}$$

If we assume, according to (7), that the mean value of $X(t)$ vanishes identically in the interval $0 \le t \le T$, the autocorrelation (13) reduces to the variance of $X(t)$ when the two instants are the same:

$$\sigma^2_{X(t)} = E\{X^2(t)\} = E\{X(t)X(t)\} = R_X(t,t). \tag{14}$$

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 (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.

## 14.3 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 $\lambda_n$ - i.e., the variances of the corresponding uncorrelated random variables $Z_n$. In this section, we solve this problem by showing that the $\phi_n(t)$ are the eigenfunctions of the autocorrelation $R_X(t_1, t_2) = E\{X(t_1)X(t_2)\}$ and that the $\lambda_n$ 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* priviledged 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 (2). This goal is achieved by writing down (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

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \phi_m(t_1)\phi_n(t_2)\, E\{Z_m Z_n\} . \tag{15}$$

Taking advantage of the statistical orthogonality of the $Z_n$, given by (11), (15) simplifies to

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty} \lambda_m \phi_m(t_1)\phi_m(t_2) . \tag{16}$$

Finally, we now want to let the $\phi_n(t)$ "disappear" from the right hand side of (16) by taking advantage of their orthonormality (4). To do so, we multiply both sides of (16) by $\phi_n(t_1)$ and then take the integral with respect to $t_1$ between 0 and T. One then gets:

$$\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)\, dt_1 =$$

$$= \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2) \int_0^T \phi_m(t_1)\phi_n(t_1)\, dt_1$$

$$= \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2)\, \delta_{mn} = \lambda_n \phi_n(t_2) , \tag{17}$$

that is

$$\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)\,dt_1 = \lambda_n\ \phi_n(t_2). \tag{18}$$

This basic result is an **integral equation**, called "of the Fredholm type" by mathematicians. Once the correlation $E\{X(t_1)X(t_2)\}$ of $X(t)$ is known, the integral equation (18) yields (upon its solution, that may not be easy at all to find analytically!) both the Karhunen-Loève eigenvalues $\lambda_n$ and the corresponding eigenfunctions $\phi_n(t)$. Readers familiar with quantum mechanics will also recognize in (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 2 and 3, and let us use the language of signal processing that will lead us directly to SETI, the main theme of this chapter.

By adding random noise to a deterministic signal one obtains what is called a "noisy signal" or, in case the signal power is much lower than the noise power, "a signal buried into the noise". The signal+noise 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 (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 advamtage ove the Fourier transform as well as over other transforms, like Wigner-Ville, Hilbert, etc.

## 14.4 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 15 years of his life (1980-1994) in investigating mathematically the connection between Special Relativity and KLT. The result was the mathematical theory of optimal telecommunications between the Earth and a relativistic spaceship either receding from the Earth or approaching it. Although this may sound like "mathematical science fiction" to some folks (whom 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, a mathematical physics book like Maccone (1994) or Maccone (1999) 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 equation (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. The task may be made much easier if one is able to reduce the Fredholm integral equation to a Volterra integral equation, as shown in Maccone (1994) 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 (2). If this variable is *discrete*, rather than continuous, then the picture changes completely. In fact, the integral equation (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, such as image processing). And huge also is the characteristic equation, i.e., the algebraic equations 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 (Montebugnoli et al., 2003) since 2007.

## 14.5 The KLT: Just a Linear Transformation in the Hilbert Space

We have explained the KL expansion (2), but we haven't yet explained what the KL Transform is! We do so in this section.

The next step towards the KLT proper is the rearrangement of the eigenvalues $\lambda_n$ 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* trasformation of axes, and nothing is easier than that! (Incidentally, this accounts for the title of Karhunen's first paper "Über Lineare Methoden in der Wahrscheinlichkeitsrechnung" - translated as "On the Linear Methods in the Calculus of Probabilities" (Karhunen, 1946) - that obviously refers to the linear character of the transformation of axes in the Hilbert space).

## 14.6 A Breakthrough in the KLT: The Final Variance Theorem

The importance of the KLT as a mathematical tool superior to the FFT has already been pointed out. However, the implementation of the KLT by a numerical code running on computers has always been a difficult problem. Both François Biraud in France (Biraud, 1983) and Bob Dixon in the USA (Dixon and Klein, 1993) failed 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 equation (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 the 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 book 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" http://en.wikipedia.org/wiki/Karhunen-Lo%C3%A8ve_theorem. 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. (16), with the traditional dummy index *n* rewritten instead of *m*. Upon replacing $t_1 = t_2 = t$, this equation becomes

$$E\left\{X^2(t)\right\} = \sum_{n=1}^{\infty} \lambda_n \phi_n^2(t) \,. \tag{19}$$

Since the eigenfunctions $\phi_n(t)$ are normalized to one, we are prompted to integrate both sides of (19) with respect to *t* between 0 and *T*, so that the integral of the square of the $\phi_n(t)$ becomes just one:

$$\int_0^T E\left\{X^2(t)\right\} dt = \sum_{n=1}^{\infty} \lambda_n \int_0^T \phi_n^2(t)\, dt = \sum_{n=1}^{\infty} \lambda \,. \tag{20}$$

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

$$\sigma^2_{X(t)} = E\left\{X^2(t)\right\} - E^2\{X(t)\} = E\left\{X^2(t)\right\}. \tag{21}$$

Replacing (21) into (20), one gets

$$\int_0^T \sigma^2_{X(t)}\, dt = \sum_{n=1}^{\infty} \lambda_n. \tag{22}$$

This formula was already given by this author as eq. (1.13) in Maccone (1994. 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 $\lambda_n$ 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 $\lambda_n$ must be some functions of the final instant *T*:

$$\lambda_n \equiv \lambda_n(T). \tag{23}$$

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:

$$\sigma^2_{X(T)} = \sum_{n=1}^{\infty} \frac{\partial \lambda_n(T)}{\partial T}. \tag{24}$$

This result we call the *Final Variance Theorem*. It is the key new result put forward in this chapter. It states that for any (either non-stationary or stationary) stochastic process $X(t)$, *the Final Variance* $\sigma^2_{X(T)}$ *is the sum of the series of the first-order partial derivatives of the eigenvalues* $\lambda_n(T)$ *with respect to the final instant T.*

Let us now consider a few particular cases of this theorem that are especially interesting.

1) In general, only the first $N$ terms of the decreasing sequence of eigenvalues will be retained as "significant" by the user, and all the other terms, from the ($N+1$)-th term onward, will be declared to be "just noise". Therefore the infinite series in (24) becomes in the practice the finite sum

$$\sigma^2_{X(T)} \approx \sum_{n=1}^{N} \frac{\partial \lambda_n(T)}{\partial T} . \tag{25}$$

In numerical simulations, however, one always wants to cut as short as possible the computation time! Therefore one might be led to consider the *first* (or *dominant*) eigenvalue only in (25), that is

$$\sigma^2_{X(T)} \approx \frac{\partial \lambda_1(T)}{\partial T} . \tag{26}$$

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 (26) instead of the $N$-term sum (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 (25) simplifies even further. In fact, by definition, the stationary processes have the *same final variance at any time, i.e. for stationary processes* $\sigma^2_X$ *is a constant*. Then (22) immediately shows that, for stationary processes only, all the KLT eigenvalues are *linear* functions of the final instant $T$:

$$\lambda_n(T) \propto T \;\; \textit{for stationary processes only} . \tag{27}$$

As a consequence, the first-order partial derivatives of all the $\lambda_n$ with respect to $T$ for stationary processes are just constants. In other words still, for stationary processes only, (25) becomes

$$\sum_{n=1}^{N} \frac{\partial \lambda_n(T)}{\partial T} \approx \text{a constant with respect to } T. \tag{28}$$

In particular, if one sticks again to the first, dominant eigenvalue only (i.e., to the roughest possible approximation), then (28) reduces to

$$\frac{\partial \lambda_1(T)}{\partial T} \approx \text{a constant with respect to } T. \tag{29}$$

In the next section we 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 nonlinear. For instance, for the well-known Brownian motion (that is to say, "the easiest of the non-stationary processes"), one has

$$\lambda_n(T) = \frac{4\,T^2}{\pi^2 (2n-1)^2} \quad (n = 1,2,...). \tag{30}$$

and so the dependence on $T$ is quadratic. For the proof, just replace the Brownian motion variance $\sigma^2_{B(t)} = t$ into (22) and perform the integration, yielding the $T^2$ directly. Of course, this is in agreement with (30), that is proven, for instance, in Maccone (2009), page 311, Appendix F, equation (F21).

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 (15.59) of Maccone (1999). And so on for more complicated processes, like the time-rescaled squared Brownian motions whose KLT is found in Maccone (2009).

## 14.7 BAM (Bordered Autocorrelation Method) to find the KLT of Stationary Processes Only

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 previously described. 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 Section 14.8, we'll 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 N*ln(N) amount of calculations required by the FFT, and that's precisely why the FFT has been preferred to the KLT for 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 the use of (29). This means to study (29) for different values of the final instant *T*, i.e., as a function of the final instant *T*.

Also, we now confine ourselves to a *stationary* $X(t)$ over a *discrete* set of instants $t = 0, \ldots, N$. In this case, the autocorrelation of $X(t)$ becomes the Toeplitz matrix (for an introduction to the research field of Toeplitz matrices, see the Wikipedia site http://en.wikipedia.org/wiki/Toeplitz_matrix) that we denote by $R_{Toeplitz}$

$$R_{Toeplitz} = \begin{bmatrix} R_X(0) & R_X(1) & R_X(2) & \ldots & \ldots & R_X(N) \\ R_X(1) & R_X(0) & R_X(1) & \ldots & \ldots & R_X(N-1) \\ R_X(2) & R_X(1) & R_X(0) & \ldots & \ldots & R_X(N-2) \\ \ldots & \ldots & \ldots & R_X(0) & \ldots & \ldots \\ R_X(N) & R_X(N-1) & \ldots & \ldots & R_X(1) & R_X(0) \end{bmatrix} \tag{31}$$

This theorem has already been proven by Bob Dixon and Mike Klein (Dixon and Klein, 1991), and will not be proven here again. We may choose *N* at will but, clearly, the higher *N*, 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'll 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. 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 value of N, we get, a new value of the first eigenvalue* $\lambda_1$ *now regarded as a function of N, i.e.* $\lambda_1(N)$. *Doing this over and over again, for how many values of which as we wish (or, more correctly, for how many values of N 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,* $\frac{\partial \lambda_1(N)}{\partial N}$,

*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 four plots below (Figures 14.1-14.4). The first plot (Figure 14.1) 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 http://en.wikipedia.org/wiki/Signal-to-noise_ratio. 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 is well known from the Wiener-Khinchin Theorem (for a concise description of this theorem, see http://en.wikipedia.org/wiki/Wiener%E2%80%93Khinchin_theorem).

Notice, however, that this procedure would not work for non-stationary *X(t)* because the Wiener-Khinchin Theorem does not apply to non-stationary processes. For non-stationary processes there are other "tricks" to compute the spectrum from the autocorrelation, like the Wigner-Ville Transform, but

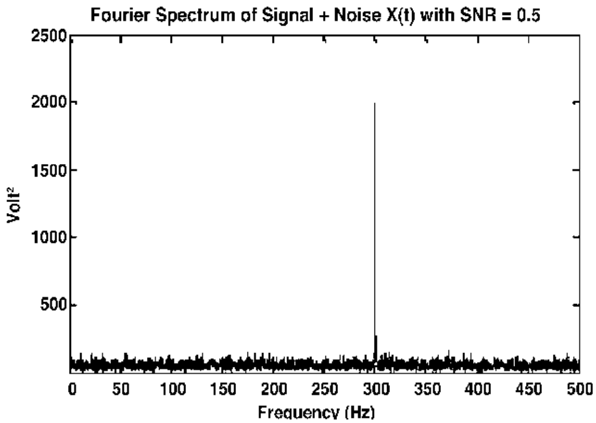



**Figure 14.1** Fourier spectrum of a pure tone (i.e., just a sinusoidal carrier) with frequency at 300 Hz buried in stationary noise with a signal-to-noise ratio of 0.5.

shall not consider them here.

The second plot (Figure 14.2) 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 (27).

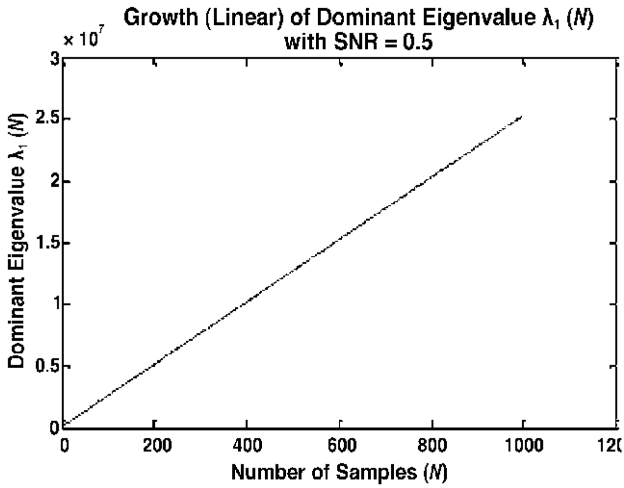



**Figure 14.2** The KLT dominant eigenvalue $\lambda_1(N)$ over N=1000 time samples, computed by virtue of the BAM, the Borderd Autocorrelation Method.

So, its derivative, $\frac{\partial \lambda_1(N)}{\partial N}$, 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 (Figure 14.3) shows such a BAM-reconstructed peak.

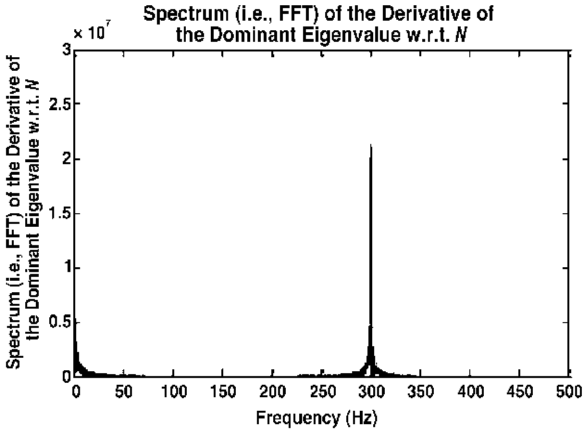



**Figure 14.3** The spectrum (i.e., the Fourier Transform) of the constant derivative of the KLT dominant eigenvalue $\lambda_1(N)$ with respect to N as given by the BAM. This is clearly a Dirac delta function, i.e., a peak, at 300 Hz, as expected.

Finally, this plot is of course identical to the Figure 14.4, showing the ordinary FFT of first KLT eigenfuction as obtained not by the BAM, but by solving the full and long system of N algebraic first-degree equations.

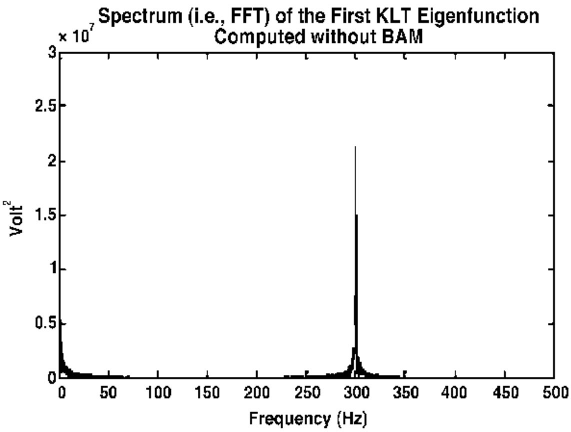



**Figure 14.4** The spectrum (i.e., the Fourier Transform) of the first KLT eigenfunction, not obtained by the BAM but rather by the very long procedure of solving the N linear algebraic equations corresponding, in discrete time, to the integral equation (18). Clearly, the result is the same as obtained in Figure 14.3 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 diffuculty of the $N^2$ calculations requested by the "straight" KLT theory.

*Let us now do the same again… but with an incredibly low SNR of 0.005.*

Poor Fourier is now turning over in his grave. Just look at the first plot below (Figure 14.5)!

No classical FFT spectrum can be identified at all for such a terribly low SNR!.

Fourier spectrum of signal + noise $X(t)$ with SNR = 0.005.

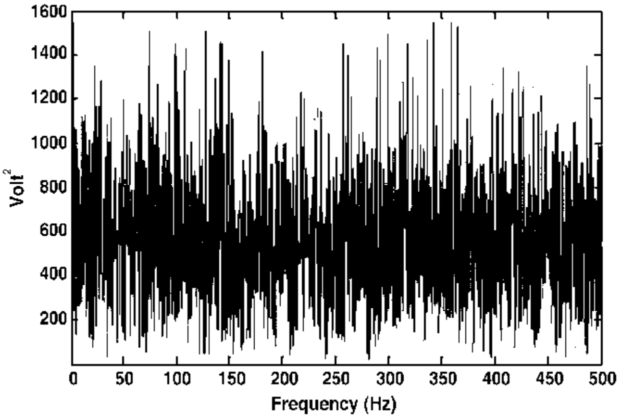



**Figure 14.5** Fourier spectrum of a pure tone (i.e., just a sinusoidal carrier) with frequency at 300 Hz buried in stationary noise with the terribly low signal-to-noise ratio of 0.005. This is clearly beyond the reach of the FFT, since we know there should just be one peak only at 300 Hz. Fourier fails at such a low SNR.

**But for the KLT… no problem!**

The next plot shows that $\lambda_1(N) \propto N$ , as predicted by our Final Variance Theorem (27).

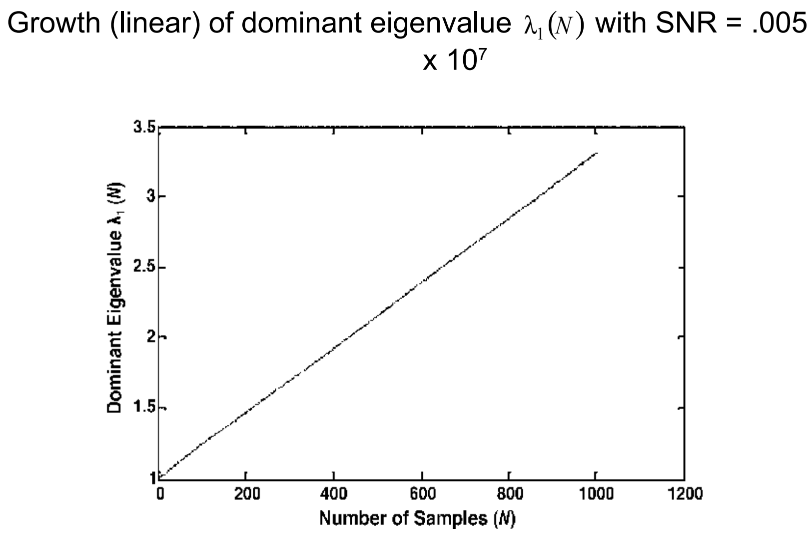



**Figure 14.6** The KLT dominant eigenvalue $\lambda_1(N)$ over N=1000 time samples, computed by virtue of the BAM, for the very low SNR=0.005.

The third plot (*KLT FAST* way via the BAM) is the *neat KLT spectrum* of the 300 Hz tone obtained by computing the FFT of the *constant* $\frac{\partial \lambda_1(N)}{\partial N}$.

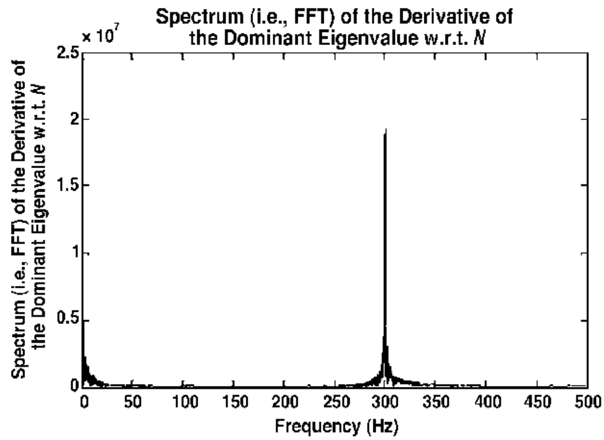



**Figure 14.7** The spectrum (i.e., the Fourier Transform) of the CONSTANT derivative of the KLT dominant eigenvalue $\lambda_1(N)$ with respect to N as given the BAM. This is a neat Dirac delta function, i.e., a peak at 300 Hz, as expected.

And this is just the same as the last plot (Figure 14.8) of the dominant KLT eigenfunction obtained by KLT SLOW way of doing $N^2$ calculations.

This proves the superior behavior of the KLT.

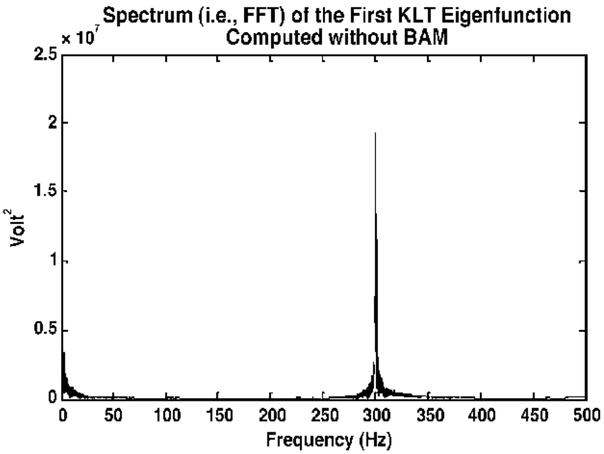



**Figure 14.8** The spectrum (i.e., the Fourier Transform) of the first KLT eigenfunction not obtained by the BAM, but rather by the very long procedure of solving the N linear algebraic equations corresponding, in discrete time, to the integral equation (18). Clearly, the result is the same as obtained in Figure 14.7 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 diffuculty of the $N^2$ calculations requested by the "straight" KLT theory.

## 14.8 Recent Developments

The numerical simulations described in the previous section were performed at Medicina during the winter 2006-7 by Francesco Schilliro` and Salvatore "Salvo" Pluchino (Schilliro` et al., 2007). 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, $\frac{\partial \lambda_1(T)}{\partial T}$, 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 $\frac{\partial \lambda_1(T)}{\partial T}$.

Why?

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 radiotelescope. 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 (Yatawatta, 2008). Unfortunately, Dr Yatawatta made two 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 phoenomena (just as 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 chapter.

## 14.9 KLT of Stationary Unitary White Noise

Before we give the analytic proof that the Fourier Transform of $\frac{\partial \lambda_1(T)}{\partial T}$ 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

$$E\{W(t_1)W(t_2)\} = \delta(t_1 - t_2). \qquad (32)$$

If one now seeks for the KLT of stationary unitary white noise, one must of course replace the autocorrelation (32) into the KLT integral equation (18), getting

$$\lambda_n \, \phi_n(t_2) = \int_0^T E\{W(t_1)W(t_2)\} \phi_n(t_1)\, dt_1$$

$$= \int_0^T \delta(t_1 - t_2)\, \phi_n(t_1)\, dt_1 = \phi_n(t_2). \tag{33}$$

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 $0 \le t \le T$:

$$\phi_n(t) \equiv W_n(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi\, n}{T} t\right). \tag{34}$$

This set of basis functions of course fulfills the orthonomality condition:

$$\int_0^T W_m(t) W_n(t)\, dt = \cdot$$

$$= \int_0^T \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi\, m}{T} t\right) \cdot \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi\, n}{T} t\right) dt = \delta_{mn}\, . \tag{35}$$

This property will be used in the next section, where we give the proof that the Fourier transform of $\dfrac{\partial \lambda_n(T)}{\partial T}$ is indeed (twice) the frequency of the unknown ET sinusoidal carrier buried into the white, cosmic noise. We conclude this section by pointing out the first analytical error made by Dr. Yatawatta in his personal communication to this author (Yatawatta, 2008): he forgot to put the square root in (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.

## 14.10 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 $\nu = \frac{\omega}{2\pi}$, that is

$$S(t) = W(t) + a\sin(\omega t). \tag{36}$$

What is the KLT of such a (signal+noise) process?

This is the central problem of SETI, of course.

To find the answer, first build up the autocorrelation of this process:

$$E\{S(t_1)S(t_2)\} = E\{W(t_1)W(t_2)\} + a^2\sin(\omega t_1)\sin(\omega t_2) \\ + a\,E\{W(t_1)\sin(\omega t_2)\} + a\,E\{W(t_2)\sin(\omega t_1)\}. \tag{37}$$

The last two terms in (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 (37) of the (signal+noise) stochastic process $S(t)$ thus becomes

$$E\{S(t_1)S(t_2)\} = E\{W(t_1)W(t_2)\} + a^2\sin(\omega t_1)\sin(\omega t_2). \tag{38}$$

In order to proceed, we now make use of the eigenfunction expansion of the autocorrelation (16), that, replaced into (38), changes it into

$$\sum_{m=1}^{\infty} \lambda_{S\,m}\, S_m(t_1)\, S_m(t_2) = \qquad \ldots.(39)$$

$$= \sum_{m=1}^{\infty} \lambda_{W\,m}\, W_m(t_1)\, W_m(t_2) + a^2\sin(\omega t_1)\sin(\omega t_2). \tag{39}$$

In the last equation, the $S_m(t)$ clearly are the (unknown) eigenfunctions of the (signal+noise) process $S(t)$, and the $\lambda_{S_m}$ are (unknown) corresponding eigenvalues. In the right-hand side, the $\lambda_{W_m}$ 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 $\lambda_{W\,m}$ 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 eigenfuctions 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 $\lambda_{S\,m}$ must not differ very much from the corresponding pure white noise eigenvalues $\lambda_{W\,m}$. In other words, the hypothesis that SNR<<1 amounts to the two approximated equations

$$\begin{cases} S_m(t) \approx W_m(t) \\ \lambda_{S\,m} \approx \lambda_{W\,m} = 1. \end{cases} \tag{40}$$

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 $0 \le t \le T$. As a consequence, the new orthonormality condition (nearly) holds:

$$\int_0^T S_m(t_1) W_n(t_1)\, dt_1 \approx \delta_{mn} \tag{41}$$

and, similarly,

$$\int_0^T S_k(t_2) W_n(t_2)\, dt_2 \approx \delta_{kn}. \tag{42}$$

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 (41) and (35) one has:

$$\sum_{n=1}^{\infty} \lambda_{S\,n}\, S_n(t_2) \approx$$

$$\approx \sum_{n=1}^{\infty} \lambda_{W\,n}\, W_n(t_2) + a^2 \sin(\omega\, t_2) \int_0^T W_n(t_1) \sin(\omega\, t_1)\, dt_1 \tag{43}$$

The good point is that the integral appearing in the right-hand side of this equation can be analytically found. In fact, replacing $W_n(t_1)$ by virtue of (34) and integrating, one gets

$$\sum_{k=1}^{\infty} \lambda_{S_k} S_k(t_2) \approx$$

$$\approx \sum_{k=1}^{\infty} \lambda_{W_k} W_k(t_2) + a^2 \sin(\omega t_2) \cdot \frac{2\sqrt{2}\pi\, n \sqrt{T} \sin(\omega T)}{\omega^2 T^2 - 4\pi^2 n^2} \tag{44}$$

We next multiply this equation by $W_n(t_2)$ and integrate with respect to $t_2$ between 0 and *T*. Because of (42) and (35), (44) becomes:

$$\approx \lambda_{W_n} + a^2 \frac{2\sqrt{2}\pi\, n \sqrt{T} \sin(\omega T)}{\omega^2 T^2 - 4\pi^2 n^2} \int_0^T W_n(t_2) \sin(\omega t_2)\, dt_2 \tag{45}$$

Again, the integral in the last equation can be analytically found (it is actually the same integral as in (43)) and so the conclusion is

$$\lambda_{S_n} \approx \lambda_{W_n} + a^2 \frac{8\pi^2 n^2 T \sin^2(\omega T)}{\left(\omega^2 T^2 - 4\pi^2 n^2\right)^2}. \tag{46}$$

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

$$\lambda_{S_n}(T) \approx \lambda_{W_n} + a^2 \frac{8\pi^2 n^2 T \sin^2(\omega T)}{\left(\omega^2 T^2 - 4\pi^2 n^2\right)^2}. \tag{47}$$

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

$$\lambda_{S1}(T) \approx \lambda_{W1} + a^2 \frac{8\pi^2\, T \sin^2(\omega\, T)}{\left(\omega^2 T^2 - 4\pi^2\right)^2} =$$

$$= 1 + a^2 \frac{8\pi^2\, T \sin^2(\omega\, T)}{\left(\omega^2 T^2 - 4\pi^2\right)^2}. \tag{48}$$

This completes our analysis of the KLT of a sinusoidal carrier buried into white, cosmic noise.

## 14.11 Analytic Proof of the BAM-KLT

We are now ready for the analytic proof of the BAM-KLT method.

Let us first rewrite (47) in the form where the pure white noise eigenvalues are replaced by 1:

$$\lambda_{S_n}(T) \approx 1 + a^2 \frac{8\pi^2\, n^2\, T \sin^2(\omega\, T)}{\left(\omega^2 T^2 - 4\pi^2\, n^2\right)^2}. \tag{49}$$

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

$$\sin^2(\omega T) = \frac{1}{2} - \frac{1}{2}\cos(2\omega T). \tag{50}$$

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

$$2\sin(\omega T)\cos(\omega T) = \sin(2\omega T). \tag{51}$$

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 (49) with respect to $T$ and just give the final result.

In conclusion, the derivative of (49) with respect to $T$ is given the following sum of three terms:

$$\frac{\partial \lambda_{S_n}(T)}{\partial T} \approx$$

$$\approx Coeff_1(T)\cdot\sin(2\omega T) + Coeff_2(T)\cdot\cos(2\omega T) + Coeff_3(T) \tag{52}$$

where the three coefficients turn out to be (after lengthy calculations)

$$\begin{cases} Coeff_1(T) = a^2 \dfrac{8\pi^2 n^2 \omega T}{\left(\omega^2 T^2 - 4\pi^2 n^2\right)^2}, \\ Coeff_2(T) = a^2 \dfrac{4\pi^2 n^2\left(3\omega^2 T^2 + 4\pi^2 n^2\right)}{\left(\omega^2 T^2 - 4\pi^2 n^2\right)^3}, \\ Coeff_3(T) = -a^2 \dfrac{4\pi^2 n^2\left(3\omega^2 T^2 + 4\pi^2 n^2\right)}{\left(\omega^2 T^2 - 4\pi^2 n^2\right)^3}. \end{cases} \tag{53}$$

But the right-hand side of (52) is nothing but… the simple Fourier series expansion of $\dfrac{\partial \lambda_{S_n}(T)}{\partial T}$.

Moreover, (52) shows that $\dfrac{\partial \lambda_{S_n}(T)}{\partial T}$ is a *periodic* function of $T$ with frequency $2\omega T$.

We conclude that: *The Fourier transform of* $\frac{\partial \lambda_{S_n}(T)}{\partial T}$

*equals twice the frequency of the buried alien sinusoidal carrier. In other words, the frequency of the alien signal is a half of the frequency found by taking the Fourier transform of* $\frac{\partial \lambda_{S_n}(T)}{\partial T}$.

And the BAM-KLT method is thus proved analytically.

## 14.12 How to Eavesdrop on Alien Chat

Following the 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 on 30 October 2008 an article by Jessica Griggs, titled "How to Eavesdrop on Alien Chat", that well summarizes the key features of the present chapter. Here are a few relevant quotes from that article:

> "ET, phone... each other? If aliens really are conversing, we are not picking up what they are saying. Now one researcher claims to have a way of tuning in to alien cellphone chatter."
>
> "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 …"
>
> "Seth Shostak at the SETI Institute in California agrees that the KLT might be the way to go … 'It is likely that aliens use the same spread-spectrum method of transmission as us on their cellphones.'"

## 14.13 Conclusions

Let us summarize the main results of this chapter.

When the stochastic process *X(t)* is stationary (i.e., it has both mean value and variance constant in time), then there are two alternative ways to compute the first KLT dominant eigenfunction (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 practical, numerical simulations of the KLT it may be much less time-consuming to choose option (2) rather than option (1).

In either case, the KLT of a given stationary process can retrieve a sinusoidal carrier out of the noise for values of the signal-to-noise ratio (SNR) that are *three orders of magnitude lower* than those that the FFT can still filter out. In other words, while the FFT (at best) can filter out signals buried in 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 30 dB improvement in sensitivity 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.

**Acknowledgements**

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 Drs Seth Shostak and 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 continue to support this author in his efforts for SETI over the years, and their help is hereby gratefully acknowledged.

## References

Biraud, F. 1983. SETI at the Nançay Radio-telescope. *Acta Astronautica*, **10**, 759-760.

Karhunen, K. 1946. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fennicae, Series A 1, *Math. Phys*, **37**, 3-79.

Loève, M. 1946. Fonctions Aléatoires de Second Ordre. *Rev. Sci*., **84:4**, 195-206.

Loève, M. 1955. *Probability Theory: Foundations, Random Sequencies*. Princeton, NJ: Van Nostrand.

Maccone, C. 1994. *Telecommunications, KLT and Relativity – Volume 1*. IPI Press, Colorado Springs, Colorado, USA. This book embodies the results of some 30 research papers published by the author about the KLT from 1980-1994 in peer-reviewed journals.

Maccone, C. 2007. Advantages of the Karhunen-Loève Transform over Fast Fourier Transform for Planetary Radar and Space Debris Detection. *Acta Astronautica*, **60**, 775-779.

Maccone, C. 2009. *Deep Space Flight and Communications – Exploiting the Sun as a Gravitational Lens*. Springer-Praxis.

Montebugnoli, S., Bortolotti, C., Caliendo, D., Cattani, A., D'Amico, N. et al. 2003. 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 International Astronautical Congress held in Bremen, Germany, September 29-October 3, 2003.

Dixon, R.S. and Klein, M. 1993. 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), pages 128-140.

## Annotated Bibiography

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.

**The KLT in Mathematics, Physics and the Theory of Relativistic Interstellar Flight, subdivided by journals:**

***Il Nuovo Cimento***

Maccone, C. 1987. Special Relativity and the Karhunen-Loève Expansion of Brownian Motion. *Nuovo Cimento*, Series B, **100**, 329-342.

***Bollettino dell'Unione Matematica Italiana***

Maccone, C. 1984. Eigenfunctions and Energy for Time-Rescaled Gaussian Processes. *Bollettino dell'Unione Matematica Italiana*, Series 6, **3-A**, 213-219.

Maccone, C. 1985. The Time-Rescaled Brownian Motion B($t^{2H}$). *Bollettino dell'Unione Matematica Italiana*, Series 6, **4-C**, 363-378.

Maccone, C. 1988. The Karhunen-Loève Expansion of the Zero-Mean Square Process of a Time-Rescaled Gaussian Process. *Bollettino dell'Unione Matematica Italiana*, Series 7, **2-A**, 221-229.

***Journal of the British Interplanetary Society***

Maccone, C. 1990. Relativistic Interstellar Flight and Genetics. *Journal of the British Interplanetary Society* **43**, 569-572.

***Acta Astronautica***

Maccone, C. 1988. Relativistic Interstellar Flight and Gaussian Noise. *Acta Astronautica,* 17:9, 1019-1027.

Maccone, C. 1990. Relativistic Interstellar Flight and Instantaneous Noise Energy. *Acta Astronautica* 2:3, 155-159.

Maccone, C. 1999. The Data Compression Problem for the "Gaia" Astrometric Satellite of ESA. *Acta Astronautica* **44**, Nos. 7-12, 375-384.

**Some important papers about the KLT for SETI**

Dixon, R.S. and Klein, M. 1993. On the Detection of Unknown Signals. Proceedings of the Third Decennial US-USSR Conference on SETI held at the University of California at Santa Cruz, August 5-9, 1991. Later published in the Astronomical Society of the Pacific (ASP) Conference Series (Seth Shostak, editor) Vol. 47 (1993), 128-140.

Maccone, C. 1990. *Karhunen-Loève Versus Fourier Transform for SETI.* Lecture Notes in Physics, Springer-Verlag, Vol. 390, 247-253. These are the Proceedings (Jean Heidmann and Mike Klein, editors) of the Third Bioastronomy Conference held in Val Cenis, Savoie, France, 18-23 June 1990.

After these seminal works were published, the importance of the KLT for SETI was finally acknowledged by the SETI Institute experts in:

Eckers, Ron Cullers, Kent, Billingham, John and Scheffer, Lou (Eds) 2002. SETI 2020. SETI Institute, page 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 transform 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:

Maccone, C. 1994. The Karhunen-Loève Transform: A Better Tool than the Fourier Transform for SETI and Relativity. *Journal of the British Interplanetary Society* **47**, 1.

### An early paper about the KLT for SETI-Italia

Montebugnoli, S. and Maccone, C. 2001. 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

Jain, A.K. 1976. A Fast Karhunen-Loève Transform for a Class of Random Processes. IEEE Trans. Commun. COM-24, 1023-1029.

### Recent papers about the KLT and BAM-KLT

Schilliro` F., Pluchino, F., Maccone, C., Montebugnoli:, S. 2007. *La KL Transform: considerazioni generali sulle metodologie di analisi ed impiego nel campo della Radioastronomia.* Technical Report (available in Italian only) for Istituto Nazionale di Astrofisica (INAF) – Istituto di Radioastronomia (IRA) – Rapporto Tecnico published in January 2007.

Maccone, C. 2007. Innovative SETI by the KLT. In: Proceedings of the "Bursts, Pulses and Flickering" Conference held at Kerastari, Greece, June 13-18, 2007 at POS (Proceedings of Science) web site, http://pos.sissa.it//archive/conferences/056/034/Dynamic2007_034.pdf

Yatawatta, Sarod, 2008. Personal communication, 17 June 2008.

### A recent paper about the KLT for Relativistic Interstellar Flight

Maccone, C. 2006. Relativistic Optimized Link by KLT. *Journal of the British Interplanetary Society* **59**, 94-98.

**The author's comprehensive book (2009) about both the Sun as a Gravitational Lens and the KLT, also for Relativistic Interstellar Flight:**

Maccone, C. 2009. *Deep Space Flight and Communications – Exploiting the Sun as a Gravitational Lens*. A 400-pages treatise about the FOCAL space mission and the KLT that embodies and updates all previously published material about FOCAL. Springer-Praxis.

# 15

# Implementing the KLT

Stelio Montebugnoli,
*Institute of Radioastronomy, INAF, Bologna, Italy*

SETI-Italia is an observing program led by the INAF-IRA (Istituto Nazionale di Astrofisica - Istituto di Radioastronomia). At present, Italy is the only European country conducting a SETI (Search for Extraterrestrial Intelligence) program. Occasional SETI searches may be conducted at the Nanacy French Kraus type radiotelescope and some artificial signals are sometime transmitted with the 70-m dish in the Ukraine. Outside Europe, continuous SETI programs are conducted in the USA, Australia and Argentina.

A Serendip IV piggyback system (coming from the University of California, Berkeley) has allowed us to conduct observations since 1998 at the Medicina VLBI 32-m antenna dish in northern Italy.

In parallel to the piggyback activities, considerable efforts were devoted to set up a fast processing system, able to compute the KLT (Karhunen-Loève Transform). This transform was proposed by Claudio Maccone some time ago (Maccone 1994; 2001) and strongly supported by Robert S. Dixon of the Ohio State Radio Observatory as early as 1993 (Dixon et al., 1993). It is a potential mathematically sound approach to SETI, and a very efficient algorithm for the detection of complex modulated signals.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_15, © Springer-Verlag Berlin Heidelberg 2011

## 15.1 The SETI-Italia Program

SETI listening programs are typically based on the most modern and sensitive radio telescopes on the planet, attempting to monitor the whole sky for radio signals, either monochromatic (i.e. “intentional”) or modulated (i.e. “possibly unintentional as ET local radio transmissions”), sent by extraterrestrial technologically advanced civilizations. Due to the inevitable rotation of both planets (the Earth and the Exoplanet), the radio signals are expected to be frequency shifted over time by the Doppler effect.

For the Italian SETI program, the Medicina 32 VLBI (very long baseline interferometry) dish, equipped with a 20 MHz bandwidth, 32 million channel spectrometer connected in piggyback mode, is used. Up to now, with the Serendip processing system, observations have been carried out in the various radio astronomical bands included in the 1.4-23.5 GHz spectral range. In this way, all the operational antenna time can be exploited at a very low cost, while it takes under control the RFI (Radio Frequency Interference) scenario in parallel to the Medicina station’s existing RFI monitoring systems.

### 15.1.1 Italian radio telescopes

The Istituto di Radioastronomia (IRA), a branch of the Istituto Nazionale di Astrofisica (INAF) (formerly Italian National Research Council, CNR), presently operates three radio telescopes and, very shortly, will add a fourth:

1 The Medicina 32 meter VLBI dish shown in Figure 15.1. Operating frequency: 1.4–23.5 GHz

2 The Noto (near Siracusa) 32 meter VLBI dish, shown in Figure 15.2, equipped with active mirror. Operating frequency: 1.4–43 GHz

3 The 30,000 m2 collecting area Northern Cross Array, a large cylindrical reflector T-shaped array. Operating frequency: 408 MHz ± 2.5 MHz (Figure 15.3)

4 The new 64 m dish in Sardenia, still under construction

The Northern Cross, a 564 x 640 mt Array equipped with 5632 dipoles, is one of the largest transit telescopes in the northern hemisphere. Its collecting area is equivalent to three football stadiums and, for this reason, it is one of the most suitable instruments to be transformed in a very large SKA (Square Kilometre Array) test bed.

A 64-m class parabolic antenna is currently under construction at San Basilio, close to Cagliari (Sardinia). It is hoped that, upon completion, this instrument will also become available for piggyback SETI observations.

**Figure 15.1** The 32-m VLBI dish in Medicina, with the Appennini hills in the background.

**Figure 15.2** The Noto 32-m VLBI dish, equipped with an active mirror that allows it to operate at $F_{max}$>33 GHz.

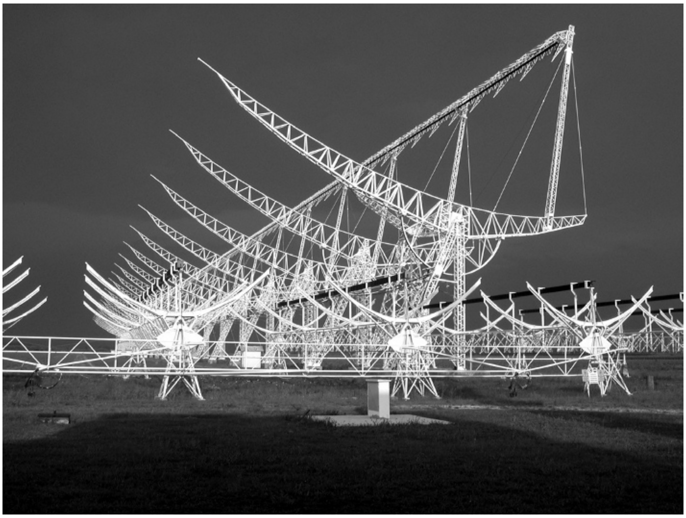


**Figure 15.3** The large T-shaped Northern Cross Array, 560 meters (East-West arm) × 640 meters (North-South arm) composed of some 5632 dipoles.

### 15.1.2 SETI activities

The SETI-Italia activities are presently performed with a 24-million channel Serendip IV high-resolution back end, connected in piggyback mode to the Medicina 32 meter VLBI. Proper software for possible ET signal detection has been successfully tested and implemented in the post processing phase. The Fast Fourier Transform (FFT) has been used so far, since it is a very efficient transform if monochromatic candidates (Dopplered signals) are searched for, in the radioastornomical 1.4-1.6 GHz band. The main bands observed during the normal telescope observations are normally:

- 1.4–1.6 GHz
- 4.8–6.5 GHz
- 8.5 GHz
- 22–23.5 GHz

within the European VLBI observation run and single dish antenna programs (Blazar, Geodynamic and Spectroscopy)

## 15.2 The KLT Approach

Up to now the basic SETI search was conducted as follows:

1 Suppose that an ET might spread out a monochromatic radio signal into space just to say "… hello, I am here!"

2 The most suitable data processing for detection such a continuous wave (CW) signal is high-resolution spectrum analysis based on FFT. This transform attempts to reconstruct signals of any type by virtue of sines and cosines only (orthonormal base functions). Although this approach works properly for CW signals, it loses its efficiency when attempting to detect wide band modulated signals embedded in the noise

Under these conditions a more general SETI post-processing algorithm needs to be implemented, based on a transform able to "understand" whether or not a complex modulated wide band signal is embedded in a very noisy environment. The KLT (Karhunen-Loève Transform, 1946) seems to be a very promising tool to detect any kind of ET signals (intentional or unintentional), and to differentiate signal from noise, because it uses an autocorrelation Matrix, always the optimal transformation base function, extracted from the signal itself.

The KLT requires a very high computational load to the processing system, because it involves the computation of the eigenvectors and eigenvalues of a large autocorrelation matrices of the data acquired in the time domain. It is based on the fact that, given any observation X(t) in the time domain, it is possible to separate the statistical component from the temporal component as follows:

$$X(t) = \sum_{n=1}^{\infty} Z_n \phi_n(t)$$

Where:

1 Zn = random variables not changing in time (similar to the sines and cosines coefficients in any Fourier series)

2 $\phi_n(t)$ = base functions, for example a set of orthonormal basis functions that represent the eigenfunctions (or eigenvectors in the Hilbert space of the square-integrable functions) of the autocorrelation of the X(t) function.

The KLT represents an optimal transform with some peculiar characteristics:

$\langle Z_n \rangle = 0$ (mean value of the coefficients Zn)

$\langle Z_n Z_m \rangle = 0$ (statistical independence from the autocorrelation point of view)

$\sigma^2_{Z_n} = \lambda_n$ ($Z_n$ coefficients variance)

$$\int_0^T \phi_m(t)\phi_n(t)dt = \delta_m = \begin{cases} 0 & \text{for} \quad m \neq n \\ 1 & \text{for} \quad m = n \end{cases},$$

(orthonormality of the base functions)

The computation of the eigenvectors/eigenvalues represents the algorithm for the KLT.

The eigenvalue of the autocorrelation matrix **A** is a scalar which corresponds to an eigenvector different from zero, leading to the expression.

For a matrix **A** of order N, we have N eigenvectors and N eigenvalues. Due to its extremely high computational complexity, and to the fact that it cannot be distributed (parallelized), at present this transform cannot be computed in real time.

### 15.2.1 The Algorithm

With the above mentioned assumptions, the "new IRA algorithm" (also called EAM, or Edging of Autocorrelation Matrix) for the detection of signals hidden in noise, is based on the following steps:

**(a) N points data acquisition** N points are acquired via an 8-12 bit A/D converter and sent to the computation block via a wideband bus.

**(b) N points autocorrelation vector computation** The N components data vector are used to compute the 2(N-1) components autocorrelation vector (symmetric in N-1).

**(c) Autocorrelation matrix computation** The (N × N) autocorrelation matrix (here N=5)

$$A_{\text{matrix}} = \begin{bmatrix} a & b & c & d & e \\ b & a & b & c & d \\ c & b & a & b & c \\ d & c & b & a & b \\ e & d & c & b & a \end{bmatrix}$$

is computed by starting from the autocorrelator vector $CORR_{vector} = [e\ d\ c\ b\ a\ b\ c\ d\ e]$

This is a Toeplitz matrix that contains the energy, given by the autocorrelation estimator, on the main diagonal. The EAM method starts with a simple consideration, i.e. every signal sample added over time into the KLT processing pipeline represents an increase at a sample in the autocorrelation vector, and then at a row and a column of the autocorrelation matrix. This equates to an increase in the order of the Toeplitz matrix, that undergoes an "edging". From an operational point of view, this equates to the emergence of a new eigenvalue in the spectrum, simultaneously to the variation of prevoius eigenvalues depending on the signal behavior.

The basic idea of the EAM method is to study the variations of only the first eigenvalue (the so called "dominant eigenvalue"), depending on the order of the autocorrelation matrix, i.e. on the edging order of it. To accomplish this, the derivative of the eigenvalues vector $\lambda(n)$ with respect to the edging order n (i.e. with respect to time) must be performed. The derivative will return (if such signals exist) every consistent signal immersed in any kind of noise. The method, if run properly, can also detect signals of very low signal-to-noise ratio (SNR).

## 15.3 Signal Analysis

We have performed some simulations with different signals, to demonstrate the behavior of the KLT employing the EAM method, and to improve it from the point of view of computational load and computer memory performance. The tests were accomplished with the following hypothesis:

1 Bandwidth: the signals were simulated with different spectral features:
   a. Monochromatic signal;
   b. Multi-monochromatic signals, tuned at close frequencies each other (i.e. FM)
   c. Signals with a gaussian spectral distribution, with different half-power band width distribution (HPBWD) values;

2 Signal-to-Noise Ratio (SNR): the signals were simulated with different SNR values, until very low values (0.005), the latest performed by averaging over the time of the autocorrelation function;

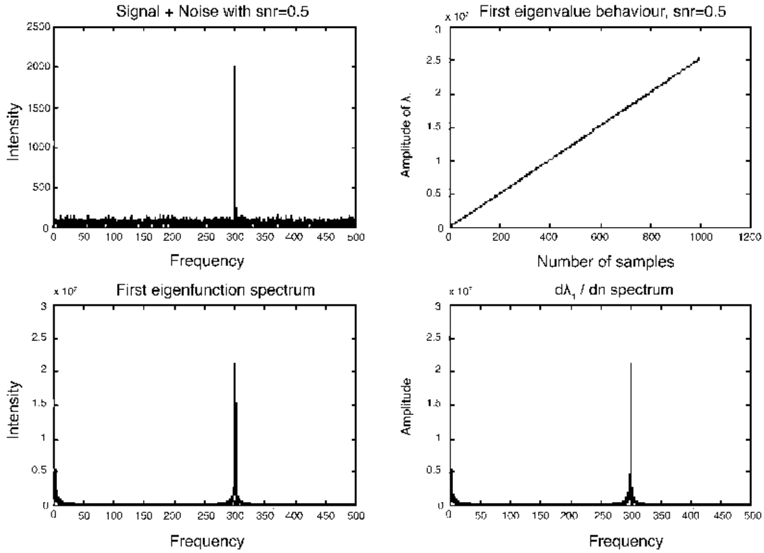



**Figure 15.4** Simulation of a signal with SNR=10, monochromatic spectrum. On the top, from the left: FFT, 1st eigenvalue; below: 1st eigenfunction, derivative of dominant eigenvalues vector over the time $\partial\lambda(n)/\partial n$ .

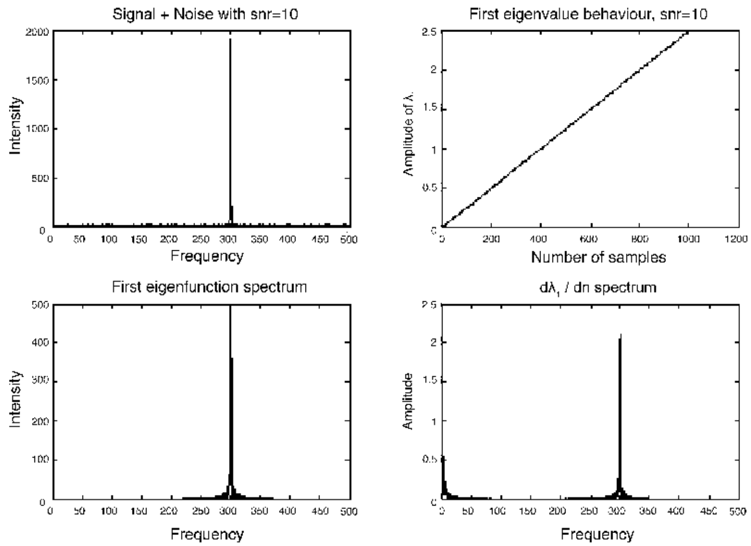



**Figure 15.5** Simulation of a signal with SNR=0.1, monocrhomatic spectrum. On the top, from the left: FFT, 1st eigenvalue; below: 1st eigenfunction, derivative of dominant eigenvalues vector over the time $\partial\lambda(n)/\partial n$ .

3 The number of eigenvalues taken into account: typically, the signals simulated were of a time duration of 1 s, then $N_{MAX} = \frac{1}{\Delta t}$, where Δt is the sampling interval;

Based on the simulations (Figures 15.4 and 15.5), we implemented a first analysis method that seems to give good results. Of course, this is a very preliminary approach and definitely needs more investigations. The Edging of Autocorrelation Matrix seems to be a good method to detect consistent signals, of any bandwidth, immersed in a high noise environment. However, it seems that the analysys of the first eigenvalue does not give us exceptional results with broadband signals, where it becomes necessary to consider other eigen-spectrum components.

## 15.4 Conclusion

Thus far, no evidence of ETI signals has been obtained from SETI observations at the Medicina radiotelescopes. However, the Serendip IV system is continuously operating for both SETI observations and RFI monitoring.

Preliminary tests were performed on the extraction of signals from the noise by use of a KLT-based algorithm. The approach we present in this chapter is much faster than traditional implementations of the KLT, because we compute only the dominant eigenvectors/eigenvalues. In addition, we accept the limitation of obtaining only a good estimation of them, rather than computing their precise values. In our approach an average of the dominant eigenvectors FFTs, ready to be further averaged per each cycle, is considered. This is necessary to get a final spectrum with the requested ***rms*** (root mean squared) noise. The sum of the averaged FFTs of the best data projection of the components on the dominant axis, seems to afford much greater detection possibilities than using only the eigenvalues alone as a detection key, as has been previous practice.

This method is, in any case, a starting point for further investigation and development of this very powerful transform for signal detection in the SETI context.

**Acknowledgements**

My Istituto Nazionale di Astrofisica colleagues Jader Monari, Salvatore Pluchino, and Francesco Schillirò contributed significantly to this research, and are hereby acknowledged as co-authors of this chapter.

## References

Dixon, R.S. et al. 1993. On the Detection of Unknown Signals. In: G. Seth Shostak (ed.) *Third Decennial US-USSR Conference on SETI*, Astronomical Society of the Pacific Conference Series, Vol. 47, 128-140.

Maccone, C. 1994. *Telecommunications, KLT and Relativity,* Volume 1, IPI Press, Colorado Springs, CO.

Maccone, C. 2001. "Innovative SETI by the KLT", Pešek Lecture 2003, paper # IAA.9.1.01 presented in Bremen at the 2003 International Astronautical Congress.

Montebugnoli, S. et al. 1999. *Medicina Radio Telescopes: The RFIs-Sentinel 2 System for RFIs Monitoring Activities*. In R.J. Cohen and W.T. Sullivan (Eds) IAU 196. Symposium Astronomical Society of the Pacific.

Werthimer, D. et al. 2000. *The Serendip IV Arecibo Sky Survey* - *A New Era in Bioastronomy.* Astronomical Society of the Pacific Conference Series, Vol. 213, 479-483.

# 16

# A Sentry on the Universe

Robert Dixon,
*Acting Director, Ohio State University Radio Observatory*

This chapter introduces the Argus radio telescope concept, a radically new approach to radio telescope design (not to be confused with The SETI League's Project Argus, a linked global array of fairly conventional radio telescopes). It represents a complete departure from designs that have been used ever since the invention of the first telescope by Galileo in 1609. It starts over from the beginning, and in doing so overcomes the legacy of Galileo. Argus can do things that are amazing and even hard to believe, when viewed in the context of previous telescopes. Here are some examples.

1 Telescopes typically must be "pointed" in a specific direction. Argus looks in all directions at once, so the concept of "pointing" does not apply.

2 Typically only one person can use a telescope ("look thru the eyepiece") at a time. Argus can be used by everyone on Earth at the same time. Translated into scientific research terms, it means that telescopes no longer have to be shared and scheduled or competed for, as is the case now.

3 Telescopes typically are large steel structures, which require precisely made mirrors or lenses, and precision machinery and control systems to "point" them. Argus has no precisely made parts, and no moving parts.

4 Telescopes typically are affected by wind, temperature, and gravity, so great effort, complication and cost is required to counteract those effects. Argus is unaffected by any of these problems.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_16, © Springer-Verlag Berlin Heidelberg 2011

5 Radio telescopes are adversely affected by radio interference from man-made (both terrestrial and satellite) and natural (such as the Sun) origins. Argus can create nulls in the directions of both fixed and moving interfering signals, and cancel them out.

6 Telescopes are always focused at infinite distance, as that is where astronomical objects are located. Argus can be focused at any closer distance, allowing observation of atmospheric effects such as lightning, or tracking Earth satellites. Also because of this focusing capability, Argus can measure the distance to terrestrial and near-Earth signals, including sources of radio interference, so that they may be identified.

7 The cost of typical telescopes is significantly dependent on the cost of labor involved in building them, and the fact that most of them are one-of-kind devices. Since the cost of labor increases with time, so do telescopes. The cost of Argus is dominated by the cost of the computing required, which is the best possible limitation today, since the cost of computing continues to drop drastically. Also, Argus takes advantage of mass production of its many small elements.

8 Typical telescopes cannot detect intermittent or transient signals that arrive from unknown directions, since they are looking in the wrong direction. In fact all the telescopes in the world taken together are looking at only a small fraction of the sky. Argus sees everything. Nothing escapes detection.

9 Argus creates an archive of every signal it has ever received, in its raw data. Since the amount of data is large, it may be that not everything can be examined in great detail in real time. But if at some later date an interesting object is discovered elsewhere (such as a supernova), then Argus has the ability to retroactively observe that object in great detail from the first day the telescope was turned on.

## 16.1 Introduction

The time has come to seriously consider a fundamentally different approach for radiotelescopes. Instead of large steel dish structures, a large number of small omnidirectional antennas can be used in an array to obtain much greater performance at lower cost. Such arrays are commonly called “phased” arrays, but that implies narrow bandwidth, so a more correct term for what is discussed here is a “timed” array.

The name Argus originated from the mythological guard-being that had 100 eyes and could look in all directions at once. This name was used for an omnidirectional, all-seeing antenna by Arthur Clarke in his novel *Imperial Earth* and by Carl Sagan in his novel *Contact*. Basically, an Argus array uses computers to combine the outputs of a large number of array elements to create a large number of beams that simultaneously cover the entire sky. An Argus array is actually a telescope, since it forms an image, whereas typical dish antennas are not telescopes at all, and are more accurately called teleradiometers.

Another name which has been applied to Argus arrays is radio camera. Note that this Argus telescope is not related to the Project Argus all-sky survey of The SETI League, described elsewhere in this book, which came later. Paul Shuch asked Robert Dixon for permission to use the same name, and it was granted.

## 16.2 Advantages of Argus over a Dish-type Antenna

Compared to a conventional dish, an Argus timed array provides many advantages, including simultaneous high-gain omnidirectional sky coverage (no scanning), high sensitivity (arbitrarily long integration time), high resolution, variable beam size and shape, low and moveable sidelobes, wide bandwidth, detection and tracking of transient and moving sources, adaptive and retroactive observations, interference rejection, and high efficiency. While the term “high-gain omnidirectional antenna” may seem self-contradictory, that is true only in the transmitting case or only if passive transmission lines are used to form multiple beams in the receiving case. In fact, information and energy are falling on any radio telescope from all directions all the time, and the vast majority of it is ignored; that is in one sense considered “good”. The apparent contradiction arises from use of the principle of conservation of energy, whereas the applicable principle is conservation of information. The larger a dish antenna is, the worse it becomes in terms of using all the energy and information that falls on it. Figure 16.1 illustrates the extremely low total efficiencies of some well-known dish-type antennas, in comparison to the Argus approach.

The sensitivity of an Argus array is the same as that of a dish having the same total collecting area and the same sensitivity receivers. In terms of cost, an Argus array is inherently less expensive than a dish since it takes advantage of mass production; has no large or moving parts; and is unaffected by gravity, sunlight or wind. It has no tight mechanical tolerances and requires no mechanical maintenance. The construction cost of a dish increases with time (since labor costs dominate), whereas the construction cost of an array decreases with time (since computing costs dominate). Hence an array must become less costly at some time, even if its other advantages are ignored.

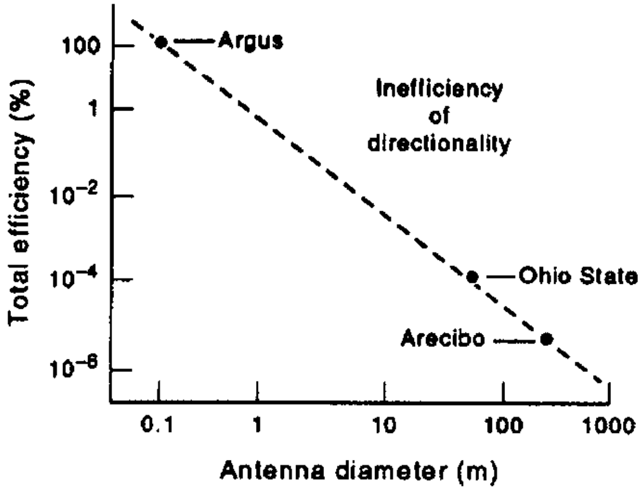



**Figure 16.1** Efficiencies of various antennas.

In terms of flexibility, an Argus array has a number of advantages. It can be easily expanded or changed in shape; its resolution can be chosen independently of its collecting area; and its resolution, beamshape, and sidelobes can be changed at will by software. One example of this is for sidelobe reduction as shown in Figures 16.2(a) and (b) (Bickmore, 1966). The main beam of an array is only sightly affected by small changes in the array, whereas the sidelobes are strongly affected by such changes. The sidelobes are also half as wide or less than the main beam. If the size of the array is change periodically and the output averaged, the sidelobes will tend to cancel. The array size can be changed by switching the outer elements on and off (by changing their weighting factors). This results in the sidelobe reduction shown in Fig. 2(b).

Argus can be self-calibrated using test transmitters located within the array. It is fault tolerant since there is no single point of failure, unlike a conventional dish that has a single signal path from feedhorn to detector.

In terms of capability, an Argus array can do many things a dish cannot do, including observe multiple objects simultaneously, track rapidly moving objects, detect transient events in unknown directions, survey the entire sky in a single integration period, receive with very wide bandwidth, observe adaptively in response to current results, and re-observe retroactively

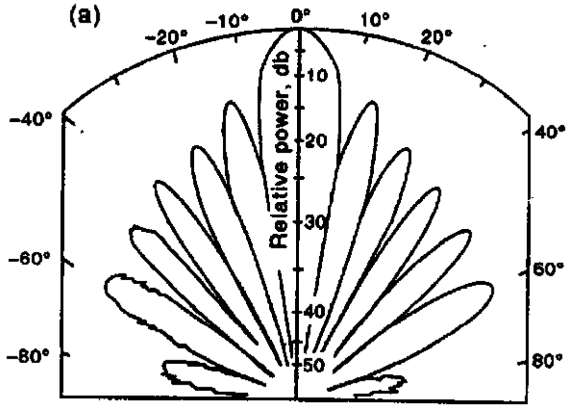



**Figure 16.2a** Classical array sidelobes.

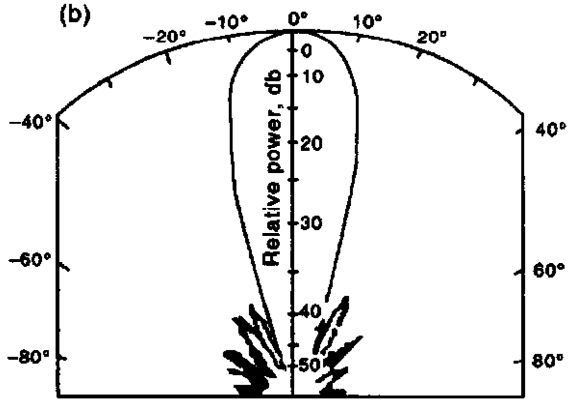



**Figure 16.2b** Ultralow sidelobes after switching the array size.

objects or events not recognized initially. The retroactive observations can be done by playing back the recorded data from the array elements, and if desired the beam and processing equipment can be re-optimized for the re-observation.

Argus has many advantages over a dish in terms of its ability to deal with radio frequency interference (RFI). The elements can be designed to have nulls at the horizon for rejection of terrestrial signals. The elements are on the ground, in contrast to the elevated feed of a dish, hence the signal strength of terrestrial signals is less. Small shield fences (rounded appropriately at the top to reduce diffraction) can be used around the elements or array if necessary for further rejection of terrestrial signals. The direction of any RFI signal is immediately known to Argus since one of its beams always points toward the RFI source, and it will be strongest in that beam. That beam will also provide a nearly noise-free version of the RFI which can be used to characterize and identify it and to blank it or cancel it in the rest of the beams. Diagnosis of RFI is immediate with no need to steer the telescope "off-source" to see if it goes away. If it is received in more than one beam, then it is known to be in a side lobe and hence be RFI.

Since each beam can be separately optimized, permanent nulls can be generated by each beam in the direction of known fixed RF1 sources. Adaptive nulls can be generated in real time as needed to deal with transient RFI. Moving RFI sources such as aircraft or spacecraft can be immediately identified as such by their movement among the beams, and henceforth tracked, predicted, and removed from the telescope output. Argus can also identify RFI sources by their distance, since it can simultaneously focus itself at all distances. A modest 64-element Argus can resolve distances out to about 3 km, whereas an Arecibo-sized Argus can do so out to 500 km. These distances would allow discrimination against almost all manmade signals.

Radio telescope design can range from the extreme of a single dish with a narrow field of view, to the other extreme of an array of many small omnidirectional antennas (Argus). Many radio telescopes are now somewhere in between, having an array of small dishes, each with a modest field of view, such as the Allen Array (Welch et al., 2009).

## 16.3 Some Other Work in this Field

Several telescopes have been built and proposed which image a small portion of the sky over a narrow bandwidth, but none have approached the general case discussed here of the entire sky at a wide bandwidth. Daishido et al. (1984, 1986) proposed a 4096 element horn array operating at 10 GHz, imaging a 9-degree field with a bandwidth of 20 MHz. The Clark Lake telescope (Erickson et al., 1982) has 720 conical helix elements, operating over the range 15-125 MHz, imaging a 6-1.5 degree field with a bandwidth of 0.15-3 MHz. NRL (Johnston et al., 1989) proposed a 20-element array of 3 m dishes, covering a 2.6-degree field at 2.7 and 8.1 GHz with bandwidths of 64 and 448 NHz. A conference was held in 1989 to discuss a Radio Schmidt telescope (Dominion Radio Astrophysical Observatory, 1991) with a "strawman" configuration of 100 12-m dishes, mapping a 1.5-degree field at 1500 MHz and other bands. Along with his coworkers, Steinberg (1983, 1991) invented the term “Radio Camera” and has written extensively on the topic. His interest is in imaging aircraft in the vicinity of airports to provide much greater detail than is now provided by radars, such as the shape of the aircraft, whether the landing gear is down, etc. His plan is to use a single nondirectional transmitter and a large number of receiving elements placed essentially randomly wherever possible throughout the airport. His camera work is contrasted from that discussed here in that it is “flash” photography rather than “available light” photography.

More recently, several radio telescopes have been designed which begin to move in the direction of the Argus design. Some of these are: The Murchison Widefield Array (Lonsdale, 2009), The Long Wavelength Array (Ellingson, 2009), the LOFAR telescope (de Vos et. al., 2009) and the Eight-meter-wavelength Transient Array (Ellingson, 2010).

## 16.4 The Argus Mark 1 Telescope

We have constructed and operated a prototype 8-element circular Argus array at 162 MHz (Bolinger, 1988). Its parameters were chosen to match

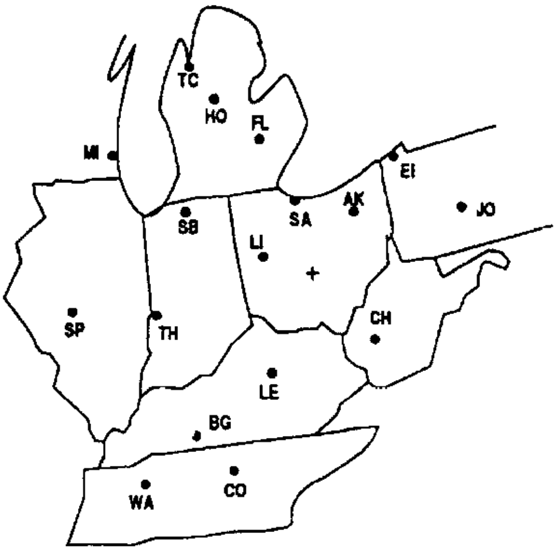
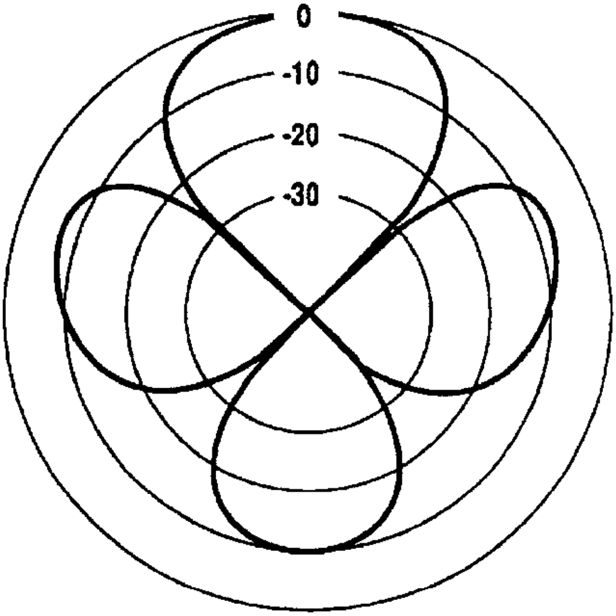



**Figure 16.3** Weather stations nearest Columbus, Ohio (shown by +).




**16.4** Theoretical beamshape of the Argus Mark I array. The radial scale is in db.

those of available radio stations so they could be used as known sources. The United States Weather Service operates many FM transmitters throughout the country, which continuously make voice announcements of weather conditions. There are hundreds on the same frequency, all at various and varying signal strengths and directions from any given location, making them ideal test signals for developing Argus beamforming techniques. Figure 16.3 shows the locations of the stations nearest our location in Columbus, Ohio. Our array was one wavelength in diameter, giving a theoretical beamwidth of about 90 degrees between nulls (see Figure 16.4). A bandwidth of 7 kHz was sampled for 1.7 ms, and then processed to form 36 simultaneous beams, equally spaced around the horizon.

Each beam was averaged over 360 samples and then its resolution was enhanced with a deconvolution method analogous to CLEAN. Plots of the received signals were made every hour (an example is shown in Figure 16.5). Note that many signals on the same frequency are clearly resolved

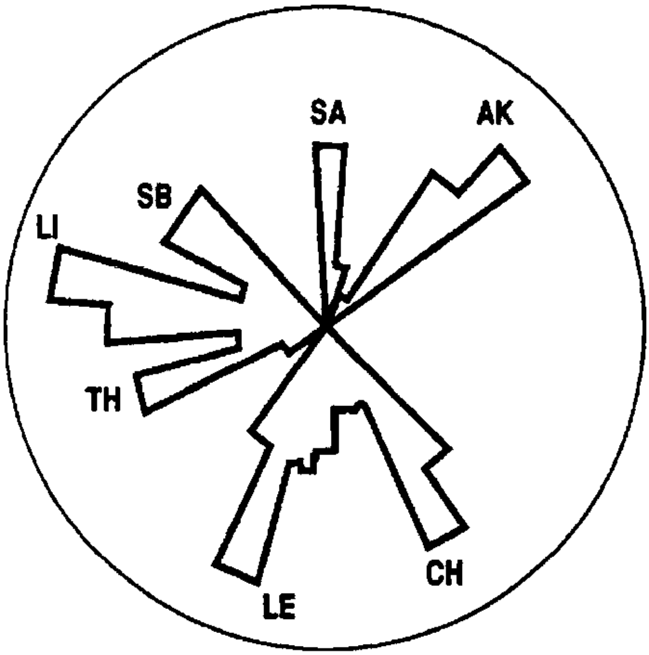



**Figure 16.5** Sample plot of received signals using the Argus Mark I array.

and that the resolution is much greater than would be expected from such a small array. By comparing the plots made over a long period of time, one can observe interesting effects as propagation changes to the various stations, and as thunderstorms move past the Argus location.

### 16.4.1 Argus array element design

The elements of a general-purpose Argus array should have hemispherical coverage, aimed straight up. They should have nulls at the horizon for rejection of terrestrial interference, have dual circular polarization, be broadband, and mass producible.

The best candidates are from the helix family. A multifilar contrawound conical helix is one antenna that can achieve these requirements. Such an antenna element design can be visualized by combining the architecture of the helices shown in Figure 16.6 (Gerst and Worden, 1966) and Figure 16.7.

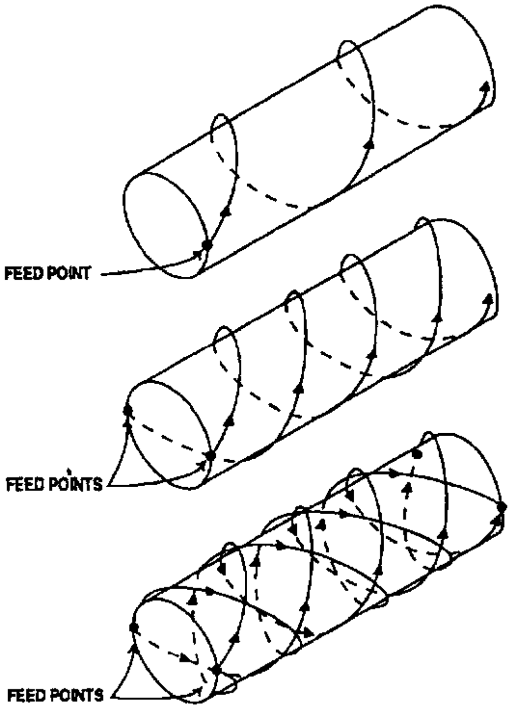



**Figure 16.6** Multifilar contrawound cylindrical helix.

**Figure 16.7** Contrawound conical helix.

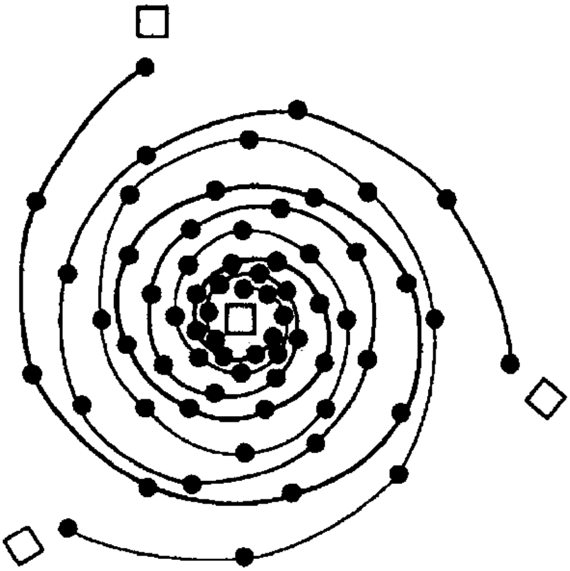


**Figure 16.8** Element locations along the arms of a multiarm logarithmic spiral. The open squares are calibration transmitters.

### 16.4.2 Argus array design

The Argus array geometry should have approximately circular symmetry (for uniform azimuth beams), not have uniform spacings (to avoid grating lobes), and be spatially and frequency (element size) tapered from the center outward (to achieve frequency independence). Placing the elements logarithmically spaced along the arms of a multiarm logarithmic spiral (Figure 16.8) is one way to achieve these requirements. To calibrate the array, small remote-controlled omnidirectional transmitters can be placed inside and near the array. In the example shown, they are located at the center of the array and at the ends of each spiral arm.

### 16.4.3 Argus computing architecture

The performance of an Argus array (as measured by its number of elements, number of beams, and bandwidth) is limited primarily by its computing power. Hence this is the most critical portion of the design. Fortunately, available computing power is rapidly increasing and its price is falling. Thus the Argus capability can only improve with time. One appropriate computing

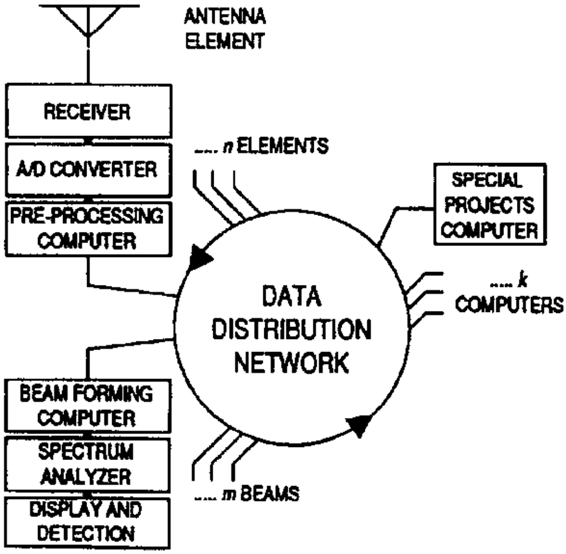



**Figure 16.9** One possible Argus computing architecture.

architecture for Argus beamforming is shown in Figure 16.9. A small computer is used at each of the n elements, which does all computations that can be done on the data coming from that element. A different set of m small computers is used to perform the calculations for each of the m beams. In general, m is much greater than n, since the array is sparse.

All the element and beam computers communicate via a general purpose or custom network. An example of a general purpose network is Ethernet, and a custom purpose network could be a conceptual token ring network. A token ring network may be viewed as a circular railroad track. As the train passes each element, the element computer places its load of data into the boxcar reserved for it, and there are n boxcars. So by the time the train reaches the beam computers, it is fully loaded. Each of the m beam computers reads the data from all of the boxcars as they pass it. As soon as the train has passed a beam computer, that computer starts its dedicated task of calculating its beam, using a pipelined approach.

Note that all of the beam computers will not necessarily (and need not) complete their calculations at the same time, and there could be a number of beam calculations proceeding through their pipelines at the same time. The only requirement is that their pipelines be able to keep accepting new data as fast as the trains arrive. There may in fact be many trains circling the track at the same time. All of the element and beam computers are dedicated, programmed, and optimized to do just one set of fixed calculations, so they can be made very fast. The element weightings used for beamforming are kept in lookup tables that are separate for each beam and can be rapidly changed as desired. In order to achieve frequency-independent beamforming, the lookup tables are two-dimensional, containing weightings for direction and for each band of frequencies.

In addition to the element and beam computers, there is another much smaller group of small computers attached to the network, each dedicated to some special project. Examples of such projects include monitoring a pulsar, tracking a spacecraft, lunar occultation, identifying RFI, calibrating the system, etc. Each special project computer is free to use whatever data is wishes and make whatever calculations it wishes, with no interference with the main computers or with each other. Hence there is no limit to the number of special projects that can occur simultaneously. One particularly important special project is to record all the element data in a compressed form for later analysis. This makes it possible to re-observe an event that occurred long ago, but was not recognized at the time. The special projects computers can be attached to the worldwide Internet, making it possible for anyone anywhere to control them and to obtain data from them.

The computational power required for an Argus array of equivalent size to a large dish is greater than can be reasonably achieved today in the microwave region. But future developments in computing will make this possible, and today modest arrays at lower frequencies are possible. Argus is limited only by the available computing power.

### 16.4.4 Argus output data example

One output of Argus is a real-time image of the whole radio sky. A circular CRT display, centered on the zenith (or transformed to the celestial pole if desired) indicates the directions of all signals being received. Signal strength is mapped into display intensity, and signal frequency is mapped into color (low frequencies toward the red, etc.). Signal polarization type and degree can be displayed with small ellipses of varying axial ratio, orientation and diameter. The integration time can be arbitrarily long, so eventually the telescope would reach its classical resolution-limited condition. For large signal-to-noise ratios, super-resolution techniques can be applied to achieve greater resolution. This would result in strong sources having small bright dots, whereas weaker ones would be more diffuse.

Such a display would show bright dots around the outer edge, representing terrestrial signals, and a line of dots along the synchronous satellite orbit (assuming their frequencies were included in the Argus coverage). Other spacecraft and all aircraft would appear as lines across the display of all colors. Aircraft can be detected by a number of modes, including their transponders, voice and data transmissions, reflection of distant terrestrial transmitters, and thermal radiation. Continuum radio sources would appear randomly scattered throughout the display, being generally white in color because of their broadband emissions, and moving slowly as the Earth turns.

Once an essentially noise-free image of the sky is obtained, a differential mode of operation can come into operation. In this mode, the telescope output displays only the differences between the "normal" sky and the current sky. This drastically reduces the amount of data to be displayed, and allows for immediate discovery of anything which has changed, appeared, or disappeared. Such discoveries could automatically be announced immediately by one of the special projects computers to everyone around the world who chose to receive such announcements, via an Internet newsgroup or mailing list.

## 16.5 The Argus Mark II Telescope

Many technical problems remain to be solved before a large general-purpose Argus array can be constructed. The most limiting factor is the computational power required for the beamforming operations. We are now looking into optimized hardware, algorithms and architectures for this. Until a larger and more general prototype is built and operational experience gained, none of the design aspects can be finalized to the point where mass production can be used to create a truly useful instrument. One of the important early choices is the frequency range. The effective aperture of a hemispherical-coverage element is lambda squared over two pi. The cost of Argus is approximately proportional to its number of elements. Hence to obtain a large collecting area at minimum cost for an initial development array it is desirable to make lambda large (i.e. use relatively low radio frequencies). But if one goes too low the elements become large and difficult to construct, and at still lower frequencies (about 30 MHz) ionospheric effects begin to occur.

A second system design choice is system bandwidth, but that choice is straightforward. The system cost is directly proportional to bandwidth. The RF portions of the system can be easily designed for large bandwidth and will be, even though the computing portion of the system may not be able to process a bandwidth that great. Then the system bandwidth is chosen to be whatever the current computing system can handle, and is expanded with time.

The computing power required for an Argus beamforming system is approximately $2B(2K + 1)NL$ multiplications per second (Brown, 1993) where B is the system bandwidth in Hertz, K is the size of the interpolation filters, N is the number of elements, and L is the number of simultaneous beams formed. For a large-scale system, with a resolution of 10' of arc, 90-degree field of view, and a bandwidth of 1 MHz, about 500 × 10 (to the twelfth) multiplications per second would be required. No single computer presently in existence can accomplish this. This does not mean that such a system is impossible, but it does mean that conventional single processor serial computing is unsuitable for implementing it. Instead, specially designed distributed computing hardware will be necessary for a large-scale radio camera.

Currently available technology could be used to implement a specialized beamforming processor capable of computing at this rate. Dedicated integrated circuits exist that can multiply at much higher rates than most general-purpose computers. Actual implementation of a radio camera telescope on this scale will have to wait for the cost of computing to fall. This leads to consideration of a more modest example, which would provide for development of designs and algorithms for larger systems, in anticipation of declining trends in computer costs.

A prototype Argus with 32 elements, a resolution of several degrees, and a bandwidth of several kHz is well within the capabilities of current computer systems, indicating that a functional prototype could be implemented without custom designed hardware to perform the imaging. Although such a prototype would have limited resolution, it would be a versatile experimental instrument. It would allow evaluation of many different geometries and many different implementations of the beamforming algorithms. It would allow various calibration schemes to be tested. Experimentation would not be limited by the computer power within the system, because the data collected could easily be transferred to different computers. The experience in radio camera technology which would be gained from this system would be invaluable when construction of larger scale radio cameras becomes economically feasible. This has led to the development of the Mark II Argus array.

The Mark II Argus array has been constructed and is now in operation. It currently has 24 elements (expandable to 32), operating in the frequency range 1200-1700 MHz. Each element is a planar Archimedean spiral antenna, having right-hand circular polarization. The beamwidth is approximately 6 degrees. This design was chosen to minimize cost and construction difficulty.

The current array is shown in Figure 16.10 and an individual element is shown in Figure 16.11. Details of the system are given in Ellingson, et. al. (2008). A complete list of Argus Mark II -related papers is located at http://argus.naapo.org/~rchilders/swe_argus_pubs/, and further relevant references are contained in the bibliographies of those papers.

Argus Mark II routinely detects the sun, the Cygnus complex, various earth satellites and occasional unknown blips. Examples are shown at http://

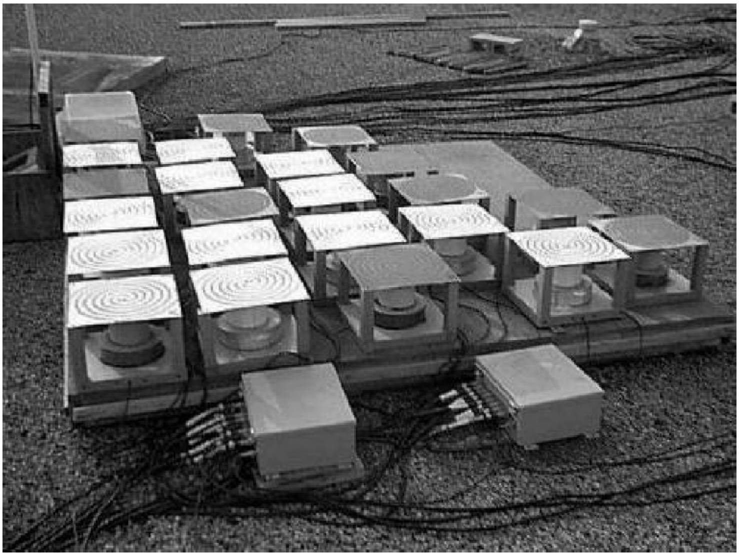


**Figure 16.10** The Argus Mark II antenna array. Elements rest on a platform covered in fine wire mesh that serves as a larger ground screen. Each of the two gray boxes in the foreground contain line amps for 8 antennas (a third box is located on the far side of the array). The base of the calibration source mast is visible in the upper left.

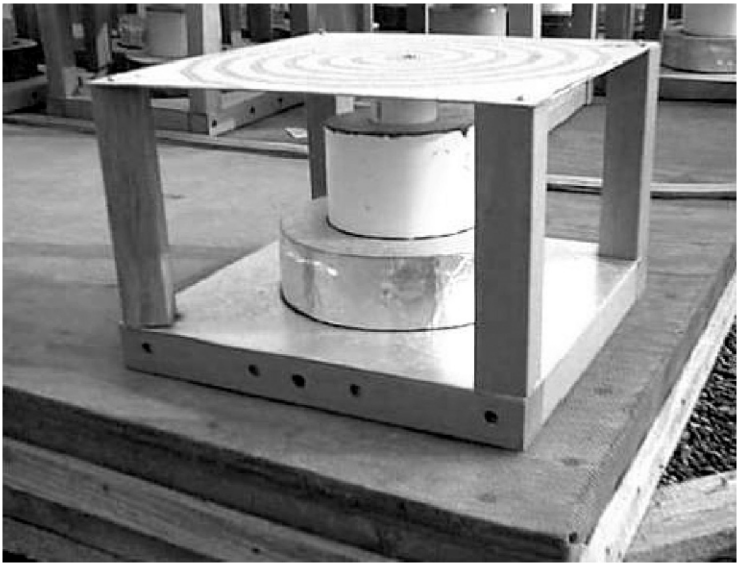


**Figure 16.11** Side view of an Argus Mark II spiral antenna element. The spiral is on a printed circuit board located at the top. A stepped ground plane is used below the spiral to achieve wide bandwidth.

argus.naapo.org/~rchilders/. The output of the Argus Mark II array is made available globally via the Internet at http://argus.naapo.org/displays.html

An overview of all these aspects, plus more in non-technical terms, is given at http://www.ohioargus.org

### 16.5.1 Some Argus Mark II signal analysis techniques

Since Argus is completely computer-driven, it lends itself to the application of many signal detection algorithms, any and all of which can be used simultaneously.

Some early investigations were made of the Karhunen-Loeve Transform (discussed elsewhere in this book), to answer the basic question of whether a signal is present in the noise or not, without having to do spectral analysis and then examine the contents of each of the spectral bins (Dixon and Klein, 1995). This is done by calculating the autocorrelation function of the data in a single beam, and examining its largest eigenvalue to see if it significantly exceeds the others. If so, a signal is present, even though at that stage nothing is known about its characteristics. The largest eigenvalue can be approximated without calculating all the others. The object of this approach is to eliminate the amount of computation required in calculating mostly empty spectral bins. The detailed characteristics of the signal can then be determined by further analysis.

An analogous approach to processing the entire array output data can be done to determine if a signal is present in any of the telescope beams in the sky, without going to the effort of computing all the beams and then examining each one for a signal. This is done by calculating the two-dimensional cross correlation function between all the element outputs and then examining its largest eigenvalue. If it significantly exceeds the other eigenvalues, then a signal is present, even though it is not known at that stage which beam it is in. Once a signal is detected, further analyses can then be done to determine its direction and characteristics.

The first example of KLT above is one dimensional, as a function of time. The second example is two-dimensional, as a function of space. It is interesting to speculate that a three-dimensional KLT might be advantageous, to simultaneously combine both types of signal detection.

### 16.5.2 Argus subarrays

Investigations have also been made into the possibility of subdividing an Argus array into smaller arrays, forming the beams of each subarray separately, and then forming the overall beams of the entire array using the subarray intermediate results (Dixon, 1997). The intention of this approach is to reduce the amount of computation required to calculate all the beams. The specific case of a uniform array with half-wavelength spacing between elements was evaluated (http://argus.naapo.org/~rchilders/swe_argus_pubs/argus_subarrays_dixon_1kt_1997.pdf). The net result is:

1 The total amount of computation required can be reduced from N(squared) to N(3/2), where N is the number of elements in the array

2 The optimum size for each subarray is N(1/2) elements

3 The process can be continued further to create subarrays of subarrays to any desired depth of nesting, to achieve even further reduction in the amount of required computation. The ultimate reduction occurs when each smallest subarray contains only two elements

4 With 2 elements per subarray and nesting by factors if two, the computation required reaches its lower limit of N log N, and is equivalent to the Fast Fourier Transform

5 However, the disadvantage of using subarrays is the partial or total loss of the ability to tailor the beam shapes and sidelobes, and to create nulls in desired directions. Nevertheless, it may be desirable to initially create all the beams in the most efficient was possible, and then use those initial results to recalculate more optimum beams dynamically for each situation

6 Since Argus is software-driven, the array can be subdivided very simply in many different ways

## 16.6 The Big Picture

It is commonly believed that humankind is basically aware of everything that goes on around us in the universe. This may seem logical, given all the telescopes in operation around the earth. But the fact is that all telescopes combined see only a tiny fraction of the universe and frequency spectrum at any one time, and as larger telescopes are built, they see even less. In our quest for ever greater detail about the trees, we are ignoring the forest. There are undoubtedly transient events occurring all the time of which we are unaware; previous examples include pulsars and supernovae. We have no global view of our electromagnetic environment, encompassing both natural and manmade signals. We have an obligation to open our eyes widely and be aware of our surroundings so we can learn more about the universe and understand the big picture. Argus will make this possible.

One fully implemented Argus array can simultaneously carry out all the observations now being done by other comparable radiotelescopes, not only for astronomy but for all scientific and commercial monitoring of the electromagnetic environment. The universality and versatility of the Argus approach, together with its riding the crest of mass-production computing, make it inevitable at some time in our future.

## References

Bickmore, R.W. 1966. Time Versus Space in Antenna theory. In R.C. Hansen (Ed.) *Microwave Scanning Antennas*, Vol. III, Academic Press, New York, p. 315.

Bolinger, J. 1988. A Simultaneous Multi-beam Phased Array Using Digital Processing Techniques. Master's thesis. The Ohio State University. Summary by Jerry Ehman at http://www.naapo.org/Argus/docs/radiocamera.htm

Brown, Stephen B 1993. Radio Camera Arrays for Radio Astronomy. Master's thesis, Ohio State University.

Daishido, T., et al. 1984. Phased Array Telescope with Large Field of View to Detect Transient Radio Sources. In J.A. Roberts (Ed.) *Indirect Imaging*, Cambridge University Press, Melbourne.

Daishido, T., Asuma, K., Ohkawa, T., Obara, H. and Komatsu, S. 1986. A Design for Digital Signal Processing in a Large Field Radio Patrol Camera. *Astrophys. Space Sci*. **118**, 467-469.

de Vos, M., et.al. 2009. The LOFAR Telescope: System Architecture and Signal Processing, *Proc IEEE*, **97: 8**, 1431-1437.

Dixon, Robert and Klein, Charles A. 1993. On the Detection of Unknown Signals. Presented at the USA-USSR Joint Conference on The Search for Extraterrestrial Life, August 8, 1993, in Santa Cruz, California. Published in Seth Shostak (Ed.) ASP Conference Series, Vol 74. Available on line at http://adsabs.harvard.edu/full/1993ASPC...47..129D

Dixon, R.S. 1997. Argus with Software and Hardware Subarrays, Square Kilometer Array Workshop, Sydney, Australia, December. http://argus.naapo.org/~rchilders/swe_argus_pubs/argus_subarrays_dixon_1kt_1997.pdf

Dominion Radio Astrophysical Observatory, Herzberg Institute of Astrophysics, and National Research Council of Canada. 1991.The Radio Schmidt Telescope: Proceedings of a Workshop held at Penticton 11-12 October 1989, 1991.

Ellingson, S.W., Hampson, G.A. and Childers, R.K. 2008. Argus: An L-Band All-Sky Astronomical Surveillance System (pdf*), IEEE Trans. Antennas & Propagation*, **56:2**, 294-302.

Ellingson, Steven, et al. 2009. The Long Wavelength Array, *Proc IEEE*, **97:8**, 1421-1430.

Ellingson, Steven, et. al. 2010. The Eight-Meter-Wavelength Transient Array, Virginia Tech University, http://www.ece.vt.edu/swe/eta/ (retrieved June 2010).

Erickson, W.C., Mahoney, M. J. and Erb, K. 1982. The Clark Lake Teepee-tee telescope. *Astrophys. J.* (Suppl. Ser.) September, 403-419.

Gerst, C.W. and Worden, R. A. 1996. Helix Antennas Take Turn for Better. *Electronics* **22**, 101.

Johnston, K.J., Fiedler, R. L. and Simon, R. S. 1989. The Fast All Sky Telescope (FAST). *Naval Research Laboratory Report*, November, 12.

Lonsdale, Colin, et al. 2009. The Murchison Widefield Array: Design Overview, *Proc IEEE*, **97: 8**.

Steinberg, B.D. 1983. *Microwave Imaging with Large Antenna Arrays: Radio Camera Principles and Techniques*. John Wiley & Sons, Inc., New York.

Steinberg, B.D. and Subbaram, H. 1991. *Microwave Imaging Techniques*. John Wiley & Sons, Inc., New York.

Welch, Jack, et. al. 2009. The Allen Telescope Array: The First Widefield, Panchromatic, Snapshot Radio Camera for Radio Astronomy and SETI, *Proc IEEE*, **97: 8**, 1438.

# 17

# Pondering the Fermi Paradox

Stephen Webb,
*University of Portsmouth, UK*

The past two decades have witnessed an unprecedented increase in the amount and quality of observational data available to astronomers and cosmologists. Orbiting observatories such as the Hubble Space Telescope and the Wilkinson Microwave Anisotropy Probe (WMAP) have peered at the universe through a variety of windows in the electromagnetic spectrum. Ground-based projects such as the Sloan Digital Sky Survey (Adelman-McCarthy et al., 2008) and the 2dF Galaxy Redshift Survey (Colless et al., 2003) have mapped the distribution of galaxies in exquisite detail. Taken together, data from these and other projects have transformed our understanding of the large-scale structure of the universe, have steadily improved our knowledge of key cosmological parameters, and have provided compelling evidence in favor of a simple cosmological model (Spergel et al., 2009). Our best understanding of the universe as of 2010 is encapsulated in the ΛCDM model, surely a pinnacle of human ingenuity, which is in agreement with all observations made to date.

The standard model of cosmology is not without its puzzles, however. For example, WMAP data tell us that 72% of the energy density of the universe is in the form of dark energy, the nature of which is unknown; the data are consistent with a small, non-zero cosmological constant but explaining its size is a major unsolved problem for theoretical physicists. A further 23% of the energy density of the universe consists of some form of dark matter, but as of 2010 we have no direct evidence for the existence of dark matter particles. Less than 5% of the universe's energy density exists in the form of baryonic matter, and even most of that is missing in the present epoch

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_17, © Springer-Verlag Berlin Heidelberg 2011

(although there are recent indications that the missing baryons are to be found in a warm–hot intergalactic medium, filaments of which form the so-called cosmic web). In short, the plethora of observational data has given rise to a successful model of cosmology – but a model that contains within it several intriguing puzzles.

Perhaps there is another conundrum lurking in those terabytes of data, another puzzle in our cosmological model: at none of the frequencies at which astronomers have been observing has any trace come to light of a signal that could not be explained in terms of natural phenomena. We see no signs that intelligent life has disturbed the universe. Isn't that rather odd? There are no signs of infrared emission from Dyson spheres (Dyson, 1960). No signs of artificial lines in stellar spectra (Whitmire and Wright, 1980). No signs of self-replicating probes (Bracewell, 1960). No signs of vehicles that are burning antimatter (Harris, 1986). No signs in light curves of the transiting of artificial objects (Arnold, 2005). Shouldn't we expect to see **some** traces of advanced extraterrestrial civilizations out there?

Of course, all those astronomical observatories mentioned above have been busy doing **astronomy** – they have not specifically been looking for intelligent life nor has the data generally been analyzed with that end in mind. However, as discussed by other contributors to this book, well over one hundred dedicated searches for extraterrestrial intelligence **have** taken place since Drake's pioneering radio search. And analyses of astronomical data with the hope of finding artificial signals **have** been carried out – the SETI@home initative, the largest volunteer computing project on the planet, is a wonderful example of that. So far, all is quiet.

Is this silence something to be expected or is it, like the dog that didn't bark in the night, worthy of consideration?

## 17.1 Paradox

The spring and summer of 1950 saw New York newspapers devote many column inches to two mysteries. First, public trash cans seemed to be disappearing. Second, the US seemed to be suffering from a spate of flying saucer sightings. At that time Enrico Fermi was working at Los Alamos and over lunch one day he and three colleagues discussed a recent *New Yorker* cartoon. The cartoon showed aliens on their home world carrying trash cans, the property of New York's Department of Sanitation, out from their flying saucer. Fermi joked that this was a reasonable theory since a single idea accounted for two quite distinct phenomena: the reports of flying saucers and the mysterious disappearance of trash cans. The talk turned to a more serious discussion of the possibility of superluminal travel; then, out of the blue, Fermi asked: "Where **is** everybody?"

Fermi was renowned amongst his peers for possessing a deep physical intuition and an ability to make rapid mental calculations. He could generate remarkably accurate estimates of quantities in situations where others struggled even to formulate an approach. Fermi's lunchtime companions immediately understood that his question – "Where **is** everybody?" – referred to extraterrestrial visitors and perhaps they also understood that Fermi's question is really rather troubling.

Fermi presumably estimated the number of advanced civilizations that might exist in the Milky Way galaxy – essentially by multiplying together the several quantities described by Drake's equation, as discussed elsewhere in this volume – and came up with a large number. He must also have been impressed by the increasing rate of change of human science and technology: the first sustained, controlled, heavier-than-air powered flight only took place in 1903 and yet less than 40 years later a V-2 rocket had become the first manmade object to achieve sub-orbital spaceflight. But 40 years is as nothing when measured in astronomical time. If a technological civilization continues to learn and progress over 400 years or 4000 years or 40 000 years … who knows what it could achieve? The Russian astrophysicist Nikolai Kardashev thought about these matters and postulated (Kardashev, 1964) that one could imagine a Type I civilization controling the energy resources of a single planet; a Type II civilization would be able to utilize all the energy of a single star; and a Type III civilization would be able to harness all the energy of a galaxy. (The Kardashev scale was purely speculative, of course, and intended merely to provide a framework for discussing such ideas. Nevertheless, perhaps it is worth mentioning that current human civilization possesses a Kardashev value that is 0.72 of a Type I civilization.) If only a tiny fraction of the billions of planets in our galaxy have given rise to civilizations then surely some of those civilizations will be older and more technologically advanced than us – possibly Type II or even Type III. That is perhaps the conclusion that Fermi reached. They should be here. At the very least we should see evidence of them or their instrumentality when we look into space.

But we don't. Where **is** everybody?

The question became known as the Fermi paradox and, in its purest form, is perhaps not too disturbing. The distances between stars are great and it is reasonable to suppose that faster-than-light travel is impossible even for the most advanced civilization. It is not surprising, then, that we don't see alien spacecraft buzzing round the Solar System. But it's not just that their craft aren't here; as mentioned above, astronomers have observed no signs of extraterrestrial intelligence anywhere. The paradox is not that their spacecraft aren't around when we might expect them to be here; it's that we neither see nor hear **any** traces of them. David Brin (Brin, 1983) called this "the Great Silence". It's this quiescence that's so disturbing.

The paradox has not been blunted since Fermi asked his question or Brin pointed out the deafening silence of the universe. If anything, it has sharpened.

On the one hand, we can point out that it seems increasingly likely that Earth is not the only abode of life. Certainly, few of the advances in the nascent science of astrobiology support the notion that life in the universe must necessarily be rare. Consider, for example, our improved understanding of planetary formation. If the once widely accepted catastrophic hypothesis had turned out to be true then planetary systems would be scarce. Instead, the leading theory of planetary formation is an updated version of the nebular hypothesis, which implies that planets are common. More than that, astronomers have actually observed other planetary systems: at the time of writing we know of 429 exoplanets, and the number increases weekly (Schneider, 2010). So if the existence of a planet is a prerequisite for life to evolve then our own galaxy contains plenty of homes. Furthermore, although we still lack a generally accepted model for how life started on Earth, we know from the existence of stromatolites that it began with almost indecent haste: primitive lifeforms may have existed on Earth as far back as the Early Archean era. It does not necessarily follow, just because life started quickly here, that life must have started with similar ease elsewhere. (As with all discussions in this arena we are in the philosophically troublesome position of having only one example – ourselves – from which to extrapolate.) Nevertheless, it seems quite unreasonable to argue that the genesis of life is miraculously rare. Moreover, the discovery of extremophiles suggests that, once life has started, it will adapt to an extraordinary range of environments. In short, we have no compelling reason to suppose that the universe is, with the sole exception of Earth, lifeless. There may be countless worlds that harbor life – and surely some of those will give rise to intelligent species that go on to develop advanced technology?

On the other hand, since Fermi's time we have learned more about how a civilization might make its presence known over interstellar distances. While the difficulties inherent in interstellar travel may mean that even extremely advanced biological species may choose not to explore the Galaxy in "person", their mind children (Moravec, 1988), who could be engineered to eliminate the intrinsic frailties of natural organisms, would have the capacity to explore – perhaps even in some way shape – the Galaxy. Perhaps biological intelligence can never give rise to non-biological superintelligence (there may be nanotechnological or computational limits of which we are unaware); but an advanced civilization would surely have the capacity to construct self-replicating Bracewell–von Neumann probes (Bracewell, 1960) that could explore the Galaxy on its behalf. Where are those probes of exploration (or conquest)? Even if interstellar travel of any description is simply too difficult or too costly, an advanced extraterrestrial civilization could **signal** its prescence – as described elsewhere in this volume, transmission in the waterhole band, or laser pulses, or even radio bridges employing a gravitational lens are all potential means of communicating over interstellar distances. And it is not as if we are deaf: in the years since Fermi asked his question our ability to search for electromagnetic signals has steadily improved by many orders of magnitude. The paradox thus remains. The question Fermi asked 60 years ago has spawned a large literature. Many dozens of papers – some humorous, some serious, most of them ingenious – have proposed ways of resolving the paradox (see Webb (2002) for a detailed discussion of the history of the paradox and a comprehensive list of references). In a brief chapter such as this it is impossible to do justice to the subtlety and range of arguments that have been deployed. I hope instead, in the next section, to give merely a flavor of the debates and to highlight the importance of the paradox to the SETI program: by addressing the paradox seriously perhaps new lines of investigation will be uncovered.

## 17.2 Pondering the Paradox

Any answer to Fermi's question derives from a perspective that is ineluctably anthropocentric. Furthermore, any response to the question is influenced by the current understanding of technology and physics. Always bearing those fundamental limitations in mind, it is possible to classify three different approaches to the paradox.

### 17.2.1 Prescence

For some people (particularly those who reject the "establishment" explanation of UFOs, crop circles and other such phenomena) there is no paradox. They argue that there is clear evidence of an extraterrestrial presence here on Earth and thus answer Fermi by saying: "**there** they are".

While this response need not long detain us, it is worth mentioning that several serious commentators have argued that the SETI program should not ignore the possibility of there being an extraterrestrial presence in our neighborhood. While it seems safe to discount tales of flying saucers and ancient astronauts (von Däniken, 1968) perhaps SETI enthusiasts should be looking for physical evidence of intelligence in our neighborhood? The proposals are numerous: the Lagrangian points L4 or L5 in the Earth–Moon system could be the abode of small Earth-observation probes (Freitas and Valdes, 1980); larger craft might be hidden in the Asteroid Belt (Papagiannis, 1978); Pluto could be the result of an astroengineering project (Stephenson, 1978); it has even been suggested that messages from extraterrestrial civilizations might be encoded in DNA (Yokoo and Oshima, 1979; Nakamura, 1986).

It goes without saying that there is no generally accepted evidence that extraterrestrials are presently in our Solar System or have been here at any time in the past. Of course humanity has explored only a vanishingly small fraction of the Solar System, and logically it is impossible to conclude that extraterrestrials are not here – but one can substitute the words "tooth fairy" for the word "extraterrestrials" in this sentence and the meaning holds. I'd suggest that most people in possession of a modicum of skepticism would argue that if extraterrestrial craft were ever here then we'd have noticed them by now. But what if the extraterrestrials did not **want** their craft to be noticed?

Any civilization sufficiently advanced to have mastered interstellar travel would certainly be advanced enough to keep their craft from human view. (Why they should want to travel all that way and subsequently hide is anyone's guess. But then alien minds, if they exist, will surely be inscrutable.) Such is the logic behind the zoo hypothesis (Ball, 1972) and the interdict hypothesis (Fogg, 1988). These ideas certainly resolve the Fermi paradox but they do so in a profoundly unattractive manner. This way lies solipsism. (It is worth noting that Baxter's planetarium hypothesis (Baxter, 2000), although seemingly even more outlandish than the zoo or interdict hypotheses, possesses the great virtue of being testable.) A more productive approach to Fermi's question is called for.

### 17.2.2 Proximity and persistence

A standard response to Fermi's question is that the galaxy is too large for exploration to take place; in other words, the distances and timescales involved effectively prohibit interstellar travel. From our vantage point – that of a civilization possessing a level of technology with which we struggle to explore even our own planetary system – this response seems reasonable. Nobody is here because there is no civilization in Earth's proximity.

Perhaps the same explanation occurred to Fermi and his colleagues. (Or perhaps not. When reading interview transcripts of physicists in the 1950s I'm often struck by the boundless optimism they displayed. Six decades on, we have a clearer appreciation of just how costly interstellar travel will be.) However, even if one accepts this explanation (and one can take issue with it: for example, would exploration by artificial, self-replicating probes really prove impossible for a technological civilization millions of years in advance of our own?) it fails to address the other aspect of the Fermi paradox: the "Great Silence". Advanced civilizations may be unwilling or unable to send out physical probes over interstellar distances, but they should be able to signal their presence over those distance. Why don't we see them? Why don't we hear them?

Earlier chapters in this volume have explained the difficulties involved in interstellar communication. For example, in order to detect a signal we first have to guess which means of communication will be used. That's perhaps not too difficult. Of course, it is impossible to know how putative extraterrestrials might think, but any civilization wishing to communicate its presence would surely appreciate that its audience will be expecting to receive modulated electromagnetic waves. Such waves are easy to generate and receive, they travel as fast as is possible, and they go where they are sent with minimal absorption or deflection. Communication channels that seem unfeasible to us with our present level of technology – modulated gravitational waves, say, or neutrino pulses, or some technology we have yet to even imagine – may be options for advanced civilizations. Nevertheless, physics is the same everywhere and so an intelligent species would presumably understand that communication by electromagnetic waves is the lowest common denominator. OK, so electromagnetic waves it is. In order to distinguish it from naturally occurring noise and other transmissions in neighboring bands the transmission would likely be a narrowband signal. (Or would it? Maybe to an extraterrestrial intelligence the transmission of very short, punctuated, broadband signals would be the obvious choice?) If we decide to look for a narrowband signal we then have to guess which of the billions of frequency channels will be used. Should we listen for radio waves at the waterhole (Cocconi and Morrison, 1959) and/or at multiples of that frequency? Or perhaps for intergalactic communication we should listen at 56.8 GHz, the peak in the cosmic microwave background, or multiples thereof (Drake and Sagan, 1973; Gott, 1995)? What about optical wavelengths using lasers (Schwartz and Townes, 1961)? Then there is the question of **where** to look: should we target individual stars with high sensitivity, or scan large areas of the sky with lesser sensitivity? There are other equally difficult questions to answer if one intends to search for signals from an extraterrestrial civilization. The volume of phase space is vast. The words "needle" and "haystack" cannot help but spring to mind.

So SETI scientists have been grappling with a many-dimensional problem over the past 50 years. And although there have been a number of radio searches in that time, and although developments such as the SETI@home project and the commencement of optical SETI have added computing power and further search frequencies to the program, one can reasonably argue that the quest remains in its infancy. The scale of the challenge is such that large SETI projects would have to run for many more decades before the lack of an observed signal might begin to appear surprising. Persistence is called for.

Nevertheless … returning to the idea behind Fermi's original question and Brin's "Great Silence", some commentators argue that the search for extraterrestrial intelligence should not be so difficult. Any civilizations out there are likely to be far more advanced than us, are likely to possess a level of technology beyond our imagining. A K2 or K3 civilization would have the capacity to create beacons – whether using radio waves or lasers or something fancier (such as using nuclear waste to alter a star's spectrum (Whitmire and Wright, 1980) or employing particle clouds to cause a star to "flash" on and off (Morrison, 1963)) – that we simply could not miss. The SETI program to date has put relatively little effort into the search for such beacons; perhaps it should do more. But why is a careful search even necessary? Shouldn't the whereabouts of such beacons be **obvious**?

If one feels the need to account for the lack of such stand-out beacons one can choose from a variety of explanations that have been proffered. The same ideas have been advanced as explanations for the failure of the SETI program to find *any* type of signal. For example, some have argued that fear of revealing one's existence is a factor. (But if interstellar distance precludes travel, would one advanced civilization really have anything to fear from another?) Perhaps, others argue, the truly advanced civilization keeps silent because it does not wish to interfere with the natural development of other species. (But even in science fiction, where such an idea is a common trope, this example of universal cosmic ethics is more often broken than observed.) Yet others argue that advanced civilizations inevitably go on to hit a technological singularity (Vinge, 1993) (But doesn't this explanation suffer a Fermi paradox of its own: where are the super-intelligences?) There is no space here to discuss the many other suggestions made in similar vein.

Of course, there could be another reason why we see no beacons and why the SETI program has, to date, found no signals: perhaps there are no intelligent civilizations out there. Perhaps we are alone in the Universe.

### 17.2.3 Peculiarity

For some people it is an arrogant and absurd notion: that a galaxy containing perhaps $4 \times 10^{11}$ stars possesses only one – our Sun – with a planet that is home to intelligent life. As mentioned earlier, we now know that planetary system formation seems to be a natural concomitant of stellar formation; and although we do not understand the mechanisms by which inanimate matter gives rise to life we have no good reason to believe that abiogenesis must be a vanishingly rare event. So we inhabit a galaxy that contains many billions of planets, and it seems a good bet that life must have started on at least some of them. How on Earth could we be unique? Nevertheless, just as some advances made since Fermi's time tend to support the notion that complex life might be common in the universe, others imply the opposite.

For example, it is often assumed that life, in order to start and then prosper, requires a rocky planet in the circumstellar habitable zone – the disk around a star in which temperatures are such that liquid water, and therefore life as we know it, can exist. (As with most assumptions in astrobiology, this one can be challenged. Places outside the traditional habitable zone – the moons of giant planets, for example – may possess liquid water. In any case, life elsewhere may well not be "as we know it".) However, it is not just the width of the habitable zone that is important: if it takes billions of years for intelligence to develop, as it did on Earth, then the planet must remain in the habitable zone all that time: this requirement of a large continuously habitable zone restricts the number of planetary systems that could be sites for life. Furthermore, the location of a star within the galaxy may also be important. Life "as we know it" requires a large level of heavy elements, which implies that the star must not be too distant from the galactic center. On the other hand, the star must not be so close to the center that it receives life-extinguishing radiation. Lineweaver and colleagues (Lineweaver et al., 2004) argue that this requirement gives rise to the notion of a galactic habitable zone, a slowly expanding annular region that they calculate is 7–9 kpc from the center of the Galaxy and is composed of stars that are between 4 to 8 billion years old. Again, this restricts the number of planetary systems that could be sites for life.

Even if intelligent life can arise only on a rocky planet in a large continuously habitable zone that orbits a star in a galactic habitable zone we are still left with a vast number of candidates. (And it should be noted that many stars in the galactic habitable zone are billions of years older than the Sun, which presumably affords plenty of time for a K2 or K3 civilization to develop.) But perhaps other conditions must be met? For example, could the architecture of a planetary system be a factor in allowing life to evolve? Our Solar System possesses a Jupiter that has a stable, nearly-circular orbit at 5.2 AU. Astronomers have found many planetary systems with "hot" Jupiters – planets with a similar mass to Jupiter and with low eccentricity, but orbiting at about 0.1 AU. Since our models of planetary formation suggest that such "hot" Jupiters could not have formed *in situ* it seems they must have formed beyond their star's ice line and then migrated to their present position. That process of migration, it has been argued (Armitage, 2003), is likely to prevent terrestrial planets forming in the continuously habitable zone (although other models have challenged this conclusion (Fogg and Nelson, 2007). Even more common than planetary systems with a "hot" Jupiter are those with an "eccentric" Jupiter, a massive planet whose orbit encroaches the habitable zone; again, it seems that such systems are unlikely to be the abode of life. On the other hand, if a planetary system has its Jupiter too far away from the star then terrestrial planets can form – but only outside the habitable zone. Does that mean we should be looking not only for a “Goldilocks” planet like Earth but one that is also accompanied by a “good” Jupiter? (Cramer (1986) provides an additional, more speculative reason why “good” Jupiters may be essential for intelligent life to evolve.)

One can speculate, too, about whether Earth itself is somehow special. Not just its location, but its particular characteristics. For example, Earth is in many ways a double planet: whereas most satellites in the Solar System possess a negligible mass compared to their parent bodies, the Moon possesses 1/81 of Earth’s mass. The importance of the Moon on the development of life here on Earth is unclear, but it is possible that it has played a key role. For example, it has been suggested that tides were a factor in getting life started, perhaps by mixing the primordial soup; without the Moon there would be no spring and neap tides, just the much smaller solar tides. More importantly, however, the Moon has played a role in stabilizing Earth’s obliquity over long timescales; the planet Mars, which lacks a large satellite, has an obliquity that appears to have changed chaotically over the past ten million years. Since even small changes in Earth’s obliquity seem to be associated with large changes in climate it is difficult to envisage how complex life-forms could have prospered had the obliquity wandered randomly between 0° and 90°. If our large Moon has indeed been a necessary factor in life’s continuing development here on Earth then it may well turn out that we are, if not rare, at the very least unusual. The point here is that the currently favored view of lunar formation is the giant impact hypothesis (Hartmann and Davis, 1975; Cameron and Ward, 1976): the Moon was formed by the off-center impact on the infant Earth of a Mars-sized object called Theia. The timing of this impact was important. Had it happened earlier, when Earth was less massive, then much of the debris from the impact would have been lost to space and the Moon would have been much smaller than it in fact is. Had it happened later, when Earth was larger, gravity would have caused a lesser amount of ejecta to reach space and again the Moon would have been much smaller than it in fact is. And even a satellite up to half the mass of the Moon – which would still be much larger in relation to its parent body than other satellites – would be insufficient to stabilize Earth’s obliquity. So whereas planetary system formation appears to be a natural concomitant of stellar formation, the formation of a large satellite may require the occurrence of a particular cataclysm.

The points discussed above are – with varying degrees of plausibility – some of the possible astronomical constraints on the emergence of intelligent life. What of biological constraints?

We lack a general theory of life. The question of how life began on Earth from non-life remains a fundamental unsolved mystery. Nevertheless, we have no good reason to suppose that life can have emerged **only** on planet Earth. Indeed, within the next decade the coming generation of space-based telescopes will be looking for potential biosignatures (see, e.g., Kaltenegger,

2009) and it is not inconceivable that we will eventually find markers that can best be explained in terms of processes involving basic life-forms. The discovery of such biosignatures would be of profound importance, of course, but that is not specifically what concerns us here. What we are discussing is not the existence of simple, unicellular life-forms but the existence of complex, multicellular, **intelligent** life-forms. Does the former lead inexorably to the latter?

As with all such discussion on the Fermi paradox we are hampered by our almost complete lack of data; our knowledge of the history of life on Earth is all we have to go on. We know that Earth formed 4.55 billion years ago; our knowledge of when life emerged is less precise, but there are reasons to suppose that life forms were present 3.8–3.5 billion years ago. The earliest eras were inimical to any form of life, so the development of the primitive, prokaryotic cell seems to have taken place relatively quickly – indeed, almost as soon as conditions allowed. The development of the much more complex eukaryotic cell – a structure possessing a nucleus and cytoskeleton; a form that lacks a rigid cell wall, thus enabling cytosis; an entity that allows for sex to take place – took far longer. Could it be that the development of complex cells from primitive cells is a difficult evolutionary step, one that is far from inevitable?

It took even longer for multicellular, eukaryotic organisms to evolve: animal fossils date back only to the Cambrian explosion, some 540 million years ago. Although it is entirely possible that soft-bodied animals were in existence before the Cambrian period (they would, after all, leave no trace in the fossil record) it is clear that complex, multicellular animals are relative newcomers to Earth. So although single-celled creatures were around almost as soon as Earth cooled, it took a further 3 billion years or so for complex creatures to develop. Why the long wait for multicellularity? Could it be that the development of multicellular creatures is an even more difficult step than the transition from prokaryotic to eukaryotic cell?

Once complex, multicellular lifeforms developed here on Earth, a quite bewildering variety of species subsequently evolved. However, although billions of complex species have lived on Earth during its history, only one has developed with the capacity to contemplate the possible existence of extraterrestrial life; only one has developed the technological ability to listen for signals from the stars and even signal its own existence to the cosmos. The steps that led to this state of affairs are not entirely clear. Certainly the development of a high order of toolmaking was critical, as must have been the development of a high level of general intelligence. Critical, too, was the development of language with a complex grammar and a form of social living in which the lessons learned by one generation could be passed on to the next. Undoubtedly there are other evolutionary steps of equal consequence. Could it be that those steps are peculiar to Earth's history, a set of accidents that is unlikely to occur elsewhere? We don't know. And it seems impossible to separate the necessary from the merely contingent.

Arguments based on the 'Rare Earth' hypothesis (Ward and Brownlee, 2000) and similar notions thus tend to address the paradox by positing that, although unicellular life may be ubiquitous, the likelihood of an intelligent, technologically advanced, communicating civilization arising from such life is extremely small. We are alone.

The notion that Earth is somehow special does not sit well with the Copernican principle, which is a principle that has served science well for hundreds of years. However, combining the Copernican principle with the Fermi paradox has led some authors to a gloomy conclusion. For example, the SETI pioneers Iosif Shklovskii and Sebastian von Hoerner ( 1978) eventually concluded that the absence of signals from technologically advanced civilizations implied that the lifetime of such civilizations is small. They argued that technologically advanced civilizations exist – Earth is not unique – but that civilizations inevitably annihilate themselves soon after discovering the destructive potential of nuclear energy. (Shklovskii and von Hoerner were writing at the height of the Cold War, a time when the optimism evident in the 1950s evaporated. Although humanity has so far negotiated the perils of nuclear war we cannot view the future with a particularly sanguine eye. We now know that advanced civilizations will have more than one technology available for self-destruction: our children, if not ourselves, will have to worry about large-scale misuses of biotechnology and nanotechnology; other technologies with the potential for destruction are on the horizon.)

The difficulty with the type of argument made by Shklovskii and von Hoerner is that it requires a **local**, one could almost say sociopolitical, cause to act independently many times in order to produce a uniform global result. A more elegant approach is to assume that some **global** cause acts to reset the clock of evolution throughout the galactic habitable zone. The phase-transition hypothesis of Ćirković and Vukotić (Ćirković, 2004; Ćirković and Vukotić, 2008) illustrates this approach nicely. Ćirković and Vukotić follow a suggestion made by Annis (1999), that a gamma-ray burst essentially sterilizes its host galaxy of complex life. (See for example Thorsett (1995) and Galante and Horvath (2007) for a discussion of the astrobiological consequences of gamma-ray bursts.) They then investigate the consequences of such a global regulation mechanism on the development of cosmic life. Other global regulation mechanisms can be postulated (see for example Vukotić and Ćirković, 2007; 2008), which could act alongside a plethora of randomly occurring local regulation mechanisms (such as meteor impacts, "Snowball Earth" events, and so on). Monte Carlo simulations of a toy model based on these ideas Vukotić (2010) suggest that an astrobiological phase transition from an essentially dead galaxy to one filled with life is unlikely to have been able to occur in our past – but it can occur in our future. In such a scenario the Fermi paradox is resolved because few if any civilizations can be far in advance of our own: we live not in a special place but at a special time.

Other scientists have invoked anthropic reasoning to resolve the paradox, arguing that the existence of intelligent extraterrestrials is unlikely. For example, Carter (1983) suggested that the development of intelligent life on Earth required a number of "hard" evolutionary steps ("hard" in the sense of being unlikely to occur in the available time). Anthropic arguments were then used (Barow and Tipler, 1986) to derive a bound that has been used to exclude the existence of extraterrestrial intelligence: the habitable zone for most Earth-like planets orbiting G-type stars will disappear before intelligent beings have a reasonable chance of evolving.

Since the publication of Carter's paper, anthropic reasoning in both physics and astrobiology has come under increasing scrutiny; it is impossible convey even a flavor of the claims and counterclaims in a chapter such as this. However, it is worth mentioning some recent work that views the anthropic principle in an interesting and slightly different light. A key conjecture of modern cosmology is the notion that our observable universe is but one of a vast number of causally disconnected universes. Furthermore, advances in string theory suggest that the laws of nature and physical constants could be different in each of those universes (Douglas and Kachru, 2007). Jaffe, Jenkins and Kimchi (2009) recently posed the question: over what range of values for parameters such as quark masses do the laws of physics allow for the existence of an observer? By allowing multiple parameters to vary at the same time they were able to find universes very different from our own that would nevertheless allow for the possibility of organic chemistry and, presumably, complex structures such as life. Could it be that our own universe is not as fine-tuned for life as has so far been supposed, that it just happens to be one of the few that are compatible with the evolution of intelligence? Could it be that, rather than inhabiting a universe whose physical laws are somehow fine-tuned to allow our existence, we inhabit a universe that is only **just** capable of accommodating intelligent observers?

These ideas are of course at the speculative extremes; but then **all** discussion involving the Fermi paradox involves speculation – that's what makes it fun. At some point, though, speculation will have to give way to fact; it seems to me increasingly likely that, if our scientific worldview is in some future time to be complete, we will need to understand not only the details of the standard models of particle physics and cosmology – we will need to fully understand our place in the cosmos. We will need to resolve the Fermi paradox.

## 17.3 Postscript

The arguments outlined above represent just a few of the many approaches people have taken in order to “resolve” the Fermi paradox. Indeed, the range of views is so large we can perhaps draw only one firm conclusion: our present understanding of the universe and our place in it is simply insufficient to know whether we are alone.

There are those who argue that the Fermi paradox **proves** we are alone, and that therefore the search for extraterrestrial intelligence is pointless. Of course, the paradox proves nothing of the sort. And since the scientific, philosophical and cultural prize at stake is immense, surely we must continue the search with vigor, imagination and increasing sophistication. What else could we do?

And yet … when one ponders the paradox one inevitably must take seriously the possibility that we are indeed be alone. This is not to say that, with the sole exception of Earth, the universe is necessarily lifeless. Myriads of planets might be home to lifeforms, with billions of species each performing a specific, interesting pattern in the dance of evolution. It is just that the particular dance pattern we find so compelling – a pattern involving conscious intelligence, with all that this entails – might exist only here.

The SETI program is 50 years old. Before another 50 years have passed, we will surely have a clearer understanding of the likelihood of other intelligences existing out there. If it turns out that we seem to be alone, as some of us believe will be the case, then what?

A recent hypothesis introduced by Peter Ward (Ward, 2009) is yet another idea that we should consider when discussing the Fermi paradox. Ward proposes that life, rather than helping to regulate some supposed system through processes involving negative feedback (the so-called “Gaia hypothesis”), inevitably consumes all resources available to it. Life, by its very nature, sows the seeds of its own destruction. He supports the argument by pointing out that, with the exception of the K–T extinction, all of the mass extinctions Earth has experienced have been caused by life itself. Indeed, it can be argued that we are currently in the midst of another mass extinction event caused by the actions of one particular lifeform. If this “Medea hypothesis” turns out to be valid, we can think of life as being a self-limiting phenomenon. Medea presents us with a pessimistic image. Except that humankind is different to every other lifeform we know about: our particular qualities have enabled us to develop a technology that may – just may – allow us to negotiate the perils that lie ahead. For some of us, the real importance of the Fermi paradox is to remind us that humankind has a duty to preserve the only known flickering of conscious intelligence in an otherwise empty universe.

## References

Adelman-McCarthy, J.K. et al. 2008. The Sixth Data Release of the Sloan Digital Sky Survey. *Astrophys J Suppl Ser*, **175**, 297–313.

Annis, J. 1999. An Astrophysical Explanation for the Great Silence. *J Brit Interplanetary Soc*, **52**, 19–22.

Armitage, P.J. 2003. A Reduced Efficiency of Terrestrial Planet Formation Following Giant Planet Migration. *Astrophys J*, **582,** L47–L50.

Arnold, L.F.A. 2005. Transit Light-curve Signatures of Artificial Objects. *Astrophys J*, **627**, 534–539.

Ball, J.A. 1972. The Zoo Hypothesis. *Icarus*, **19**, 347–349.

Barrow, J.D. and Tipler, F.J. 1986. *The Anthropic Cosmological Principle*. Oxford: Clarendon.

Baxter, S. 2000. The Planetarium Hypothesis: A Resolution of the Fermi Paradox. *J Brit Interplanetary Soc*, **54**, 210–216.

Bracewell, R.N. 1960. Communication from Superior Galactic Communities. *Nature*, **186**, 670–671.

Brin, G.D. 1983. The "Great Silence": The Controversy Concerning Extraterrestrial Intelligent Life. *Quarterly J Royal Astronomical Soc*, **24**, 283–309.

Cameron, A.G.W. and Ward, W.R. 1976. The Origin of the Moon. *Abstracts of the Lunar and Planetary Science Conference*, Vol 7, 120.

Carter, B. 1983. The Anthropic Principle and its Implications for Biological Evolution. *Phil Trans R Soc London,* **A310**, 347–363.

Ćirković, M.M. 2004. On the Temporal Aspect of the Drake Equation and SETI. *Astrobiology*, **4**, 225–231

Ćirković, M.M. and Vukotić B. 2008. Astrobiological Phase Transition: Towards Resolution of Fermi's Paradox. *Origin of Life and Evolution of the Biosphere*, **38**, 535–547.

Cocconi, G. and Morrison, P. 1959. Searching for Interstellar Communications. *Nature*, **184**, 844–846.

Colless, M. et al. 2003. The 2dF Galaxy Redshift Survey: Final Data Release. In eprint archive astro-ph/0306581

Cramer, J.G . 1986. The Pump of Evolution. *Analog*, **106**, 124–127.

Douglas, M. and Kachru, S. 2007. Flux Compactification. *Rev Mod Phys*, **79**, 733–796.

Drake, F.D. and Sagan, C. 1973. Interstellar Radio Communication and the Frequency Selection Problem. *Nature*, **245**, 257–258.

Dyson, F.J. 1960. Search for Artificial Stellar Sources of Infra-red Radiation. *Science*, **131**, 1667–1668.

Fogg, M.J. 1988. Temporal Aspects of the Interaction among the First Galactic Civilizations: The "Interdict Hypothesis". *Icarus*, **69**, 370–384.

Fogg, M.J. and Nelson, R.P. 2007. On the Formation of Terrestrial Planets in Hot-Jupiter Systems. *Astron Astrophys*, **461**, 1195–1208.

Freitas, R.A. Jr and Valdes, F. 1980. A Search for Natural or Artificial Objects Located at the Earth–Moon Libration Points. *Icarus*, **42**, 442–447.

Galante, D. and Horvath, J.E. 2007. Biological Effects of Gamma-ray Bursts: Distances for Severe Damage on the Biota. *Int J Astrobiology*, **6**, 19–26.

Gott, J.R. 1995. Cosmological SETI Frequency Standards. In B. Zuckerman and M.H. Hart (eds) *Extraterrestrials: Where are They?* Cambridge: CUP.

Harris, M.J. 1986. On the Detectability of Antimatter Propulsion Spacecraft. *Astrophys Space Sci*, **123**, 297–303.

Hartmann, W.K. and Davis, D.R. 1975. Satellite-sized Planetesimals and Lunar Origin. *Icarus*, **24**, 504–514.

Jaffe, R.L., Jenkins, A. and Kimchi, I. 2009. Quark Masses: An Environmental Impact Assessment. *Phys Rev* D, **79**, 065014.

Kaltenegger, L. 2009. Deciphering Spectral Fingerprints of Biomarkers on Exo-Earths. Symposium on the Search for Life in the Universe, May 4–7, Baltimore, STSI.

Kardashev, N.S. 1964. Transmission of Information by Extraterrestrial Civilizations. *Soviet Astronomy*, **8**, 217–221.

Lineweaver, C.H., Fenner, Y. and Gibson, B.K. 2004. The Galactic Habitable Zone and the Age Distribution of Complex Life in the Milky Way. *Science*, **303**, 59–62.

Moravec, H. 1988. *Mind Children.* Cambridge, MA: Harvard University Press.

Morrison, P. 1964. In W. Sullivan (ed.) *We Are Not Alone*, London: Pelican, 254.

Nakamura, H. 1986. SV40 DNA: A Message from Epsilon Eridani? *Acta Astronoautica*, **13**, 573–578

Papagiannis, M.D. 1978. Are We All Alone, Or Could They Be in the Asteroid Belt? *Quarterly J Royal Astronomical Soc*, **19**, 236–251.

Schneider, J. 2010. *The Extrasolar Planets Encyclopaedia*. Retrieved 27 January 2010, from: http://exoplanet.eu

Schwartz, R.N. and Townes, C.H. 1961. Interstellar and Interplanetary Communication by Optical Masers. *Nature*, **190**, 205–208

Spergel, D.N. et al. 2009. Five-year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. *Astrophys J Suppl Ser*, **180**, 330–376.

Stephenson, D.G. 1978. Extraterrestrial Cultures within the Solar System? *Quarterly J Royal Astronomical Soc*, **19**, 277–281.

Thorsett, S.E. 1995. Terrestrial Implications of Cosmological Gamma-ray Bursts. *Astrophys J Lett*, **444**, L53–L55.

Vinge, V. 1993. *VISION-21*. NASA Lewis Research Center.

von Däniken, E. 1968. *Chariots of the Gods.* New York: Bantam.

von Hoerner, S. 1978. Where Is Everybody? *Naturwissenschaften*, **65**, 553–557.

Vukotić, B. 2010. The Set of Habitable Planets and Astrobiological Regulation Mechanisms. *Int J Astrobiology*, in press.

Vukotić, B. and Ćirković, M.M. 2007. On the Timescale Forcing in Astrobiology. *Astron J*, **175**, 45–50.

Vukotić, B. and Ćirković, M.M. 2008. Neocatastrophism and the Milky Way Astrobiological Landscape. *Serb Astron J*, **176**, 71–79.

Ward, P. 2009. *The Medea Hypothesis: Is Life on Earth Ultimately Self-Destructive?* Princeton: Princeton University Press.

Ward, P.D. and Brownlee, D. 2000. *Rare Earth: Why Complex Life Is Uncommon in the Universe.* New York: Springer.

Webb, S. 2002. *Where Is Everybody?* New York: Copernicus.

Whitmire, D.P. and Wright, D.P. 1980. Nuclear Waste Spectrum as Evidence of Technological Extraterrestrial Civilizations. *Icarus*, **42**, 149–156

Yokoo, H. and Oshima, T. 1979. Is Bacteriophage Phi X174 DNA a Message from an Extraterrestrial Intelligence? *Icarus,* **38**, 148–153.

# Part III: The Spirit of SETI Future

# 18

# Focusing the Galactic Internet

Claudio Maccone,
*Technical Director, Scientific Space Exploration,*
*International Academy of Astronautics*

The gravitational lens of the Sun is an astrophysical phenomenon predicted by Einstein's general theory of relativity. It implies that, if we can send a probe along any radial direction away from the Sun up to the minimal distance of 550 AU and beyond, the Sun's mass will act as a huge magnifying lens, letting us "see" detailed radio maps of whatever may lie on its other side even at very large distances. This author's recent book (Maccone, 2009) studies such future FOCAL space missions to 550 AU and beyond.

In this chapter, however, we want to study another possibility entirely: how to implement future interstellar radio links between the Solar System and any future interstellar probe, by utilizing the gravitational lens of the Sun as a huge antenna. In particular, we study the Bit Error Rate (BER) across interstellar distances, with and without using the gravitational lens effect of the Sun. Only by exploiting the Sun as a gravitational lens will we be able to communicate with our own probes (or with nearby aliens) across the vast distances to even the nearest stars in the Galaxy, and that at a reasonable Bit Error Rate

Furthermore, we study the "radio bridge" between the Sun and any other star that is made up by the two gravitational lenses of both the Sun and that Star. The alignment for this radio bridge to work is very strict, but the power saving is enormous, due to the two stars' lenses contributions to the overall antenna gain of the system.

We provide five numerical examples of this situations.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_18, 

The conclusion is that a radio interstellar communications network can indeed be built if the gravitational lenses of all stars involved are exploited.

Perhaps one or more advanced Galactic Civilizations already built such a network. If so, we really are now “Focusing the Galactic Internet”.

## 18.1 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 (Einstein, 1936), but his work was virtually forgotten until 1964, when Sydney Liebes of Stanford University (Liebes, 1964)) 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 (von Eshleman, 1979). 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 radio astronomer 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 (Drake, 1987), as did Nathan “Chip” Cohen of Boston University (Cohen, 1987). Non-technical descriptions of the topic were also given by them in their popular books (Drake and Sobel, 1992; Cohen, 1988).

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 by the author in Turin, Italy. The relevant Proceedings were published in the *Journal of the British Interplanetary Society* (Maccone, 1994). Meanwhile, on May 20, 1993 the author also submitted a formal Proposal to the European Space Agency (ESA) to fund the space mission design (Maccone, 1993). The optimal direction of space to launch the FOCAL spacecraft was also discussed by Jean Heidmann of Paris Meudon Observatory and the author (Heidmann and Maccone, 1994), 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 (Maccone, 1995), 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* (Maccone 2002), 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 400-page and comprehensive book by the author, entitled *Deep Space Flight and Communications – Exploiting the Sun as a Gravitational Lens* (Maccone, 2009), was published. This book embodies all the previous material published about the FOCAL space mission and updates it.

## 18.2 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

$$\alpha(r) = \frac{4GM_{Sun}}{c^2 r}. \qquad (1)$$

Figure 18.1 shows the basic geometry of the Sun gravitational lens with the various parameters in the game

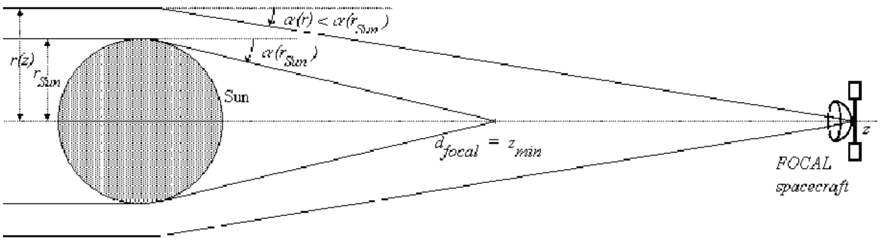



**Figure 18.1** Geometry of the Sun gravitational lens with the minimal focal length of 550 AU (= 3.17 light days = 13.75 times beyond Pluto's orbit) and the FOCAL spacecraft position beyond the minimal focal length.

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 $\alpha$ occurs for those rays just grazing the Sun surface, i.e., for $r = r_{Sun}$. This yields the inequality

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

with

$$\alpha(r_{Sun}) = \frac{4GM_{Sun}}{c^2 r_{Sun}}. \tag{3}$$

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

$$\tan(\alpha(r_{Sun})) = \frac{r_{Sun}}{d_{focal}}. \tag{4}$$

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:

$$\alpha(r_{Sun}) \approx \frac{r_{Sun}}{d_{focal}}. \tag{5}$$

Eliminating the angle $\alpha(r_{Sun})$ between equations (3) and (5), and then solving for the minimal focal distance $d_{focal}$, one gets

$$d_{focal} \approx \frac{r_{Sun}}{\alpha(r_{Sun})} = \frac{r_{Sun}}{\frac{4GM_{Sun}}{c^2 r_{Sun}}} = \frac{c^2 r_{Sun}^2}{4GM_{Sun}}. \tag{6}$$

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

$$r_{Schwarzschild} = \frac{2GM_{Sun}}{c^2} \tag{7}$$

yielding

$$d_{focal} \approx \frac{r_{Sun}}{\alpha(r_{Sun})} = \frac{r_{Sun}}{\frac{4GM_{Sun}}{c^2 r_{Sun}}} = \frac{r_{Sun}^2}{2\, r_{Schwarzschild}}. \tag{8}$$

Numerically, one finds

$$d_{focal} \cong 542 \text{ AU} \approx 550 \text{ AU} \approx 3.171 \text{ light days}. \tag{9}$$

*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 problem of Corona fluctuations and related distortions is currently being studied by Von Eshleman and colleagues at Stanford University (please refer to Chapter 6 of Maccone, 2009).

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 page 55 of Maccone (2002).

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 focusing star center? The answer is given by the formula

$$d_{focal} = \frac{r_{star}^2}{\frac{4GM_{star}}{c^2} - \frac{r_{star}^2}{D_{source}}}. \tag{10}$$

This is the key to gravitational focusing for a pair of stars, and may well be the key to SETI in finding extraterrestrial civilizations. It could also be considered for the magnification of a certain source by any star that is perfectly aligned with that source and the Earth: the latter would then be in the same situation as the FOCAL spacecraft except, of course, it is located much further out than 550 AU with respect to the focussing, intermediate star. Finally, notice that equation (10) reduces to equation (6) in the limit $D_{source} \to \infty$, i.e., (6) is the special case of (10) for light rays approaching the focusing star from an infinite distance.

## 18.3 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 focusing of light rays is caused by the gravitational field of the Sun. The answer to such a question is provided by the technical notion of "antenna gain", 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 $\lambda$ or the frequency $\nu$ by virtue of the equation (see, for instance, Orta et al.(1992), in Maccone (1992), or Kraus (1966) in particular Chapter 6, p. 117, equation (6-241)):

$$G_{antenna} = \frac{4\pi \, A_{effective}}{\lambda^2}. \tag{11}$$

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

$$A_{effective} = \frac{A_{physical}}{2} = \frac{\pi \, r_{antenna}^2}{2}. \tag{12}$$

Substituting this back into (11) yields the antenna gain as a function of the antenna radius and of the observed frequency :

$$G_{antenna} = \frac{4\pi\, A_{effective}}{\lambda^2} = \frac{2\pi\, A_{physical}}{\lambda^2} =$$

$$= \frac{2\pi^2\, r_{antenna}^2}{\lambda^2} = \frac{2\pi^2\, r_{antenna}^2}{c^2} \cdot \nu^2 . \tag{13}$$

*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

$$G_{Sun} = 4\pi^2 \frac{r_{Schwarzschild}}{\lambda}\; G_{Sun} = 4\pi^2 \frac{r_{Schwarzschild}}{\lambda} . \tag{14}$$

or, invoking the expression (1.2-7) of the Schwarzschild radius

$$G_{Sun} = \frac{8\pi^2\, G\, M_{Sun}}{c^2} \cdot \frac{1}{\lambda} = \frac{8\pi^2\, G\, M_{Sun}}{c^3} \cdot \nu \cdot \tag{15}$$

The mathematical proof of equation (14) is difficult to achieve. The author, unsatisfied with the treatment of this key topic given in Einstein (1936), Eshleman (1979) and Orta et al. (1992), turned to three engineers at 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), but also of the focal distance for rays originated from a source at finite distance, equation (10). Their proof is fully described in Maccone (1992), and is based on the aperture method used to study the propagation of electromagnetic waves, rather than on ray optics.

Using the words of these three authors' own abstract, they have "computed the radiation pattern of the [spacecraft] Antenna+Sun system, which has an extremely high directivity. It has been observed that the focal region of the lens for an incoming plane wave is a half line parallel to the propagation direction starting at a point [550 AU] whose position is related to the blocking effect of the Sun disk (Figure 18.1). 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 approximantion".

One is now able to compute the Total Gain of the Antenna+Sun system, that is simply obtained by multiplying the two equations yielding the spacecraft gain proportional to $\nu^3$ and the Sun gain proportional to $\nu$ :

$$G_{Total} = G_{Sun} \cdot G_{antenna} = \frac{16\pi^4 G M_{Sun} r_{antenna}^2}{c^5} \cdot \nu^3 . \quad (16)$$

Since the total gain increases with the *cube* of the observed frequency, it favors electromagnetic radiation in the microwave region of the spectrum. Table 18.1 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, had it been built jointly by ESA and NASA as planned before 1988. However, Quasat was abandoned by 1990 due to lack of funding. The definition of dB is of course:

$$\text{N dB} = 10\,\text{Log}_{10}\text{N} = 10\ln\text{N}/\ln 10$$

| **Line** | **Neutral hydrogen** | | *OH* **radical** | | $H_2O$ |
|---|---|---|---|---|---|
| **Frequency ν** | **1420 MHz** | **327 MHz** | **1.6 GHz** | **5 GHz** | **22 GHz** |
| Wavelength λ | 21 cm | 92 cm | 18 cm | 6 cm | 1.35 cm |
| S/C Antenna Beamwidth | 1.231 deg | 5.348 deg | 1.092 deg | 0.350 deg | 0.080 deg |
| Sun Gain | 57.4 dB | 51.0 dB | 57.9 dB | 62.9 dB | 69.3 dB |
| 12-meter Antenna S/C Gain | 42.0 dB | 29.3 dB | 43.1 dB | 53.0 dB | 65.8 dB |
| **Combined Sun + S/C Gain** | **99.5 dB** | 80.3 dB | 101.0 dB | 115.9 dB | 135.1 dB |

**Table 18.1** Table showing the gain of the Sun's lens alone,the gain of a 12-meter spacecraft (S/C) antenna and the combined gain of the Sun+S/C Antenna system the at five selected frequencies important in radioastronomy.

## 18.4 The Radio Link

The goal of this chapter is to prove that only by exploiting the Sun as a gravitational lens will we be able to have reliable telecommunication links across large interstellar distances. In other words, a direct link between different star systems, even if held by virtue of the largest radiotelescopes on Earth, will not be feasible across distances of the order or thousand of light years or more. We want to show that only a FOCAL mission in the direction from Earth opposite to that target star system will ensure a reliable telecommunication link across thousands of light years. Namely, we prove that Bit Error Rate (or BER, see site http://www.en.wikipedia.org/wiki/Bit_error_rate) will be unacceptable already at the distance of Alpha Centauri unless we resort to a supporting FOCAL space mission in the opposite direction from the Sun.

In order to face these problems mathematically, we must first understand the radio link among any two stars, and we think that no neater treatment of this subject exists than the book *Radio Astronomy* by the late professor John D. Kraus of Ohio State University (Kraus, 1986), that we follow hereafter.

Consider a radio transmitter that radiates a Power $P_t$ isotropically and uniformly over a bandwidth $B_t$. Then, at a distance $r$ it produces a flux density given by

$$\frac{P_t}{B_t \, 4\pi \, r^2} \, . \tag{17}$$

A receiving antenna of effective aperture $A_{er}$ at a distance $r$ can collect a power given by (17) multiplied by both the effective aperture of the receiving antenna and its bandwidth, namely the received power $P_r$ is given by

$$P_r = \frac{P_t}{B_t \, 4\pi \, r^2} A_{er} \, B_r . \tag{18}$$

It is assumed that the receiving bandwidth $B_r$ is smaller or, at best (in the "matched bandwidths" case) equal to the transmitting bandwidth $B_t$, that is $B_r \le B_t$.

So far, we have been talking about an isotropic radiator. But let us now assume that the transmitting antenna has a directivity $D$, that is an antenna gain in the sense of (11):

$$D = \frac{4\pi \, A_{et}}{\lambda^2} . \tag{19}$$

The received power $P_r$ is then increased by just such a factor due to the directivity of the transmitting antenna, and so (18) must now be replaced by a new equation where the right-hand side is multiplied by such a increased factor, that is

$$P_r = \frac{4\pi\ A_{et}}{\lambda^2} \cdot \frac{P_t}{B_t\ 4\pi\ r^2} A_{er}\ B_r. \tag{20}$$

Rearranging a little, this becomes

$$P_r = \frac{P_t\ A_{et}\ A_{et}}{r^2\ \lambda^2} \cdot \frac{B_r}{B_t}.. \tag{21}$$

This is the ***received signal power*** expression. For the matched bandwidths case, i.e., for $B_r = B_t$, this is called the Friis transmission formula, since it was first published back in 1946 by the American radio engineer Harald T. Friis (1893-1976) of the Bell Labs. In space missions, we of course know exactly both $B_t$ and $B_r$ and so we construct the spacecraft in such a way the two *bands match exactly*, i.e., $B_r = B_t$. So, for the case of telecommunications with a spacecraft (but not necessarily for the SETI case), we may well assume the matched bandwidths and have (21) reducing to

$$P_r = \frac{P_t\ A_{et}\ A_{et}}{r^2\ \lambda^2}. \tag{22}$$

Let us now rewrite (22) in such a way that we may take into account the gains (i.e., directionalities) of both the transmitting and receiving antennae, that is, in agreement with (11)

$$\begin{cases} G_t = \dfrac{4\pi\ A_{et}}{\lambda^2} \\ G_r = \dfrac{4\pi\ A_{er}}{\lambda^2}. \end{cases} \quad \text{that is} \quad \begin{cases} A_{et} = \dfrac{G_t\ \lambda^2}{4\pi} \\ A_{er} = \dfrac{G_r\ \lambda^2}{4\pi}. \end{cases} \tag{23}$$

Replacing the last two expressions into (22), we find that (22) is turned into

$$P_r = \frac{P_t\ G_t\ G_r}{(4\pi)^2\ r^2} \cdot \lambda^2.$$

This may finally be rewritten in the more traditional form

$$P_r = \frac{P_t\, G_t\, G_r}{L(r,\lambda)} \; . \tag{24}$$

if one defines

$$L(r,\lambda) = (4\pi)^2 \cdot \frac{r^2}{\lambda^2} \tag{25}$$

is the Path Loss (or path attenuation), i.e., the reduction in power density (attenuation) of the electromagnetic waves as they propagates through space. Path loss is a major component in the analysis and design of the link budget of a telecommunication system, see the site
http://www.en.wikipedia.org/wiki/Path_loss

## 18.5 Bit Error Rate for an "Ordinary" Direct Link with a Probe at the Alpha Centauri Distance

In this section we first define the Bit Error Rate (BER). Then, by virtue of a numerical example, we show that, even at the distance of the nearest star (Alpha Cen. at 4.37 AU) the telecommunications would be impossible by the ordinary powers available today for interplanetary space flight. But in the next section we shall show that the telecommunications would become feasibe if we could take advantage of the magnification provided by the Sun's gravity lens, i.e., if we would send out to 550 AU a FOCAL relay spacecraft for each target star system that we wish to communicate with. And this is the key new result presented in this chapter.

So, let us start by defining the Bit Error Rate or BER. In telecommunication theory an *error ratio* is the ratio of the number of bits, elements, characters, or blocks incorrectly received to the total number of bits, elements, characters, or blocks sent during a specified time interval. Among these error rations, the most commonly encountered ratio is the *bit error ratio* (BER) - also called *bit error rate* – that is, the number of erroneous bits received divided by the total number of bits transmitted (Wikipedia site: http://www.en.wikipedia.org/wiki/Bit_error_rate). It is shown that the likelihood of a bit misinterpretation

$$p_e = p(0 \mid 1)p_1 + p(1 \mid 0)p_0. \tag{26}$$

(believing that we have received a 0 while it was a 1 or the other way round) is basically given by the "complementary error function" or erfc($x$) as follows

$$BER(d,\nu,P_t) = \frac{1}{2} erfc\left(\sqrt{\frac{E_b(d,\nu,P_t)}{N_0}}\right) \tag{27}$$

In this equation one has:

1 $d$ = distance between the transmitting station on Earth and the receiving antenna in space. For instance, this could be the antenna of a precursor interstellar space probe that was sent out to some light years away.

2 ν = frequency of the electromagnetic waves used in the telecommunication link. The higher this frequency, the better it is, since the photons are then more energetic (E = h ν). In today's practice, however, the highest ν for spacecraft links (like the link of the Cassini probe, now at Saturn) are the ones in the Ka band, that is: $\nu_{Ka} \approx 32$ GHz.

3 $P_t$ is the power in watts transmitted by the Earth antenna, typically a NASA Deep Space Network antenna 70 meters in diameter.

4 The complementary error function *ercf(x)* is defined by the integral

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \tag{28}$$

(for more maths, see the relevant Wikipedia site: http://www.en.wikipedia.org/wiki/Complementary_error_function).

5 $E_b(d, \nu, P_t)$ is the received energy per bit, that is the ratio

$$E_b(d,\nu,P_t) = \frac{P_r(d,\nu,P_t)}{\text{Bit\_rate}} \tag{29}$$

6 Finally, $N_0$ is given by the Boltzmann's constant $k$ multiplied by the noise temperature of space far away from the Sun and from any other star. This "empty space noise temperature" might be assumed to equal, say, to 100 K.

This is the analytical structure of the MathCad code that this author wrote to yield the BER. Let us now consider the input values that he used in practice:

1 Suppose that a human space probe has reached the Alpha Cen. system at 4.37 light year distance from the Sun: then, $d$=4.37 light years.

2 Suppose also that the transmitting antenna from the Earth is a typical NASA Deep Space Network (DSN) antenna having a diameter of 70 meters (like those at Goldstone, Madrid and Canberra), and assume that its efficiency is about 50%.

3 Suppose that the receiving antenna aboard the spacecraft is 12 meters in diameter (it might be an inflatable space antenna, as we supposed in Maccone (2009) for the FOCAL spacecraft) and assume a 50% efficiency.

4 Suppose that the link frequency is the Ka band (i.e., 32 GHz), as for the Cassini highest frequency.

5 Suppose that the bit rate is 32 kbps = 32000 bit/second. This is the bit rate of ESA's Rosetta interplanetary spacecraft now on its way to a comet.

6 And finally *(this is the most important input assumption) suppose that the transmitting power* $P_t$ *is moderate: just 40 watts.*

Then:

1 The gain of the transmitting NASA DSN antenna (at this Ka frequency) is about 84 dB.

2 The gain of the spacecraft antenna is about 69 dB.

3 The path loss at the distance of Alpha Cen is 395 dB (a very high indeed path loss with respect to today's interplanetary missions, of course).

4 The power received by the spacecraft at that distance is $2.90 \times 10^{-23}$ watt.

5 The received energy per bit (lowered by the noise temperature of the space in between the Sun and Alpha Cen) is $1.3 \times 10^{-37}$ joule.

6 *And finally the BER is 0.49, i.e., there is a 50% probability of ERRORS in the telecommunications between the Earth and the probe at Alpha Cen. if we use such a small transmitting power !*

In other words, if these are the telecommunication links between the Earth and our probe at Alpha Cen., then this precursor interstellar mission is *worthless.*

The key point in this example is that, for all calculations, (24) and (25) were used *without taking the gain of the Sun gravity lens into account, because this was a direct link and not a FOCAL mission.*

## 18.6 BER at the Alpha Centauri Distance Enhanced by the Magnification Provided by The Sun's Gravity Lens (FOCAL)

The disappointing BER results of the previous section are totally reversed, however, if we suppose that a FOCAL space mission has been previously sent out to 550 AU in the direction opposite to Alpha Cen. so that we now have the MAGNIFICATION of the Sun's Gravity Lens playing in the game (Figures 18.2, 18.3 and 18.4).

Mathematically, this means that we must introduce a third multiplicative gain at the numerator of (24): the Sun's Gravity Lens GAIN, given by (14) where the Schwarzschild radius of the Sun is given by (7).

This new gain is huge at the Ka band frequency:

$$G_{Sun}(\nu_{Ka}) = 12444837 \sim 70 \text{ dB} \tag{30}$$

and so the received power (24) at Alpha Cen, with the usual Earth-transmitted power of just 40 watts becomes

$$P_r = 2.9 \times 10^{-23} \text{ watts} \tag{31}$$

and the relevant BER becomes absolutely acceptable:

$$BER = 0.000000526387845 \tag{32}$$

*This should convince anybody that the FOCAL space mission is indispensable to keep the link at interstellar distances equal or higher than Alpha Cen.*

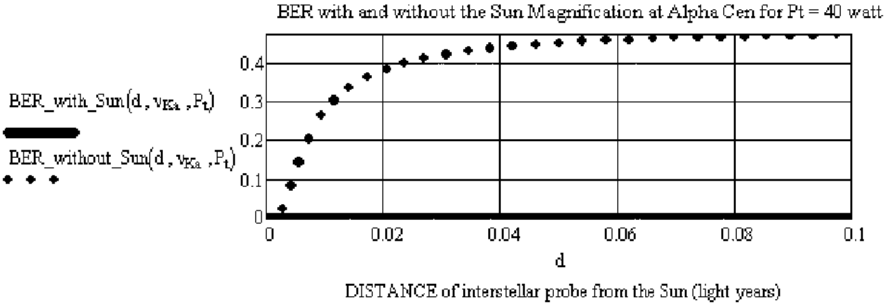



**Figure 18.2** The Bit Error Rate (BER) (upper curve) tends immediately to the 50% value (BER = 0.5) even at moderate distances from the Sun (0 to 0.1 light years) for a 40 watt transmission from a DSN antenna that is a DIRECT transmission, i.e., without using the Sun's Magnifying Lens. On the contrary (bottom line) the BER keeps staying at zero value (perfect communications!) if the FOCAL space mission is made, so as the Sun's magnifying action is made to work.

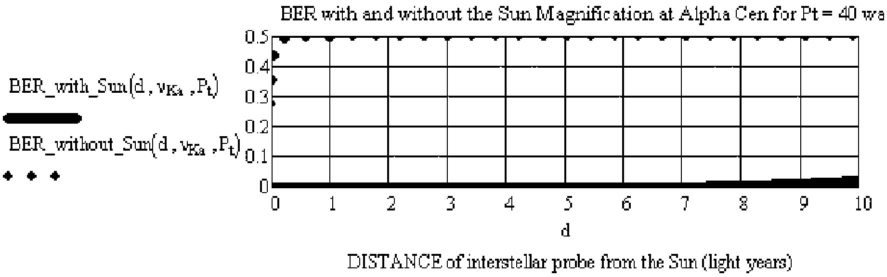



**Figure 18.3** Same as in Figure 18.2, but for probe distances up to 10 light years. We see that at about 9 light years away the BER curve starts increasing slowly.

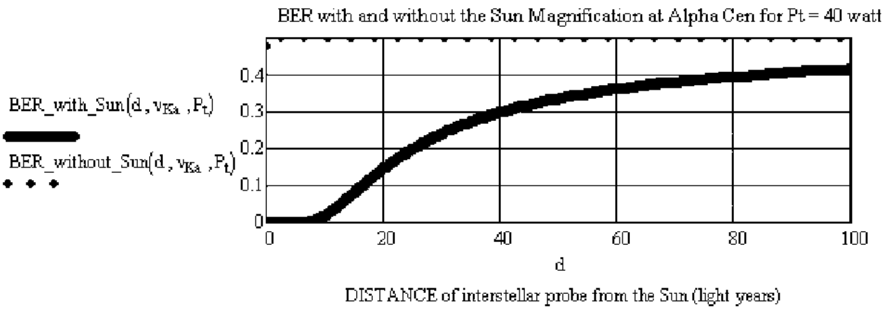



**Figure 18.4** Same as in Figure 18.3, but for probe distances up to 100 light years. We see that, from 9 light years onward, the Sun-BER increases, reaching the dangerous level of 40% (Sun-BER = 0.4) at about 100 light years. Namely, at 100 light years even the Sun's lens cannot cope with this very low transmitted power of 40 watt.

## 18.7 The Fantastic "Radio Bridge" Between the Sun and Alpha Cen A by using Two Matched Gravitational Lenses

In this section, we provide one more new result: we define the radio bridge between the Sun and Alpha Cen A by using *both* stars as gravitational lenses!

In other words, suppose that in the future we will be able to send a probe to Alpha Cen A, and suppose that we succeed in placing this probe just on the other side of Alpha Cen A with respect to the Sun and at the minimal focal distance typical of Alpha Cen A. This distance is *not* 550 AU, obviously, because both the radius and the mass of Alpha Cen A are different (actually slightly higher) than the values for our Sun:

$$\begin{cases} r_{Alpha\_Cen\_A} = 1.227\ r_{Sun} \\ M_{Alpha\_Cen\_A} = 1.100\ M_{Sun} \end{cases} \tag{33}$$

Replacing these values into (6) (obviously rewritten for Alpha Cen A), the relevant minimal focal distance is found

$$d_{focal\_Alpha\_Cen\_A} \approx \frac{c^2 r^2_{Alpha\_Cen\_A}}{4GM_{Alpha\_Cen\_A}} \approx 749\ \text{AU}.. \tag{34}$$

The Schwarzschild radius for Alpha Cen A is given by

$$r_{Schwarzschild\_Alpha\_Cen\_A} = \frac{2GM_{Alpha\_Cen\_A}}{c^2} = 3.248\ \text{km} \tag{35}$$

And so the gain, provided by (14), turns out to equal

$$G_{Alpha\_Cen\_A}(\nu_{Ka}) = 4\pi^2 \frac{r_{Schwarzschild\_Alpha\_Cen\_A}}{\lambda_{Ka}} =$$

$$= 13689321\ . \tag{36}$$

That is

$$G_{Alpha\_Cen\_A}(\nu_{Ka}) \approx 71\ \text{dB}. \tag{37}$$

Incidentally, we chose Alpha Cen A, and not B or C, because it has the highest mass, and so the highest gain, in the whole Alpha Cen. triple system. The future telecommunications between the Sun and the Alpha Cen. system are thus optimized by selecting Alpha Cen A as the star on the other side of which to place a FOCAL spacecraft at the minimal distance of 750 AU. That FOCAL spacecraft would then easily relay its data anywhere within the Alpha Cen. system.

Having found the Alpha Cen A gain (37) we are now able to write the new equation corresponding to (24) for the Sun-Alpha Cen bridge. In fact, we must now put at the numerator of (24) three gains:

1 the Sun gain at 32 GHz;

2 the Alpha Cen A gain at 32 GHz; and

3 the 12-meter FOCAL antenna gain at 32 GHz raised to the square because there are two such 12-meter antennas: one at 550 AU from the Sun and one at 749 AU from Alpha Cen A, and they must be perfectly aligned with the axis passing thru both the Sun and Alpha Cen A.

Thus, the received power given by (24) now reads

$$P_r = \frac{P_t\, G_{Sun}\, G_{Alpha\_Cen\_A} \left(G_{12\_meter\_antenna\_at\_Ka}\right)^2}{L(r,\lambda)} \tag{38}$$

where obviously $r$ equals 4.37 light years and $\lambda$ corresponds to a 32 GHz frequency.

Let us now go back to the BER and replace (38) instead of (24) in the long chain of calculations described in Section 18.6. Since the received power $P_r$ has now changed, clearly both (29) and (27) yield different numerical results. But now:

1 The link frequency has been fixed at 32 GHz (Ka band), and so no longer is an independent variable in the game.

2 Also the distance $d$ has been fixed (it is the distance of Alpha Cen A) and so it is no longer an independent variable in the game.

3 It follows that, in (29) and (27) the only variable to be free to vary is now the transmitted power, $P_t$.

Let us rephrase the last sentence in different terms. Practically, we are now studying the BER as a function of the transmitted power $P_t$ only and, physically, this mean that:

(a) We start by inputting very low transmission powers in watts, and find out that the BER is an awful 50%, i.e., the telecommunication between the Sun and Alpha Cen are totally disrupted. This is of course because the energy per bit is so much lower than the empty space noise temperature.

(b) We then increase the transmitted power, at a certain point the BER starts getting smaller than 50%. And so it gets smaller and smaller until the transmitted power is so high that the BER gets down to zero and the telecommunications are just perfect (Figure 18.5).

(c) But the surprise is that… for the Sun-Alpha Cen direct radio bridge exploiting both the two gravitational lenses, this minimum transmitted power is incredibly… small! Actually it equals just less than $10^{-4}$ watts, i.e., one tenth of a milliwatt is enough to have perfect communication between the Sun and Alpha Cen. through two 12-meter FOCAL spacecraft antennas. How is that possible?

(d) Well, that is the "miracle' given to humanity by the gravitational lenses to both explore the universe and keep the link with other stars! Just remember that, in 2009, the discovery of the first extrasolar planet in the Andromeda galaxy (M31) was announced because of the gravitational lens caused by something in between!

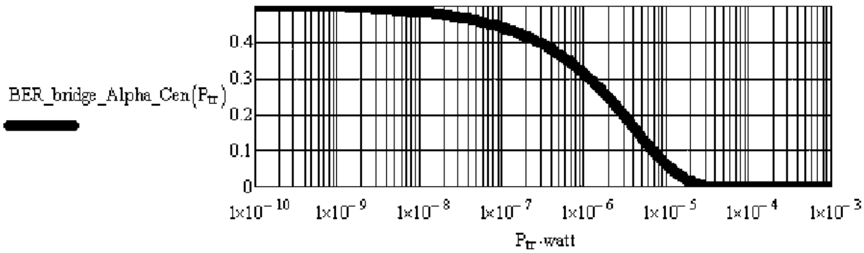



**Figure 18.5** BER for the double-gravitational-lens system giving the radio bridge between the Sun and Alpha Cen A. In other words, there are two gravitational lenses in the game here: the Sun one and the Alpha Cen A one, and two 12-meter FOCAL spacecraft are assumed to have been put along the two-star axis on opposite sides at or beyond the minimal focal distances of 550 AU and 749 AU, respectively. This radio bridge has an overall gain so high that a miserable $10^{-4}$ watt transmitting power is sufficient to let the BER get down to zero, i.e., to have perfect telecommunications! Fantastico!

## 18.8 The "Radio ridge" between the Sun and Barnard's Star, using the Gravitational Lenses of Both

The next closest star to the Sun beyond the triple Alpha Cen. system is Barnard's star (see, for instance, http://www.en.wikipedia.org/wiki/Barnard's_Star). Let us now repeat for the gravitational lens of Barnard's star the same calculations that we did in the previous section for Alpha Cen A. Then one has:

$$\begin{cases} d_{Barnard} = 5.98 \text{ light years} \\ r_{Barnard} = 0.17\, r_{Sun} \\ M_{Barnard} = 0.16\, M_{Sun}. \end{cases} \tag{39}$$

Barnard's star is thus just a small red star, that is actually "passing by" the Sun right now and is not known to have planets around it. As a consequence of the numbers listed in (39), one infers that

$$\begin{cases} d_{focal\_Barnard} = 98 \text{ AU} \\ r_{Schwarschild\_Barnard} = 0.47 \text{ km} \\ G_{Barnard}(\nu_{\text{Ka}}) = 1991174. \end{cases} \tag{40}$$

Especially the gain is important to us:

$$G_{Barnard}(\nu_{\text{Ka}}) = 63 \text{ dB}. \tag{41}$$

We replace this into the Bernard's star equivalent of (38), again supposing that two 12-meter FOCAL spacecraft antennas are placed along the Sun-Barnard straight line at or beyond 550 AU and 100 AU, respectively. The result is the new graph of the BER as a function of the transmitted power only as in Figure 18.6.

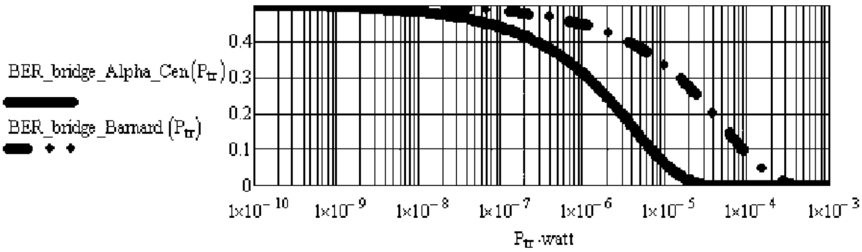



**Figure 18.6** BER for the double-gravitational-lens of the radio bridge between the Sun and Alpha Cen A (lower curve) plus the same curve for the radio bridge between the Sun and Barnard's star (upper curve): for it, $10^{-3}$ watt are needed to keep the BER down to zero, because the gain of Barnard's star is so small when compared to that of Alpha Cen A**.**

## 18.9 The "Radio Bridge" between the Sun and Sirius-A uisng the Gravitational Lenses of Both

The next star we want to consider is Sirius A. This is because Sirius A is a massive bluish star, and so it is completely different from both Alpha Cen A (a Sun-like star) and Barnard's star (a small red star). Data may again be taken from the Wikipedia site: http://www.en.wikipedia.org/wiki/Sirius and one gets:

$$\begin{cases} d_{Sirius\_A} = 8.6 \text{ light years} \\ r_{Sirius\_A} = 1.711\ r_{Sun} \\ M_{Sirius\_A} = 2.02\ M_{Sun}. \end{cases} \tag{42}$$

From these data one gets:

$$\begin{cases} d_{focal\_Sirius\_A} = 793 \text{ AU} \\ r_{Schwarschold\_Sirius\_A} = 5.96 \text{ km} \\ G_{Sirius\_A}(\nu_{Ka}) = 251385723 \end{cases} \tag{43}$$

The important thing is of course the gain:

$$G_{Sirius\_A}(\nu_{Ka}) = 74 \text{ dB}. \tag{44}$$

Then, one replaces this into the Sirius A equivalent of (38), again supposing that two 12-meter FOCAL spacecraft antennas are placed along the Sun-Sirius A straight line at or beyond 550 AU and 793 AU, respectively. The result is the new graph of the BER as a function of the transmitted power only, as in Figure 18.7.

## 18.10 The "Radio Bridge" between the Sun and Another Sun-like Star Located at the Galactic Bulge, using the Gravitational Lenses of Both

Tempted by the suggestion to increase the distance of the second star more and more, and then see what our calculations yield, we now imagine that the second star is Sun-like (i.e. that it has the same radius and mass exactly as the Sun) but is located… inside the Galactic Bulge! Namely 26,000 light years away, according to the Wikipedia "Milky Way Galaxy" site: http://www.en.wikipedia.org/wiki/Milky_Way. So, the equivalent of (24) and (38) now becomes

$$P_r = \frac{P_t\,(G_{Sun\_at\_Ka})^2\,(G_{12\_meter\_antenna\_at\_Ka})^2}{L(r,\lambda)}. \tag{45}$$

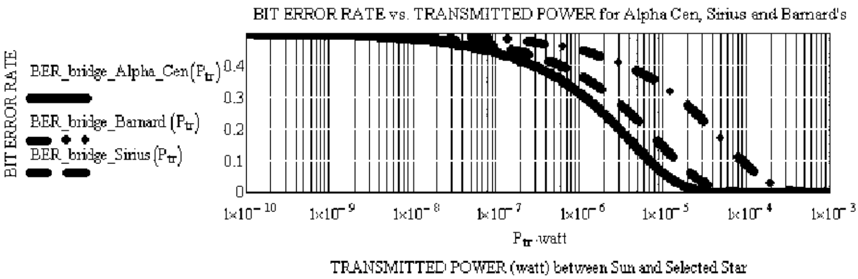



**Figure 18.7** BER for the double-gravitational-lens of the radio bridge between the Sun and Alpha Cen A (lowest curve) plus the same curve for the radio bridge between the Sun and Barnard's star (middle curve), plus the same curve of the radio bridge between the Sun and Sirius A (upper curve). From this last curve we see that only $10^{-4}$ watt are needed to keep the BER down to zero, because the gain of Sirius A is so big when compared the gain of the Barnard's star that it "jumps closer to Alpha Cen A's gain" even if Sirius A is so much more further out than the Barnard's star! In other word, the star's gain and the size combined matter even more than its distance!

and the plot of the BER vs. transmitted power is shown in Figure 18.8 as the new curve at the far right of the previous three curves of Alpha Cen A, Barnard's star, and Sirius A. The new curve showing the BER of a Sun-like star at the Galactic Bulge is naturally much to the right of the previous three stellar curves inasmuch as the bulge distance of 26,000 light years is so much higher than the distances of the three mentioned nearby stars (all less then 10 light years away from our Sun). The horizontal axis scale is much higher now, since the last BER curve gets to zero only for transmitted power of about 1000 watt.

## 18.11 The "Radio Bridge" between the Sun and Another Sun-like Star Located Inside the Andromeda Galaxy (M 31) using the Gravitational Lenses of Both

We conclude this chapter by calculating the radio bridge between the Sun and another Sun… in Andromeda! The distance is now 2.5 million light year, but the bridge would still work if the transmitted power was higher than about $10^7$ watts = 10 Megawatt. This is shown by the new curve on the far right in Figure 18.9. Perhaps this idea is not as "crazy" as it might appear, considering that recently (June 2009) the first extrasolar planet in

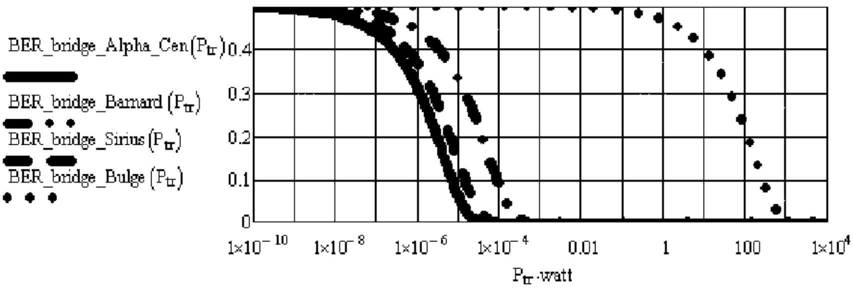



**Figure 18.8.** BER for the double-gravitational-lens of the radio bridge between the Sun and Alpha Cen A (lowest curve), plus the same curve for the radio bridge between the Sun and Barnard's star, plus the same curve of the radio bridge between the Sun and Sirius A. In addition, to the far right we now have the new curve showing the BER for a radio bridge between the Sun and another Sun (identical in mass and size) located inside the Galactic Bulge at a distance of 26,000 light years. The radio bridge between these two Suns works, and their two gravitational lenses work perfectly (i.e., BER = 0) if the transmitted power is higher than about 1000 watts.

the Andromeda galaxy was announced to have been discovered, just by gravitational lensing. See, for instance, the web site:
http://www.redorbit.com/.../possible_planet_found_outside_our_galaxy/index.html

## 18.11 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:

1 the science related to the mission;

2 the propulsion tradeoffs to get there in the least possible time; and

3 the optimization of the telecommunication link.

Nevertheless, it plainly appears that the Sun focus at 550 AU is the next most important target that humankind must reach, in order to be prepared for the successive and more difficult task of achieving the interstellar flight. In particular, we proved that the FOCAL mission only will allow us to *extend* our telecommunications with spacecraft located in space at least at the distance of Alpha Centauri or higher. This we did by resorting to the notion of Bit Error Rate, that would be zero or nearly zero only if we build up radio bridges between the Solar System and the destination target, whether that would be a spacecraft or another star system.

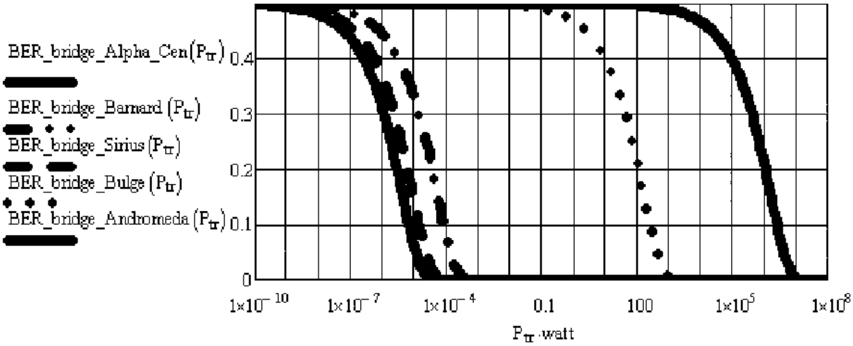



**Figure 18.9** The same four BER curves as shown in Figure 18.8, plus the new curve appearing here on the far right: this is the BER curve of the radio bridge between the Sun and another Sun just the same but located somewhere in the Andromeda Galaxy M 31. Notice that this radio bridge would work fine (i.e., with BER = 0) if the transmitting power was at least $10^7$ watt = 10 Megawatt.

We also put forward the new notion of a radio bridge between any two stars by exploiting the gravitational lenses of both and setting up two FOCAL relays on opposite sides of the two stars.

Then the powers requested for the transmissions are enormously reduced.

Perhaps Advanced Extraterrestrial Civilizations made that to work already.

If so, we have proved that Focusing the Galactic Internet is indeed a physical reality.

**Acknowledgement**

The author is indebted to many colleagues for conversations and suggestions but, in particular, he would like to thank Paul Gilster for maintaining his terrific Centauri Dreams web site: http://www.centauri-dreams.org.

## References

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

Cohen, N. 1987. The Pro's and Con's of Gravitational Lenses in CETI. In *Proceedings of the Bioastronomy International Conference* held in Balatonfüred, Hungary, June 22-27, 1987, G. Marx editor, page 395.

Derosa L. and Maccone, C. 2007. Propulsion Tradeoffs for a Mission to Alpha Centauri. *Acta Astronautica*, **60**, 711-718.

Drake, F. 1987. Stars as Gravitational Lenses. In *Proceedings of the Bioastronomy International Conference* held in Balatonfüred, Hungary, June 22-27, 1987, G. Marx editor, pages 391-394.

Drake, F. and Sobel, D. 1992. *Is Anyone Out There?* Delacorte Press, New York, in particular pp. 230-234.

Einstein, A. 1936. Lens-like Action of a Star by the Deviation of Light in the Gravitational Field. *Science*, **84**, 506-507.

Heidmann, J. and Maccone, C. 1994. AstroSail and FOCAL: Two extraSolar System Missions to the Sun's Gravitational Focuses. *Acta Astronautica*, **35**, 409-410.

Kraus, John D. 1966. *Radio Astronomy*, 2nd ed. Cygnus-Quasar Books, Powell, Ohio.

Liebes, S. Jr. 1964. Gravitational Lenses. *Physical Review*, **133**, B835-B844.

Maccone, C. 1994. *Telecommunications, KLT and Relativity*, Vol. 1. IPI Press, Colorado Springs, CO, USA.

Maccone, C. 2002. *The Sun as a Gravitational Lens: Proposed Space Missions*, 3rd edition, IPI Press, Colorado Springs, USA.

Maccone, C. 2009. *Deep Space Flight and Communications – Exploiting the Sun as a Gravitational Lens*. A 400-page treatise about the FOCAL space mission that embodies and updates all previously published material about FOCAL. Praxis Publishing Ltd, Chichester, UK.

Maccone, C. 1992/1994. 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* held in Turin, Italy, June 18, 1992, and *Journal of the British Interplanetary Society*, **47**, 45-52.

Maccone, C. 1993. 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) reconsidered by ESA within the "Horizon 2000 Plus" space missions plan.

Maccone, C. 1995. The SETISAIL Project. In G. Seth Shostak (ed.), *Astronomical Society of the Pacific Conference Series*, Vol. 74, 407-417.

Maccone, C. 2005. Radioactive Decay to Propel Relativistic Interstellar Probes along a Rectilinear Hyperbolic Motion (Rindler Spacetime). *Acta Astronautica*, 57, 59-64.

Maccone, C. 2006. Relativistic Optimized Link by KLT. *JBIS*, **59**, 94-98.

Orta, R., Savi, P. and Tascone, R. 1992/1994. Analysis of Gravitational Lens Antennas. In C. Maccone (ed.) *Proceedings of the International Conference on Space Missions and Astrodynamics* held in Turin, Italy, June 18, 1992, *Journal of the British Interplanetary Society*, **47**, 53-56.

von Eshleman, R. 1979. Gravitational Lens of the Sun: Its Potential for Observations and Communications over Interstellar Distances. *Science*, **205**, 1133-1135.

# 19

# SETI in Science Fiction

Stephen Baxter,
*British Interplanetary Society*

For more than a century, authors have imagined contact with the alien and its consequences. As a science fiction (SF) author, I am pleased to review in this chapter various concepts of extraterrestrial intelligence (ETI) and its role in human destiny as explored in science fiction. In particular, modern fiction based on the scenarios envisaged in SETI (Search for Extraterrestrial Intelligence) methodologies is considered in detail. SF has bequeathed a bank of thought experiments that might be used to guide the development of future strategies and policies.

Carl Sagan's *Contact* (1985) is perhaps the most familiar dramatization of the SETI paradigm in science fiction. But SETI fiction began much earlier.

In *The Cassiopeia Affair* (Zerwick and Brown, 1968), American radio astronomers detect a pulsed signal at the hydrogen frequency coming from a star in Cassiopeia 30 light years away. The sequence length is a multiple of two primes ($41 \times 41$); when the data is set out in a grid pattern information is revealed about the sending ETI. The data is taken directly to the US President by a senior scientist who uses the argument of the Drake equation (Michaud, 2007) to explain the message's provenance and significance. There is some discussion about a possible reply, which might provoke some kind of attack, or incite a contact that will lead to culture shock; though the information stays under national control, a decision about replying is left to a global consensus. The President is enthused enough to announce the message's existence publicly and use it to justify a plan for a world government – but hawkish factions attempt to discredit the signal. The book is optimistic in

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_19, 

terms of post-detection implications: the very existence of such a message will inspire us to better ourselves.

Thus from just a few years after its emergence in about 1960, fictional works have been inspired directly by modern SETI, including the underlying logic and assumptions of the Drake equation, the reception of microwave signals at 'universally significant' frequencies, the encoding of messages in simple mathematical forms, dilemmas concerning response, and complex post-contact implications. But these works draw on a much wider and older tradition of science fiction. Dating from both before and after 1960, there have been many works of SF depicting detection and contact events much more exotic than through radio astronomy – and many outcomes for mankind other than simply beneficent.

Of course SF works are pieces of commercial fiction; a story of alien invasion may be told primarily for entertainment, satiric, comic or other literary purposes, rather than to reflect scientific plausibility. Nevertheless the best "hard" SF — that is, fiction which respects scientific knowledge, method and boundaries — gains its very rhetorical force as an exploration of the possible, and has given us many imaginative yet plausible depictions of the alien and human interaction with it. Certainly SF has informed the public imagination concerning ETI, and SF is the nearest we have to the "war gaming" of specific detection and contact scenarios. This literature is the subject of this essay.

This necessarily selective survey is intended to highlight relevant works and themes. Useful surveys of the subject include the relevant chapter of Dick (1996) and relevant essays in Clute and Nicholls (1993). Reviews of the genre as a whole include Aldiss and Wingrove (1986), James and Mendlesohn (2003) and Roberts (2006).

Titles of novels, motion pictures and television series are generally given in italics; titles of shorter works are generally given in quotation marks.

## 19.1 Defining the Alien

Science fiction is a product of post-Renaissance western secular and scientific culture. The modern genre could be regarded as fully formed with the publication of H.G. Wells' first novel *The Time Machine* in 1895 (though the label science fiction, or 'scientifiction', was not coined until 1926 by Hugo Gernsback).

The genre, however, has deeper roots in various forms of literature dating back to antiquity, and from as early as the second century AD (Lucian's *True History*) many of these predecessor works have featured speculations on alien beings. (See Guthke, 1983, for a useful review of this background, and for more complete references to early works cited in this section.) It was, however, with the publication of Copernicus' (1543) and Galileo's (1610) seminal astronomical works that speculations about the plurality of worlds could for the first time be based not on faith and metaphysics but on data; from now on other worlds were imagined as planets like Earth suspended in space, as opposed to multiple copies of an Aristotelian geocentric cosmos.

The new Copernican universe was soon explored in fiction. In Johannes Kepler's *Somnium* (1634) the lunar folks' lives are shaped by the known conditions of the Moon's surface, particularly the long days and nights. Kepler's purpose is to deal with the fraught philosophical and theological implications of the emerging cosmology, but this is clearly a precursor of modern "hard" SF, using scientific understanding and analogical reasoning to speculate about life on another world. Mankind's first contact with the new breed of post-Copernican aliens was made by the protagonist of Bishop Francis Godwin's *The Man in the Moone* (1638), who was flown to the moon by migrating geese. The lunar giants were Christians, and were portrayed as superior to mankind, and happier.

It is understandable that in the centuries after Copernicus fiction writers at first saw the possibilities of alien life through the prism of Christian theology, though wider philosophical issues were explored, often satirically. Gradually, however, following the work of such thinkers as Newton and Laplace and the emergence of modern scientific thinking, this gave way to more secular and scientifically realistic explorations.

By the end of the 19th century key foundations for recognisably modern imaginings of alien life and intelligence were in place. The successes of spectroscopy from about 1860 had confirmed the Newtonian view that the Universe is everywhere composed of the same elements and obeys the same physical laws. Advances in astronomy were providing details of the environments offered for life by other worlds, such as Percival Lowell's hypotheses about civilization on an arid Mars. Finally Darwin's theory of evolution, published in 1859, provided a mechanism with which to speculate realistically about the nature of life as shaped by exotic environments. An explosion of fiction about other worlds and extraterrestrial life followed.

Thus the French astronomer Camille Flammarion's *Recits de l'infini* (1872) included tales of intelligences embodied in forms dictated by other evolutions, such as sentient plants. J.H. Rosny the Elder, also French, wrote of living minerals in *Les xipehuz* ('The Shapes') (1887), and ferromagnetic life in *La Mort de la Terre* ('The Death of Earth') (1910). Works of this period were often influenced by Laplace's "nebular hypothesis" of the formation of the solar system (1796–1827), which predicted that worlds further from the Sun should be older. In Percy Greg's *Across the Zodiac* (1880) a more ancient race of Martians is portrayed as advanced but essentially decadent; Greg's travellers return home confirmed in their view of their own superiority.

All these works were overshadowed in stature and lasting influence by two books, both about Mars, which encouraged no such complacency.

*The War of the Worlds* by H.G. Wells (1897) and *Two Planets* by Kurd Lasswitz (1897) both made their first appearance in 1897, and both derived from Lowell's work. In Wells's *The War of the Worlds*, technologically advanced Lowellian Martians launch a dramatic invasion of Earth: the War of the Worlds "isn't a war … any more than there's war between men and ants" (Book 2, Chapter 7).

Lasswitz's Martians, however, are humanoid. Ethically, morally and technologically advanced, they embody Lasswitz's post-Kantian belief in the betterment of mankind through science and education. But humanity's response is complex and fearful, foreshadowing decades of explorations of the implication of alien contact: "The friendship of the Martians appears dangerous to me, their enmity appears disastrous" (Chapter 14).

Wells's and Lasswitz's complementary works did much to define the notion of ETI in modern SF, including:

- the nature of the alien and how it is determined by evolution in differing environments;
- speculation about its capability and its apprehension of the universe;
- our fears about its hostility or hopes for its benevolence; and
- our wondering at our future in an inhabited universe.

Indeed these are the concerns that direct SETI searches and inform the consideration of post-contact implications.

## 19.2 Space Opera

From this beginning, hordes of aliens swarmed through the pages of the popular fiction of the 20th century.

Mars continued to be a favoured setting (Baxter, 2004), such as in Burroughs's prototypical *Barsoom* series (Burroughs, 1912). A late return to a Burroughs-like Mars was Bradbury's wistful *The Martian Chronicles* (Bradbury, 1950).

On a wider scale the tradition of "space opera" (a term coined in 1941), of a universe crowded with aliens as an arena for gaudy adventures, was developed in such works as Smith's *The Skylark of Space* (1928) and its many sequels (Smith, 1928), and van Vogt's *The Voyage of the Space Beagle* (van Vogt, 1950). White's very popular *Hospital Station* series (White, 1962) features attempts to supply medicinal help across species boundaries. The tradition continues to the present day with the George Lucas *Star Wars* movie franchise (1977 onwards), the TV series *Babylon 5* (1994–99) and *Battlestar Galactica* (2004–present), and in prose in such works, some very thoughtful, as the 'Vorkosigan' series (1986–present) (McMaster Bujold, 1986), the 'Polity' series (2001-present) (Asher, 2001) and the 'Culture' series (1987-present) (Banks, 1987).

In many of these works the aliens could be quasi-humanoid, though the more malevolent beings often borrowed attributes from reptiles, octopuses or insects. These depictions reach back to 19th-century ideas that convergent evolution could mandate a human form on other worlds. The visions of media franchises like *Star Wars* and *Star Trek* (1966–present) are surely defined by constraints of budget as well as imagination. But even under such constraints intriguing ideas can be explored, such as the hive-like nature of the Borg of *Star Trek.*

These colourful genre creations have had a fundamental educative role in establishing our place in the Universe in the public mind, and have been hugely influential in their impact on young imaginations. Meanwhile, however, a more thoughtful literature has evolved.

## 19.3 The Nature of the Other

### 19.3.1 Other evolutions

In 1934 a significant addition to the growing canon of Martian literature was *A Martian Odyssey* (Weinbaum, 1934) about a bird-like alien called Tweel who lives in a weird but self-consistent ecosystem, including silicon-based life forms that produce bricks as waste. This story would prove to be a founder of a sub-genre of works that, trying to adhere to scientific plausibility, used evolutionary principles to show adaptations to exotic environments (see Cohen and Stewart, 2002, for a review). Thus Clement's influential novel *Mission of Gravity* (Clement, 1954) featured creatures adapted to a high-gravity, fast-spinning world. Clarke's 'A Meeting with Medusa' (Clarke, 1971) imagined inflated alien life forms drifting in the clouds of Jupiter. Wells was once again a precursor (Wells, 1897). Wells, a student of Huxley, was a powerful fabulist of natural selection, and his Martians were a vision of mankind's evolutionary future.

Some writers tried to construct complete ecologies, including the *Helliconia* trilogy (Aldiss, 1982), about life adapted to a planet with "long seasons" thanks to its elliptical orbit in a double star system, and *The Legacy of Heorot* and its sequels by Niven et al. (1987). Herbert's *Dune* and its sequels (Herbert, 1965) about life on a desert world, gave a powerful impression of a rich ecology without, however, addressing many basic questions (such as what the sandworms ate).

Most alien fiction depicts complex life forms (that is, multicellular or equivalent), presumably because this offers more speculative and dramatic possibilities. However, microbial life on Mars featured in *The Secret of Life* (McAuley, 2001); in Benford's *The Martian Race* (Benford, 1999), microbial networks spawn intelligence on that planet.

Some writers have spun visions of aliens shaped by different social systems rather than (or as well as) by different physical environments. The engaging aliens of Niven's 'Known Space' series (Niven, 1970) tend to be dominated by one social characteristic: his puppeteers are super-cautious; his cat-like Kzinti are super-aggressive. Beginning with Wells' *The First Men in the Moon* (Wells, 1901), there have been many explorations of alien hive societies (Baxter, 2004). Perhaps because of the anti-individual aspects of hives, giant insectile creatures are the opponents of humanity in such works as Heinlein's *Starship Troopers* (Heinlein, 1959) and *Ender's Game* (Card, 1985). But in Silverberg's novel *The Queen of Springtime* (Silverberg, 1989) the hives of the hjjks, human-sized and intelligent, are places of love and life. Meanwhile the Tines of Vinge's *A Fire Upon the Deep* (Vinge, 1992) are another popular example of collective intelligence.

### 19.3.2 The outer limits

While many fictional aliens are, in terms of their consciousness and motivation, little more than distorted representations of humanity, many attempts have been made to depict the genuinely alien. Lindsay's remarkable novel *Voyage to Arcturus* (Lindsay, 1920) depicts creatures with senses other than ours, with skin colours reflecting their emotions, and communicating by telepathy. In *The Game-Players of Titan* (Dick, 1963), to silicon-based, methane-breathing slugs from Titan we are "stunted, alien creatures, warped by enormous forces into miserably malformed, distorted shapes" (Chapter 17).

Aliens may not be as we are with regard to fundamental parameters; that is, they may not be finite, mortal creatures of planetbound chemistry. Lem's novel *Solaris* (Lem, 1961) is about a living ocean, a solipsistic planetary consciousness with which, through sheer scale differences, it proves impossible to achieve contact. Clarke's 'Out of the Sun' (Clarke, 1958) featured life formed of tangles of magnetic flux on the surface of our Sun. *Sundiver* (Brin, 1980) explored the idea of life inside the Sun, and the author's own *Ring* (Baxter, 1991-1997) posited creatures of dark matter infesting stars' gravity wells. *Dragon's Egg* (Forward, 1980) was about tiny, fast-living creatures based on a chemistry of the neutron-rich nuclei of the surface of a neutron star; the author's own *Flux* (Baxter, 1991-1997) featured life inside such a star.

Hoyle imagined intelligence arising away from solid bodies altogether, in an interstellar cloud (Hoyle, 1957). See also *Sunborn* (Benford, 2004), about life in tangled magnetic fields on the edge of the Solar System. There have been many depictions of living planets and stars, dating back to Stapledon's *Star Maker* (Stapledon, 1937) and even earlier. In *Aftermath* (Sheffield, 1998), Alpha Centauri impossibly goes supernova – but this is in fact a manifestation of a new order of life altogether, as one living star reaches out to another.

In the author's own *Exultant* (Baxter, 2003–2006) and Cohen and Stewart's *Wheelers* (2000), life infests the universe on every scale in space and time, right back to the first moments after the Big Bang when living things fed off the energies of inflation. Perhaps the ultimate novel environment is to imagine aliens in a universe whose very physical laws are different from our own, for example nuclear forces (Asimov, 1972) or the gravitational constant (Baxter, 1991–1997).

Some fictions show ETIs so different from us that we will forever be mutually incomprehensible. *The Embedding* (Watson, 1973) explored issues of language. An idea recently explored by Schroeder (2003) is that of communication horizons. Perhaps among creatures that are simply too divergent, no symbolic communication is possible.

### 19.3.3 Big and smart

What if the difference between us and ETI is quantitative as well as qualitative?

Fictional explorations of close encounters of humanity with true super-intelligences are rare – perhaps because readers do not like being humbled – but are not reassuring. In Heinlein's story 'Goldfish Bowl' (Heinlein, 1942) we encounter aliens as far above us as we are above fish in a tank, and as uncaring: "The human race had reached its highest point – the point at which it began to be aware that it was not the highest race, and the knowledge was death to it, one way or the other – the mere knowledge alone …" A superior intelligence may, however, be benevolent. In *The Inferno* (Hoyle and Hoyle, 1973), godlike aliens save the Earth from a Galaxy centre explosion: "It was as if a man should hold up a hand to shield a moth as it flew near a candle" (Chapter 12).

There have been many depictions of communities of life more or less organized on galactic or even larger scales. Vinge (1992) sketched a galactic geography of rings of differentiated mentation, with innumerable races linked by a kind of galactic internet. In Egan's *Incandescence* (2008) the Galaxy is occupied by several overlapping "panspermia", islands of life types. But ultimately there is a technological merger; the Galaxy becomes a kind of computing substrate shared by many species.

A subset of depictions of superior races are tales of "big dumb objects" (Clute and Nicholls, 1993) of engineering sometimes on a very large scale, which we, dwarfed, may or may not comprehend. In the 1956 movie *Forbidden*

*Planet* astronauts explore a 20-mile-wide cubical machine left buried inside a planet by the long-dead alien Krell. In Clarke's *Rendezvous with Rama* (1973) a kilometers-long spacecraft sails through the Solar System; human astronauts explore the interior, but do not come close to understanding its true purpose. *Eon* (Bear, 1985) features Rama's conceptual descendant, a hollowed-out asteroid that contains seven chambers; the last is infinitely long inside. Reed's *Marrow* (2000) features a spaceship of unknown origin the size of a gas giant; as it circumnavigates the Galaxy its crew explores its interior. *Ringworld* (Niven, 1970) is a ring-shaped habitat around a star. Shaw (1975) trumps the Ringworld with a Dyson sphere with the surface area of 600 million Earths.

Sometimes alien civilizations leave behind Universe-spanning artefacts, or they may shape the destiny of the cosmos as a whole, as if the Universe itself is an artefact. In Pohl's *Gateway* and its sequels (1977), the Heechee have left behind a network of wormhole-like transportation systems. But the Heechee are hiding in fear of a more powerful race that is believed to be destroying the Universe in order to rebuild it with physical constants adjusted more to their liking.

Thus SF works have explored a wide phase space of possible and apparently plausible forms of alien life and ETI.

### 19.3.4 The Fermi Paradox

Some modern fiction appears to have been inspired by studies of the Fermi Paradox – and conversely the study of the Paradox has been enriched by fictional explorations. A popular example is the "Prime Directive" of *Star Trek*, as a Fermi solution an example of an "interdict hypothesis" (Webb, 2002), which requires members of a galactic Federation not to meddle in the affairs of indigenous civilizations before they discover a faster-than-light warp drive. Solutions of the Paradox require systematic constraints on the behaviour of *all* ETIs, such as domination by lethal destroyers as in *The Forge of God* (Bear, 1987): "We've been sitting in our tree chirping like foolish birds for over a century now, wondering why no other birds answered. The galactic skies are full of hawks, that's why" (Chapter 46). In Brin's "Uplift War" series (Brin, 1980), all intelligent aliens are part of a vast intergalactic society based on the fact that all of them (save humans) have been nurtured into intelligence by a process of uplift, in a chain started by the vanished Progenitors. This society systematically hides from non-uplifted species. See also the author's own *Manifold* series (Baxter, 1998–2000), exploring various solutions to the Paradox; in parallel universes we are alone, or we find ourselves in a Galaxy repeatedly sterilized by core explosions, or we are in a cosmos engineered by hidden builders.

## 19.4 ETI, Humanity and Religion

Our fiction about alien life and ETI must always reflect our humanity. Depictions of the alien are often rooted in older mythologies of the Other, including legends of angels and demons, witches, elves and trolls, and more generally our own deeper psyches. Sometimes modern SF draws consciously on this back catalogue of imagery; the Aliens of Ridley Scott's movie and its sequels (1979–1997) are nightmares of motherhood, and the *E.T.* of Spielberg's movie (1982) is like a human toddler. Meanwhile, stories of alien invasion draw on a tradition of apocalyptic literature dating back to the Book of Revelation.

As noted in Section 19.2, the idea of alien intelligence has thrown up deep theological questions since the age of Copernicus. Fiction writers have continued to explore these implications to the present day; perhaps in some deep sense SF recasts aliens and demons into the spaces between the stars.

Ancient theological questions were dramatized directly in Bradbury's 1949 story "The Man" (Bradbury, 1996), in which successive incarnations of Christ are pursued from world to alien world, and in Lewis's "Cosmic Trilogy" (Lewis, 1938), a Christian allegory in which Earth is the only "fallen" planet. Clarke's "The Star" (Clarke, 1955) is a poignant story of the significance of a supernova, studied by a Jesuit astronaut-scientist: "What was the need to give these people to the fire, that the symbol of their passing might shine above Bethlehem?" In some works we are converted to alien religions (Silverberg, 1970; Heinlein, 1961). Stapledon's *Star Maker* (1937) is effectively an alternate theology, in which a kind of god uses creation as an experimental laboratory.

The classic depiction of a collision between faith and the reality of the alien is surely *A Case of Conscience* (Blish, 1958), in which a Jesuit priest must deal with the theological implications of the discovery on the planet Lithia of reptilian ETI who know nothing of God; he decides the world is the work of the devil and must be exorcised. Russell's *The Sparrow* covered similar ground (Russell, 1997).

## 19.5 Paranoia vs. Pronoia

Fears of aggression from ETI, yet hopes for its benevolence, date back at least as far as the polar-opposite books of Wells and Lasswitz (Wells, 1897; Lasswitz, 1897).

While many of the aliens depicted in space opera were routinely malevolent, the possibilities of peaceful contact could be depicted, such as in

Gallun's "Old Faithful" (Gallun, 1934) in which human and Martian learn to cooperate despite extreme differences in biology. In the story "First Contact" (Leinster, 1945), an Earth ship bumps into aliens in the Crab Nebula: "Failure to be suspicious might doom the human race – and a peaceful exchange of the fruits of civilisation would be the greatest benefit imaginable."

From around 1950 the Cold War generation found its fears reflected in a wave of malevolent-alien books such as Heinlein's *The Puppet Masters* (Heinlein, 1951) and movies including George Pal's 1953 adaptation of *The War of the Worlds*, *The Thing* (1951) about an alien trapped under Arctic ice, and *It Came from Outer Space* (1953) about shapeshifting aliens. The theme has continued to the present day with movies like Emmerich's *Independence Day* (1996) and Spielberg's 2005 version of *The War of the Worlds*, which reflects the post-9/11 uncertainty of the 21st century. These media representations with their powerful special effects did much to cement the idea of the hostile alien in the popular imagination.

The UFO phenomenon spawned its own tales of covert alien interference, notably the TV series *The X-files* (1993–2002). In *Earth vs. the Flying Saucers* (1956) UFO paranoia is fulfilled in a battle over Washington DC. Spielberg's *Close Encounters of the Third Kind* (1977) is a benevolent expression of the UFO mythos.

Invasion movies are generally less convincing in terms of motive: what could the aliens want that would be worth interstellar aggression? Some compelling suggestions have however been made. Wells's Martians wanted our warmer wetter world, and to use us as livestock. In Niven and Pournelle's *Footfall* (Niven and Pournelle, 1985) the invaders are migrants fleeing a ruined world. The Moties (Niven and Pournelle, 1974) are driven to aggressive expansion by a quirk of their biology which renders them unable to control their own population growth.

The aggression may be systematic on a galactic scale, such as in Saberhagen's "Berserker" series about lifeless killing machines (Saberhagen, 1967). The six-book "Galactic Centre" series (Benford, 1977–96) [77] describes a Galaxy riven by constant conflict between organic beings and sentient machines. In the "Time Odyssey" (Clarke and Baxter, 2004–2008), an advanced culture "culls" wasteful young civilizations in order to preserve cosmic resources for the very far future; the aliens believe they are being benevolent. Bear's *The Forge of God* (1987) and its sequel *Anvil of Stars* (1992) are perhaps the ultimate expression of such paranoid visions, portraying world-destroyers hiding from their victims – a Galaxy of lies and genocide.

Some writers argue with the assumptions of militaristic fiction. Vietnam veteran Joe Haldeman's *The Forever War* (Haldeman, 1974) was a response to the xenophobia of Heinlein's *Starship Troopers.* Card (1985) followed up the genocide of his own *Ender's Game* (1977) with the redemption of *Speaker for the Dead* (1986).

Meanwhile, visions like Lasswitz's are ancestral to dreams of friendly alien contact. Wells's own 1937 novel *Star Begotten: A Biological Fantasia* (Wells, 1937) is a kind of response to *The War of the Worlds*. Martians, or aliens of another sort, might be meddling with the destiny of humanity by tinkering with our genomes using "cosmic rays". There have been many explorations of complex cultural and biological interactions between human and alien, including the *Faded Sun* trilogy (Cherryh 1978–9), with perhaps a limit being reached in the sexual intimacies of the *Xenogenesis* trilogy (Butler, 1987–9).

Perhaps the most sublime expression of our longing for peaceful alien contact is the idea that ETI may come to help us survive our problems in the short term, and in the long term even lead us to a destiny among the stars. This is foreshadowed in Lasswitz (1897). In the 1951 movie *The Day the Earth Stood Still*, the alien Klaatu attempts to quell human violence. The "Federation" of *Star Trek* is a familiar example of what Bracewell (1975) referred to as a "Galactic Club". The Vegans in *Contact* (Sagan, 1985) call to give us short-term help, and a glimpse of the galactic society of which we might one day be a part.

This reaches an ultimate expression in Stapledon's *Star Maker* (1937), in which a joined cosmic mind sets out in search of the Star Maker, the architect of reality. Stapledon's ideas heavily influenced Arthur C. Clarke. In his *Childhood's End* (Clarke, 1953), a near-future first contact by the alien "Overlords" leads to a shepherding of the children of mankind into a transcendent form of intelligence called the "Overmind". *The City and the Stars* (Clarke, 1956) depicts a kind of conclusion of such stories, set in a universe shaped by a billion years of pan-cosmic projects. *2001: A Space Odyssey*, a reworking of these ideas (1968), perhaps the best-known and best-loved SF story of the 20th century, tells of the intervention of alien technology in the destiny of the "man-apes", our remote ancestors, and ultimately transcendence in our own time. One cannot deny the emotional power of such visions, as a secular response to the loss of religious certainties in the wake of Copernicus and Darwin.

But not everybody might be eager to join such a community. Like Hoyle's Black Cloud (Hoyle, 1957), the 'Eater' of Benford's novel of the same name (Benford, 2000) is an ancient solitary intelligence encoded into the magnetic field of a small black hole. It appears never to have engaged in cooperation even with others of its own kind; contact with it cannot lead to us joining any "Galactic Club".

And any benevolent gesture may backfire. This is shown even in Clarke's *2001* cycle. *3001* (1997) is a saga of resistance against the Monolith-builders, who after all turned australopithecines into wielders of weapons.

## 19.6 SETI Fiction

This section is a discussion of fictional scenarios relevant to modern SETI: that is, they depict the detection of ETI in the near future in the form of the reception of signals, the discovery of artefacts, or direct contact (or a combination).

### 19.6.1 Signals

As noted in Section 19.1, since 1960 some works have been directly inspired by the SETI paradigm. In Gunn's *The Listeners* (Gunn, 1972), SETI astronomers pick up a signal coming from the star Capella, 45 light years away, which contains echoes of early Earth broadcasts. There turns out to be a dot-dash message embedded in the "noise". Gunn was a contact-optimist who argued that malevolence is illogical: "The only thing worth sending from star to star is information, and the certain profit from such an exchange far outweighs the uncertain advantage from any other kind of behaviour" (p.126 of the 2004 Benbella edition). We reply, and during the 90 years' wait for a further response, mankind focuses on the exchange and peace breaks out. The book is a poetic exploration of the motivation of SETI workers.

*Contact* (Sagan, 1985) is perhaps the most faithful dramatisation of the SETI paradigm, as to be expected from a figure so closely associated with the field. Ellie Arroway's detection of a Message from Vega follows the procedural model later set out in the field's 1989 Declaration of Principles (Michaud, 2007), including verification and responsible reporting. The Message itself is a prime-number grid-pattern manual for fabricating a Machine that, when constructed, whisks Ellie to the centre of the Galaxy, where she encounters an hierarchy of helpful intelligences who have "cultivated" the Galaxy (Chapter 20).

Sagan portrays a dramatic cultural reaction to the Message. The mature religious reaction is to see the Message as a sign of a greater God who spans a Galaxy: "It … makes God very big" (Chapter 22). The fundamental secular reaction is one of hope, a proof-of-existence that we don't have to destroy ourselves: "For decades, young people had tried not to think too carefully about tomorrow. Now, there might be a benign future after all" (Chapter 7).

In Sagan's novel some believe that the evidence of ETI "would require several generations to be 'decoded' and properly assimilated …" (Chapter 22). Time provides a cushion against culture shock. But the speed of assimilation of alien information must depend on its decipherability and the ingenuity with which it is packaged. In *A for Andromeda* (Hoyle and Elliot, 1962), a signal is first picked up by a new radio telescope in the UK which is more powerful than its competitors – so that the UK and its scientists have a temporary monopoly on access. There is a tussle for control between the scientists, government ministries and foreign espionage agents. The message turns out to be an instruction kit for building a kind of super-computer; this then begins a dialogue with mankind, thereby overcoming translation and distance difficulties. Maverick scientist John Fleming repeatedly advocates a policy of "killing' the computer and its products, calling it "an intellectual fifth column from another world" (Chapter 5). Fleming turns out to be right. In the sequel (1962) the AI tries to exterminate mankind.

In *The Hercules Text* (McDevitt, 1986) a signal is received from what turns out to be an artificial pulsar in the intergalactic void. The aliens' motivation appears to be simple loneliness; they are a group mind, so effectively a single individual, alone in the dark. Rather as in *Andromeda* a "self-initiating program" (Chapter 6) is found in the data stream. Though there is no purposeful malevolence as in *Andromeda* the information contained in the data is destabilising on a number of levels: the theological – do the aliens have souls?; the technological – advanced medical methods offer the possibility of immortality; and the political – a Star Wars-like missile shield becomes possible, bringing the threat of a pre-emptive strike. In the end the hero, a bureaucrat with a conscience, ensures that the last copy of the data is hidden until such time as we mature enough to use it responsibly. Thus policy questions linger even after the contact incident is over. A similar scenario is the author's own "Turing's Apples" (Baxter, 2008).

*SETI* (Fichman, 1990) is a fast-paced juvenile variant on these themes, in which the ETI signal is a direct reply to a ham-radio signal sent out by the young hero when aged six. A crazed deputy director of the Jet Propulsion Laboratory tries to steal priority. Kube-McDowell's *Emprise* (1985) interestingly describes the impact of a signal reception in a near future in which mankind's ability to deal with the incident is almost nullified by a global crash. While the reception of radio-frequency messages has often been dramatized, there are examples of messages received from other parts of the spectrum, such as the naked-eye optical signal received in the author's own 'Eagle Song' (Baxter, 2008).

### 19.6.2 Direct contact

A number of apparently plausible direct contact scenarios have been developed by the fiction writers, often based on slower-than-light technology.

Perhaps the best thought out modern scenario is *Footfall* (Niven and Pournelle, 1985), which shows a slower-than-light interstellar invasion made with modern technology or feasible extrapolations. SF writers are summoned to advise the US President, who notes that "they're the only experts we have" (Chapter 17). Niven and Pournelle carefully outline a war waged from the sky. Rather than land immediately as in Wells's novel, the "fithp" use the advantage of the "high ground" of space, by knocking out space-based resources such as communications satellites, setting off high-altitude nuclear explosions to induce destructive electromagnetic pulses, and using kinetic-energy weapons to take out military installations and key infrastructure items such as dams, harbors and transport nodes. A "footfall", a 4000-Mt a meteoroid strike, triggers a tsunami. More extreme possibilities for interstellar war include the use of relativistic-speed kinetic weapons (Pellegrino and Zebrowski, 1995), implanted matter-antimatter bombs (Bear, 1987), and the destabilizing of the sun (Clarke and Baxter, 2004–2008).

Alternatively, a more subtle interference may be attempted, as in the post-Roswell saga of alien manipulation *The X-files* (1993–2002), or *Invasion of the Body Snatchers* (1956) about a takeover by alien replicas of humans. In Bear's *The Forge of God* (1987) "planet-eaters" test us with a campaign of disinformation and confusion. Interestingly the ETI, knowing our psychology, even manipulates our expectation of positive SETI outcomes: "We thought the arrival of something like you would change us all. You've taken advantage of that" (Chapter 30).

Aside from these relatively "standard" scenarios of visitors in spacecraft, contact with more exotic beings has been imagined. The first detection of the 'Black Cloud' of Hoyle's 1957 novel is by optical astronomers, and then through gravitational perturbations. The scientists recognize they have a "psychological block" when trying to weigh up the possibility of intelligence versus naturalistic explanations for the Cloud's behavior. After the incident, one copy of the "code" by which the Cloud can be contacted is retained, and a dilemma is faced about whether this code should be destroyed or disseminated.

Even benevolently intended "invasions" may have adverse effects. The arrival of the Overlords in Clarke's *Childhood's End* (1954) appears to fulfil contact-optimist dreams in that we are saved from poverty, ignorance and war. But the very first human reaction, by a space engineer seeing his project suddenly rendered futile, is despair. In the face of the aliens' vastly superior knowledge our creative arts and sciences decline, and religions are undermined by the Overlords' historical evidence about the founders. Though he was a believer in a cosmic destiny for mankind Clarke expressed disquiet about the impact of contact with a higher ETI: "It might be better, in the long run, for us to acquire knowledge by our own efforts, rather than be spoon-fed" (Clarke, 1992).

### 19.6.3 Artefacts

First contact might come with the discovery of an artefact of ETI, rather than ETI itself, on Earth or a near-space location (see Chapter 33 of Shklovskii and Sagan, 1966; and Benford, 1999 for reviews).

A familiar speculation is that an ETI artefact might be discovered in the course of an archaeological excavation, or recorded in an historical document. In Kneale's BBC serial *Quatermass and the Pit* (1958–9), an apparent unexploded bomb under London turns out, rather thrillingly, to be a Martian spaceship, buried since prehistory; it has influenced human development, and is activated when dug up. The same idea was explored in King's *The Tommyknockers* (King, 1987).

The "Life Probe" of McCollum's (1983) novel is a smart probe of the kind suggested by Bracewell (1960), coming in search of civilizations with faster-than-light technology. An initial detection is mistaken for a deep space weapons test. It is pointed out that revealing the probe's existence to the public would itself be an event detectable to the monitoring probe itself, and therefore would give information to a potentially hostile visitor.

The depositing of long-lasting artefacts in some stable location might be a rational communication strategy for a short-lived ETI. The most famous alien artefacts in SF are of this type: the Monoliths of Clarke's *2001: A Space Odyssey* (1968). Theories sketched for the purpose of the Monolith found on the Moon include: a supply cache, a shrine, a survey marker, a tomb, a science instrument. A null hypothesis is that it is a natural formation. It turns out to be an alarm system, revealing humanity's ability to cross space.

The power of even a mute artefact to evoke strong psychological and cultural reactions should, perhaps, not be underestimated. Thus in *2001* the lunar Monolith's mere existence provokes a philosophical revolution: "Here was the proof, beyond all shadow of doubt, that [man's] was not the only intelligence that the universe had brought forth" (Chapter 12).

### 19.6.4 Analysis

This brief survey can be seen to highlight a number of common features of ETI detection and contact events as imagined by the science fiction writers. These can be seen as a contribution to the discussion on SETI, its strategies and possible outcomes that has been ongoing for five decades.

While the standard SETI scenario, of the receipt of a coded microwave signal from a distant star, has been dramatized a number of times, fiction writers have developed many variant scenarios. For instance such a signal if put together with sufficient ingenuity may be easily decoded and unwrapped; assimilation and consequent impact may be rapid. Other kinds of detection than through formal SETI searches may occur, perhaps accidentally.

The contacting ETI may be similar to us in fundamental senses but much more exotic possibilities have been imagined. A variety of motivations for an ETI to attempt contact has been imagined, including: an end to loneliness; a legacy; the rescue or uplift of mankind; an invitation to join some shared project or organization; a request for help.

Control of the contact event may be attempted at a variety of levels up to the international. However, there seems general skepticism of the efficacy of decision-making above the national level (i.e., not the UN). Depending on circumstances individuals with local control may wield a good deal of power.

In handling the contact event there is often tension between groups with different interpretations of the problem and different needs in dealing with it. Scientists must work in a cross-disciplinary manner, and the political or military need for control of information often clashes with the scientists' need for openness.

The modern public is expected to be well informed in advance of the general nature of an ETI contact and possible implications. A wide variety of social implications of contact events have been sketched, ranging from panic and demoralization to optimism and uplift. Responding to a contact is often a subject of intense debate. Even releasing the news to the public, if human communications are being monitored by the ETI, may be seen as sending a signal of a kind.

Even after the contact event is over, there may be policy decisions to be made, such as to retain or destroy the legacy of the event and/or the possibility of resuming contact.

Direct contact, even by means of slower-than-light travel, has often been described. Motivations include: migration; appropriation of resources or enslavement; extermination of a potential threat. Many plausible artefact discovery scenarios have been described.

Above all, we may underestimate the strangeness of what awaits us after a detection or contact. The cognitive challenge required to comprehend the activities and products of a non-human intelligence, not fitting our basic philosophical categories of either "natural" or "human-made", is not to be underestimated.

## 19.7 Conclusion

In the century since Wells and Lasswitz defined the modern concept of the alien, science fiction writers have speculated on the nature of ETI and the possibilities and implications of human contact with it. Since 1960, many of these works have been inspired by the SETI paradigm or offer alternatives to it. While SF is primarily fiction and is meant to entertain, the more thoughtful of such works may serve as a bank of thought experiments to assist in the development of future SETI strategies and policies.

Of course, genuine contact with ETI may be nothing like any of the dreams of SF. But until contact is made, as Dick noted (1996, p. 266), "Science can as yet add nothing to the question of the physical, mental, and moral nature of intelligence beyond the earth … For that, the speculations of science fiction … are as valid as anything science can suggest."

**Acknowledgements**

The author is grateful to Greg Bear, Gregory Benford and David Brin for comments on a draft of this chapter. The final version is based on the paper 'Imagining the Alien: The Portrayal of Extraterrestrial Intelligence and SETI in Science Fiction', previously published in *The Journal of the British Interplanetary Society,* **62**,131–138 (2009).

## References

Aldiss, B. 1982. *Helliconia Spring*. Atheneum, New York.

Aldiss, B. and Wingrove, D. 1986. *Trillion Year Spree: History of Science Fiction*. Gollancz, London.

Asher, N. 2001. *Gridlinked*. Tor Science Fiction, New York.

Asimov, I. 1972. *The Gods Themselves*. New York

Banks, I. M. 1987. *Consider Phlebas*. Macmillan, London.

Baxter, S. 1991–97. The Xeelee Sequence: *Raft, Flux, Timelike Infinity, Ring, Vacuum Diagrams*. London.

Baxter, S. 1998–2000. The Manifold Sequence: *Time, Space, Origin*. London.

Baxter, S. 2003–2006. Destiny's Children: *Coalescent, Exultant, Transcendent*. London..

Baxter, S. 2004. *The Hunters of Pangaea*. Nesfa Press, Boston.

Baxter, S. 2004. *Genetic Politics: Hive Minds in Science and Science Fiction*. Vector no. 235.

Baxter, S. 2008. Turing's Apples. In J. Strahan (Ed.) *Eclipse Two*. New York.

Baxter, S. 2008. Eagle Song. In P. Crowther (Ed.) *Postscripts 15*. Harrogate.

Bear, G. 1985. *Eon*. New York.

Bear, G. 1987. *The Forge of God*. New York.

Benford, G. 1977. *In the Ocean of Night*. New York.

Benford, G. 1999. *The Martian Race*. Aspect, New York.

Benford, G. 1999. Alien Technology. In B. Bova (Ed.) *Are We Alone in the Cosmos*. New York.

Benford, G. 2000. *Eater*. New York.

Benford, G. 2004. *Sunborn*. New York.

Blish, J. 1958. *A Case of Conscience*. New York.

Bracewell, R. 1975. *The Galactic Club: Intelligent Life in Outer Space*. San Francisco.

Bracewell, R. 1960. Communications from Superior Galactic Communities. *Nature* **186**, 670–71.

Bradbury, R. 1950. *The Martian Chronicles*. Bantam, New York.

Bradbury, R. 1996. The Man. In *S is for Space*. New York.

Brin, D. 1980. *Sundiver*. New York.

Burroughs, E.R. 1912. *A Princess of Mars*. New York. Penguin Classic Edition, 2007.

Butler, O. 1987. *Dawn: Xenogenesis I*. New York.

Card, O.S. 1985. *Ender's Game*. New York.

Cherryh, C. 1979. *Kesrith*. New York.

Clarke, A.C. 1954. *Childhood's End*. London.

Clarke, A.C. 1955. "*The Star"*. Infinity Science Fiction, New York.

Clarke, A.C. 1956. *The City and the Stars*. London.

Clarke, A.C. 1958. Out of the Sun. *If Magazine*, New York.

Clarke, A.C. 1968. *2001: A Space Odyssey*. London.

Clarke, A.C. 1971. 'A Meeting with Medusa'. *Playboy*, New York.

Clarke, A.C. 1973. *Rendezvous with Rama*. London.

Clarke, A.C. 1992. "Greetings, Carbon-based Bipeds!" *Life Magazine*, New York.

Clarke, A.C. and Baxter, S. 2004–2008. A Time Odyssey: *Time's Eye, Sunstorm, Firstborn*. New York.

Clement, H. 1954 *A Mission of Gravity*. Pyramid Publications, New York.

Clute, J. and Nicholls, P. 1993. *The Encyclopaedia of Science Fiction*. Orbit, London.

Cohen, J. and Stewart, I. 2000. *Wheelers*. New York.

Cohen, J. and Stewart, I. 2002. *Evolving the Alien: The Science of Extraterrestrial Life*. Ebury Press, London.

Dick, P.K. 1963. *The Game-Players of Titan*. New York.

Dick, S.J. 1996. *The Biological Universe: The Twentieth Centure Extraterrestrial Life Debate and the Limits of Science*. Cambridge University Press, Cambridge, UK.

Egan, G. 2008. *Incandescence*. London.

Fichman, F. 1990. *SETI.* New York.

Forward, R. 1980. *Dragon's Egg*. New York.

Gallun, R. 1934. Old Faithful. *Astounding Stories*, New York.

Gunn, J. 1972. *The Listeners*. New York.

Guthke, K. 1983. *Der Mythos der Neuzeit*. Berlin. Translated 1993 as *The Last Frontier*, Cornell University Press.

Haldeman, J. 1974. *The Forever War*. New York.

Heinlein, R. 1942. Goldfish Bowl. *Astounding Stories*, New York.

Heinlein, R. 1951. *The Puppet Masters*. New York.

Heinlein, R. 1959. *Starship Troopers*. New York.

Heinlein, R. 1961. *Stranger in a Strange Land*. New York.

Herbert, F. 1965. *Dune*. Chilton, New York.

Hoyle, F. 1957. *The Black Cloud*. London.

Hoyle, F. and Elliot, J. 1962. *A for Andromeda*. London.

Hoyle, F. and Hoyle G. 1973. *The Inferno*. London.

James, E. and Mendlesohn, F. 2003. *The Cambridge Companion to Science Fiction*. Cambridge University Press, Cambridge, UK.

King, S. 1987. *The Tommyknockers*. New York.

Kube-McDowell, M. 1985. *Emprise*. New York.

Lasswitz, K. 1897. *Auf Zwei Planeten*. Berlin. (Translated 1971 as *Two Planets,* Carbondale Southern Illinois University Press).

Leinster, M. 1945. First Contact. *Astounding Stories*, New York.

Lem, S. 1961. *Solaris*. Warsaw.

Lewis, C.S. 1938. *The Silent Planet*. London.

Lindsay, D. 1920. *A Voyage to Arcturus*. London.

McAuley, P. 2001. *The Secret of Life*. Tor Science, London.

McCollum, M. 1983.. *Life Probe*. New York.

McDevitt, J. 1986. *The Hercules Text*. New York.

McMaster, Bujold L. 1986. *Shards of Honor*. Guild American Books, New York.

Niven, L. 1970. Ringworld. New York.

Niven, L. and Pournelle, J. 1974. *The Mote in God's Eye*. New York.

Niven, L. and Pournelle, J. 1985. *Footfall.* New York.

Niven, L., Pournelle, J. and Barnes, S. 1987. *The Legacy of Heorot*. Simon and Schuster, New York.

Pellegrino, C. and Zebrowski, G. 1995. *The Killing Star*. New York.

Pohl, F. 1977. *Gateway*. New York.

Reed, R. 2000. *Marrow*. New York.

Roberts, A. 2006. *The History of Science Fiction*. Palgrave, London.

Russell, M.D. 1997. *The Sparrow*. New York.

Saberhagen, F. 1967. *Berserker*. New York.

Sagan, C. 1985. *Contact*. Simon & Schuster, New York.

Schroeder, K. 2003. *Permanence*. New York.

Shaw, B. 1975. *Orbitsville*. London.

Shklovskii, I. and Sagan, C. 1966. *Intelligent Life in the Universe*. New York.

Sheffield, C. 1998. *Aftermath*. New York.

Silverberg, R. 1970. *Downward to the Earth*. New York.

Silverberg, R. 1989. *The Queen of Springtime*. London.

Smith, E.E. 1928. *The Skylark of Space*. New York.

Stapledon, O. 1937. *Star Maker*. London.

van Vogt, A.E. 1950. *The Voyage of the Space Beagle*. Signet, New York.

Vinge, V. 1992. *A Fire Upon the Deep*. New York.

Watson, I. 1973. *The Embedding*. London.

Webb, S. 2002. *Where Is Everybody*. New York.

Weinbaum, S. 1934. 'A Martian Odyssey'. Wonder Stories, New York.

Wells, H.G. 1897. *The War of the Worlds*. London. Penguin Classics Edition, 2005.

Wells, H.G. 1901. The First Men in the Moon. London.

Wells, H.G. 1937. *Star Begotten: A Biological Fantasia*. London.

White, J. 1962. *Hospital Station*. New York.

Zerwick, C. and Brown, H. 1968. *The Cassiopeia Affair*. Doubleday, New York.

# 20

# What's Past is Prologue: Future Messages of Cosmic Evolution

Douglas A. Vakoch,
*Director of Interstellar Message Composition, SETI Institute, Mountain View, CA*

> … what's past is prologue, what to come
>
> In yours and my discharge.
>
> William Shakespeare, *The Tempest*, Act II, Scene I

Even before the first search for extraterrestrial intelligence (SETI) experiment was conducted, people have been pondering what reply we might send if some day we discover an extraterrestrial civilization. Some have suggested that the United Nations would be the international body of choice for deciding such a question and that would seem one appropriate starting point. The challenge that the international SETI community has faced is gaining a space on the already full agenda of the United Nations; indeed, the preface to the existing SETI protocols endorsed by the International Academy of Astronautics (IAA) and the International Institute of Space Law explicitly acknowledges the difficulty of gaining the attention of the United Nations.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_20, © Springer-Verlag Berlin Heidelberg 2011

If some day we detect direct evidence of extraterrestrial intelligence, all that may well change, but what are we do to in the meantime?

There is a natural alternative to the United Nations – a group whose discussions over the past decades already puts it in a position to recommend a coherent, consistent message that reflects broad-based, international consensus: the scientific community. To be clear, a solely scientific account of ourselves would not capture the depth and breadth of human experience. For precisely that reason, over the past several years the IAA through its Interstellar Message Construction Study Group, in conjunction with the SETI Institute, has organized a series of workshops and conferences bringing together scholars from a range of disciplines – including the arts, music, humanities, theology, and law – aimed at identifying some of the many voices that should be represented in a comprehensive reply from Earth.

But a reply message representing contemporary society would also surely include some of our scientific accounts of the world and of ourselves. And perhaps the most all-encompassing such story has evolution as its central theme.

## 20.1 Understanding our Origins

The view that the universe is in flux is an ancient one. Contemporary scientific understandings of evolution are multifaceted, aimed at understanding multiple transitions and developments (Swimme and Berry, 1992). How did heavy atoms originate from lighter ones? How did life arise from inert matter? How did consciousness and culture evolve from the biological world?

Our messages to other worlds might start by telling this evolutionary epic, and in the process, describe something about our own place in the universe. Indeed, we humans bear witness to the process of evolution in the very composition of our bodies. The calcium that gives solidity to our bones, the iron that lets our blood carry oxygen to our brains, the sodium and potassium that make possible the transmission of impulses along our nerves, all of these elements were formed inside a star that had its own birth and life and death, hurling its remains outward in a supernova explosion billions of years ago. As Steven Dick (2009, p. 25) summarizes the significance of this epic, "Cosmic evolution provides the proper universal context for biological evolution, revealing that the latter is only a small part of the bigger picture, in which everything is evolving, including life and culture."

Eric Chaisson has examined the evolutionary epic in several books, using a framework that has remained largely constant over a quarter of a century of his writing. From his *Cosmic Dawn: The Origins of Matter and Life* (1981) to his *Epic of Evolution: Seven Ages of the Cosmos* (2006), he has identified seven periods, with a recent version (Chaisson 2006) including the following epochs: particle, galactic, stellar, planetary, chemical, biological, and cultural. In Chaisson's (2006, p. xiii) view, this evolutionary account provides more than a scientifically accurate story:

> As sentient beings, we humans now reflect back on the matter of the Universe that gave us life. And what we find is a natural history, a universal history, a rich and abiding story of our origins that is nothing less than an epic of creation as understood by modern science—a coherent *weltgeschichte* that people of all cultures can adopt as currently true as truth can be.

A similar view is expressed by historian Cynthia Stokes Brown in her *Big History: From the Big Bang to the Present* [2007, p. xi], when she writes: "Within the last fifty years the scientific community has established a verifiable, and largely verified, account of the origins of our universe – of where we came from, how we got here, and where we may be going. This is a creation story for our time – for a world built on the discoveries of modern science, a world of jet travel, heart transplants, and the worldwide Internet."

At the beginning of the 21st century, scientific inquiry has become a primary means by which we attempt to understand the universe and our place within it. As a result, science naturally provides the foundation for evolution to become a central "myth" of our time.

To speak of evolution as being a myth – perhaps the preeminent scientific myth of the past century – does not refute the facts of evolution; to call evolution a myth is not to dismiss its scientific accuracy or adequacy. Instead, the term "myth" points toward the impact that this evolutionary perspective has on understanding our place in the cosmos. Barbara Sproul [1979, pp. 2–3] captures this meaning of myth in her description of traditional creation myths: "Not only are creation myths the most comprehensive of mythic statements, addressing themselves to the widest range of questions of meaning, but they are also the most profound. They deal with first causes, the essences of what their cultures perceive reality to be." Across a wide range of disciplines, from biology to cosmology and beyond, many scientists today attempt to perceive reality by studying cosmic evolution.

## 20.2 Essentials of Evolution

There are two distinctive features of evolutionary explanations of nature: they involve both change and historical embeddedness. Though there may be *constant* laws of nature,[1] they are manifested through *transformations* of the stuff of the universe: things change. Moreover, these changes build upon the past. The natural order we see today depends, at least in part, on historical circumstances. The dramatic changes in climate that led to the Cretaceous extinctions provided an opening for the proliferation of mammalian life in the Cenozoic era. We, as human beings, fundamentally embody both change and history. By characterizing ourselves in these terms in an interstellar message, we capture not only some of our fundamental biological attributes, but also some of the core dimensions of our contemporary self-understanding.

Of course, we might argue that to describe the origin and development of galactic structures, of planetary systems, of life, of civilization and technology, all under the generic name of evolution is to blur critical distinctions about the varied mechanisms responsible for such diverse phenomena. Admittedly, the processes of mutation, recombination, and natural selection in biological evolution are markedly different from the gravitational processes involved, or example, in the formation of planetary systems (Lupisella, 2009). Nevertheless, Eric Chaisson (2001, p. 214) defends the use of the term "evolution" for such varied processes:

> Given the powerful underlying phenomenon of change quite naturally everywhere, evolution itself should not be a disciplinary word exclusive to only one field of science, but rather an interdisciplinary word that helps connect often disparate fields of scientific scholarship … [N]eo-Darwinism, which has largely appropriated the term for itself, becomes but a special case (with powerful value-added features) within the much wider purview of cosmic evolution.

Perhaps the very fact that we do use the term "evolution" to describe such radically different processes is evidence of the mythic power that evolution has for organizing our self-understanding. But that may well change.

---

[1] To be sure, not all cosmologists would maintain that physical constants remain constant over time. Nevertheless, some constraints might be assumed. For example, Paul Davies, Tamara Davis, and Charles Lineweaver [2002] suggest that though there is evidence that the fine-structure constant may be increasing slowly over cosmological timescales, one might test which constants could be variable without violating the second law of thermodynamics.

## 20.3 Evolution's Importance, For Now

We should not expect this myth of evolution to remain as central to our self-understanding in future generations as it is now, an idea suggested by the continuation of Cynthia Stokes Brown's (2007, p. xi) above comment about the contemporary scientific account of our origins: "This is a creation story for our time – for a world built on the discoveries of modern science, a world of jet travel, heart transplants, and the worldwide Internet. *This world will not last forever* [emphasis added], but while it does, this is our story."

Brown's prediction that "[t]his world [as understood in evolutionary terms] will not last forever" is relevant for interstellar communication. Too often those who discuss interstellar communication take a simplistic view that somehow the message we send can reflect value-free scientific concepts; if we can only identify the "right" concepts, it is often assumed, we might communicate information that will be as obviously relevant to another technological civilization as it is to us. As Lupisella (2009) has argued, cosmic evolution itself may be value-laden. Indeed, we should be cautious in assuming that our messages – even messages about potentially widespread evolutionary processes – could ever objectively mirror the nature of reality in itself, independent of the culture from which our scientific understanding arises.

Nevertheless, we might well expect that evolutionary concepts – so central to our contemporary self-understanding – will also be embedded into the worldviews of scientifically literate extraterrestrials. After all, extraterrestrials too will have evolved in the same galaxy as we; if they are astronomically curious enough to seek out other civilizations, they would plausibly have come to understand the nature and history of our shared neighborhood in the universe.

Perhaps so. But this is quite different from holding an evolutionary perspective as especially *central* to their self-understanding. On the contrary, we might well expect that evolutionary processes will lose their central place in extraterrestrial self-understanding precisely because they become so commonplace: because they are taken for granted. Part of the impact of evolutionary accounts of our origins comes from the relative *novelty* of understanding ourselves, not as a fixed species in a static cosmos, but as a mutable species in an evolving universe.

Brian Swimme and Thomas Berry (1992, pp. 2 –3) contrast this sense of transformation inherent in an evolutionary perspective with the cyclical notion of time characteristic of earlier views:

> The most significant change in the twentieth century, it seems, is our passage from a sense of cosmos to a sense of cosmogenesis. From the beginning of human consciousness, the ever-renewing seasonal sequence, with its death

> and rebirth cycles, has impinged most powerfully upon human thought. This orientation in consciousness has characterized every previous human culture up to our own. During the modern period, and especially in the twentieth century, we have moved from that dominant spatial mode of consciousness, where time is experienced in ever-renewing seasonal cycles, to a dominant time-developmental mode of consciousness, where time is experienced as an evolutionary sequence of irreversible transformations.

In the same way that scientists continue to recognize seasonal sequences, so too we should not be surprised to find advanced extraterrestrials who remain aware of evolutionary principles, even though these beings may have ways of understanding the cosmos that they value even more. But the fact that evolution is important for humankind's self-understanding now could provide an important foundation for introducing ourselves to denizens of another world.

## 20.4 Evolutionary Voyages

What would an interstellar message with an evolutionary theme look like? For concrete examples of such messages that were already sent into space, we need look no further than the interstellar recording attached to two Voyager spacecraft, launched by NASA in 1977 – arguably the richest portrayal of life on Earth thus far intentionally sent into space.

In addition to over 100 images, greetings in 55 languages, and music from around the world, the message also included two components specifically intended to indicate the evolution of humankind. The first is a diagram of vertebrate evolution drawn by artist Jon Lomberg (1978). The images on the Voyager recording were presented in a manner that would enhance their intelligibility, drawing on a range of innovative approaches developed by Lomberg for communicating pictorial content to extraterrestrial intelligence (Lemarchand and Lomberg 2010, Lomberg, 1974, 1978). As an example, Lomberg sought to draw links between the various pictures, so his diagram of vertebrate evolution includes sketches of several animals featured in other photographs that appear on the recording. Though there is no absolute indication of progress in the diagram, we might infer an implicit indication that humankind is at the apex of the evolutionary process on Earth, given that the human couple is located at the top of the diagram.

The second evolutionary message included in the Voyager recording is a 12-minute sequence of selected sounds of Earth. In April 1977, Jon Lomberg drafted an extensive proposal for the sounds he thought should be included.

Among his recommendations was an emphasis on evolution (Lemarchand and Lomberg 2010):

> I propose that the sounds be ordered to reflect in a general way the evolution of life on the planet. After our opening, that identifies us, the sequence of sounds should move through natural sounds (mostly sounds), sounds of non-human life, sounds of human life, and sounds of society. Darwinian evolution may not operate [on] all worlds, but there is some chance that it does, and surely an arrangement that is ordered according to some possibly common principle has a better chance of being decoded than a random ordering.

Ann Druyan (1978, p. 153), who coordinated this part of the project, describes a similar rationale for ordering the sounds: "I felt that it would be most informative to arrange them chronologically. We took many liberties within that very broad structure, but the fundamental direction of the montage is evolutionary: from the geological through the biological into the technological."

The sound sequence begins with a musical rendition of part of Johannes Kepler's *Harmonica Mundi*, a 16th-century instantiation of the "music of the spheres." To evoke a sense of the geological activity in Earth's early history, next follow sounds of volcanoes, earthquakes, thunder, and mudpots. To highlight the centrality of water in life's development on our planet, next we hears sounds of a rainstorm – and again thunder – followed by crashing surf and the gentler lapping of ocean waves on a shore. As the sound of flowing water continues, the vocalizations of a sampling of Earth's varied lifeforms is added: first crickets and frogs, then a variety of birds. Then comes a selection of mammals: the distinctive voice of a hyena, elephants trumpeting, and a chimpanzee calling. The first sequence of non-human animal life ends with the sounds of a windy evening, punctuated by the plaintive howl of a wild dog.

To signal humankind's entrance into the world, next come footsteps intertwined with heartbeats, then laughter, underscored by the crackling of a fire, and the first sounds of language. Then we hear the development of stone tools, with the chip and click of flint on flint, followed by sounds of these tools used for scraping and for cleaving wood. Domestication is indicated by the barking of a dog, soon followed by the bleating of sheep. Increased sophistication of tool use is signaled by sounds of hammer against metal in a blacksmith's shop, the sawing of wood and hammering of nails, moving into the sounds of a riveter and tractor.

Advances in communication and transportation are intermixed, with Morse code overlaid on the sounds of a ship's fog horn, followed by progressively faster means of transportation – starting with a lengthy recording of horse and cart, moving on to fully mechanized locomotion, first through the characteristic chug and whistle of a train, on to the uncertain ignition of an internal combustion engine, ending with the flyby of an airplane and the *Saturn 5* engine launching humans toward the Moon.

The evolutionary sound montage ends with a sequence moving from human intimacy to technology and exploration, with a chaste kiss introducing the cry of a child and the succor of his mother, the crackle and fizz of a speeded-up recording of an electroencephalogram, and finally, the regular beat of pulsar CP1133, located some 600 light-years from Earth.

As Druyan (1978, p. 150) explains, listeners on different worlds might have quite different experiences: "The twelve-minute sound essay was conceived for two audiences: the human and the extraterrestrial. In the former, we hoped to evoke smiles of recognition, and in the latter, a sense of the variety of auditory experiences that are part of life on Earth."

## 20.5 Chemical Evolution

In an age when science is becoming increasingly specialized, attempts to find bridges between disciplines are rare, but not absent. One such infrequent but important example of looking for transdisciplinary connections is Stephen F. Mason's (1991) *Chemical Evolution: Origin of the Elements, Molecules, and Living Systems*. He begins by examining the historical context of 19th-century chemistry, then considers cosmic evolution in chemical terms, with topics ranging from stellar nucleosynthesis and the interstellar medium; to the evolution of the solar system and its planets, meteors, and comets; to the energetics of living systems. Mason (1991, p. viii) emphasizes the value of understanding our origins in chemical terms: "Surveys of the principal discoveries in the fields divergent from nineteenth-century chemical science, in cosmochemistry, geochemistry, biochemistry, and molecular biology, restore some coherence and provide a wider chemical view of the world, particularly when set in an evolutionary context."

*Maps of Time: An Introduction to Big History*, by David Christian (2004), expands the range of chemical phenomena relevant to an evolutionary understanding. Christian illustrates additional ways that chemistry can help tell about the origins and development of our world and our civilizations. At one level, we can describe the chemical evolution of the universe – in a cosmic scale, and as well as more locally – as providing the substrate from which life, and eventually intelligence, arose. But we can also describe the ongoing evolution of Earths' civilizations through artifacts created (Gräslund 1987, Heizer 1962), as well as environmental changes induced – all using the basic vocabulary of chemistry.

But how, precisely, might such chemical concepts be conveyed in interstellar messages?

## 20.6 Encoding Chemistry

To create a message that may intelligible to an independently evolved civilization, we attempt to identify basic principles that human and extraterrestrial civilizations are likely to share. The most frequently proposed set of universals is derived from mathematical and scientific principles. Why rely on such principles as a foundation for interstellar discourse? Because, it is typically argued, any civilization having a technology capable of contact at interstellar distances will also need to know some of the same fundamental principles of mathematics and science that humans know in order to construct this technology. To build a radio transmitter, for example, it seems reasonable that an extraterrestrial would need to know at least some basic math and science.

Among the most frequently proposed sets of universals are those related to principles of chemistry. As we consider the intentional messages that have already been sent from Earth, we see that chemical principles

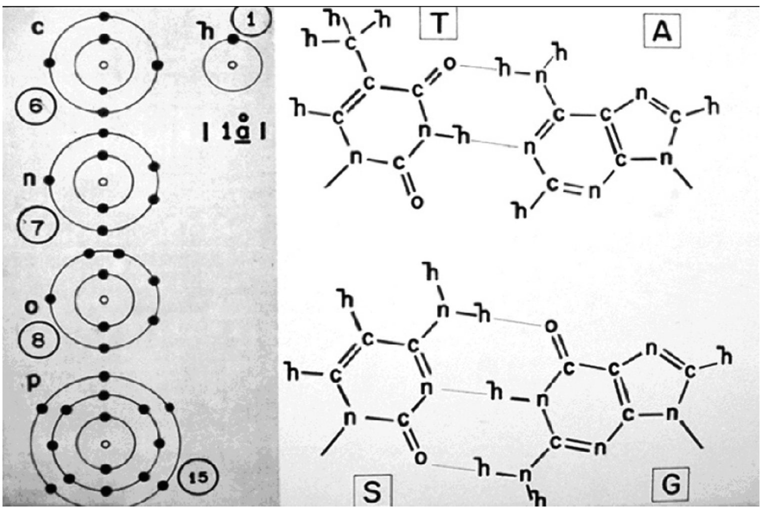



**Figure 20.1** Depictions of the atomic structure of several elements central to life on Earth, individually (on the left) and as found in deoxyribonucleic acid (DNA; on the right). From the *Voyager* Interstellar Recording. Credit: Frank D. Drake.

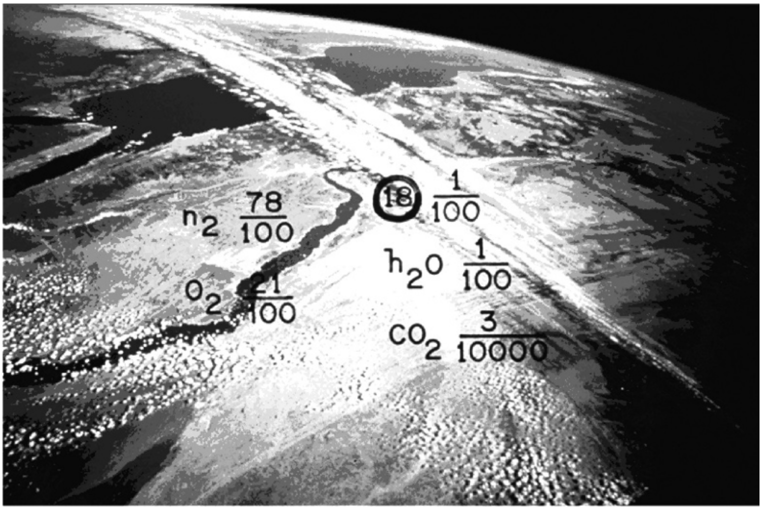



**Figure 20.2** Depicting the chemical composition of the Earth's atmosphere in terms of elements basic to life on Earth, introduced in Figure 20.1. From the *Voyager* Interstellar Recording. Credit: NASA.

are typically presupposed from the earliest stages. For example, the two *Pioneer* spacecraft, launched by NASA in 1972, include engraved plaques that depict the hyperfine transition of hydrogen, which then provides a unit of time and distance for other parts of the message (Sagan, Saltzman Sagan, and Drake, 1972). Similarly, the *Voyager* recording, mentioned earlier, contains schematic diagrams of atomic and molecular structures, with special emphasis on the structure of the DNA molecule as the biochemical foundation for life on Earth (Figure 20.1) (Drake, 1978). Likewise, the Earth's atmosphere is described in the Voyager recording in terms of its chemical composition (Figure 20.2).

A similar emphasis on chemical principles is found in proposed and actual interstellar messages that could be conveyed by radio broadcasts rather than by space probes. In 1961, following the first contemporary conference on communication with extraterrestrials, the meeting's organizer, Frank Drake, sent a message consisting of 551 ones and zeros to each of the participants (Drake and Sobel, 1992). His instructions indicated that, when properly formatted, the recipients would find a message from a hypothetical extraterrestrial civilization. Among the mathematical, scientific, and pictorial information included in this first message by Drake, there were schematic representations of two elements central to the biochemistry of this hypothetical civilization: oxygen and carbon. When Drake in 1974 transmitted an actual message from the world's largest radio telescope and radar facility, located near Arecibo, Puerto Rico, he began his message with a basic numbering system in binary digits, quickly followed by an identification of elements central to life on Earth, providing a numerical description of the structure of DNA in terms of its basic chemical constituents (Staff of the National Astronomy and Ionosphere Center, 1975).

While these early messages tend to combine mathematical, scientific, and pictorial information in the same message, Carl DeVito and Richard Oehrle (1990) propose messages that would give chemistry an even more central role. After introducing basic characteristics of set theory in their interstellar message, they describe the set that corresponds to the naturally occurring chemical elements, portrayed as a two-dimensional array that humans know as the Periodic Table of Elements. Moreover, this team of a mathematician and linguist describe chemical principles through a series of chemical reactions, introducing such concepts as volume, mass, and temperature.

Douglas Vakoch (1998c) emphasizes the value of starting with more direct representations of chemical elements and molecules. He suggests that we communicate concepts related to specific elements and molecules by transmitting signals at frequencies that mimic the emission spectra of the chemical constituents. For example, to communicate that hydrogren

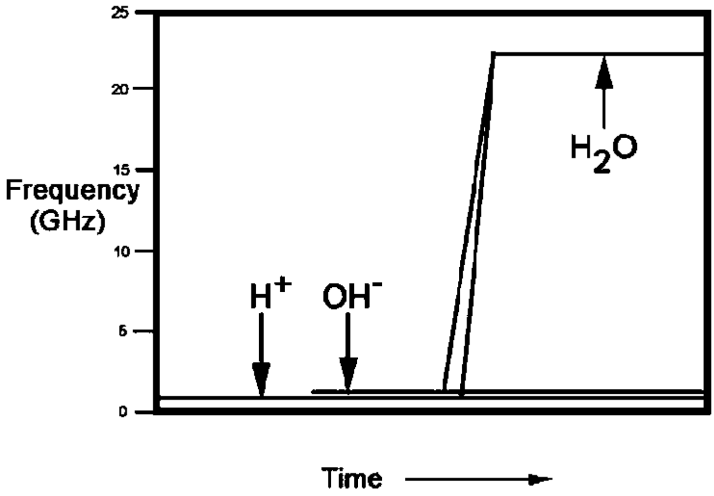



**Figure 20.3** An iconic approach to communicating that hydrogen and hydroxyl ions combine to form water. From Vakoch (2008a).

and a hydroxyl ion combine to form water, we might transmit signals at frequencies associated with the emission spectra of the reactants and the final product (see Figure 20.3). As with earlier proposals to use "magic frequencies" such as the hydrogen line as a base frequency for a search, we might transmit slightly to one side of these characteristic emission lines so our signals are not lost in naturally occurring background radiation. While Vakoch recognizes that such an approach lends itself especially well to only a circumscribed range of phenomena, he is concerned that it may be difficult to encode *any* information in a way that will be understandable to extraterrestrial intelligence. Once some information content – even a small amount – can be conveyed in any format, more general principles about the multiple formats used in other parts of the message might be conveyed through redundant encoding, providing a key to the multiplicity of ways that humans have of representing phenomena (Vakoch, 1998a).

For example, consider the multiple ways we have of describing chemical concepts. As we have just seen, we might transmit chemical information through two-dimensional images or through signals that mimic emission spectra. Alternatively, we might highlight the three-dimensional structure of molecules (Vakoch, 2000). If the recipients understand the connections between these multiple representations of the same or related concepts, we will have succeeded in introducing formats for describing phenomena that may be quite alien to extraterrestrials. In this case, we might communicate our ways of describing the three-dimensional structure of objects ranging in size from molecules to galaxies. Such a message would provide anchors for expressing other representations, using scientific objects that extraterrestrials may also have experience modeling. If we can find a shared means of representing objects ranging from the microscopic to the macroscopic – on scales ranging from angstroms (for molecules) to light-years (for galaxies) – we would be well-positioned to describe objects of intermediate size that extraterrestrials would never have seen before, such as the human body (Vakoch, 2000).

## 20.7 Encoding Culture

Having established a basic vocabulary of chemistry, an interstellar message could then go on to describe biological and even cultural phenomena, either based on or by analogy to such chemical processes. For example, given the evidence that individuals tend to be more altruistic toward closely related individuals, consistent with kin selection, we might iconically show the fate of individuals when threatened by predators – with closely related individuals (as shown by shared genetic material, described in chemical terms) more likely to survive than unrelated individuals (Vakoch, 2008a).

Alternatively, the same approach could be used to provide accounts of group selection, which Howard Bloom argues in *The Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century* is found throughout the evolutionary epic. Bloom (2000, p. 4) suggests that networking and cooperation are evident in evolutionary processes at multiple times and levels of complexity, from before the origins of life to the advent of the Internet and beyond: "Such a need to cooperate would have been necessary long ago to make a global brain and a planetary nervous system possible."

Given arguments for "memetic" rather than genetic transmission of cultural practices, we might imagine an extension of this approach to describe practices not apparently reducible simply to biology. For example, Susan Blackmore (1999, p. 167) proposes that a memetic version of altruism may underlie vegetarianism:

> I suggest that vegetarianism succeeds as a meme because we all want to be like the *nice* people who care about animals, and we copy them. Not everyone will get infected by this meme; some like meat too much and others have sets of memes that are not very compatible with this one. Nevertheless, it does quite well. Vegetarianism is a memetically spread altruistic fashion.

We might also use the vocabulary of chemistry to explain to extraterrestrial intelligence some of the consequences of our cultural evolution about which we are less proud (Vakoch, 2007). After describing the evolution of the Earth's atmosphere, hydrosphere, and lithosphere, as they have occurred largely independently of major human intervention, we might then describe how the Industrial Revolution caused such significant changes to the environment.

To restrict our descriptions of terrestrial culture to its chemical manifestations would however, be unnecessarily restrictive. As Lupisella (2009, p. 322) notes, it can be "helpful to think about culture as the *collective manifestation of value* – where value is that which is valuable to 'sufficiently complex' agents, from which meaning, purpose, ethics, and aesthetics can be derived." As we have seen, we might begin to communicate the biological underpinnings of altruism in terms of the chemical basis of our genetics. But even if we restrict ourselves to the broad category of altruistic acts, multiple mechanisms have been proposed, each with its own explanatory framework. For example, notions of reciprocal altruism might be encoded into interstellar messages using game theory (Vakoch, 2001, 2002), providing a mathematical expression of concepts related to fairness (Fehr and Gächter 2000), a concept within the purview of ethicists. Similarly, if we can communicate some basic numerical concepts in interstellar messages, we may have the foundation needed to begin expressing some aspects of the human aesthetic experience (Vakoch, 2004a, 2004b).

## 20.8 What Can We Offer?

SETI scientists typically assume that extraterrestrial civilizations are much longer lived than terrestrial civilization. That is, the average lifetime of extraterrestrial civilizations, as measured in the time they are actively seeking to make contact with other civilizations, is assumed to be much longer than the time that humans have had the technology and motivation to communicate at interstellar distances. Without this assumption, it is statistically improbable that extraterrestrial and human civilizations will exist sufficiently close enough to one another in time and space to make contact. If we make contact at all, we can expect to be the junior partner in the conversation.[2]

This presumed asymmetry in the lifetimes of extraterrestrial and terrestrial civilizations raises the question, "What would humans have to say that would be of interest to much older civilizations?" Typically it has been assumed that more long-lived civilizations will also be more technologically and scientifically advanced. If so, then humans are unlikely to be able to teach extraterrestrial civilizations much in these realms, at least assuming that there is a convergence of technological developments and scientific discoveries across civilizations, with more advanced civilizations attaining an understanding that encompasses and surpasses that of less advanced civilizations.

Following the above line of reasoning, even if humankind is much younger than extraterrestrial civilizations, we may nevertheless be in possession of information that could be of significant scientific interest to intelligent beings on other worlds: information about the longevity of our own civilization, as well as factors that threaten our continued existence as a species.

As we attempt to assess the likelihood that SETI will succeed, one of the most elusive variables to quantify is the lifetime of technological civilizations. By beginning a serious program in active SETI – transmitting evidence of our existence to other civilizations – we could provide at least one data point to scientists on other worlds attempting to make this same estimate

[2] For a discussion of another possibility—that the burden of transmitting may lie with the less advanced civilization—see Vakoch's [2005] "Expanding Human Presence beyond the Solar System through Active SETI: On the Prerequisites for Legal Relations with Extraterrestrial Intelligence."

of the lifetime of civilizations with both the capacity and the willingness to make their existence known to other forms of intelligence.

Although more advanced civilizations may be able to glean some information about the threats to our survival as a species by monitoring atmospheric changes and unintentional leakage radiation from Earth, intentional messages describing the social, political, and ecological factors that contribute to the instability of our planet may provide a rare glimpse into the cultures of a young civilization that has some insight into the threats it faces. Whether or not we continue such transmissions over the millennia would be informative to sociologists and psychologists beyond Earth, potentially providing greater insights into the critical years during which civilizations attempt to make the transition to becoming long-lived civilizations themselves. Whether or not humankind succeeds, such messages from Earth could be useful to extraterrestrials attempting to understand better the factors that contribute to the lifetimes of other civilizations. [3]

## 20.9 Acting on the Environment

In a strikingly different approach to communicating to extraterrestrials the environmental challenges facing contemporary humankind, we might focus not only on the physical manifestations of environmental problems, but also on the humans who contribute to these problems. One step toward creating a language to describe human behavior was proposed by Vakoch (2006a), who noted that scientific explanations of human behavior typically have significantly more limited predictive ability than physicists and chemists are used to. Rather than being able to identify with great precision the antecedents and consequences of any particular person's behavior, psychologists are typically content to predict the behaviors of *groups* of individuals, even when such predictions account for only a modest amount of the total variance in the behaviors between individuals. While physicists may achieve considerable accuracy in determining the trajectory of a billiard ball of particular mass when it is hit at a specific point with a specific force, psychologists must typically remain content to provide probabilistic accounts of human behavior.

Even in the rare interstellar messages that have addressed human behavior, differences between individuals have typically been neglected. As an example, Freudenthal (1960) devotes one section of his book *Lingua Cosmica*, or

[3] For additional ways that older, more technologically civilizations may benefit from learning about humanity, see Vakoch's [2006b] "To Err is Human … and of Interest to ET?" and [2007] "A Shadow of Ourselves."

*LINCOS*, to an examination of human behavior. However, as he describes the actions of humans under a variety of contexts, he makes no attempt to provide a consistent account of the behaviors of the particular individual identified in his "mini-plays" as Human A, for example. That is, Human A may act in one scenario in a manner that is completely inconsistent with the same person in a different scenario. Such an arbitrary pairing of names of individuals with the actions of particular actors captures some of the variety of human behaviors. But it fails to show that any *particular* individual may have stable dispositions to act in certain ways across situations. In short, Freudenthal's approach fails to describe stable personality characteristics, sometimes known as traits.

Contemporary psychological research has shown, however, that personality plays an important role in determining differences between actors. For example, Fraj and Martinez (2006) examined the relationship between environmental behavior and the five personality factors identified by Costa and McRae (1992): neuroticism, extroversion, openness, conscientiousness, and agreeableness. Fraj and Martinez (2006) found that conscientious individuals were more likely to purchase ecological products or to switch products for ecological reasons. On the other hand, extroverted and agreeable people were more likely either to join an environmental group or to attend an ecological conference. In both cases, their findings readily translate into interstellar messages that show individuals interacting in a range of contexts. As might be expected, individuals who seek out and enjoy being with others are more likely to engage in environmental concerns in settings that require such cooperative, extroverted action, and dispositionally conscientious individuals are likely to be conscientious about their buying patterns. Such probabilistic accounts may provide a foundation for communicating the correlates and even causes of behaviors related to environmental changes.

## 20.10 A Message to Terrestrial Intelligence

Typically, we imagine the benefits of interstellar communication in terms of what we might gain *from extraterrestrial intelligence* as a result of such an exchange. But we might also consider ways we might benefit by transmitting messages, even if we never receive a reply. What might we gain, for example, by grappling with the challenges of describing our understanding of our place in the universe in terms of the epic of evolution – whether or not anyone beyond Earth ever hears us?

In an interstellar message that describes life on Earth in terms of its diversity over the ages, we would be compelled to describe the cataclysmic changes that have occurred periodically throughout the history of our planet – changes such as those that marked the transition from the Mesozoic Era to the Cenozoic Era. In addition, we would be invited to reflect on our own role in creating cataclysms of comparable scale.

Some have argued that we are in the midst of another great extinction – this one due to human intervention (e.g., Swimme and Berry, 1992). As we have expanded our scientific understanding of and mastery over the physical world, we have also significantly drawn upon the Earth's limited resources, we have taxed the terrestrial environment with byproducts of our industrial progress, and we have reduced the Earth's biodiversity. Perhaps the strongest argument for not undertaking a serious program of active SETI – in which we would transmit messages *de novo*, rather than merely listening for messages from other intelligence – is that we do not expect humankind to survive long enough to receive a reply.

Such a program of transmitting messages to other civilizations, however, would make a strong statement here on Earth – a statement that we do expect to be around hundreds or thousands of years from now to receive a reply. And even if we are not – that is, either *not around*, or simply *not listening* any longer, such an experiment could be of significant value to SETI scientists living around distant stars. Indeed, a transmission project undertaken for the benefit of extraterrestrial civilizations would be consistent with a "cosmocentric ethic" (Lupisella and Logsdon, 1997), which may provide an ethical foundation for future transmissions from Earth (Vakoch, 2005).

Although the focus of SETI is on making contact with intelligence beyond Earth, the exercise of portraying ourselves in interstellar messages provides us with an opportunity to cultivate greater intelligence on our own planet. Few things are more critical to approach more intelligently than the environmental problems that threaten the very existence of human civilization as we know it. Even considering only the climate changes we can anticipate due to greenhouse gas emissions, in the coming decades we should expect environmental effects such as extreme weather events, rising sea levels, and environmental degradation, as well as threats to health due to thermal stress, microbial proliferation, changes in infectious diseases, diminished food sources, and increased poverty (McMichael, Woodruff, and Hales, 2006).

By focusing interstellar messages on the ecological challenges we face, we provide a forum for discussing critical issues in a way that is both concrete yet not excessively aversive. The challenge of taking environmental issues seriously is that we need to find a way to make issues like global warming seem sufficiently immediate to attend to, without evoking such strong negative emotional reactions that people avoid the discussions altogether (Lorenzoni et al., 2006).

## 20.11 Agreeing to Disagree

Since the 1980s, a protocol developed by the IAA SETI Committee, in consultation with the International Institute of Space Law, provides guidance about appropriate actions following the detection of an extraterrestrial civilization. This protocol recommends that any response from humankind to a signal from extraterrestrials should represent a consensus. Differences of opinion, in contrast, should be minimized in interstellar messages according to this guideline.

A markedly different approach is proposed in the Dialogic Model (Vakoch, 1998b), which advocates transmitting messages that highlight different perspectives, in an attempt to reflect the reality of the current human condition in which there are significant differences of viewpoint between groups and even between individuals within relatively homogeneous groups. Vakoch (1998b) argues that to minimize such differences would neglect some of the most important information that humankind could convey: the diversity of our views. As Lupisella (2009) notes:

> Cultural diversity, and perhaps diversity in general, may have practical benefits (e.g. having a wide variety to choose from as needed), but diversity may be a value in its own right, an end unto itself—worth pursuing for its own sake. Given the potential for quite diverse life forms throughout the universe, diversity may have broad cosmic significance beyond our own aesthetic appreciation.

In any interstellar message that would attempt to describe the environmental challenges that humankind faces in order to survive in the coming decades, alternative perspectives must be acknowledged. Not only do multiple accounts of humankind's role in the current environmental situation portray a diversity of views in contemporary society, but openly discussing these differing accounts may also have a salutary effect by providing a forum for the ongoing dialogue between individuals and groups with divergent perspectives. As we consider the potential value of transmitting messages to extraterrestrial intelligence, we should remember the value of the *process* of deciding on the content of an interstellar message, regardless of whether that message is ever received by an extraterrestrial civilization, or even whether it is ever transmitted.

## 20.12 The Evolution of the Evolutionary Epic

But in an interstellar message based on the evolutionary epic, how much of the content should reflect evolution on a galactic, stellar, or planetary scale, and how much should reflect the idiosyncrasies of our planet's biological and cultural histories? Both are important, but for different reasons.

As we convey physical accounts about, say, the mechanics of galactic structure and the dynamics of planetary formation, we have an opportunity to make a link between basic principles of mathematics and physics and an external reality shared by humans and extraterrestrials. Even if humans and extraterrestrials have a common commitment to modeling ever more accurately the nature of physical reality, there is no guarantee, however, that these models of reality will necessarily be obviously commensurable (Vakoch, 1998a). Peter Barker (1982) has articulated the challenges of terrestrial scientists from different times and cultures understanding one another; how much more difficult might it be for terrestrial and extraterrestrial scientists to understand one another, given they have evolved in different environments? The differing evolutionary histories of independently evolved species may indeed affect the goals that scientists pursue on different worlds. As philosopher Nicholas Rescher (1985) argues, an aquatic intelligence may have a very sophisticated science of hydrodynamics, because its survival and flourishing depends on it. But it may be lacking in some concepts fundamental to land-based civilizations.

This view of scientific progress contrasts with a standard view of linear progress typically assumed – often implicitly – in SETI circles (Vakoch, 2008b). In this standard view, more advanced civilizations have passed through the same stages as less advanced civilizations on other worlds. If more advanced civilizations want to make themselves understood, it is argued, they will start with the principles that would surely be understood by less advanced civilizations. But, the skeptic might ask, is it so obvious which principles those would be, and even if the principles are widely known, is the conceptual apparatus for describing these principles universal?

Perhaps an analogy of mountain climbing will help clarify the issue. Science progresses, we might argue, in the same way that a mountain climber progresses toward the peak of a mountain.[4] Not all climbers will progress as far; novice climbers may only make it part way up the mountain. But as these neophytes become more skilled, they will be able to progress to greater altitudes, all the while pointed toward their goal: the highest point of the mountain.

In this analogy, the scientist is akin to the climber, progressing step-by-step toward ever clearer understanding of the nature of reality as it really is, symbolized by the mountaintop. There may, indeed, by times when the scientist/climber diverts from the path, but in the long run, the interplay of theory and experiment ensures that the successful scientist – the one who

[4] For a similar analysis, see Lupisella's [n.d.] "Increasing Verisimilitude as the Goal of Science."

makes progress in ascending the mountain – will find the right path. A more sophisticated scientist/climber, having ascended higher, could look back and even leave pointers for the less experienced scientist/climber, potentially providing clues that might speed up the ascent of the less experienced.

As we apply this metaphor to interstellar communication, we assume – by necessity – that we are the less experienced climber. In the thirteen-plus billion year history of our galaxy, on purely statistical grounds, it is highly improbable that any civilization we contact by radio signals or brief laser pulses will be as technologically youthful as we are. Less advanced civilizations will not have the capacity to communicate at interstellar distances. And if the typical age of an extraterrestrial civilization is as short as ours, then the number of civilizations that exist at any one time will be very small – and the few that exist will be located far from one another.

But to continue the analogy, what if we and the extraterrestrial scientist/climber ascend different sides of the mountain? How then could the more advanced civilization point the way up a path it did not take? Or even more pessimistically, what if the human and extraterrestrial scientists/climbers are ascending different mountains – both getting continually closer to their respective goals of increasingly comprehensive understanding of the universe, but each headed toward a different mountain peak, providing a perspective on a different aspect of the universe? If Rescher (1985) is right, then science may take varied forms on varied worlds.

While the possibility of multiple directions of progress does little to reassure us of easy interspecies communication, it does open the possibility of learning much, if ever we do establish contact. Indeed, the possible plurality of sciences on different worlds may provide a sense of reassurance that even a civilization as young as ours might contribute substantially in an interstellar exchange. Even beyond accounts of the challenges we face simply to survive, our scientific and cultural accomplishments could be of considerable interest on other worlds. By Steven Dick's (2009) analysis of the history of ideas about cosmic evolution, even on Earth there was no inherent necessity that scientists would come to understand the universe in specifically evolutionary terms: "The humble and sporadic origins of the idea of cosmic evolution demonstrate that it did not have to become what is now the leading overarching principle of twentieth century astronomy." If in fact there is not one single path of scientific progress taken by all civilizations, but different paths depending on each species' idiosyncratic environment as well as its unique evolutionary and cultural histories, then even our relatively primitive accounts of the universe may provide novel insights to extraterrestrials.

Just as the anthropologists and historians of Earth are interested in the development of other cultures' ways of understanding the world about them, so too might extraterrestrial intelligence be interested in the specific trajectory that our elaboration of the evolutionary epic has taken. Though we tend to value our most recent scientific understanding most highly, assuming this most accurately reflects the nature of reality, historians of science on another world may not be especially interested in learning about the models most widely accepted in the early 21st century. Instead, extraterrestrial historians may be more intrigued by the entire history of the idea of cosmic evolution (Dick, 2009) or the ideas of particular scientists whose models have in some ways now been superceded, for example, those of Harlow Shapley (Palmeri, 2009). If, as many have argued, the most readily comprehensible parts of an interstellar message are those addressing scientific topics, then a history of human theories of cosmic evolution may provide one of the most accessible ways to introduce other civilizations to terrestrial historical and cultural concepts.

## 20.13 “What’s Past is Prologue”

When William Shakespeare (1604/1999) wrote in *The Tempest* that “what’s past is prologue, what to come / In yours and my discharge,” he reminds us that history often sets the stage for our present actions. His observation could well serve as counsel for the content of our first interstellar messages to another civilization. What could be more fitting to introduce ourselves in an exchange that could continue for millennia than to recount our own evolutionary history up to the present, along with our hope that our actions in sending such a message might yield a reply that will be heard by future generations of humankind?

We might, of course, describe our evolutionary origins in other than chemical terms, which was the focus of this chapter. We might describe the dynamics of galactic formation, for example, with some of the same basic concepts of physics through which we can analyze the evolution of locomotion in terrestrial life (e.g., Radinsky, 1987). Astronomers on other worlds, we might argue, would be as likely to share basic principles of physics with humans as they are likely to have concepts of chemistry in common. Indeed, Freudenthal’s (1960) interstellar language *LINCOS* – perhaps the most sophisticated language for cosmic discourse yet developed on Earth – gives concepts from physics a central place.

Regardless of whether we choose a language based on principles of chemistry, physics, or something else, as we ponder what we might say in transmissions to other worlds – should we choose some day to transmit evidence of our existence in a serious fashion – it would be very fitting if our messages reflected some of the very processes of the universe that ultimately led to the origin and evolution of ourselves as a species attempting to make contact with other worlds.

Yes, we humans are more than merely biological creatures. We appreciate beauty, we struggle with ethical conflicts, and we strive to make sense of our purpose in the universe, asking questions that science cannot answer. And yet, our sense of aesthetics, our moral sensibilities, and our search for meaning may themselves be intricately connected to the fabric of the cosmos (Lupisella, 2009). It would seem fitting, then, if our first exchange with sentient beings on other worlds started by explaining that we too recognize our origins in the early universe when hydrogen and helium were created; that our life's breath requires the oxygen first released from Earth's oceans some two billion years ago; and that as we have learned to trace the history of the elements that make up our bodies and that give rise to our consciousness, we have discovered an evolutionary creation myth that helps us start to understand our place in the cosmos. And that they, the recipients of this message, living on a distant planet, may well be interested in hearing it.

**Acknowledgement**

This chapter is adapted from Douglas A. Vakoch's (2009) Encoding Our Origins: Communicating the Evolutionary Epic in Interstellar Messages. In S.J. Dick and M. Lupisella (Eds) *Cosmos and Culture: Cultural Evolution in a Cosmic Context*. Washington, DC: NASA History Series, 415-439.

## References

Barker, P. 1982. Omilinguals. In N.D. Smith (Ed.) *Philosophers Look at Science Fiction*. Chicago: Nelson-Hall, 75 –85.

Blackmore, S. 1999. *The Meme Machine*. Oxford: Oxford University Press.

Bloom, H. 2000. *Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century*. New York: John Wiley & Sons Inc.

Brown, C.S. 2007. *Big History: From the Big Bang to the Present*. New York: The New Press.

Chaisson, E. 1981. *Cosmic Dawn: The Origins of Matter and Life*. Boston: Little, Brown and Company.

Chaisson, E. 2001. *Cosmic Evolution: The Rise of Complexity in Nature*. Cambridge: Harvard University Press.

Chaisson, E. 2006. *Epic of Evolution: Seven Ages of the Cosmos*. New York: Columbia University Press.

Christian, D. 2004. *Maps of Time: An Introduction to Big History*. Berkeley: University of California Press.

Costa, P.T., Jr., and McCrae, R.R. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Davies, P.C.W.,. Davis, T.M. and Lineweaver, C.H. 2002. Cosmology: Black Holes Constrain Varying Constants. *Nature* **418**, 602 –603.

DeVito, C.L. and Oehrle, R.T. 1990. A Language Based on the Fundamental Facts of Science. *Journal of the British Interplanetary Society* **43**, 561 – 568.

Dick, S. J. 2009. "Cosmic Evolution: History, Culture, and Human Destiny". In S.J. Dick, and M. Lupisella (Eds) *Cosmos and Culture: Cultural Evolution in a Cosmic Context*. Washington, DC: NASA History Series, 25 –59.

Drake, F. 1978. The Foundations of the Voyager Record. In C. Sagan (Ed.) *Murmurs of Earth: The Voyager Interstellar Record*. New York: Random House, 45 –70.

Drake, F. and Sobel, D. 1992. *Is Anyone Out There?: The Scientific Search for Extraterrestrial Intelligence*. New York: Delacorte.

Druyan, A. 1978. The Sounds of Earth. In C. Sagan (ed.) *Murmurs of Earth: The Voyager Interstellar Record*. New York: Random House, 149 –160.

Fehr, E. and S. Gächter (2000) "Fairness and Retaliation—The Economics of Reciprocity", *American Economic Review*, (90), pp. 980-994.

Fraj, E. and Martinez, E. 2006. Influence of Personality on Ecological Consumer Behaviour, *Journal of Consumer Behaviour* **5**, 167 –181.

Freudenthal, H. 1960. *LINCOS: Design of a Language for Cosmic Intercourse. Part I*. Amsterdam: North-Holland.

Gräslund, B. 1987. *The Birth of Prehistoric Chronology: Dating Methods and Dating Systems in Nineteenth-century Scandinavian Archaeology*. Cambridge: Cambridge University Press.

Heizer, R.F. 1962. The Background of Thomsen's Three-age System. *Technology and Culture* **3:3**, 259 –266.

Lemarchand, G.A. and Lomberg, J. 2010. Communication among Interstellar Intelligent Species: A Search for Universal Cognitive Maps. In D.A, Vakoch (Ed.) *Communication with Extraterrestrial Intelligence*. Albany, NY: State University of New York Press.

Lomberg, J. 1974. *Some Thoughts on Art, Extraterrestrials, and the Nature of Beauty*. Mimeo publication by the National Air and Space Museum, Washington, D.C.: Smithsonian Institution.

Lomberg, J. 1978. Pictures of Earth. In C. Sagan (Ed.) *Murmurs of Earth: The Voyager Interstellar Record*. New York: Random House, pp. 71 –121.

Lorenzoni, I. et al. 2006. Cross-National Comparisons of Image Associations with 'Global Warming' and 'Climate Change' among Laypeople in the United States of America and Great Britain. *Journal of Risk Research* **9:3**, 265 –281.

Lupisella, M. (n.d.) "Increasing Verisimilitude as the Goal of Science," Unpublished manuscript.

Lupisella, M. 2009. Cosmocultural Evolution: The Co-evolution of Culture and Cosmos and the Creation of Cosmic Value. In S.J. Dick, and M. Lupisella (Eds) *Cosmos and Culture: Cultural Evolution in a Cosmic Context*. Washington, DC: NASA History Series, 321 –359.

Lupisella, M. and Logsdon, J. 1997. Do We Need a Cosmocentric Ethic? Paper presented at the 48th International Astronautical Congress, Turin, Italy.

Mason, S.F. 1991. *Chemical Evolution: Origin of the Elements, Molecules, and Living Systems*. Oxford: Clarendon Press.

McMichael, A. J., Woodruff, R.E. and Hales, S. 2006. Climate Change and Human Health: Present and Future Risks. *The Lancet* **367**, 859 –869.

Palmeri, J. 2009. Bringing Cosmos to Culture: Harlow Shapley and the Uses of Cosmic Evolution. In S.J. Dick, and M. Lupisella (Eds) *Cosmos and Culture: Cultural Evolution in a Cosmic Context*. Washington, DC: NASA History Series.

Radinsky, L.B. 1987. *The Evolution of Vertebrate Design*. Chicago: The University of Chicago Press.

Rescher, N. 1985. Extraterrestrial Science. In E. Regis, Jr. (Ed) *Extraterrestrials: Science and Alien Intelligence*. Cambridge: Cambridge University Press, 83 –116.

Sagan, C., Saltzman Sagan, L. and Drake, F. 1972. A Message from Earth. *Science* **175**, 881 –884.

Shakespeare, W. 1604/1999. The Tempest. Ed. Peter Holland, New York: Penguin.

Sproul, B.C. 1979. *Primal Myths: Creating the World*. New York, Harper & Row.

Staff of the National Astronomy and Ionosphere Center. 1975. The Arecibo Message of November 1974. *Icarus* **26**, 462 –466.

Swimme, B. and Berry, T. 1992. *From the Primordial Flaring Forth to the Ecozoic Era—A Celebration of the Unfolding of the Cosmos*. New York, HarperCollins.

Vakoch, D.A. 1998a. Constructing Messages to Extraterrestrials: An Exosemiotic Perspective. *Acta Astronautica* **42**, 697 –704.

Vakoch, D.A. 1998b. The Dialogic Model: Representing Human Diversity in Messages to Extraterrestrials. *Acta Astronautica*, 42, 705 –710.

Vakoch, D.A. 1998c. Signs of Life beyond Earth: A Semiotic Analysis of Interstellar Messages. *Leonardo* **31**, 313 –319.

Vakoch, D.A. 2000. Three-dimensional Messages for Interstellar Communication. In G.A. Lemarchand and K. Meech (Eds) *Bioastronomy 99: A New Era in Bioastronomy*. San Francisco: Astronomical Society of the Pacific, 623 –627.

Vakoch, D.A. 2001. Altruism as the Key to Interstellar Communication. *Research News & Opportunities in Science and Theology* **2:3**, 2, 16.

Vakoch, D.A. 2002. Encoding Altruism. *Galileo*, **52**, 30 –36.

Vakoch, D.A. 2003. "Le Sacrifice de Soi: Une Composition Interstellaire. *Anomalie Digital_Arts* **4**, 201 –202.

Vakoch, D.A. 2004a. The Art and Science of Interstellar Message Composition. *Leonardo* **37**, 33 –34.

Vakoch, D.A. 2004b. To the Stars, Silently. *Leonardo* **37**, 265.

Vakoch, D.A. 2005. Expanding Human Presence beyond the Solar System through Active SETI: On the Prerequisites for Legal Relations with Extraterrestrial Intelligence. *Proceedings of the Forty-eighth Colloquium on the Law of Outer Space: 17-21 October 2005, Fukuoka, Japan* **48**, 149 –155.

Vakoch, D.A. 2006a. Describing the Probabilistic Nature of Human Behavior in Interstellar Messages. Paper presented at the SETI II Interdisciplinary Connections Review Meeting, 57th International Astronautical Congress, Valencia, Spain.

Vakoch, D. 2006b. To Err is Human … and of Interest to ET?" *Explorer* **3:2**, 16 –17.

Vakoch, D. 2007. A Shadow of Ourselves. *Explorer* **4:4**, 16 –17.

Vakoch, D.A. 2008a. Representing Culture in Interstellar Messages. *Acta Astronautica* **63**, 657 –664.

Vakoch, D. 2008b. Technology as a Manifestation of Intelligence: Does Shared Technology Imply Shared Science and Mathematics? *Astrobiology* **8:2**, 391.

# 21

# METI: Messaging to ExtraTerrestrial Intelligence

Alexander L. Zaitsev,
*Chief Scientist, Institute of Radio Engineering and Electronics, Russian Academy of Sciences*

Perhaps, after 50 years of listening to nothing but cosmic static, it is time to recognize that the time has come for humankind to take the lead in helping to end the Great Silence. Could it be that the future of SETI lies not in receiving, but rather in transmitting? In this chapter we introduce Messaging to Extra-Terrestrial Intelligence (METI) as a complementary science to SETI observations.

METI represents a cardinally new kind of human activity. Some argue that the Search for Extra-Terrestrial Intelligence (SETI) also is cardinally new. Yes, it is new, but not cardinally so, because people have always passively surveyed the heavens in the hope of detecting something unknown, whether natural or artificial. However, a purposeful effort directed toward converting terrestrial civilizations into the object of detection by possible extraterrestrial civilizations, which is the focus of METI, is a substantially new activity.

The scientific program known as SETI endeavors as its main goal to search for any kind of electromagnetic radiation *from* aliens. In contrast, METI's main goal is to create and to send intelligent messages from humans *to* aliens. SETI scientists sometimes ask whether Active SETI (as METI is sometimes called) makes sense. Would it be reasonable, in the context of ensuring SETI success, to transmit messages with the express intention of attracting ETI's attention, thus eliciting a response?

Although this question is a valid one, the overall goal of METI is much broader: to overcome the Great Silence in the Universe by conveying to

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_21, 

ETI the long-awaited news: "You are not alone!" Indeed, since a basic tenet of SETI science is the tacit assumption that civilizations transmitting interstellar messages do indeed exist, the scientists who are involved in SETI should unavoidably accept that messaging to ETI is a reasonable and fully complementary activity.

## 21.1 Brief History of METI

SETI scientists have now been listening to the cosmos for 50 years, hoping to detect artificial signals in the Universe. Unfortunately, their efforts to date have failed to produce positive results. There are a number of possible reasons for this, which have been discussed and analyzed in previous chapters of this book. As compared to SETI, METI is in an advantageous position. Indeed, in the act of composing a dedicated message and sending it to a properly selected star in our galaxy, METI scientists have created a tangible result. One can consider the launching of intelligent radio message into space as the first stone set in building **a** *radio bridge* between terrestrial and reputed extraterrestrial civilizations. Once this stone is in place, establishing contact depends only upon whether *they* will discover our Message and send a recognizable response.

Both the first interstellar messages and the first experiments on search for aliens' signals are associated with the name of Frank Drake. In 1972 Drake, along with Carl Sagan and others, developed the «Pioneer Plaques» [1] and affixed them to two interplanetary space probes. Then, in 1977, he and his colleagues produced two «Voyager Golden Records»[2], disks which were placed on two spacecraft which ultimately flew outside of our Solar system (see Figure 21.1).

The Arecibo Message, the first deliberate interstellar radio message, was also created by Drake and Sagan. It was transmitted on November 16, 1974 using a radar telescope with an antenna diameter of 1000 feet (305 m) and a transmitter with a mean power of 500 kW at a wavelength 12.6 cm. Radio messages of four other projects, Cosmic Call 1 (1999), Teen Age Message (2001), Cosmic Call 2 (2003) and A Message From Earth (2008), were later transmitted into space using the Evpatoria Planetary Radar (Figure 21.2).

Thus, during the entire history of terrestrial civilizations, only five interstellar radio messages (IRMs) have been beamed into interstellar space. Table 21.1 lists these five IRMs in order of the dates of the first transmission

---

1 Pioneer Plaque; http://grin.hq.nasa.gov/ABSTRACTS/GPN-2000-001623.html

2 Voyager Record; http://voyager.jpl.nasa.gov/spacecraft/goldenrec.html

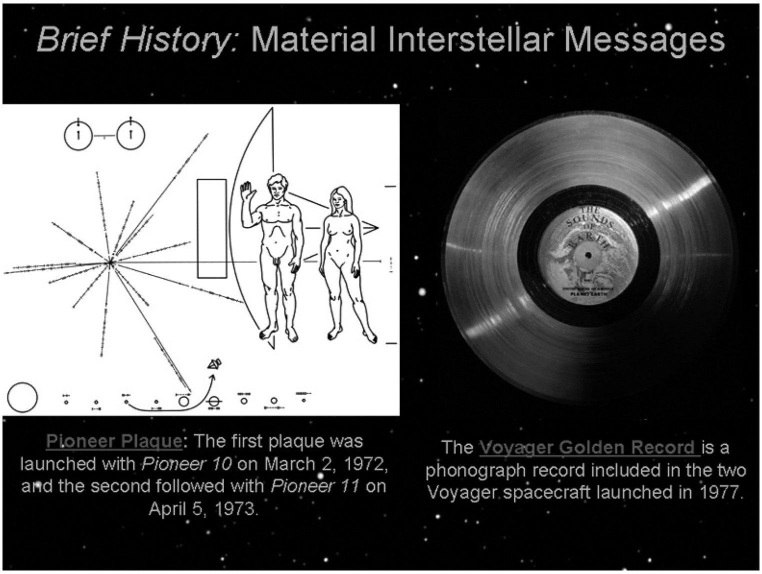



**Figure 21.1** The first interstellar messages sent beyond the Solar System onboard the American spacecraft *Pioneer* 10 and 11 and *Voyager* 1 and 2.


**Figure 21.2** The first interstellar radio messages.

session (notice that the overall number of the sessions is 17), the symbols T and E, respectively, the total duration of all sessions of each IRM and the total energy transmitted. The later figure correlates to the range of possible detection.

| **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's First IRM (digital) | First multi-page IRM | First digital and analog IRM | First International IRM | First Collective IRM |
| Authors | Drake, Sagan, Issacman, et al | Chafer, Dutil, Dumas, Braastad, Zaitsev, et al | Pshenichner, Filippova, Gindilis, Zaitsev, et al | Chafer, Dutil, Dumas, Braastad, Zaitsev, et 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 |

**Table 21.1** Interstellar radio messages transmitted from Earth.

The Arecibo Message was constructed of 1679 bits. It was sent to the globular cluster M13. The content and structure of the message have been repeatedly described in various books (e.g., Sagan et al., 1978) and on the Web[3], and thus 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 (Zaitsev and Ignatov, 1999). 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 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.

[3] Arecibo Message; http://en.wikipedia.org/wiki/Arecibo_message

In 2001, I was involved in the development and transmission of a Teen Age Message[4] to six Sun-like stars. For the first and, regretfully, the last time, the transmitted message consisted of three separate sub-messages: 1) a 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 (Zaitsev, 2008a). The digital part consisted of 28 binary images with a total size of 648220 bits.

IRM Cosmic Call 2 was sent to 5 Sun-like stars in 2003 (Braastad and Zaitsev, 2003). It was the first truly international interstellar radio message, 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) was prepared and sent from Evpatoria in October 2008[5]. 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 the article "Project METI@home: Messages to ETI from Home"[6].

Standing slightly outside of the main stream, there are two more IRMs: "Across the Universe 2008"[7] and "Hello From Earth 2009"[8] which were transmitted to the space using 70-m radio dishes of the NASA JPL Deep Space Network, located in Robledo (Spain) and Canberra (Australia). The first of the above-mentioned IRMs was critically discussed in Zaitsev

[4] 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

[5] A Message From Earth, http://en.wikipedia.org/wiki/A_Message_From_Earth

[6] Alexander L. Zaitsev. Project METI@home: Messages to ETI from home, http://www.cplire.ru/html/ra&sr/irm/METI@home.html

[7] NASA Beatles Transmission, http://www.nasa.gov/home/hqnews/2008/jan/HQ_08032_NASA_Beatles.html

[8] Hello From Earth, http://en.wikipedia.org/wiki/HELLO_FROM_EARTH

(2008a), the second message also had its drawbacks. I consider their main defect to be insufficient scientific and technical justification.

The summary shown as Table 21.2 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.

| 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 |

**Table 21.2** Details of the 17 sessions of the conducted IRM transmission.

The last column of Table 21.2 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. I consider triggering such a communications revolution to be one of the most worthy applications of the united intellect of human civilization!

## 21.2 Interdependence of SETI and METI

Our present SETI activitiy, 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 21.3).


**Figure 21.3** Searching for intelligent signals from the cosmos that come from the *past*, both temporally and positionally.

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.

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 21.4).

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".


**Figure 21.4** Any transmission of intelligent signals to prospective extraterrestrial civilizations is directed to the *future*, in both time and position.

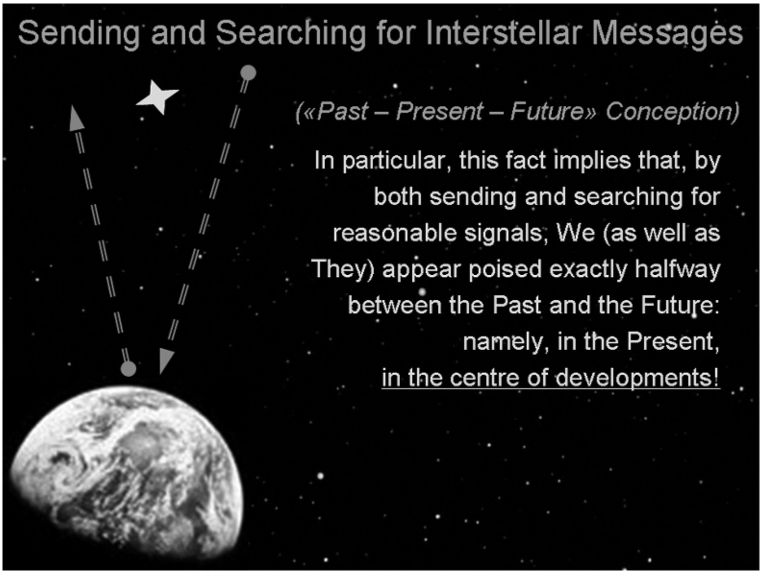



**Figure 21.5** A truly advanced civilization will perform both sending of and searching for artificial radio signals.

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 21.5). 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, unsurpassable deed!

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

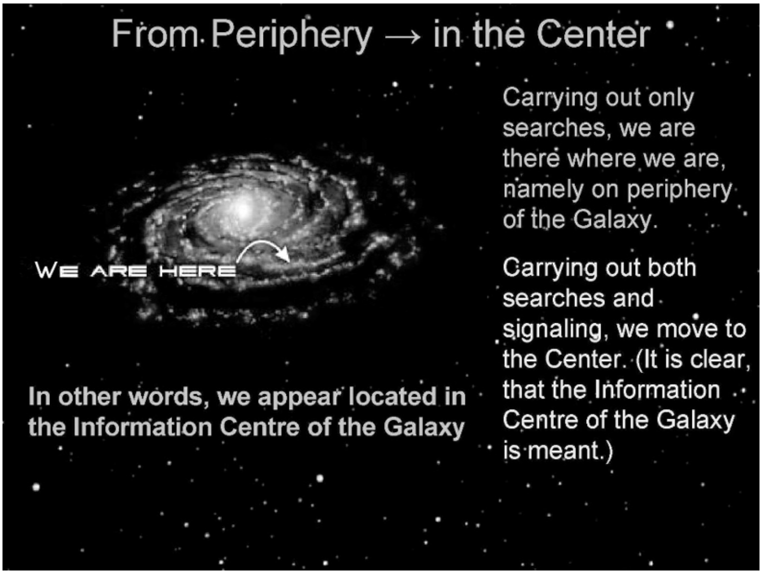



**Figure** 21.6 The concept of the "Information Centre of the Galaxy" is applied to a socially mature, advanced civilization that performs transmitting as well as searching for interstellar radio messages.

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 redirectinging 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 the17th 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 center”, as it appears at the center of the events that move it in the information space from the periphery of our Galaxy to its center (Figure 21.6).

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 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

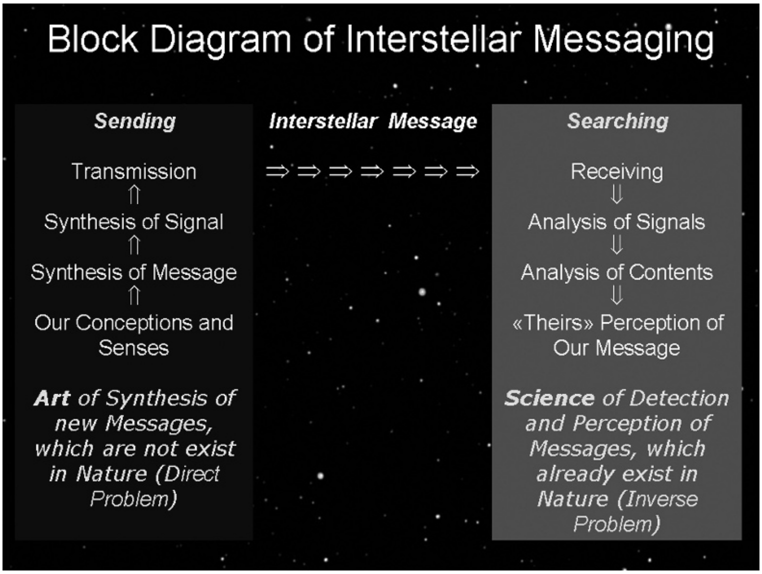



**Figure 21.7** Block diagram of sending and searching for intelligent signals in the Universe.

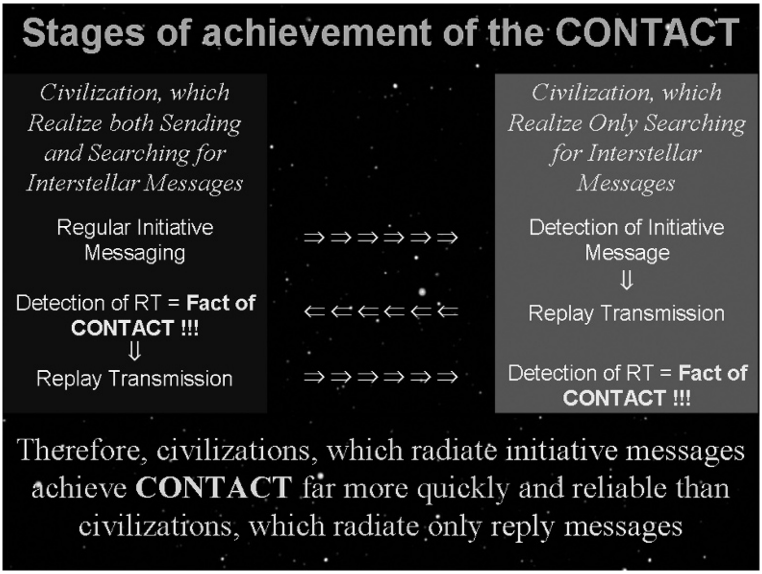



**Figure 21.8** Comparison of two civilizations, one of which is both transmitting and receiving, and one which is only receiving.

exists in Nature (Figure 21.7). 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 transmits 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 acknowledgment is received, it will be possible to say that contact has been established (Figure 21.8).

## 21.3 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 (Tarter, 1986). 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"[9] 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).

[9] Alexander Zaitsev and Richard Braastad. METI Art, http://www.cplire.ru/html/ra&sr/irm/METI_Art.html

When we consider the problem of METI, we formulate the same type of questions that were considered in SETI[10]:

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 (Zaitsev, 2008c), namely:

1 Presumable targets for sending *and* searching

2 Synchronization of sending and searching

3 Optimum frequency bands

4 Polarization

5 Power of transmitted and received 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 dangers of sending and receiving messages

---

[10] Alexander L. Zaitsev. Transforming SETI to METI, http://www.cplire.ru/html/ra&sr/irm/metitran.html

Next, we present our vision of how to solve these novel and controversial problems. We understand that our approach should be critically analyzed 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:

### 21.3.1 Presumed 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 sunlike star, 51 Pegasus (Mayor and Queloz, 1995). 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 center or anti-center 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"[11] – 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.

### 21.3.2 Synchronization of sending and searching

The problem of time synchronization between our transmission and an alien civilization's searches or, in the case of SETI, between an alien's transmission and our searches, is vitally important. Peter Makovetsky estimated[12] 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 neighbor 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 neighboring galaxies. These can also be used for the time synchronization of messaging and searching.

### 21.3.3. Optimal frequency bands

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

[11] Л. С. Марочник и Л. М. Мухин. Галактический «пояс жизни». В сборнике «Проблема поиска жизни во Вселенной», М.: Наука, 1986, стр. 41–46.

[12] П. В. Маковецкий. Новая Лебедя – синхросигнал для внеземных цивилизаций. АЖ, 1977, т. 54, № 2, с. 449-451.

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 centimeters. 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 optimum wavelength may even take on one or more "magic" values. As an example, it is likely that the wavelength 3.36 cm (which equals 21 cm / 2π), is recognized as significant 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 (the number of radians in a circle)[13]. This relationship is evident regardless of the measuring or counting system used, because it is derived from physical, observable constants that transcend culture. Other such relationships exist in nature, which may suggest interesting "magic wavelengths" to explore.

### 21.3.4 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!

### 21.3.5 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[14] shows the rates of transmission of information

---

[13] П. В. Маковецкий. О структуре позывных внеземных цивилизаций. АЖ, 1976, т. 53, № 1, стр. 222-224.

[14] Alexander L. Zaitsev. Limitations on Volume of Interstellar Radio Messages; http://www.cplire.ru/html/ra&sr/irm/limitations.html

for selected METI experiments conducted at 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 transmission 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 $m^2/K$. 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[15].

### 21.3.6 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 (Zaitsev, 2008a).

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 analyzer consisting of two million channels with a bin width 1 Hz. This allowed scientists to analyze a bandwidth on the order of 2 MHz

[15] SKA – Square Kilometer Array; http://www.skatelescope.org

in a real time, on-line mode, and on the order of 2 GHz of spectrum in post-processing (off-line) mode[16].

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 analyzers[17]. Such a modulation scheme, with a format well known and widely used on Earth, is frequency modulation (FM).

### 21.3.7 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 behavior of frequency of the radiated signal:

1 the frequency is constant over time;

2 the frequency jumps chaotically between several fixed values; or

3 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,

---

16 Project Phoenix General Overview; http://www.seti.org/Page.aspx?pid=583

17 А. Л. Зайцев. Язык радиопосланий к другим цивилизациям. Доклад на конференции «Джордано Бруно и современность», февраль 2000, ГАИШ МГУ. Вестник SETI, № 2/19, 2002, стр. 73–82; http://lnfm1.sai.msu.ru/SETI/koi/bulletin/19/articles/1.html

and interplanetary space. Applied to METI, this method, called radio sounding (Phinney and Anderson, 1968) 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 behavior 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 21.3 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[18].

[18] Г. М. Идлис. В поисках истины. М. Издательство, 2004.

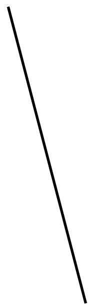
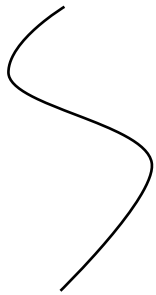
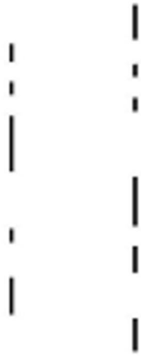
| Parameter | Three types of modulation | | |
|---|---|---|---|
| Type | 1 Constant | 2 Continuous | 3 Discrete |
| Author (Earth site) | Radio Engineer | Artist | Scientist |
| Language | «Nature» | «Emotion» | «Logic» |
| Information | Absent | Analog | Digital |
| Sonogram of transmitting signals (X axis is horizontal, Y axis is vertical) |  |  |  |
| Analyst ***(Alien)*** | Astrophysicist | Art Critic | Linguist |

**Table 21.3** Spectral languages for messaging to ETI.

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 baricenter of the Solar System (or to the center 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.

The third part of the message consists of discrete frequency shifts, digital dataflow showing constituents of our logic (algorithms, theories, etc.) and representing cumulative knowledge about ourselves and the world around us.

In Table 21.3, 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 analyzed by art critics; the third part by linguists, logicians, and behavioral 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.

### 21.3.8 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. 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 (see footnotes 3, 4 and 5; see also Zaitsev and Ignatov, 1999, Zaitsev, 2008a). We would like to emphasize that, at present, neither has a standard procedure of the 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 newfound 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*.

### 21.3.9 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" (Zaitsev, 2008b). which sought an answer to the question as to whether METI makes sense. There, we analyzed 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[19].

The paper was extensively discussed in blogs where more than 90 comments were posted, and in the SETI League's site[20]. 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" $f_m$. After taking into account this factor, Drake's classical formula now assumes the following form:

$$\boldsymbol{N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times f_m \times L},$$

where $\boldsymbol{f_m}$ is the fraction of the communicative civilizations indeed conducting systematic transmission of purposeful interstellar messages.

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.

---

19 А. Л. Зайцев. Парадокс SETI. Бюллетень САО, т. 60–61, стр.. 226–229; http://fire.relarn.ru/126/paradox.htm

20 Paul Gilster. SETI's Paradox and the Great Silence; http://www.setileague.org/editor/silence.htm

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 then one percent: these data follow from the review of Jill Tarter published in the recently released “SETI-2020” collection of papers (Ekers et al., 2003). 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 to it”.*

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.

### 21.3.10 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[21]:

[21] Radar Astronomy, http://en.wikipedia.org/wiki/Radar_astronomy

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[22] 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 21.9 in the ecliptic plane[23].

The total area of the sky illuminated by these transmissions is about 0.022 steradians (sr), or about $2 \cdot 10^{-3}$ (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 21.10).

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[24] that the probability of detecting the radar astronomy transmissions by a hostile super-civilization is $(2000 \times 450) \approx$ 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*, (Zaitsev, 2008b), 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[25]. For this reason, we conclude that all the on-going

[22] The San Marino Scale, http://iaaseti.org/smiscale.htm

[23] Д. А. Чураков. Анализ работы планетных радаров применительно к SETI и METI. Журнал радиоэлектроники, № 3, 2009, http://jre.cplire.ru/jre/mar09/index.html

[24] Alexander L. Zaitsev. Detection Probability of Terrestrial Radio Signals by a Hostile Super-civilization, http://arxiv.org/abs/0804.2754

[25] Asteroid and Comet Impact Hazard, http://impact.arc.nasa.gov/index.cfm

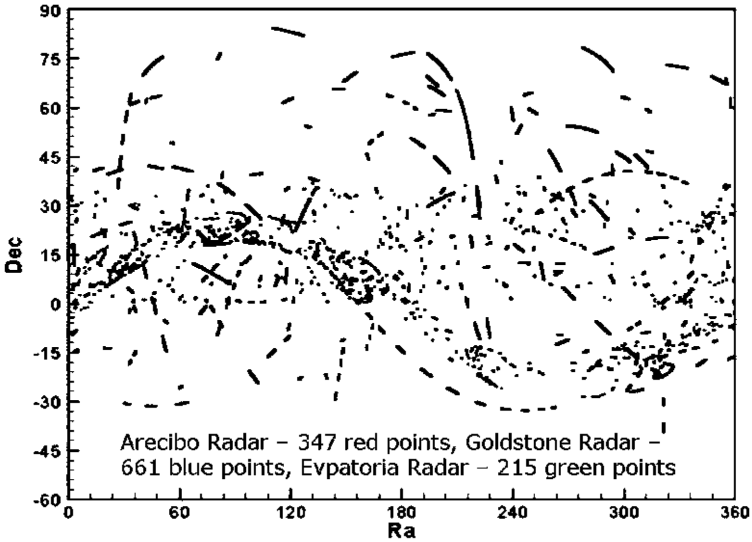



**Figure 21.9** Illumination of the sky by the overall radiation emitted during radar observations of celestial bodies.

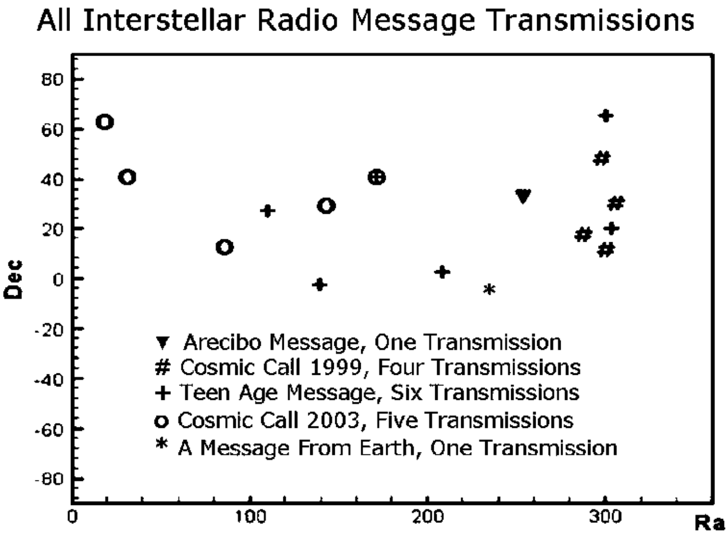



**Figure 21.10** 17 sessions of radiation of interstellar radio messages.

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 (Sagan et al., 1978), 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 centers, 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 (Carrigan, 2006). 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.

## 21.4 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 extra-terrestrials 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 can approach zero."

So, we can formulate the following two versions of the thesis implied by the SETI Paradox: "Only those who are trying to overcome the Great Silence, deserve to hear the Voice of the Universe" and "Only through the sharing of information between communicating civilizations will the Universe, in due course, find its Voice."

**Acknowledgements**

The author is sincerely grateful to numerous colleagues from Russia, the USA., the Ukraine, and Western Europe, with whom he has worked on the development and transmission of interstellar radio messages. My colleagues who work now in the USA, Sergei Kopeikin (University of Missouri-Columbia) and Ludmilla Kolokolova (University of Maryland), provided great improvement to the text by editing my initial English translation. I highly appreciate their help. Comments to the text by Sergey Kopejkin and Stefane Dumas (Canada) also were a valuable source for the improvement of this chapter. Many thanks to my numerous colleagues from Russia, the USA., Canada, and Western Europe, electronic correspondence with whom, and the pleasure of direct dialogue at seminars and conferences, undoubtedly have contributed to the creation of this chapter.

## References

Braastad, Richard and Zaitsev, Alexander. 2003. Synthesis and Transmission of Cosmic Call 2003 Interstellar Radio Message; http://www.cplire.ru/html/ra&sr/irm/CosmicCall-2003/index.html

Carrigan, Richard A. (Jr.). 2006. Do Potential SETI Signals Need to be Decontaminated? *Acta Astronautica* **58:2**, 112–117.

Ekers R.D., Billingham J., Cullers D.K., Schefer L.K. and Zajdel T.T. (Eds) 2003. *SETI 2020: A Roadmap for the Search for Extraterrestrial Intelligenc*e. SETI Press.

Mayor M. and Queloz, D. 1995. A Jupiter-mass Companion to a Solar-type Star. *Nature* **378**, 355–359; http://www.nature.com/nature/journal/v378/n6555/abs/378355a0.html

Phinney, R.A. and Anderson, D.L. On the Radio Occultation Method for Studying Planetary Atmospheres. *J. Gephys. Res*. **73**, 1819–1927.

Sagan, C., Drake, F., Druyan, A., Ferris, T., Lomberg, J. and Salzman, L. 1978. *Murmurs of Earth*. Random House, 1978.

Tarter, Jill C. 1986. 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).

Zaitsev, Alexander L. 2008a. The First Musical Interstellar Radio Message. *Journal of Communications Technology and Electronics* **53:9**, 1107–1113, http://springerlink.com/content/m62151781m500p16/?p=a9f198a3a40a488dbe3d3e84bbfbbda1&pi=11

Zaitsev, Alexander L. 2008b. The SETI Paradox. http://arxiv.org/abs/physics/0611283

Zaitsev, Alexander L. 2008c. Sending and Searching for Interstellar Messages. *Acta Astronautica*, **63**, Issues 5-6, Sept 2008, pp. 614–617, http://arxiv.org/abs/0711.2368

Zaitsev, A. and Ignatov, Sergey P. 1999. Report on Cosmic Call 1999; http://www.cplire.ru/html/ra&sr/irm/report-1999.html

# 22

# A Contrarian Perspective on Altruism: The Dangers of First Contact

David Brin

## 22.1 Altruism in the Natural World: Advantage and Satiation

The key word in the title of this chapter – Altruism – generally conveys certain assumptions. The first of these is that altruism – a selfless imperative to assist others without expectation of reward – is likely to be a valued attribute among advanced technological civilizations. Moreover, in the SETI context, it implies that humanity should strive to display this attribute in communicating with extraterrestrial life forms that may be 1E8 to1E9 years ahead of us in development. Finally, one topic much under discussion within the SETI community – how to craft and send a deliberate message from Earth into space – is based on the supposition that we can dismiss any substantial likelihood that transmitting will expose humanity and our world to danger. Are all of these assumptions warranted? Or do they reflect the personal inclinations and wishes of a narrow group, arising from a particular culture and era? Given the potentially overwhelming implications of contact, we may be wise to reflect upon the full range of possible outcomes, not only those we yearn for. I, for one, would feel more confident in the inevitability of alien altruism if that beneficent trait appeared more often in nature.

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_22, © Springer-Verlag Berlin Heidelberg 2011 

John Alcock (Alcock, 2001), shows that reciprocal altruism between related individuals occurs in many species; the real question concerns altruism between unrelated individuals and groups. It helps to divide generous behavior into two categories: "pragmatic cooperation" and "pure altruism". Biologists consider reproductive fitness to be the coin of the evolutionary realm. They study how this coin is spent in games like The Prisoner's Dilemma, which many animal species seem fully capable of playing. In simulations involving various kinds of rewards, you quickly get clear examples of cooperation and/or competition, depending on a pre-set payoff matrix. Emergent strategies like cheating, stealing, trust-building and honesty also appear. A basic concept of quid pro quo seems to manifest even among "lower" animal species.

In contrast to pragmatic cooperation, the purest form of altruism – in which individuals sacrifice advantage to benefit others without hope of recompense – does not at first appear to have anything to do with a cost/benefits game matrix. That is, until you include the "payoff" of genetic reproductive success. Then we see that the greatest and most prevalent forms of personal sacrifice – e.g., a mother for her child – fall elegantly into place. An uncle who risks his life to save a nephew benefits by helping his close gene pool to thrive. Biologists have documented extensively a basic fact: that selfless generosity occurs less often, and with decreasing intensity, as individuals grow more distantly related. This may seem a cold-blooded way to view something that we idealize as a noble quality. But shall we ignore scientific results? Especially results that shine revealing light on the very thing we desire?

Moreover, science acknowledges important exceptions to this curve (relating generosity to genetic payoff). We have all seen well-publicized examples in which mothers of one species adopt and nurse surrogate offspring from another. Dolphins have pushed human castaways toward boats or islands. And today, upon hearing word that sea creatures are stranded on some shore, modern people are frequently known to drop everything and race down to the beach... with the same alacrity and eagerness that their ancestors would have shown, upon hearing the same news.

Pause for a moment and consider that final example – human beings racing toward stranded whales. The vigor and speed of that response has remained constant. Today, their aim is to gently rescue rare, precious creatures. During most of our past, people hearing the same news would have hurried to the shore with a different purpose in mind... lunch.

The difference is clearly based on two transformations – education and satiation. We now know more about cetaceans and can thus identify with them far better. But above all, we no longer need their flesh to feed our hungry young. Satiation appears to be a critical element in the rising environmental movement, in the drive to include others within the protection of law, and in elevating altruism above other ideals that our ancestors considered paramount – like tribal patriotism and glory-at-arms.

Satiation seems important, as does a strong cultural drive toward valuing altruism as an admired goal. There are also aspects to altruism about which an idealist may not want to know. It has long been known that groups and animals and humans will – under certain circumstances – find ways to ensure that generosity is a widely exhibited trait, by either overtly or subtly reproving or disciplining those who behave selfishly. Ernst Fehr and Simon Gachter have carefully examined "altruistic punishment". Simple and clearly realistic game rules result in players ganging up – en masse – on defectors who play selfishly or fail to meet minimal standards of cooperation or beneficence. This occurs even when the act of punishing the defector adds costs and no benefits to the other players, and when any resulting altered behavior will help some other, later team, not themselves (Fehr and Gachter, 2002). We can all recognize the emotional drive that appears under certain circumstances, when we resent discourteous or selfish public behavior. The impulse to punish such behavior appears to have roots that go deeper than human nature.

Is this "true" alruism? Is it possible that we need to divide up this word and recognize that it represents a wide range of possible definitions and variants? Some of these variants may be crucially different in their effects during a contact situation. They may also merit quite different styles of representation in any message or interstellar art that is meant to convey our hopes and wishes to the stars.

Let us summarize up to this point:

1 Nature indicates that both pragmatic cooperation and selfless altruism occur in largely predictable ways, having to do with either quid pro quo payoff or reproductive success.

2 Interestingly, the fall-off curve for altruism appears quite similar to the curve of likelihood that two groups can cross-infect each other with disease. Both events happen in direct proportion to the degree of shared genetic heritage. The less related that two groups are, the less frequently they appear to be mutually generous or mutually infectious.

3 This fall-off curve bodes ill for the likelihood of interplanetary altruism, even as it bodes well for the likelihood that we could survive interplanetary disease.

4 Even what we recognize as altruistic behavior can have certain callous or game-based aspects that we should not ignore simply out of aesthetic puritanism.

5 Nevertheless, it is worth noting special anomalies, such as dolphin and human compassion for the strange and unrelated. These exceptions, and

a few others, seem to leap right off the genetic-relatedness curve, having no apparent "game" benefit. Here the driving force appears to be abstract sympathy, unleashed by full bellies and brains that are capable of seeing enlightened self interest in the long-term survival of an entire world.

Clearly, while remaining painfully aware of facts 1–4, we must invest in the hope offered by point 5.

What, then, can we conclude about extraterrestrial altruism?

Why, nothing, of course. We are exploring new territory. Any conclusions that we draw – either from nature or our inner wishes – should be taken as tentative, in a spirit of willing uncertainty.

Nevertheless, it is wise to bear nature in mind, as a de facto ground state for our discussions[1]. What biologists seem to be telling us, is that evolution does not predispose living creatures toward truly selfless altruism any more than it does toward esthetics. True, these are properties that humans have recently come to cherish. We may be doing so because that is what advanced creatures always and automatically do at this point in their rise. This idea – that sophistication and beneficence go hand-in-hand – appears to be the assumption of many SETI optimists.

On the other hand, our bent for altruism may instead be a quirky outcome – an "emergent property" –of our background as a species of already gregarious, exogamous and cooperative apes. For contrast, consider what kind of moral systems you might expect to arise if lions independently developed sapience. Or solitary and suspicious tigers? Bears are omnivores, like ourselves, and yet their consistent habit of male-perpetrated infanticide seems deeply rooted. Metaursine moralists might later view this inherited tendency as an unsavory sin and attempt to cure it by preaching restraint. Or else, perhaps they would rationalize and sacralize it, writing great literature to portray and justify the beauty of their way, just as we romanticize many of our own most emotion-laden traits[2].

Is our present fixation on "altruism" – in a strange twist – somewhat chauvinistic and humanocentric?

That ironic possibility is something to bear in mind. Please do not misconstrue. I heartily approve of altruism and try to live my life guided by

---

1 Are biologists too cynical to see something that seems obvious to SETI researchers? Is this why the SETI community (as opposed to the quite separate field of exobiology) appears largely made up of physical scientists? Perhaps they know something we do not. We might be wise to invite more of them into the tent.

2 Anyone who doubts that intolerant or even murderous habits can be romanticized should study the religious rites of the ancient Aztecs and Carthaginians. If we are capable of rationalizing and even exalting brutally un-altruistic behaviors, might advanced extraterrestrials also be capable also of such feats of mental legerdermain? Especially if their evolutionary backgrounds predispose them?

this rising standard. I certainly have no intention to denigrate an enthusiasm for self-improvement. To the contrary, I have often demonstrated my own idealistic yearnings for "otherness". As a stage in our development, this admirable trend may save us all. Nevertheless, scientific honesty warns against extrapolating any trend into a natural law. That is teleology – perceiving a plan, or cause-and-effect, where there may only be coincidence and happenstance.

And yet, even if it is largely absent from the natural world, that fact alone does not render pure altruism irrelevant. I just mentioned emergent properties. Complexity theory teaches that new forms of order arise as systems gain intricacy. It may be no accident that the most complex society created by the most complex species on Earth has elevated altruism from a rare phenomenon to an ideal – something to be striven toward across the present and into future years. Furthermore, in another ironic twist, it is entirely by these recent, higher standards that we now project a higher level of altruism upon those we hope to find more advanced than ourselves.

## 22.2 The Power of Thought Experimentation

In a strange kind of conservatism, SETI researchers have long striven to sever all links to the long tradition of science fiction, with its vast variety of contemplations about First Contact, ranging from high-end gedankeneksperiments to B-movie drivel. One can understand that this reflex has some basis in self-preservation, during an era when ridicule can be used to undermine your grant proposal. Above all, any talk of "danger" from first contact tends to be dismissed as sensationalism, conjuring up lurid images of pop-eyed invaders with jaws dripping formic acid. Hardly the stuff of serious science in the 1970s, 1980s and 1990s.

And yet, doesn't this aversion give Hollywood entirely too much power over our thought processes? To draw premature conclusions, and exclude a huge trove of plausible scenarios, seems inordinately unwise, especially when the asymmetry is so great between positive and negative consequences.

For this reason – in a spirit of cordial, contrarian questioning – let me offer to play devil's advocate. I intend to suggest that it may be foolish for us to beam any messages from this planet until we know a lot more. To do so will be like ignorant children, screaming "Hello!" at the top of their lungs, in the middle of a dark, unknown jungle.

## 22.3 Fools Rush In...

Interstellar space may hold only the wise, grandfather types predicted by Cornell-based SETI founders Frank Drake and Carl Sagan. Kindly ancient ones may welcome us into their advanced, pacific civilization. On the other hand, consider our own practical experience over the last 6000 years, when various human cultures have collided with each other here on Earth. In history, “first contact” has seldom been gentle and benign. At best, cultural values were shaken, requiring painful readjustments. At worst, the outcome was often genocide.

In other words, altruism appears to have been as rare for intra-human first-contact experiences as it is between animal species. Yes, that may change. We may yet become a civilization that lives and works under codes such as the famous “Prime Directive”. Even if this is not now in our nature, we may choose to change that nature, turning ourselves into truly noble beings. This is our ambition and hope for the future. Still, it is wise to remember our context and our past.

Bearing this history in mind, SETI pioneer Phil Morrison said: “I share the idea of caution before any reply.”

Elsewhere in this book we have discussed the “Great Silence”, also called the Fermi Paradox – the mystery of why the nearby regions of our galaxy appear to be rather quiet, emptier of living voices than many of us expected when the SETI era began. I will readily concede that half a century without a clear signal proves nothing about absence. What it does imply is either some degree of scarcity or else a reticence on the part of aliens to broadcast at the maximum levels achievable by highly-advanced technological cultures. This reticence to broadcast at full strength – a lack of the Giant Beacons once predicted by Drake and colleagues – should be at least somewhat worrisome. Especially to those among us who feel an urge to shout.

Those who contemplate the Great Silence have listed a wide range of possible explanations for this strange state of quiet (in more detail than I have room for here). Not all of these reasons are pessimistic. Some may be benign, raising the possibility that patience and perseverance will eventually bring success.

On the other hand, there seem to be numerous plausible ways that our Galaxy may be hazardous. These begin with natural phenomena. Supernovas, comet swarms and giant molecular clouds are among just a few of the natural threats that “life-worlds” like Earth have to survive before they can bring forth technological civilizations. One explanation: we may be among the few survivors to reach this phase.

There are also unnatural ways the universe could turn unfriendly. For example, suppose some earlier species unleashed a wave of irresponsible colonization across the galaxy, sweeping like a prairie fire, leaving over-exploited worlds and ravaged ecospheres in its wake. Malevolence is not required, only shortsightedness and unsustainable appetites across many millennia, (a trait that is completely consistent with the behavior of the one sapient species currently known.) If such an unfortunate interstellar ecological disaster happened, our Earth might be among the few life-worlds to have escaped. That, too, could explain why we don't hear anybody.

Again let me emphasize, no single explanation has any great weight of evidence for being true. All merit study.

In this chapter, I want to narrow the focus onto Contact itself – the day we actually learn we aren't alone. What dangers should we consider during the following days and months? What possibilities should we keep in mind while seeking neighbors among the stars?

## 22.4 Physical and Biological Contact

The first question has to be, will First Contact be made in person? Or will it be a mere exchange of greetings and information by radio? It is the latter scenario most SETI scholars predict. But let's start by briefly considering dangers that might arise if we met alien beings face to face.

For starters, we can almost certainly eliminate the obvious – direct conquest by some interstellar empire. While many scientists believe various forms of interstellar travel will someday be possible, nearly all spurn the idea of armadas filled with enslaving conquerors, swooping down from the sky.

For one thing, why invade us now, when we can fight back? Why not come during the several billion years that Earth was prime real estate, but had no technological civilization to defend it? The temporal coincidence implicit in most sci fi invasion films makes them absurd on that basis alone.

Then there are the economics of interstellar travel. Even if star flight proves plausible, it is likely to remain an expensive proposition. Bulk natural resources won't be worth the shipping costs. Information-based commodities, such as inventions, cultural works and genetic codes are far more transportable. Such commodities might be given away, traded or stolen. But even in the last category, the thieves will most likely use subtle or surreptitious means rather than brute force.

Of course invaders might not come for plunder but to colonize. Even here though, most physicists and science fiction writers agree the prospect is farfetched. "Just how do you maintain an invading army at the end of a supply line several light-years long?" one might ask. Conquerors would have to live off the land, at least until they altered Earth's biosphere to suit their needs – a difficult undertaking while they're being harried by determined guerrillas. Despite its prevalence in cheap movie melodramas, invasion may seem the least likely of dangers from outer space.

But other, more plausible hazards might arise from physical contact. Suppose a single alien starship decelerates into our Solar System, say on the folding wings of a great light-sail or behind a super-efficient antimatter engine. Presumably we would send welcomers to say hello. Or their emissaries may come down to meet us. Let's further suppose they show no signs of weaponry and appear to be on a genuine mission of peace.

In that case, one of the most fearsome possibilities for us to worry about would be disease. Until our recent AIDS epidemic, the concept of plague had grown strange to modern westerners. Yet history shows that infection was a major element in countless first-contacts between human cultures. Often, it played a crucial role. Anthropologist Alfred W. Crosby points out that the European conquest of the Americas and Oceanea was facilitated by such Eurasian diseases as measles and smallpox – sometimes introduced intentionally, but more often quite inadvertently and, ironically often, quite soon after both sides shook hands over treaties of friendship!

Some claim alien physiologies would be too incompatible ... that extraterrestrial parasites would be unable to prey upon human organisms and our organisms would certainly fail against our guests. But there is wide disagreement about this among biologists.

Stanley Miller, one of the premier experts on the origins of life, has a different opinion. Miller now believes that biological chemistry throughout the universe involves the same small set of amino acids and nucleic bases Earth lifeforms use. Those chemicals happen to be the most stable, the best at forming the complex structures of enzymes and proteins.

On the other hand, arguing from earthly experience, it seems that cross-infection follows a curve not too dissimilar to that of interspecies altruism! The more genetically remote a given species is from us, the less likely it is to transmit a lethal agent to us. A lot of the most lethal agents (e.g., HIV, monkey B virus) seem to have started off in other primates, albeit in modified form. But as you move away on the genetic continuum, these events are fewer. Once you leave mammals, you have parrot fever and various flu viruses from birds, little or nothing from amphibians, reptiles or fish. Insects, which make up most of the eukaryotic biomass of the planet, serve as carriers for a few things like malaria, but these are more incidental vectors than hosts. If you assume that ET is very far from us genetically, the likelihood of cross-infection seems pretty low.

In other words, there is no clear consensus about the danger from Space Bugs. Nevertheless, even dismissing scenarios such as H.G. Wells's *War of the Worlds*, we would be fools not to at least bear human history in mind, before some handsome alien steps down the ramp and offers his hand.

Suppose our extraterrestrial guests pass successfully through quarantine. There are still reasons to be nervous. For example, how are we to guarantee their safety? Would you risk letting alien tourists walk unguarded down our city streets? Ninety nine percent of the population may welcome them gladly. But most people also liked John Lennon. Human diversity is one of our treasures. Alas, it also means our mad fringe will be a persistent danger to visitors from space. This may be hard for guests to understand if they come from a homogeneous, uniform society. [3]

In the past, several human societies found themselves plunged into calamitous wars against European powers, precipitated by the actions of a few local hot-heads, acting against the wishes of wise and cautious local chiefs. This will be a source of danger in any future contact situation, as well. Of that you can be sure.

## 22.5 Non-biological Probes

Some scientists, such as the late engineer and SETI scholar Bernard Oliver, long held that interstellar travel by living organisms is too uneconomical ever to be practical. While I disagree, it hardly matters. Even if we eliminate that entire set of possibilities, it turns out that there are plenty of dangerous scenarios that do not involve direct physical contact between organic beings.

What about space probes? Following the lead of the British Planetary Society, NASA has already commissioned preliminary studies of a survey device which might be sent toward Alpha Centauri within our lifetimes, carrying sophisticated cybernetic systems that (it is hoped) will border on human intelligence. If such probes seem possible for us to dispatch within decades, some advanced civilization would surely come up with even better plans. Perhaps machine emissaries capable of making copies of themselves at each new arrival point, using local materials to multiply and then speed many duplicates onward, unhampered by the weight of onboard life-support systems.

[3] What about diversity among the extraterrestrials themselves? In both SETI and science fiction, we tend to envision each type as uniform in characteristics, with little variation -- a bad habit that is related to the evils of racism, sexism and stereotyping others by class. It is, in fact, quite possible that the first exemplars of communicating aliens that we meet may be atypical. Moreover, they may have reasons not to convey this fact to us. How do you know whether you're dealing with a council of elders that have high tolerance and a low fear level, or an "alienated alien teenager"… or for that matter an autonomous "PDA" buried in the tracking software for an advanced radio or optical telescope. Keep in mind our SETI program, which gives "first crack" at looking for signals to thousands of unvetted amateurs. Another reason for caution.

Simple propagation algorithms show that – based on reasonable assumptions for ship speed and rebuild times – a single self-reproducing probe might create enough progeny to visit every star in the Galaxy within less than five million years. A mere heartbeat in the life of our cosmos. It's generally thought that such "Von Neumann Self-Replicating Probes" would be programmed to be friendly. But this is only an assumption. Might such probes turn out to be dangerous?

Physicist and Nebula Award winning novelist Gregory Benford points out that all "self-replicating" systems – such as living things – are controlled by programs of internal information containing their design, and plans for the fabrication of new copies. These plans inevitably suffer changes in time – called mutations. Life relies on mutation to drive variation and evolution. But mutation also means no species will adhere forever to its original program. The same would hold for any probe emissaries sent forth by curious aliens.

If such a probe arrived in our Solar System, in what condition would its programming be? Some of Benford's fiction, along with those of Fred Saberhagen and others, portrays the dread possibility of "deadly probes" – deliberately or accidently programmed to destructively home-in on new civilizations soon after they become detectable by their radio transmissions. Such horrible "berserker" machines may seem garish, even sensational, and nobody claims they are particularly likely. Still, they are in no way inconsistent with natural law. Indeed, they are quite consistent with the observed state of silence.

They remind us to consider just how unwise it may be to shout in a jungle, before we have any idea what's out there.

## 22.6 Propagation as Information

We have only touched lightly on the range of possible outcomes and drawbacks from direct physical contact between ourselves and extraterrestrials. But let us move on, putting aside that category for now (it is highly unpopular among SETI enthusiasts) and concentrating instead on what most scholars consider the more likely eventuality – communication with other worlds solely via radio or light waves, exchanging only information.

Only information? Surely, no harm can come to either side from such an encounter!

Well, actually, we shouldn't be too blithely certain about that. One has only to look again at the history of first contact between human cultures to see how much pain sometimes came about not from conquest or disease, but when one civilization encountered another's ideas. What are some of the mistakes we might make, if ever we encounter someone out there with something to say?

What if a government manages to slap a TOP SECRET classification on the discovery, sequestering knowledge of contact for the benefit of some group or nation here on Earth. We cannot know for certain that this hasn't already happened! Just because an idea has been worked to death in bad dramas doesn't mean that it's completely impossible. America's NSA (National Security Agency) is just one group already possessing far more sophisticated listening apparatus than all of the world's SETI teams put together. If SETI discovers a point source in some portion of the sky next week, can we know for certain that the NSA did not pick it up first, perhaps many years ago?

A chief argument against this paranoid scenario is that the intelligence community seems neutral – even mildly supportive – toward SETI, implying they're not worried about secrets being uncovered by those civilian astronomers. Still, it's worth considering what the consequences might be, if extraterrestrial life were first discovered not by independent searchers, but by one of the security agencies, or by the intelligence service of a hostile power.

One could imagine how information from the stars might be used in unfortunate ways if access were restricted to a narrow group. At the minimum, it would deprive the rest of us of a startling and wonderful experience which we, as taxpayers, paid for. Clearly, from the success of many popular science fiction "contact" films, people in our civilization feel positively toward the search for otherworldly life, and would resent being coddled, or cut off from full participation in such a momentous event.

Many SETI scholars do worry about this possibility, and a consensus has spread among them that information about alien life is nobody's "property" – save, perhaps, all of mankind. An unofficial but influential "SETI Protocol" has been signed or initialed by most of the first-rank workers in this field, accepting general principles of accountability and openness. Sequestration of information is a clear danger to be guarded against. But now – in the spirit of contrarian criticism – I want to turn around and warn about the opposite trend, the growing assumption that absolutely everything about First Contact should automatically and unquestionably be released right away, into the direct spotlight of mass media.

This extreme, too, could cause severe problems. Take, for instance, the way the press turns some events into "media circuses". During the early phases of a discovery – while scientists are still trying to verify that it's "contact" and not some fluke or natural phenomenon – premature media attention could do great harm. What if a mistake was made?

I am reminded of the events surrounding detection of the first pulsar, which was initially thought to be an interstellar beacon because of its uncannily regular radio pulsation. If there had been an Internet back then, perhaps that false alarm might have aborted the entire SETI enterprise! How many false alarms can a program survive before it turns into a laughing stock? For this reason, we must expect some caution while responsible researchers triple check and discreetly seek verification from colleagues around the world.

Also, we must remember, researchers are people, with families and obligations. Their employers – for instance, NASA – may have operational rules and internal procedures that scientists are expected to follow, before any public announcement is made. It would be unfair to shout "coverup!" just because a little bureaucratic paperwork delays the big press conference by a few days.

This may mean the first announcement won't be made by responsible, careful scientists, but by a person on the periphery, perhaps a lurker in the rumor loop, someone with an appetite for headlines. Those who grab the front pages may not be the ones most qualified or deserving to represent us during the critical stages of First Contact.

Let's take the matter further. Say contact has been verified, to the best of our scientists' abilities. Miraculously, nobody leaked prematurely or tried to steal their thunder. They've cross checked, fulfilled their institutional requirements, and are now ready to release the good news. Might there be some justification for delaying the announcement for just a little longer? Or to limit the amount of knowledge released? (Perhaps excluding specific location and frequency information.) Yes, I am about to question one of the core tenets of the "SETI Protocol". But do hear me out.

We should recall that it is only very recently that a few cultures began ascribing to the notion of freely exchanging ideas. Throughout history, nearly every tribe or nation held instead to the more traditional notion – that some concepts are too dangerous (or valuable) to be let loose among common folk. Were all those cultures entirely wrong to believe this?

I happen to believe they were! I hold to my own culture's central tenet that openness is good. The best way to protect people from bad ideas is to let them experience the entire range of human concepts, so they can learn for themselves to judge wheat from chaff. Clearly, the SETI Protocol is based entirely on this premise. Indeed, the Protocol is clearly a wager that we have the toxicity question figured correctly, and others did not.

Let me state again that I agree with the maturity worldview. My life revolves around it and I approved back when a few of us were deliberating the SETI Protocol, line by line. But then, honesty compels me also to admit I might be wrong. My culture's central assumption could be mistaken. Every other human culture may have been right instead, when they posited that ideas are inherently dangerous. It is the height of arrogance not to at least ponder this possibility, instead of simply assuming that a very recent set of upstart principles is automatically and obviously true.

In his famous book, *The Selfish Gene*, Oxford scientist Richard Dawkins made this idea of toxic or infectious information look startlingly plausible. He coined a word, “meme”, to stand for an idea which catches the attention of a person hearing or reading it ... and intrigues that person enough to make him want to tell someone else about it. And then she passes it on to someone else. And so on. It sounds like what goes on every day, as people talk to other people about what interests them, spreading everything from useful knowledge to acrid rumors.

It also sounds a lot like the way we catch and pass on the common cold, passing it from host to host with our sneezes!

Dawkins made the interesting case that “memes” behave very much like our “genes”. In other words, successful information replicates (makes copies of itself) whether via the coding mechanisms in a cell’s DNA or via the connected words communicating an idea. Dawkins pointed to how eager we sometimes are to persuade others to share our opinions, and to the tenacity with which some people fight for their beliefs.

This is not the place to go into Dawkins’s fascinating idea in detail. (Though you’ll notice I’ve already “infected” you with the concept of “memes”. In some of you it will take root, you’ll go look it up, and tell others. So it is with all interesting ideas, whether they’re true or not.) Still, we are led to speculate about several rather chilling and dangerous scenarios that could come about the day after information about First Contact is finally announced.

For instance, what will the news of contact do to people? Some suggest it will inevitably lead to mass hysteria and alienation – even riots and suicide – as paranoia and xenophobia (fear of outsiders) takes hold. This hoary sci fi cliché – which drives a story plot by assuming the worst – has even appeared even in some high quality speculations, like *2001: A Space Odyssey.*

SETI scholars take the opposite view, conveyed aptly in another film, *Contact,* in which humanity is portrayed accepting the news from outer space with commendable reflection, awe and humility, eager to put our petty Earthly struggles into perspective. (Should contact be made by the natives of my homeland – California – the first question asked of any visitors would probably be: *“Say, groovy gentlebeings, have you got any new cuisine?”*)

In truth, we’ll most likely see every possible reaction. Panic and calm, mysticism and reason, hope and despair. Each combination will mirror the heart of a different human being, or a different segment of the population. This may or may not be dangerous, but it certainly does promise interesting times, soon after the announcement is made.

What if an ambiguous message from the stars seems to verify or validate the cherished belief meme of some group on Earth? For instance, imagine that, after transcription of the messages, a star-and-crescent symbol appears repeatedly on our alien correspondents’ interstellar letterhead, and this is taken by some to mean that the aliens are Muslims? Or that some ET name happens to translate similar to a central myth figure of an obscure Christian sect? Or that hive-like beings express uncomprehending contempt for democracy? If two-way communication takes decades, even centuries, it may be hard to ask our new friends to clarify their meaning in time to make a difference in the resulting confusion.

This is serious. Once upon a time, wars were fought over differing interpretations of a single line or word of scripture. Or even a smudge, as in the row over *homo ousias*. We like to think such pettiness lies behind us. But then, we also thought that epidemic was an obsolete word, for a brief innocent while. We ought to be prepared for the inevitable likelihood that individuals and groups on Earth will seek any advantage they can from the first messages from the stars, whatever form those messages take.

How much worse might these problems be, if the extraterrestrials are responding to an ill-considered message of our own? Whether they do so inadvertently, or out of deliberate malice, it will be within the power of alien communicators to use words and symbols in unhelpful ways. History suggests caution.

Which brings up the inevitable question: "How do we decide who will speak for us?" Will every nation, sect, and religious group begin casting its own pleadings, threats, and dogmas skyward, almost the instant that contact is announced? Probably. One thing our alien friends are certain to learn about us right away is just how undisciplined a species we are.

That's only the truth, after all.

But let's return again to the topic of dangerous ideas. Is it possible that we may be the infectious ones? Before dismissing the idea out of hand, consider that the apparent silence out there could have any number of possible reasons. We who are so new to understanding the depth and potential of syntactical information flow – are we the best judges of what is possible, let alone dangerous to others?

Would it really hurt to spend a little while advancing our knowledge in those areas, before ecstatically and impulsively shouting (or "sneezing") in all directions? How about those wonders of technology we hope to acquire, once we begin learning under the remote tutelage of our wise, beneficent predecessors? There has been talk about solving many of the problems that dog us – e.g., energy crises, disease and unsafe transportation – by sharing solutions that were discovered long ago by others out there. They might even know answers to biological and sociological quandaries which today threaten our very survival.

For now, let's put aside the interesting philosophical question of whether we'd be better off earning our rightful place, instead of becoming dependent on technological crumbs, like beggars at a banquet. That is a serious question, but I don't expect it to receive a congenial hearing here. Suppose we do start receiving a wad of generous schematics for all sorts of wonders. What if they are technologies we're not ready for? Like a simple way to make antimatter, using common household materials and wall current? Ninety nine point nine percent of the population may behave responsibly and refrain from blowing us up. The remaining 0.1% would kill us all.

A SETI manager who would take great care to quarantine actual visitors may feel uncomfortable with the proposition that data need also to be checked. But can a case be made for putting a buffer between the main SETI receiving facility and the rest of the world, so both time and geography will give us a chance to pause and evaluate each part of the message before committing ourselves irrevocably?

Many westerners believe in the free competition of ideas – letting the fittest survive in open argument. We tend, quite rightly, to see any attempt to restrict that openness as a direct threat. And yet, there may be ways, quite conceivable ways, in which information from the stars could prove harmful, as in "virus" computer codes which infect a mainframe or microcomputer, proceeding to gobble up memory space, ruin data, and then spread to other hosts. So far, most inimical programs have proved fairly primitive – nothing compared to the voracious, computereating monsters depicted in some science fiction stories. And yet, those stories were correct in predicting computer viruses in the first place. And they are getting more sophisticated all the time.

A software "invader" needn't be intentional. On Earth there are endless stories of programs interfering destructively with other programs. What, then, of sophisticated code from an alien culture, taken in through our antennas and suddenly introduced into a data-handling system for which it wasn't designed? Any message from the stars is likely to include error correction modules, designed to repair damage done to the message during transit through the dust and plasma of interstellar space. Once the code is embedded in an active computing medium, such modules would "wake up" – much like a hibernating animal aroused from sleep – and would then begin using available computing resources to restore the integrity and function of the message.

As bizarre as this concept may sound at first, it isn't science fiction. Far from it. This is how the world's best information specialists say they would design any complex code meant to beam at the stars! (Consider how each of these dangers should be considered in the opposite direction, as we prepare potential messages to transmit. Our own coding assumptions may have unexpected side effects when they enter the medium of an alien information system.)

Under normal circumstances, an extraterrestrial message may be completely harmless. But what is "normal" for alien software? There is no guarantee such a program won't inadvertently take over more of an unfamiliar host system than anyone ever imagined. This accident might be made even worse if the program suffered "mutation" in transit.

## 22.7 Giving It All Away

Today, SETI scientists worry far more about lurid headlines (...SCHOLARS THINK ET PROGRAMS MIGHT EAT US!!...) than about warding off infection by self-replicating alien software. And they are right. After all, nobody believes virus codes really represent a high probability hazard to us or our civilization. But the wrong type of publicity, even misquoted, is a sure way to see your grant slashed. With that, far more imminent danger always looming nearby, it's no wonder that talk of potential hazards from First Contact rates far down most researchers' list of priorities.

And yet, is it wise to go into this enterprise simply assuming there's no danger at all? That's called "success-oriented planning", and it was used extensively by the US Space Shuttle Program. Need I say more?[4]

Consider the Intermediate Contact Scenario – in which those we encounter by radio are too far away to meet physically, but near enough that two-way communication is a practical possibility. (By this I mean that you might cast forth a question and expect that you, or your grandchild, may hear a reply.) Let's further assume the scholars are right, and First Contact will be made with an older, utterly benign civilization, completely uninterested in harming us. Furthermore, say they loose no dreaded plagues upon us, either physical or informational – either genes or memes – and none of the ideas or technology we receive are beyond our ability or wisdom to handle.

Assume further that competing powers on Earth don't conspire to withhold bits of the message for their own advantage, nor vie with each other to influence our faraway friends. Let's say we manage to appoint a proper committee to speak for Earth while, at the same time, allowance is made for the melange of other human voices that will inevitably cast forth, outside all official channels. ("It's often that way with bright, impatient young species," the Ancient Ones might say. "We'll negotiate with your committee, and happily set up cosmic pen pals for the rest of you.")

Finally, let's assume the news that we aren't alone affects us in all the right ways. That it causes us to reflect on our lives and to grow closer, deeper in our understanding of ourselves and the Universe. That we do not wind

[4] Success-oriented planning is actually the most reasonable thing to do in many cases, where there isn't a large asymmetry or irreversibility in the payoff matrix. First Contact with an unknown life form does not meet the criterion, however. Potential downsides of failure are immense and irreversible. This makes success-oriented planning truly irresponsible.

up feeling cowed or intimidated or shamed by having to be saved, instead of managing it ourselves. This is the classical Contact Scenario, a glowing prospect which many consider the most likely result of verified discovery of extraterrestrials.

Actually, I agree. It is the most likely result... one of many reasons why I enthusiastically support SETI. But now, even after making every one of those blithe assumptions, can we relax at last? Are we ready to enjoy and celebrate First Contact in complete safety?

We are not!

For even in a civilized setting, life can still be dangerous if you don't know the rules. (Don't believe me? Try investing in Wall Street without any experience!)[5]

What, after all, is the most common peaceful enterprise of human beings? Commerce, of course. And what is likely to be the main – perhaps the only – commodity of commerce on an interstellar scale?

Again, it will almost certainly be information. Not the malign, dangerous information we spoke of earlier, but useful information, neat inventions and brilliant innovations and even – especially – art and literature. Anything novel and original. Whatever's fresh and new.

How will most of you respond if the first thing we're asked by aliens is, "Send us your music and your art!" The *Voyager* spacecraft carry disk recordings of samples of Earth culture, along with graphic instructions on how to read the information. In the spirit of the United Nations, it simply never occurred to any of the people planning this gesture that the album should have carried a price tag, as well.

It's all very well to speak of altruism, and of the joys of free exchange. But we should always remember that is a very recent concept in human affairs. Quid pro quo is a more venerable theme. Throughout human history, in most of our daily lives, and even among the higher animals, the real rule for civilized relations is not "be generous". It is "be fair". And make no mistake, there is a difference!

Nice as they may be, our extraterrestrials will almost certainly engage in trade. And their stock in trade will be information. We may seek from them the answers to our ultimate questions. They, in turn, may reply, "Great. We've got some answers. But surely you have something to offer in exchange?"

What can we offer? All we may have is ourselves – our art, our music, our books and drama. Forget physical resources. The true wealth of humanity

[5] The most effective con artists are the least rapacious-seeming folks you will probably ever have the misfortune to meet. Kenneth Galbraith once said that we experience big financial cons about every 20 years, because we let our guard down. We can afford several-year setbacks every 20 years. What we can't afford is a millennia-scale setback, simply because we didn't argue about something for a while before responding.

lies in our culture. That is what we have to trade. It is our treasure. And it is also the very first thing we are likely to beam to the stars, in gigabytes, within days after First Contact! Given the spirit of the times, and our ecstatic enthusiasm for contact, it's what would seem only natural as we eagerly seek to "share with" (or impress) our newfound neighbors. And that very admirable rush to share – proving our altruism in an orgy of transmission – might turn out to be the worst mistake of all time.

They may be nice. They may operate under rules we would call fair. But nobody expects to pay for a free gift! It could be that history will speak of no worse traitors to humanity than those who, with all the best intentions, cast out to the skies our very heritage, asking nothing in return, thereby impoverishing us all.

Let me reiterate this point. Nature is mostly tooth-and-claw.

At the opposite end are some glimmers of genuine altruism, exhibited by dolphins now and then, an occasional dog, plus a large number of recent human beings who want to be much better than they are. Our great opportunity for improvement shines at this end of the spectrum. I hope we make it. But as yet there is no guarantee. There is hardly even a trend.

What is more firmly based in both nature and human experience is something that lies midway along the spectrum: our concept of fairness in dealing with each other on a basis of quid pro quo.

Many animals seem to understand the basic notion of exchanging favors, tit-for-tat, making a deal. Unlike pure altruism, pragmatic cooperation stands on much firmer ground, rooted firmly in observed nature, halfway between predation and total beneficence. Moreover, one can easily imagine how to portray fair trade in a message. There is every chance that intelligent aliens will understand this concept, even if they find "altruism" incomprehensible.

Because of this, let me humbly suggest that a fair and open approach based on cautious quid pro quo should be our central theme as we take measured steps toward Contact, while all the time remembering that we are new and small and weak in a vast Universe. If aliens truly are benignly altruistic, they will forgive us this precaution, this vestige of pragmatic self-interest. Noble beings will bear in mind our recent difficult experience. They will understand.

## 22.8 Already Too Late?

Is it already too late? A long-held truism maintains that the Earth has been extremely noisy in the radio spectrum, especially since the end of World War II, with the advent of television broadcasting and continental missile-detection radars. So noisy that any thought of reticence or patient listening is already moot.

If the Galaxy really is a dangerous "jungle", predators have already picked up "I Love Lucy" – so we might as well shout as loudly as possible, in hope of also meeting the best people out there. This supposition – which always reeked of rationalization – has lately been questioned by experts such as Seth Shostak, who calculate that it would take a very large and carefully-aimed antenna receiver to pick out signs of technology in our Solar System's emanation-spectrum, from more than a dozen or two light years away. The modulated portions probably stand out from the background far less than we thought. The sole exception would be deliberately-beamed messages, which pack a lot of signal energy into a narrow beam area.

Until recently, the one well-known intentional "message" was cast forth from Arecibo many years ago by one of the teams affiliated with Frank Drake, in the direction of a distant cluster of stars. With that target an innocuous distance away – tens of thousands of light years – the act was more a symbol of faith in the SETI enterprise (or else a "stunt', depending on your view) than a serious attempt to attract attention. Drake's group, despite their enthusiasm, had the maturity to refrain from doing anything more, or taking upon themselves a decision that belonged properly to all humankind.

This wise reticence has been broken in the past few years. Russian astronomer Alexander Zaitsev has more recently beamed forth a handful of interstellar messages, including pictorial and musical transmissions, from the Evpatoria radio telescope in the Ukraine. Another group in Brazil claims to have sent forth some narrow-casts. We can certainly expect more such unilateral spasms in future years, as radio equipment becomes cheaper and available to pseudo-scientists – with or without academic credentials – who lack the patience or scientific courtesy to respect the wishes of others. History shows that people can rationalize anything, when it offers their only hope for self importance.

The consensus of all major SETI research groups, however, as reported by American space lawyer Patricia Sterns, is to follow policy guidelines developed by the SETI Committee of the International Academy of Astronautics and the International Institute of Space Law. These protocols discourage intentional transmissions targeted at extraterrestrials unless preceded by broadly based international discussion.

Falling under a completely separate category are endeavors that do not try to unilaterally impose a point of view upon the rest of humanity. Examples are conferences which engage in discussions of the scientific and esthetic aspects of interstellar communication, aim to explore things that we have always taken for granted, using the imagined viewpoint of alien outsiders to gain fresh perspective on fundamentals that may be shared across the cosmos. This is a valuable undertaking that falls under the general rubric that Albert Einstein called gedankenexperiment, or thought experimentation, and can help broaden our thinking even in the absence of First Contact. Indeed, this overlaps strongly with the venerable tradition of high-end science fiction, which contains a plethora of deeply-related scenarios.

Nor is it improper or impatient to create exemplars of contact art while we wait. The best example of such art, which served the purpose of exciting human imaginations without taking untoward risk, was the "calling card" placed aboard each of the *Voyager* spacecraft, back in the 1970s (Sagan, 1978). These symbolic gestures did not appreciably increase our detectability cross-section. Moreover, no one can deny that the salutary and inspirational value of the *Voyager* exercise far exceeded its modest cost.

Just thinking ahead can have benefits that pay wonderful dividends. The conceptual foundations that are being laid down may someday prove invaluable, should Contact come – as it probably will – by complete surprise.

## 22.9 Gambling Our Posterity

This chapter has been, I freely admit, a lot of fun to write. Despite many years spent professionally contemplating the notion of alien life, in a myriad variations, I personally don't expect Contact to happen in my lifetime. When it does, I hope and predict that our grandchildren will be a whole lot wiser and far better able to deal with it than you or I. Our top priority should not be rushing toward Contact, but preparing our heirs to be ready for it.

A parallel might be the way we sometimes screen our calls, listening to messages instead of answering right away. What we almost never do (past the age of 12), is just punch random numbers into the phone, jabbering at anyone who happens to be out there, telling them our names and where we live. We certainly don't go roaming about, shouting, in the darkest part of an unknown town.

Optimistic scholars may be right that we have nothing to fear from that eventual encounter with wise beings from the stars. Still, we cannot be reminded often enough to look back on our own history of contact among humans here on Earth, a litany of dire cautionary tales. We are, all of us, descended – only a few generations back – from folk who suffered horribly because they weren't ready for the challenges brought on by new vices, new technologies, new diseases, new ideas, new opportunities, new people. And those ancestors were the lucky survivors! Many peoples and cultures – including every species of hominids other than our own – left no descendants at all.

How ironic that this reminder should come from someone who is a dedicated believer in the new! Ironic, and yet somehow apropos. For I would rather bet on a horse that I know – human improvability and progress – than on salvation from some hypothetical super-beings high above. We have tried that route, countless times before, and the lesson has always been that we should rely (mostly) on ourselves.

In this chapter, I've only touched on just a few of the dangers conceived by various gloomy thinkers and writers over the years. I could go on, but a complete listing isn't necessary. What matters is the lesson, one of circumspection and caution. The worst mistake of first contact, made throughout history by individuals on both sides of every new encounter, has been the unfortunate habit of making assumptions. It often proved fatal. Let us hope it is a habit that we, or our grandchildren, manage to break. If so, we may pass a crucial test when the time comes to meet and greet beings from the stars.

## References

Alcock, John. 2001. *Animal Behavior: An Evolutionary Approach*, 7th edition. Sinauer Assoc.

Fehr, Ernst and Gachter, Simon. 2002. Altruistic Punishment in Humans. *Nature* **415**, 137–140.

Sagan, Carl et. al. 1978. *Murmurs of Earth: The Voyager Interstellar Record*. New York: Random House.

# 23

# *L*: How Long Do They Last?

Seth Shostak,
*Senior Astronomer, SETI Institute*

The Drake equation, a commonly-used starting point for discussions about the likelihood of finding extraterrestrial intelligence, is now nearly a half-century old. It dates from 1961, a year after Frank Drake made his pioneering radio search for artificial signals from other worlds. That search, dubbed *Project Ozma*, was a 200-hour scrutiny of two nearby, Sun-like stars for transmissions spectrally situated near the 1420 MHz line of neutral hydrogen, and was conducted with an 85-foot antenna at the National Radio Astronomy Observatory in Green Bank, West Virginia (Drake, 1960; Kellermann and Seielstad, 1986). These efforts to find easy evidence of intelligence in other star systems provoked considerable public interest, including a major article in *Saturday Review* (Lear, 1960).

As a sequel, Drake organized a two-day conference a year later searching for sentience in the Galaxy – the so-called Green Bank Conference. The invitees comprised approximately 10 astronomers, biologists, and technical specialists. As a conference agenda, Drake composed a simple, linear equation (Drake, 1965) for estimating the number of galactic civilizations that are sending signals we could detect. The last term in this famous formula is *L*, the lifetime of a signaling society. *L* is *sui generis* among the equation's factors for two reasons:

1 It is dependent on sociology, not astronomy or biology (the only other term that is similar in this regard is $f_c$, the fraction of intelligent species that develop a technical civilization).

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_23, © Springer-Verlag Berlin Heidelberg 2011

2 It is arguably the term that we know, and perhaps *can* know, least about. At a conference in 1971, Carl Sagan noted that in trying to evaluate the terms of the Drake Equation: "We are faced … with very difficult problems of extrapolating from, in some cases, only one example and in the case of *L*, from no examples at all. When we make estimates we cannot pretend that these values are reliable." (Sagan, 1973).

[This is a daunting caveat. It has not, however, squelched speculation on the value of *L*. The fact that these estimates *are* speculative can be gauged by the degree to which they differ. In a compilation by Steven Dick, published estimates for *L* range over five orders of magnitude (Dick, 1996).

Clearly, the chances of finding a signal with SETI experiments depend strongly on the value of *L*. As an example, if the invention of nuclear weapons is always nearly simultaneous with the development of radio and laser technology (as is the case for *Homo sapiens*), then it is seductive to argue that when a species is technically mature enough to make its presence known from afar, it is also ripe for effecting its own destruction. In that case, *L* might be only a few centuries or less, and the opportunity for intercepting a signal is very limited. Having some inkling of what *L* might be – even if that estimate has an uncertainty of a magnitude or two – is significant in motivating (or perhaps demoralizing) those seeking evidence of intelligence elsewhere.

The other reason for considering the value of *L*, quite independent of SETI, is that as a matter of self-interest, it's clearly of consequence to know if our species – or at least our culture – can reasonably hope for a long future.

In this chapter, we will consider some of the suggestions made, primarily with sociological arguments, for a short *L*, and then ask if – even granting a society the good fortune to escape self-destruction – what would be the limits on *L* imposed by external factors. It will be our contention that, in fact, the short-term threats posed by our own activities might be rendered ineffective, and that, on the basis of our own likely future, *L* could be $>10^6$ years.

## 23.1 Relevance to SETI

The Drake Equation estimates the number of contemporary signaling societies as the product of the rate at which they are born and *L*, their lifetime in a transmitting state. In gauging what value the former might have, computed as the product of all six factors preceding *L*, the 1961 Green Bank Conference attendees estimated that it was of order unity. In other words, detectable galactic civilizations were believed to arise at the rate of approximately one per year. (Other estimates, as compiled by Dick, are not always this sanguine, and dip as low as $10^{-3}$ (Dick, 1996)).

The most sensitive SETI experiments, so-called targeted searches, carefully examine plausible, individual star systems. Project Phoenix, the most comprehensive radio search of this type, spent a decade observing somewhat less than 1000 stellar targets at microwave frequencies (Tarter, 1997). Assuming even the optimistic Green Bank estimate for the rate at which technical societies are born, $L$ would need to be of order $10^8$ years for Project Phoenix to have booked a success, assuming that all stars are equally likely to shelter intelligence. In the coming decades, new radio telescopes will be able to extend the target list by three orders of magnitude. Even so, in order for this far larger search to have a high probability of detecting a signal, $L$ must have a value approaching $10^4$–$10^5$ years (Shostak, 2004). Ergo, if $L$ is very significantly less than this, the chance that targeted searches of the foreseeable future will uncover extraterrestrial sentience should be rated as small.

Before proceeding to consider estimates of $L$, we note some restrictions on its relevance to SETI.

1 It is clearly dependent on the technology used for searching. If, for example, societies eventually abandon high-powered radio transmissions in favor of optical communication links, the value of $L_{radio}$ could be short, but $L_{optical}$ might be long. As a current and possibly important illustrative example, the switch from television broadcasting to delivery of content via optical fibers or direct satellite broadcast could greatly reduce our visibility to SETI projects on other worlds, no matter how long-lived our technology. Similarly, communication modes based on physics or technology that are beyond our ken or current ability to easily detect would each have their own values for $L$ in the Drake Equation.

2 Such considerations imply that estimates of $L$ based on the length of time that a society survives (and thrives) beyond technological puberty might be sociologically interesting but irrelevant to SETI. Advanced societies could be "there", but not in a broadcasting state, as defined by our current abilities to find them.

3 In either of the above cases, the estimated value for the average technological lifetime could overestimate the chances of making a detection. There are other scenarios in which the lifetime of a civilization will be an *underestimate* of $L$ as germane to SETI. It has been noted (Soter, 2005) that there have been dozens of major civilizations in the history of humankind (e.g., the Mycenaean, Roman, etc.), and these have a typical longevity of 400 years. The relevant value for $L$ is not the average lifetime of any of these civilized epochs, but their sum. Note that this might be substantially different than the species lifetime, as these high-level periods could be intermittent.

4 Another circumstance in which the lifetime of a technological species underestimates the time during which it might be detected by a SETI experiment is if that society constructs transmitting hardware that outlives its makers. This idea was famously exploited in the film *Forbidden Planet*, in which the Krell, the erstwhile inhabitants of a distant world, constructed self-repairing apparatus that continued to function long after they were gone. It was also implicit in the movie *Alien*, in which a transmitting beacon attracts visitors to a planet populated only by eggs.

5 The Drake Equation assumes that each transmitting society arises independently, and forever remains in its natal star system. If interstellar colonization is practical and sometimes undertaken, this assumption would be violated, and many transmitting sites might eventually derive from a single instance of a technological society. The extreme extension of this idea would be the colonization of the entire Galaxy by a small number of civilizations (possibly even one), a circumstance not accurately gauged by the Drake Equation (We note a variation on this scenario known as *panspermia*, in which simple life is widely dispersed throughout the Galaxy via rocks kicked off planets by impacts. This would greatly change one of the least known terms of the Equation, $f_L$, the probability that a suitable planet will evolve life.)

Having noted these limits to *L*'s applicability to SETI, we consider what estimates have been made.

## 23.2 *L* is Small

In all previous attempts to estimate *L*, researchers have tried to extrapolate the one technological society we know, our own. We have been transmitting powerful, high frequency signals – the type that our own SETI experiments could find if they were coming from another star system – since the Second World War. The one example of a technological society we have has a value for *L*, so far, of about 60 years.

Almost every approach to *L* has been an effort to extrapolate from this limited baseline to predict the long-term consequences of our own activities. And most of these analyses have focused on catastrophe: how long will it be before we do ourselves in via nuclear war, pollution, destruction of the environment, exhaustion of our energy and mineral resources, or just having too many children? The long-term outlook for a society in which progress is both accelerating and, in some measure, frightening (viz: the brouhaha over stem cell research) suggests to many that for human society, *L* may be short.

Twenty five years ago, Sebastian von Hoerner considered many of the critical societal factors that could end technological society on our planet, and concluded that Armageddon was just over the horizon, less than a century hence (von Hoerner, 1975). Most of von Hoerner's dystopian view was driven by a 2% per annum population growth. Aside from the obvious crush of humanity, this growth, if unabated, would provoke an unsustainable pressure on food and energy reserves. And, von Hoerner cautioned, despite the optimistic scenario often portrayed in fiction, interstellar colonization cannot hope to solve the problems created by a rapidly swelling population. When, in 1972, von Hoerner wrote his treatise, the doubling time for the world's human inhabitants population was 35 years. In fact, and as was earlier pointed out (von Foerster et al., 1960), the population growth at this time was actually *hyperexponential*, with the rate of increase itself increasing. A straightforward calculation shows that this growth would lead to an infinite population by the year 2027, surely an untenable (and uncomfortable) situation.

Since exponential – let alone hyperexponential – increase will quickly outrun every resource, von Hoerner's simple point is that this growth will inevitably break down, either because we voluntarily put on the reproductive brakes, or because of external circumstance. He notes that the food supply is a critical resource that imposes a stringent limit in the face of a rapid swelling of population. If every square inch of land were planted with wheat, humanity would still starve by about 2025.

That's more or less the date at which von Hoerner figured we run out of energy. But while substitutes for fossil fuels can be found, he pointed out that the generation of waste heat – from whatever source we exploit to power our lifestyle – will set a strict limit on our activities. He assumed that we cannot have an average global temperature rise of more than ~1 C without severe climate change, and this sets a limit on energy consumption that's about 300 times greater than the world-wide total in 1972. With population growth at the levels of that time, we will hit this limit by 2054, even if we find all the oil we ever need. In fact, China and India, with more than one-third of the planet's population, now consume, per capita, approximately one-tenth and one-thirtieth the amount of energy used in the United States. Consequently, and assuming that one succeeds in raising the living standards of most of the world's peoples to parity with the US, von Hoerner overstated the amount of time remaining until the waste energy limit is reached. It's interesting to note that he anticipated the threat of global warming without anticipating its proximate cause – greenhouse gases.

Finally, using an argument based on a simple probability calculation, and noting that there is a greater chance of a fatal misstep every time a new weapons system comes on-line, von Hoerner was led to expect devastating nuclear war within 40 to 80 years.

It was a one-two-three punch, leading to a societal knockout. Having delineated the problem, von Hoerner treated a solution that many people assume is both obvious and effective: the expansion of our civilization into space. We should simply get the majority of humankind off the planet. He shuts this idea down immediately, by pointing out that even with the population growth of the 1970s, we would have to send 200,000 people a day to the launch pads to prevent the indefinite swelling of human protoplasm on Earth.

Since, as von Hoerner states "medicine will always come before nuclear engineering," population pressures will always precede any ability for interstellar travel, and the problem of short lifetime that he predicts for us – one or two centuries at best – will also apply to extraterrestrials.

A similar conclusion, predicated on a somewhat different analysis, was reached by Lemarchand (2004). He first pointed out that Sagan (1980) defined a technological adolescent age when a society has the ability to exterminate itself, and then makes the barely controversial statement that we've entered such a period. Lemarchand then tried to estimate how long we'll be in this precarious position before reaching a more stable, safer, technologically mature age. To do so, he appealed to historical timescales for major societal transitions, and noted that these are typically a century or so. For example, the world population began a sharp rise in growth *rate* in about 1960 which is predicted to abate by 2050, a century later. He pointed out that the time required for the world-wide shift to democratic governments is similarly a century or so. On the basis of such long-term societal transitions, Lemarchand figured that our situation is precarious for the next 150–200 years, and therefore unless we change our social behavior, we "have a high probability of becoming extinct" within that interval.

Such somber predictions have become less frequent in recent years, largely as a result of sociological developments. The growth in population that was the principal driver of von Hoerner's analysis has lessened. It was then above 2% per year. It is now approximately half that (see Figure 23.1, and http://www.census.gov/ipc/www/world.html) and is projected to drop by another factor of two by 2050. This suggests that the total world population will reach a peak of about 9 billion at mid-century, and may decline after that. The apocalyptic scenarios predicted by von Hoerner, driven by hyperexponential growth, seem to have been written out of the 21st century script, at least.

The other development that has served to rescue humanity, at least temporarily, has been the end of the Cold War. In Figure 23.2 are reproduced the readings of the Bulletin of Atomic Scientists' "Doomsday Clock." Note that we are somewhat farther from the apocalyptic hour of midnight than previously. Of course, the possibility of nuclear war may vary strongly on short time scales, so the current lessening of menace might be only a

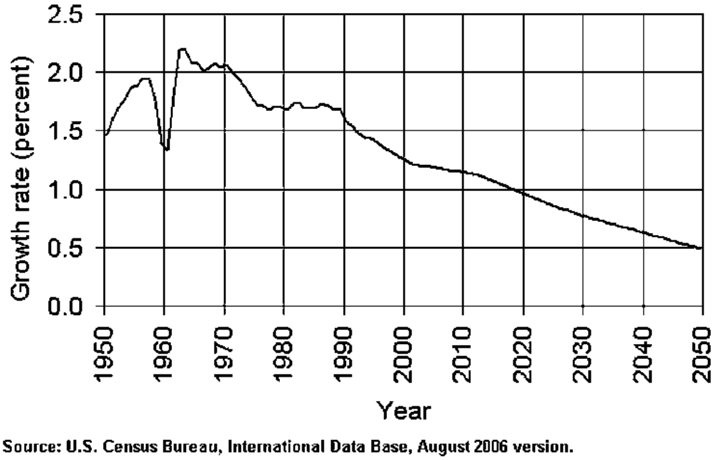



**Figure 23.1** World Population Growth Rates: 1950–2050.

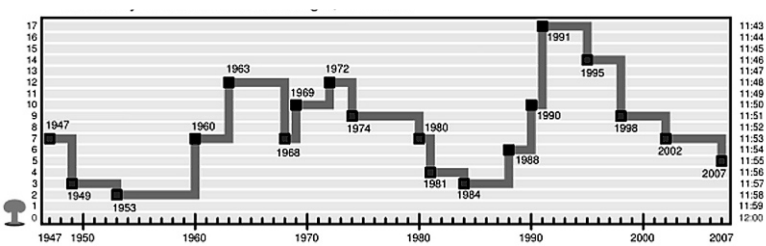



**Figure 23.2** Domesday Clock: Minutes to Midnight 1947–2007.

temporary respite, and such catastrophe need only happen once to vindicate those who suggest that our society is doomed to a brief future.

Nonetheless, it is plain that we live in dangerous times, arguably the most perilous since the emergence of *Homo sapiens*. But could we really wipe out humanity *entirely*? The greatest catastrophes in recorded history were the epidemics of Black Plague more than half a millennium ago, a pestilence that killed about a third of Europe's inhabitants. Nonetheless, those terrible events barely register as a blip on the growth of world population. Nuclear war, and even nuclear winter, in light of various analyses, would seem to be less than 100% efficient in destroying all humanity.

In short, while many have suggested that we are doomed to destruction within a handful of generations thanks to our own activities, these arguments are scarcely hermetic. Other scenarios deserve our consideration.

## 23.3 *L* is Large

As noted, it has been fashionable to project a dystopian future in which our species snuffs itself out only a few centuries after developing the means for interstellar communication. This unfortunate future has been generalized to societies on other worlds, and *L* is thereby estimated to be small (less than a millennium.)

However, in this section we will hypothesize that the apocalyptic scenario of death at our own hand might be stayed. If so, would that guarantee a large value for *L*? Or are there other factors that would rapidly eliminate our species despite good behavior? Put another way, are there non-societal constraints that will keep us, or them, from being technologically active for at least tens or hundreds of millennia?

One possible limit is the biological lifetime of our species. This doesn't seem to enforce a low value for *L*, however. When considering the fate of long-lived habitants of Earth, we find that while individual species have typical lifetimes of $10^6$ years, some orders and classes (trilobites, sharks, cockroaches, and even dinosaurs) have survived for $10^8$ years or more. Successful species are often opportunists and generalists, eschewing narrow ecological niches. They are geographically widespread, and can make use of a variety of resources.

*Homo sapiens* is, of course, both generalist and widespread. There is no obvious, compelling biological argument why humans could not last many millions of years.

What about various cosmic catastrophes? Our predecessors survived ice ages, but what about a future asteroid collision, such as eliminated the majority of species 65 million years ago? While we might be vulnerable to such destruction today, systematic observing programs are currently increasing our knowledge and surveillance of these dangerous projectiles (Morrison, 2006). With advance warning, they could be diverted.

Global warming and other near-term environmental concerns may be serious, but are at least tractable and amenable to cure. The long-term external factors that might limit Earth's habitability, for instance the massive distortion of the carbon dioxide cycle by the warming Sun or the eventual death throes of our host star, are at minimum $10^8$ years in the future, and more likely $10^9$ years (Caldeira and Kasting, 1992). In addition, they can, at least in principle, be circumvented with engineering fixes in the former case and emigration to a nearby star in the latter.

Astronomer Ray Norris has considered two potentially lethal phenomena from far beyond the Solar System that might set upper bounds on the durability of life, intelligent or otherwise (Norris, 1999). According to Norris, nearby supernovae should severely afflict a planet at intervals of roughly 200 million years. Lethal gamma-ray bursters are expected to sterilize a planet about as often. The fact that there has been an unbroken reign of life on Earth for 4 billion years, or 20 times the mean interval between occurrence of these catastrophes can, according to Norris, have only two explanations. Either (1) we are extraordinarily fortunate, beating enormously long odds against destruction (which would mean that we are likely alone in the Galaxy, and therefore *L* is of no particular interest) or (2) the estimate of 200 million years between deadly events is wrong. Assuming the latter explanation (which, and with all deference to Norris, seems more plausible), then we can better estimate that the average interval between such explosive disasters is at least 4 billion years – the duration of life on this planet so far. And since there would be variation in this number, a typical civilization could last for billions of years. Norris's argument boils down to noting that 4 billion years of uninterrupted life on Earth implies that natural catastrophes don't set a severe limit on *L*.

Neither biology nor cosmic interference seem to mandate low values for *L*. But non-destructive technology might introduce complications that could change the ground rules of our existence. We might deliberately modify our species, or threaten its role on Earth by introducing a manufactured competitor.

To begin with, there is a general expectation that we will, sometime in the 21st century, begin to direct our biological future, and disrupt the march of Darwinian evolution. One can easily envision two major developments in this regard: (1) the manipulation of human DNA to produce individuals that have greater talent and are free of inherited disease (and perhaps eventually, the malaise of mortality), and (2) the development of implantable, technical aids to improve body performance, for example eyesight, hearing, and thought.

Such improvements to our species might not affect *L*. But taking these developments to their logical – and some would say inevitable – conclusion, we might expect the creation of true, artificial sentience: thinking machines (Moravec, 1999). If this occurs, the curtain of human dominance on this planet is likely to drop quickly. The improvement of machines, which can, after all, proceed as a Lemarckian rather than Darwinian process, is much faster than biological evolution. Digital electronics currently enjoy an exponential growth in functionality, doubling in capability per unit cost each 18 months, a phenomenon known as Moore's Law (Moore, 1965). At this rate, an artificial intelligence device that is equivalent to the brain power of a single person, will improve to the point of outstripping the cerebral capability of the entire human population within 50 years. The possibility of this rapid change-over from wet, biological brains to dry, technological ones has led Ray Kurzweil and others to speculate about an impending "singularity" in the history of our species (Kurzweil, 2006).

Would such developments, which seem imminent to some, produce a large or small value for *L*? There is no convincing answer we can offer to this question. On the one hand, a machine-run society might be less aggressive, and therefore less susceptible to certain types of self-ruin. It might remain "communicative" for long periods of time, resulting in large *L* according to the definition of this term implicit in the Drake Equation. On the other hand, intelligent machines might be in less need of, or have less desire for, the sort of communication that would make them detectable at a distance, shortening *L*.

Such musings, while interesting, are also highly speculative. In truth, and quite obviously, we cannot predict what artificial sentience – successors to our own species – might do.

In addition to the disruption that such technical developments might provoke, the value of *L* might be affected by a SETI detection itself. Some SETI practitioners have argued that the ability to communicate with other societies might forge durable civilizations by promoting interstellar intercourse and a transfer of knowledge and social norms that are useful for long-term survival (Billingham et. al., 1979). Indeed, even a single instance of contact between two star systems would surely encourage substantial effort to find more, thus quickly fostering a growing communications web throughout the Galaxy. This has led these investigators to speculate that *L* might be as great as $10^9$ years, although one could be justifiably suspicious of an argument by SETI researchers that their endeavors can reward humanity with a billion years of continued existence.

The major technical developments described above could end the unchallenged reign of *Homo sapiens*, although not necessarily the visibility of earthly intelligence. Biological engineering removes our species from the slow and uncertain path of Darwinian evolution. Success in developing artificial intelligence would replace living beings as the prime repository of intelligence in its home star system. The third – exchange of information with an existing galactic club – might be initially disruptive (consider the 18th century encounters between James Cook and the South Sea Islanders) but could ultimately prove transformational in a positive sense. All three of these possible developments are wild cards in the assessment of *L*.

We have seen that – absent species suicide, and excepting the unpredictable consequences of either re-engineering humankind or exchanging information with other cosmic societies – there is no short-term limit to human existence Ergo, it seems that estimating L really does boil down to guessing what society will do to itself.

However, there is another approach to the problem – a meta analysis that circumvents the uncertainty of all the detailed phenomena that could wipe us out. This approach is that taken by physicist J. Richard Gott (1993), who has estimated the species lifetime of *Homo sapiens* with a calculation largely independent of the socio-economic factors considered by von Hoerner and the technology speculation given above. His method was supposedly inspired by a visit to the Berlin Wall, during which he wondered how much longer that onerous edifice would remain standing. The approach itself is described as an application of the Copernican principle. Copernicus was the first to demonstrate that our spatial position in the universe is unremarkable. The analogous assumption made by Gott is that the person inquiring about the duration of humankind as a species is not special in *time*. While sometimes gracefully monikered as the Copernican principle applied to time, this approach is also known as the principle of indifference, a term derived from probability theory (Keynes, 1921).

Gott's argument, in its simplest formulation, is as follows: Suppose that the total lifetime of *Homo sapiens* is $L_s$ (species lifetime, not technological lifetime). Then if the probability of being alive today is equally distributed from the origin of the species until its demise, then we can trivially say with 95% probability that we are living somewhere between 2.5% and 97.5% of the distance along the span of time $L_s$. Thanks to the dusty labors of paleontologists we know that *Homo sapiens* has already strutted across Earth's stage for 150,000 years (Lewin, 1997). This means that $L_s$ must range from $150{,}000/0.975 < L_s < 150{,}000/0.025$, or that we have a future as a species (and possibly as a technological species) ranging from 3800 to 6.0 million years.

One notable refinement to this argument derives from the obvious observation that, in a world with a rising birth rate, the assumption that one's chance of being born is uniform in time during our species' first 150,000 years (and more important, will remain uniform in the future) is clearly wrong. There are many more births today than, say, 10,000 years ago. If we take the more reasonable approach of assuming equal probability of "birth order" – that is, an equal chance that our name would appear at any place in the complete list of human births – then we can recalculate $L_s$ as follows. Define the total number of humans who will ever live as $N_o$, and the number that have been born so far as $n$. Since the chance is the same of appearing anywhere in the birth list, we can say that we are in the last 95% of all humans to be born if $n/N_o > 0.05$. The total number of humans to have lived so far is estimated to be ~100 billion (Haub, 1995), and to be in accord with the likely circumstance that we haven't won the lottery and appeared by chance in the first 5% of the birth list, this means that the total that will ever live is 2000 billion, with 95% probability.

How long will it take for this number of souls to strut Earth's stage? If we stabilize our planet's population at 15 billion, and extend human lifetimes to 100 years, this total will be reached in another 13,000 years If we manage to subdue our more destructive impulses, and perhaps colonize nearby space, we might dramatically increase our population rather than merely stabilizing it, and this number will become shorter. In either case, this simple reckoning suggests that the majority of our species' lifetime is over, but that our technological lifetime has just begun, in sharp contradiction to the shorter, more pessimistic estimates based on socio-economic factors.

## 23.4 Passing Through a Bottleneck?

We have seen that, if the dismal, albeit trendy, apocalyptic scenarios of war, environmental degradation and short-term cosmic threats can be thwarted, our future might be anything from thousands to million of years. However, even with this sunnier prognosis, there is little doubt that – sooner or later – we will be obliged to move at least some of our population into space. The Earth, being spherical, has the minimum surface area for its mass. Resources – both the obvious ones such as arable land, as well as the less obvious ones, such as platinum – are finite, and in many cases already scarce. So, putting aside the possibility that, by engineering our own successors or joining the "galactic club" we may introduce a major discontinuity in the story of *Homo sapiens*, there's one reasonably reliable expectation we can have for our activities of the next hundred years: the expansion of habitat to the nearby, extraterrestrial realms of the Solar System. This settlement of a new frontier could have a telling, and salubrious effect on the earthly value for *L*.

We have visited the moon, and our mechanical proxies have landed on Mars. Both worlds could be colonized, and in the case of Mars, made more amenable to life (Wood, 2007). That this will happen is less a question of "if" than "when." While the initial colonies will be small, historical analogs suggest that within a century they will have populations measured in the tens of thousands or more.

The carrying capacity of these nearby bodies is limited. However, their populations could be dwarfed by the numbers of humans living in orbit. More than two decades ago, Gerald O'Neill (1977) and Thomas Heppenheimer (1979) described in detail how we could build artificial habitats in space:

slowly-rotating aluminum cylinders, having diameters of several kilometers, that could house entire villages and towns. Their prediction was that by the 1990s, millions of Earthlings would be living in these space habitats. That hasn't yet happened, but not because it's technically impractical. Rather, at the moment, building such artificial cities in orbit is economically and politically impractical.

In the somewhat longer view, perhaps one to two centuries hence, we can consider colonizing the larger bodies of the asteroid belt.

While the exact time scale of these projects is subject to the vagaries of political will, one can conservatively foresee that within two centuries, at most, enough of us will be off the planet – in O'Neill colonies, on the moon and Mars, and burrowed into the asteroids – and that total annihilation of human society will be as impossible as the total annihilation of Earth's ants. We will be dispersed, and dispersal is the ultimate insurance policy for survival. Modest colonization will inoculate us against self destruction. It might be possible to exterminate all the individuals in one habitat, but not the entire populace of all habitats.

A similar bottleneck – during which a civilization has dangerous weapons, but is still confined to a small chunk of real estate – will presumably be encountered by most intelligent, technologically developed species. Since the time scale for getting through the bottleneck is small, one or two hundred years, many societies will manage to do so. In this view, the doomsday scenarios so popular in the literature, and which have been so influential in estimating low values for $L$, are unrealistically pessimistic.

## 23.5 Conclusion

It seems that a reasonable alternative to the various doomsday scenarios that foretell our own destruction is the possibility that humankind is passing through a "bottleneck." The development of powerful weapons and the pressures of a rapidly growing population have produced this constriction. But this risk is short-lived compared with the time scale of human evolution.

Clearly, estimates of low $L$ are reactions to social developments associated with the bottleneck that any society will enter once it has developed sufficient technology. But as the bottleneck is short, many – possibly even most – civilizations will pass through. Once dispersed and no longer vulnerable to total annihilation, they might, like some other species, remain viable for $\sim 10^8$ years or more. However, we note that there are three possible near-term developments that might affect this scenario in unpredictable ways:

1 The use of genetic manipulation to re-engineer the species.

2 The development of machine intelligence.

3 Communication with other galactic societies.

Setting these aside, we argue that the suggestions that $L$ is short ($< 10^3$ years) are unduly pessimistic, and suggest that the very technology that threatens us will soon alter our situation such that extinction of our species becomes impossible. The less threatening future that lies beyond the bottleneck becomes attainable by our (and their) dispersal into nearby space. This accords with a view of a Galaxy that hosts long-lived civilizations, societies that may have established mutual communication networks, and in so doing, brought many worlds to the technological level of the most accomplished member.

## References

Billingham, J., Oliver, B.M. and Wolfe, J.H. 1979. A Review of the Theory of Interstellar Communication. *Acta Astronautica* **6**, 47–57.

Caldeira, K. and Kasting, J.F. 1992. The Life Span of the Biosphere Revisited. *Nature* **360**, 721.

Dick, Steven J. 1996. *The Biological Universe: The Twentieth-Century Extraterrestrial Life Debate and the Limits of Science*. Cambridge: Cambridge University Press.

Drake, F. 1960. How Can We Detect Radio Transmissions from Distant Planetary Systems. *Sky and Telescope* **39**, 140.

Drake, F. 1965. The radio search for intelligent extraterrestrial life In G. Mamikunian and M.H. Briggs (Eds) *Current Aspects of Exobiology*, Pergamon Press, 324.

Gott, J. Richard III. 1993. Implications of the Copernican Principle for our Future Prospects. *Nature* **363**, 315–319.

Haub, Carl. 1995. How Many People Have Ever Lived on Earth? *Population Today* **2**, 4.

Heppenheimer, T.A. 1979. *Toward Distant Suns*. Harrisburg: Stackpole Books.

Kellermann, K.I. and Seielstad, G.A. (Eds). 1986. *The Search for Extraterrestrial Intelligence*. Proceedings of an NRAO Workshop held at the NRAO, Green Bank, West Virginia, May 20–22, 1985, NRAO/AUI.

Keynes, John Maynard. 1921. *Treatise on Probability*. New York: Macmillan.

Kurzweil, Ray. 2006. *The Singularity is Near*. New York: Penguin.

Lear, John. 1960. The Search for Intelligent Life on Other Planets. *Saturday Review*, January 2, 39.

Lemarchand, Guillermo. 2004. The Technological Adolescent Age Transition: A Boundary to Estimate the Last Factor of the Drake Equation. In R.P. Norris and F.H. Stootman (Eds), *Bioastronomy 2002: Life Among the Stars*. IAU Symposium 213. San Francisco: Astron. Society of the Pacific.

Lewin, Roger. 1997. It's A Jungle Out There. *New Scientist* 5, 30–34.

Moore, Gordon E. 1965. Cramming More Components onto Integrated Circuits. *Electronics* **8:8**, 82–85.

Moravec, Hans. 1999. *Robot: Mere Machine to Transcendent Mind*. Oxford: Oxford University Press.

Morrison, David. 2006. Asteroid and Comet Impacts: The Ultimate Environmental Catastrophe. *Phil. Trans. of the Royal Society A* **264**, 2041–2054.

Norris, R.P. 1999. How Old is ET? *Acta Astronautica* **47**, 731

O'Neill, Gerard K. (1977). The High Frontier: Human Colonies in Space. New York: William Morrow & Company.

Sagan, Carl. 1973. The Lifetimes of Technical Civilizations. In G. Sagan (Ed.) *Communications with Extraterrestrial Intelligence*, MIT Press, p. 147.

Sagan, Carl. 1980. *Cosmos*. New York: Random House.

Shostak, Seth. 2004. When Will We Detect the Extraterrestrials? *Acta Astronautica* **55**, 753.

Soter, Steven. 2005. SETI and the Cosmic Quarantine Hypothesis. *Astrobiology Magazine*, www.astrobio.net/news/modules.php?op=modload&name=News&file=article&sid=1745

Tarter, J.C. 1997. Results from Project Phoenix: Looking Up from Down Under. In C.B. Cosmovici, S. Bowyer, and D. Werthimer (Eds) *Astronomical and Biochemical Origins and the Search for Life in the Universe*. Bologna: Editrice Compositori, 633.

von Foerster, H., Mora, P. and Amiot, L. 1960. Doomsday: Friday, 13 November, AD 2026. *Science*, **132**, 1291–1295.

von Hoerner, Sebastian. 1975. Population Explosion and Interstellar Expansion. *Journal of the British Interplanetary Society*, **28**, 691.

Wood, Lowell.2007. As reported in http://www.space.com/scienceastronomy/070623_mars_terraform.html

# 24

# What Will They Look Like?

Jack Cohen

## 24.1 What *Won't* They Look Like?

This is the problem, met eyeball to eyeball (if, indeed, *they* have eyeballs at all!). We can easily see what they won't look like: they won't look like people, or indeed vertebrates – or insects, or echinoderms, or coelenterates… But I called my early lectures – and my book – "*What Does a Martian Look Like?*" and that meant that we could cheat: a beginning could be made with Martians… Possibly, on the one hand, they're bacteria like ours, infected from Earth – or it may be that bacteria from early Mars infected Earth, so that **we** are the Martians! But that's a trick, only applicable to Martians. How can we tell what real aliens look like?

*"What does a Martian look like?"* was scrawled at the bottom of a British Association for the Advancement of Science lecture-title-list by my first boss. He was impressed by my copies of *Astounding*, *Galaxy,* and *Fantasy & Science-Fiction*, scattered over the work bench between the microscopes. We saw the possibilities straight away and, indeed, it was nearly the most popular lecture for the first three years of the Schools Lecture Service run by the Brit AAS (the most popular was *Science in the Detection of Crime*). In an attempt to damp down requests (I did it 35 times in the second year) we decided to change its title to *The Possibilities of Life on Other Planets.* That seemed to let in all those who'd been put off by the gimmicky nature of

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_24, © Springer-Verlag Berlin Heidelberg 2011

the first title. It became number 1! POLOOP, as it was affectionately known, was given more than 360 times from 1960 to 1985!

In that lecture, I could claim – and I did – that the parochials of life on this planet would not be duplicated elsewhere. I still make that claim. Even if we had DNA-based life, the mutations would all be different, selected differently by different geography at least, so that the creatures would all look different from Earth beasts. Even if the play was the same – photosynthesis, flight, grass and trees, carnivores and herbivores, fur and feathers – and the carnivores' eyes faced forward and the herbivores' eyes faced outward – the characters would be different. We would have different body-plans, and different variations.

There were famous biologists then – Conrad Waddington was the best known, I suppose – who believed that the highest life form on any planet would resemble Waddington. His arguments were specious, at best. Certainly they didn't resemble the arguments of a Christian like Simon Conway Morris, a famous biologist of today: Simon believes that convergences will all lead toward the "ideal" highest form – and that such highest forms will resemble each other because they are all modelled on God.

Those of us who don't believe in an anthropomorphic god, however, have problems with this convergence. We don't have trouble with convergence generally – ichthyosaurs look like dolphins look like sharks, because big aquatic piscivores need to be that shape for hydrodynamic reasons – but there are problems with convergence to the human shape. All very well to say, as Waddington did, that two eyes, two arms, etc., are to the good, that the overall size fits in to the biosphere so we can find food and there aren't too many creatures that can easily eat us, that being up on two legs frees the arms/hands/fingers for clever tool-making and, above all, that having a large brain is a pre-requisite for being a top creature. But this starts from terrestrial, and from eukaryote, and from *having* eyes…

Let's start with the origins of life, and see where that gets us as to what they'll look like.

## 24.2 The Origins of Life

Let us suppose – the evidence is moderate, but getting better all the time – that there are many "aqueous" planets out there, like Earth was, and let's follow them through the kind of chemistry that Earth must have had.

But before we leap in, let's get three arguments out of the way. For a start, they are not "Earth-like planets", as Earth is *now*! What are Earth-like planets, then? They are planets that are like Earth has been for most of its 4000-million year history: three-quarters covered with sea, in the sea microbes but not much more, atmosphere generally reducing: nitrogen, carbon dioxide and monoxide, methane, some ammonia, sulphur dioxide, water vapour, perhaps hydrogen, a whiff of oxygen from time to time. Earth got photosynthesis, and oxygen levels went up drastically, about 1000 million years ago (Mya). There had been a couple of previous occasions, when much of the oxidisable ores and gases had been oxidised, when oxygen had replaced CO2 and methane in the atmosphere, and Earth had frozen – because CO2 is a better greenhouse gas. Relaxing out of this frozen-Earth scenario, photosynthesis in the seas made sure that oxygen levels went up – but it was still 550Mya before we had a range of complicated animals, and nothing very interesting on the shores yet!

The second argument to get out of the way concerns the substrate for life. The general consensus is that a whole planet with boiling seas, cooling down, has a great variety of milieux in which life will start, given time. There are some 30 good "stories", ranging from Graham Cairns-Smith's beginning with clays and hitch-hiking into carbon-based life, through Gunther Wachtershauser's iron/sulphide surfaces with attached peptides (primordial pizza rather than primordial soup) which come to duplicate biological, cellular oxidation/reduction chemistry, through Stuart Kauffman's auto-catalytic loops and Doron Lancet's mesobiosis, which make chemical sense and which lead into primitive life forms. We don't have to ask "When does life start?' any more than having to ask "When does a baby start?": it moves from chemistry through recursion and auto-catalysis into mesobiosis, which matures into life – perhaps via an RNA-world, perhaps via a world of peptides, and/or sugars… and/or chemical families we haven't come across because our kinds of life don't use them. So, life will likely start on any aqueous planet. We will be able to check this with Europa, possibly Titan, unless they have been infected from Earth and just have boring old nucleic acids, peptides and carbohydrates!

The third argument is the most pertinent. Some people seem to get this right away and say "of course"; there are still some folk with whom I am having on-going discussions. It starts "Let's imagine that we have a time machine and can go back and 'start Earth again'!". That is, put the dice in the cup again for another throw… Now, one of several things might happen. Because one of the origins-of-life scenarios is *much* more probable than any of the others, the same origin happens again; alternatively, several may be about equivalent in likelihood, in which case another form of life will appear. Perhaps, several people have suggested, something else altogether will happen, not Life As We Know it a bit; I want to forget this for now. Just consider the possibilities like what happened the first time: we get life, but even if it starts out the same, it very soon becomes different. Why? Because, at least, of differences in geography: e.g., this particular little patch of ferrous pizza was inundated with fresh water the first time around, but not the second. Or, because of chemistry: there was a surplus of phenylalanine (an amino-acid) first time around, more tyrosine next time. Or the first little wiggler went off SW instead of NE. Most persuasively (but I don't see why it's more persuasive...) the mutations in the genetic material were all different, because random. So we get a different drunkard's walk of evolution, starting with creatures of much the same grades and kinds – because you have to start simple, and there's probably less ways of being simple – and working. So, whether we started out the same Origin again, or somewhat – or indeed *very* – different… we soon have a different starting-point. Now, what happens from here?

## 24.3 Universals and Parochials

Divergence, both from where they start and how they continue. Nothing, so far as we know, holds them to any particular path. We don't have any evidence about the structure of early life on Earth, a little evidence about its chemistry. The chemical evidence is of two kinds. In very early rocks, there are a few simple organic compounds that would not be expected from non-organic chemistry. And, a whole new kind of evidence, some of the protein enzymes that are common to the most primitive life-forms of today are boiling-water resistant, as if they were evolved in very high temperatures. There are, today, several life forms whose habitat is boiling-water or – in the depths of the sea, where pressure permits much higher temperatures without boiling, there are creatures whose systems can withstand temperatures as high as 140$^{O}$C. It has been suggested that these are not-very-modified descendants of original life-forms, all the rest of us having adapted to the chilliness of the regular oceans. There were, quite early on – 3500Mya – aggregations of ancient algae and oddities, layered piles just like the stromatolites we find today in Australian shallow seas – and there are fossil stromatolites right through the ages.

A word here about primitiveness. "How clever you are, teacher, to have maintained that Amoeba for thousands of millions of years… while the rest of us were all evolving!" Indeed not. The amoeba culture I have at home has many more generations from primitive life forms than I do; they have 240 chromosomes, I have 46. It looks cutely primitive, but I don't know. Think about the ape-like ancestor of Man. Man is not, of course, descended from present-day apes (or amoebas); they are our cousins, not our ancestors. Indeed, the ape-like ancestor of man might exactly as properly be called the man-like ancestor of the apes! We both diverged from that creature. There isn't a pure line, starting with those boiling-water creatures and coming up to Man, with all the other creatures branching from it: a Christmas-tree version of evolution. We might even go as far as Christopher Dobell, and talk not only of the amoeba-like ancestor of man but of the man-like ancestor of the amoeba! In some biochemical aspects I'm sure that we're more like that common ancestor than the amoeba is!

So, divergence, from the earliest creatures. And, on another planet – or a re-run of life on this one – divergence away from the particular patterns that life took here. Does that mean that we cannot argue from these different evolutionary trees one-to-another? Yes, in every detail, it does.

But, in generalities, perhaps not. There are vast categories into which life-forms can be divided, and this is not – totally – ancestry-dependant. Many of this planet's creatures employ photosynthesis for their energy needs; we call them plants, more or less, as distinct from animals, that eat plants. Among single-celled creatures, there are many that can swap lifestyles, being plants or eating plants. There are also many animals, from flatworms and corals to giant clams, that keep algae in their tissues and live on them, much as plants live on their chloroplasts – which were once separate living organisms. Now here's the logical jump. We might suppose this swap from plant to animal and back again having happened many times on Earth, that it would happen similarly on a re-run Earth or a parallel Earth, another aqueous planet. If so, and this is the bit of logic I love, we can argue from *our* evolutionary tree, in general terms, to others: first to the re-run Earth, then as a next logical step to other aqueous planets' evolutionary trees.

Let's see where that gets us. Think first of photosynthesis. Had there been but one kind of photosynthesis on Earth, that might have been a lucky happenstance. But in fact there are three common forms: chlorophyll, violet bacteria and green bacteria – and several other oddities and chemosyntheses that involve light. (The photosynthetic mechanism of violet bacteria is as efficient as chlorophyll in some circumstances, and efforts are being made to incorporate it into solar arrays.) There must have been many, perhaps hundreds of other "attempts" at photosynthesis down the ages, which have been lost in time. And equally, on a re-run Earth. And, equally, on a parallel aqueous planet.

So, our aqueous alien planet has probably got an oxygen atmosphere (oxygen's the easiest, by breaking up water, but sulphur is a real possibility). And, by a similar argument, there are flying creatures in it, given time (but note how much time!): bees, birds, bats, even a few fishes invented flight separately. And fur: bumblebees have it so they can warm up for flight, several plants have it to protect from frost; the mammals as a Class have it, to maintain their constant body heat. Sexuality has appeared and disappeared in all of the great branches of the animal and vegetable evolutionary trees, so we would expect to find varieties of sexuality on a re-run Earth, or on our alien planet.

So much for the Universals, then. Photosynthesis, flight, fur and sex – "the four Fs," if you'll pardon my misspelling of the first and accept an expletive for the fourth – can stand in for all the commonalities in evolution, all those tricks, events, functions like eyes and limbs, ears and electromagnetic senses, bloods and blood pigments, exo- and endoskeletons, nervous systems with a head end…

Let's look at a few more of the tricks that were adopted by various evolutionary lines on this planet, before turning to the converse: cases where this planet has it and no other – not even the re-run Earth – will. Perhaps the greatest evolutionary advance on this planet was the aggregation of several bacterial-grade creatures to form eukaryotes, creatures with nuclei in their cells, "evolutionary symbiosis". And, as by now you've come to expect, *different* aggregates produced the red algae, the different kinds of fungi, animals and plants, and various unicellular parasites. So that will happen, variously as it did here, on a re-run or on another planet. And the sponges, perhaps the nematodes, became cellular in different ways from the rest of us. And, of course, intelligence appeared in several branches of the animal kingdom: dimly, in the insects, but fairly brightly in the crustaceans (some mantis shrimps are as bright as rats), very brightly among the molluscan octopuses and squids, some fishes and reptiles but mostly in the birds – especially parrots (not owls, who are pretty dim!) and of course the mammals. Mammals have made a speciality of intelligence, culminating (we would like to believe) in we apes. Intelligence, then, will appear on the re-run Earth – and on alien planets – but it will likely only be the kind of intelligence seen in the mantis shrimp or the cat: learning cleverly as it goes through life, but not sharing with other creatures even of the same species.

For extelligence, present in a few mammals like wild dogs and meerkats, chimps and – especially – people, who share their knowledge and (for people), build a social capital into future generations, who knows? Is it just in one evolutionary line, the mammals? Are we justified in finding several kinds of extelligence and then arguing to the re-run and the alien? The insects don't do anything like it. No, just one case! I doubt we will find extelligent aliens, on a re-run Earth or on our alien planet.

I think, indeed, that extelligence is an example of the other great class of evolutionary events, the Parochials ("only in this parish…"). Because of the ubiquitous morphological and physiological divergences, we have thousands of kinds of creatures. In the Burgess Shale, 570 Mya, there was nearly as big a spread of marine creatures as we have now. Our ancestors were there, perhaps *Pikeia,* and there were many other representatives of different phyla and classes. The later, 400Mya–300Mya, divergences on land were at least as dramatic. But, on a re-run Earth, there won't be the same divergences. Equivalent, yes of course (unless there's a move towards a case where one species, in all its stages, takes over; see *The War with the Chtorr*, by David Gerrold, where we exploited that idea), but totally different. No arthropods, no molluscs, no vertebrates. So the vertebrate plan is a parochial, this planet and no other. Molluscs only here. Echinoderms nowhere else. All parochials.

And of course, vertebrate, mammal, primate… all parochial. So much for Star Trek "aliens"! We certainly won't find an extelligent creature with airway-crossing-foodway (because that fishy ancestor that came out of the sea had ventral lungs, unlike many other contemporary fishes), or with five-fingered hands/feet ( because that fishy ancestor that came out of the sea had eight-membered fins, unlike many other contemporary fishes) or with excretory and genital system confused (because that fishy ancestor that came out of the sea had them sharing tubes, unlike many other contemporary fishes) – so we have coughs, and ''dirty" books. Other extelligences, if there are any, will find those puzzling.

So, we are nearly all parochial. What does that say about what aliens look like?

## 24.4 What Do Aliens Look Like?

It lets Star Trek aliens out of the window, as far as realism goes. But they're not intended to be real, biologically. They're intended not to be part of the immediate action, but to be bystanders with attitude; dramatics not biology! Similarly Galactic what-have-you's, Babylon 5; the made-up-people are to put the characters into different groups, with different interests; black, brown and Eskimo are too close to human sensitivities. The creature in *Independence Day* is quite interesting: tentacles, yes, but we can't tell if the things hung about it are appendages, clothes, weapons… or none of the above. *Starship Troopers* has copies of Earthly insects, not very original. A few of the Star Wars creatures are interesting, but most have vertebrate affiliations that take them away from likelihood – just as the dissection of the putative "alien" from 1947 Roswell showed it to be humanoid and therefore *not* alien!

Probably the first "aliens" we will encounter will be alien machines. Interstellar distances are so great, and the uncertainties of time so large, that it's a good idea to send a patient machine rather than a life-form – we think! Unless there really is a cute way of getting there really fast…

So, if not a machine, then what might we encounter? Well, if we go to their planet, and they're not extelligent – the likely scenario, I suppose – we might find almost anything. We won't find anything resembling Earth parochials, that's for sure: no vertebrates, no molluscs, particularly no dinosaurs – unless an alien race took pity on their imminent demise, and had lofted them abroad to another home (Annie McCaffrey was caught by a contract for a *Dinosaur Planet*, and that was the only way we could work it, "real world").

So, intelligence common, but extelligence very rare at best. But there are just so many stars out there, that even at the longest of odds SETI might pick up… it has to be *extelligence*, one intelligent creature can't do the inter-stellar communication bit. It might be insectile, that is to say with the individuals much less autonomous than people, and building a technological culture for thousands of generations.

So, let's exercise our imaginations to the full, and imagine a terribly unlikely creature. Perhaps a strangely backboned water creature with stubby limbs ventures out on land, there are plenty of insects to eat. In the interests of becoming independent of water, it gets cleidoic eggs, then these rest longer and longer in the female; suddenly the shells are lost, the eggs get tiny, and the mother becomes an incubator – this saves a lot of genetic information, all those enzymes for all the different temperatures of development. Making mother an incubator is a lot cheaper in information terms – and the eggs run away with her too! It does mean making her warm, though – constant temperature, expensive, but a nice trick because it makes the brain a good deal more reliable. These mammaloids diversified, and one gang went up into the trees (which had themselves diversified to be all branchy rather than Christmas-tree-like – plenty of lianas too). Here they got the hand/eye tricks pretty good, so that when they were forced down on to the plains one option was to go bipedal, releasing the hands for fine tricks like tools. They got bigger, diversified again; several of them specialised in running, lost their hair – became tribal, so that hunters could share the prey, gatherers could divvy up... the rest is history...! An extelligence, but not one that consciously addressed the stars; accidentally, a sphere of radio waves went out into its galaxy and it listened for others, it listened…

*Editor's note:* You will notice that there are no illustrations in this chapter. The reader is encouraged to exercise his or her imagination, as Professor Cohen has done.

## References

Cohen, J. 1979. *Reproduction.* Butterworths.

Cohen, J. 1989. *The Privileged Ape: Cultural Capital in the Making of Man***.** In V. Serebriakoff (Ed.) 'Frontiers of Thought' series. Carnforth, Lancs: Parthenon Press.

Cohen, J. and Stewart, I. 1994. *The Collapse of Chaos: Simple Laws in a Complex World*. New York: Penguin, Viking.

Cohen, J. and Stewart, I. 2002. *Evolving the Alien: The Science of Extra-terrestrial Life*. London: Ebury Press.

Cohen, J. and Stewart, I. 2003. *What Does a Martian Look Like? The Science of Extra-terrestrial Life.* New York: John Wiley &Sons Inc. (US version of *Evolving the Alien*).

Cohen, J. and Stewart, I. 2003. *Heaven.* New York: Warner-Aspect.

Pratchett, T., Stewart, I. and Cohen, J. 1999. *The Science of Discworld*. London: Ebury Press.

Pratchett, T, Stewart, I. and Cohen, J. 2000. *The Science of Discworld 2: The Globe.* London**:** Ebury Press, Random House.

Stewart, I. and Cohen, J. 1997. *Figments of Reality: The Origins of the Curious Mind.* Cambridge: Cambridge University Press.

Stewart, I. and Cohen, J. 2000. *Wheelers.* New York: Warner-Aspect.

# 25

# Being Technological

Kathryn Denning,
Associate Professor, *Department of Anthropology, and Science and Technology Studies Program, York University, Toronto*

SETI's essential premises involve evolution in multiple domains: cosmology, biology, culture, and technology. Comparatively little has been written about the last of these, technology, in relation to SETI's targets, but it is a crucial variable, and well worth deep examination. In particular, it would seem prudent to consider carefully our assumptions about hypothetical extraterrestrial societies which have developed technology that SETI could detect, or which could detect, at interstellar distances, the existence of intelligent life on Earth. This chapter contributes to that effort by reflecting upon our habits of projecting terracentric assumptions onto hypothetical worlds, exploring dominant narratives about technological development, and presenting varied philosophical theories about the nature of technology. It highlights the cultural aspects of technology here on Earth, particularly their role in the development of radio technology.

## 25.1 Introduction: Ideas About Technology in SETI

This inquiry is motivated by the many noteworthy assumptions and inferences about extraterrestrial technology made in the SETI literature.[1]

[1] I have discussed these previously, in Denning 2010(2006) and Denning 2005

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7_25, © Springer-Verlag Berlin Heidelberg 2011

These have included, for example, the following:

- If sufficient intelligence is present, technology will arise and evolve, as a result of natural selection, and in some cases that technology will be detectable by us.
- Since we have ourselves only recently achieved detectability, if we ever detect an extraterrestrial signal, the originating society will very likely have been detectable for longer than we have been, will be older than we are, will have been developing longer than we have, and thus will have technology superior to ours.
- Since they are more advanced than we are, ETIs will use the most logical and energy-efficient transmission methods possible.
- ET societies which have technology detectable by SETI will also have weapons of mass destruction/the capability of auto-extinction.
- ET societies which have technology detectable by SETI will colonize planets beyond their home world.
- An ET society which is detectable for longer than a brief period must have adapted to its technology's inherent potential for destruction by becoming peaceful, or colonizing other worlds, or dominating its neighbors.

These conjectures have a substantial influence on major debates in SETI today, on topics ranging from optimal search methods to the advisability of Active SETI. And, should there ever be a SETI detection, these theories will surely affect our concept of what or whom we have detected, and thus our collective response. Thus, as I have argued previously, it is worthwhile to think deeply about the underlying assumptions concerning technology (Denning, 2010(2006)).

However, this is no small task, for technology is an integral part of our human experience, and there is also a vast literature on the subject. Therefore, this chapter asks more questions than it answers. I address the stories we tell to explain the development of technology on Earth, the ways in which technology is less than logical, and technology's agency in our world. I then consider what it really means to have radio technology, by working through some aspects of radio technology on Earth. Throughout, I argue that the best place to begin is with the familiar, and so I turn first to our much-spied-upon neighbor: Mars.

## 25.2 The Veil and the Screen: Mars and US

If we turn the clock back a century or so, we arrive in the time of Percival Lowell and his famous “canals of Mars”.[2] Lowell cheerfully assumed a parallel between atoms and Martians; though neither had been directly witnessed, each could be inferred from things which had been observed. Just as the behavior of chemical compounds indicated the existence of atoms too tiny to be seen, so Martian canals implied Martians too distant to be seen. In turn, from the planet-spanning characteristics of that canal system, Lowell proceeded to infer much about the Martians: that they were not only intelligent but also wise, globally united, peaceful, and cooperative, and possessed of technology superior to our own (Lowell, 1985; 1906). Even though Lowell’s detection of extraterrestrial technology was an error (Sheehan, 1988), it is easy to sympathize with his excited speculations. Indeed, if scientists ever again think they see evidence of technology on another world, theories about its makers will surely follow in mere moments. The question “*What is it*?” will quickly be succeeded by “*What are they*?”

Today, of course, we know more about Mars, and so Lowell’s observations of canals and claims about Martians have been relegated to history, just as earlier claims for Lunarians – men on the Moon – faded away as humanity’s sight grew stronger(Sheenan, 1988, Chapter 4). The frontiers of science can be understood as a veil, always moving away from us as our knowledge expands. But a veil does not merely conceal reality. It also gives us a screen upon which we project our ideas of what lies behind it. And those ideas are inevitably derived from our understanding of our own world.

Lowell’s case is just one of many in humanity’s venerable history of projecting our own cultural and technological concerns onto other worlds, in science and in fiction. Mars has been a screen for many such tales (Guthke, 1990). For example, in *War of the World*s, H.G. Wells very specifically drew from the British experience of colonizing other lands and fear of being invaded by other Europeans, and from the emerging theory of evolution (Fitting, 2000), and also from 19th-century ideas of technology, from war machines to telescopes. He didn’t portray the Martians as being *biologically* like us, but he did imagine that they were *technologically* like us, just more advanced. Just as we humans turned our telescopes to gaze at Mars, so Wells imagined the Martians gazing back through similar instruments (Wells, 2003 (1898); Crossley, 2004, p. 83).

Other contemporary fiction used Mars to express different concerns about technology. For example, the Bolshevik author Alexander Bogdanov, writing in 1908, used Mars as the site of a communist world. Bogdanov wrote breathlessly of the building of the Great Canals and the social

[2] For an excellent overview, see Dick 1998.

engineering required to produce them, the massive factories, the relationship of the Martian populace with industry and technology… and, naturally, the essential importance of socialism to the successful Martian way of life (Bogdanov, 1984 (1908)).

One might suggest that we should disentangle science fiction from the scientific activity of striving to know Mars for its own sake… but we cannot entirely. The division between scientist and science fiction author is often fuzzy in practice; one person often plays both roles, both roles require imagination concerning the next frontier, and bold hypotheses and engaging storylines are not so different, particularly in the earliest stages of an inquiry. And indeed, rather like Lowell, Bogdanov and Wells did more than merely project simple self-portraits onto what was then the fairly blank screen of Mars – they used Mars as an allegory (Crossley, 2004), a narrative space in which to explore Earthly concerns, and explain their theories about technology and society. (Both Wells and Bogdanov were influential scholars of history and politics, in addition to being authors of science fiction.)

It is easy enough, now, for us to understand what was happening with these books about Mars from a century ago. But perhaps we can use this example to help us consider what is happening in *our* thinking *now*, about what lies beyond the veil which floats at the far edge of our knowledge today. Our views of other worlds, our characterizations of their inhabitants, and our speculations about technology on those worlds are certainly influenced by our own culture today. We cannot eliminate that influence by stepping entirely outside our own culture, species, or history, but we can and should nonetheless ask, What ideas about technology and society are *we* projecting onto distant worlds? What are *our* hopes and fears about technology? How complete is our understanding of technology on *Earth*? Such a process might help us assemble a better composite picture of the essential truths about the nature of technology. One place to begin is with our tales about technology.

## 25.3 The Stories We Tell About Technology

As human beings, we often tend to regard our species as the necessary, inevitable outcome of biological evolution. We often regard our technology in a similar way.

Historian George Basalla outlined some traditional Western ways of thinking about the evolution of technology (Basalla, 1988). Premises have included these ideas: first, that technological development is discontinuous, and invention results from the heroic efforts of gifted individuals; second, that necessity and utility drive the development of technology; third, that technology develops in a logical progression.

Basalla points out that the first view was effectively refuted by historians in the early twentieth century – e.g., the steam engine was not spontaneously developed by James Watt, but was the cumulative result of many prior inventions, plus some insight (Basalla, 1988, p. 21). However, the second isn't the whole story either, for necessity and utility alone cannot explain the tremendous diversity of human objects and technologies. In fact, Basalla contends that invention is often the mother of necessity, rather than the other way around. For example, the automobile "was not developed in response to some grave international horse crisis or horse shortage" (Basalla, 1988, p. 6). Rather, the automobile was initially a toy for the wealthy, and only slowly became widespread enough that cities changed form and made motorized transport actually necessary. In short, the invention created the need for the technology. As for the third view, Basalla argues that each form of technology does emerge from antecedents – a new thing is always based on an existing thing (Basalla, 1988, p. 45) – but that we should not assume that certain inventions are inevitable, or that the "stream of made things is entirely self-generating" (Basalla, 1988, p. 62). Moreover, he notes, "A talented inventor and a likely antecedent are necessary, but not sufficient conditions to create an innovation with wide social and technological repercussions". Some new technologies take hold, and others lapse into obscurity; the difference lies not just in their utility, but in a host of cultural factors.

Overall, then, serendipity and a hero's flashes of inspiration or genius are still occasionally recognized in accounts of technological development, but the overall story tends to be that one invention led to another, which led to another, in a logical trajectory of progress. We have been encouraged to believe that technologies are created and adopted in rational, logical, scientifically progressive, and economically useful ways. The larger cultural context is left out, incidents of technological stasis or retrogression aren't mentioned, the inventions which have no descendants today are forgotten, and rather little thought is given to technology that might just as easily have been developed or widely adopted, but never was.

This latter point is a particularly interesting one in the historical disciplines: we tend to pay little attention to what isn't here. But why do some forms of technology disappear, while others thrive? Why are some kinds of technology never developed at all? Why did things turn out *this* way, and not *that* way? We cannot assume that it is simply because the surviving form of technology is superior, because history includes many instances where one form of a technology has triumphed in the market over other forms which were just as good or better. Understanding why our society has some technologies, but not others, is akin to understanding why we have some forms of knowledge but not others. This domain of "agnotology" – the study of why we don't know what we don't know – reminds us that culture and politics shape knowledge at every turn.[3]

But what *are* the relevant cultural factors? They range from the fairly localized to forces that span centuries and sweep continents.

On the localized front, we can apply the concept of the "lash-up" – or the necessary intersection of multiple cultural elements necessary to bring an object into the world and keep it here… each element being interdependent with the others (Molotch, 2005, Chapter 1). The example of the electrical toaster is mundane but useful. As Molotch remarks,

> "It does not just sear bread, but presupposes a pricing mechanism for home amperage, government standards for electric devices, producers and shopkeepers who smell a profit, and people's various sentiments about the safety of electrical current and what a breakfast, nutritionally and socially, ought to be…. There is a global system that yields a toaster's raw materials, governments that protect its patents, a labor force to work at the right price, and a dump ready to absorb it in the end." (Molotch, 2005, p. 1)

But there's more to consider: the toaster is a regional phenomenon. British and North American kitchens generally have one, but most of humanity's kitchens don't. Why? The key here is that although toaster technology is useful, its prevalence and distribution is not a result of technical logic or economic rationality; it's also a matter of people's tastes, and those are hardly objective, self-evident, or predictable (Molotch, 2005, p. 3).

On a much broader level, we must consider the role of the capitalist economic system in the production of new technology. Although civilizations have emerged in many times and places, and have shared features such as farming, social stratification, warfare, elaborate religion, urbanization, and specialization of labor (Trigger, 2003), it is nonetheless true that most human societies have had fairly stable collections of technologies for centuries or millennia. Ours, today, is different: as Basalla notes, "No other cultures have been as preoccupied with the cultivation, production, diffusion, and legal control of new machines, tools, devices, and processes as Western culture has been since the eighteenth century." (Basalla, 1988, p. 124). But *why* do we now have such an enormous diversity of technology, with new forms emerging at an accelerating pace? Kurzweil, for example, has plotted this accelerated development of computer-related technologies (Kurzweil, 2005)… but why is this happening? It is obviously interrelated with the global capitalist economy, and the way that the consumer market rewards new technologies; the system is predicated, as Marx said, on the "constant revolutionizing of production" (Marx, 1959 (1850), p. 324). In turn, the

3 www.stanford.edu/dept/HPS/AgnotologyConference.html

global capitalist system emerged not only from the *technologies* of the Industrial Revolution, including the steam engine and textile machinery, but from the *social conditions* present in Europe in the 18th and 19th centuries, and the available natural resources.[4] Further, the political structure of these societies provided economic incentives for inventors, with the result that we live today in a world with an "obsession with technological novelty that is without precedent." (Basalla, 1988, p. 124).

It's clear that our technology obsession is intertwined with our belief systems – with our ideas about what is good, and what humanity is meant to do. Many associate technological innovation with progress itself, i.e., with the improvement of humanity. Novelty is seen as good *in itself*. Now connect that to the notion that we're journeying towards a future golden age, and to the idea that nature is here for us to use: this triad is a reasonable summary of the Renaissance beliefs which spawned modern technological culture (Basalla, 1988, p. 132). Not one of those three beliefs has been universal in human civilizations, and so their intersection was certainly not preordained or inevitable. Nor were the subsequent imperialisms, wars, and military sponsorship of technological development. And now, it has been suggested by some scholars that the modern affinity for technology is, in effect, an "unconscious religion", with a central belief that improving our technology will redeem us, perfect us and bring us salvation (Gerrie, 2005).

Not everyone feels this affinity, of course, or buys into that proposition that technology will save us. Many are the technoskeptics who are deeply suspicious of technology's backlashes. Even the traditional Maya creation epic, the *Popul Vuh*, has a section where grinding stones, the pots and griddles, even the stones from their fireplaces, fight back against and destroy those who used them.[5] The Luddites destroyed machines which threatened their livelihoods, many a novelist has created a dystopia born of the union of our technological brilliance and hubris, and each major new invention, from the atomic bomb to genetic engineering, certainly has its critics. Scholars of human history like Ronald Wright explain the accumulating evidence that civilizations collapse because their technologies contain traps that undo them… and observe that global warming and global hunger suggest that we, today, are not exempt from this trend (Wright, 2004).

What, then, is the true story about technology? I think the only sensible answer is this: the true story is the one that is big enough to contain all the others, a story that explores all technology's dimensions. Obviously, this poses a problem in a discussion of this length, a problem which can only be solved by being highly selective. Therefore, in the following sections, I

---

[4] Many commentators have written about this, including Heilbroner 1967. See Diamond 1997 for related discussion about natural resources.

[5] www.mesoweb.com/publications/Christenson/PopolVuh.pdf

endeavour only to highlight two aspects of our relationship to technology here on Earth. First, I discuss the ways in which our technology is not as simply logical as we sometimes suppose. Second, I describe some ways in which technology doesn't merely do our bidding, but also exerts its own influence upon us.

## 25.4 Technology Is Not Merely (Or Not Even) Logical

Technological determinism says that technological development follows a progressive course, from less to more complex, through a series of necessary stages, and that societies must adapt to their technology. Constructivism, on the other hand, says that it isn't that simple; there are always multiple technological options, technology is flexible, and social contingencies determine which paths technology takes (Feenburg, 2004 (1992)). Most scholars of technology now emphasize the latter perspective.

A classic, well-known example would be that of the computer keyboard: the QWERTY layout we all know is a legacy from the electrical typewriter, in turn derived from the mechanical typewriter. It is not ideal as a computer input device in terms of efficiency or ergonomics, but this is no surprise, since the QWERTY board was designed to prevent jamming in the mechanical typewriter. In turn, it's interesting that in the initial decades of typewriting, there were over 500 types of typewriter in existence[6], and there were unquestionably many factors other than efficiency which eventually conspired to make the QWERTY design the standard, and the prototype for the subsequent electric typewriter and computer keyboard. The convention was conserved for social reasons – that is, because users were accustomed to it.

But the matter of society and technology goes much deeper, and gets much stranger than that, because any given technology can be used in vastly different ways, according to the cultural setting. An interesting illustration is the case of the Kaliai people of West New Britain, in Melanesia, and their response to the 20th-century introduction of Western technology, including telephones, radios, and televisions (Lattas, 2000).

When the Kaliai encountered these objects for the first time, they grasped not the physics, but the principle that these things had the power to cross distances, and to access the invisible world, which the Kaliai considered to be a realm of power. So they made copies of these technological objects out

[6] Royal Ontario Museum, Toronto, Ontario, Typewriters exhibition, August 2007.

of things like discarded tin cans, and used them to communicate with the spirit world and with the dead (Lattas, 2000).

Of course, this case illustrates some processes that occur when a new technology is introduced into a culture, but it's also a powerful demonstration of the different uses for which a technology can be adapted. For 20th-century Westerners, the technologies were instruments for shaping *their* reality: seeing the invisible, hearing the inaudible, communicating across space, and making present what was absent. But for the Kaliai people, these technologies were used to reimagine *their* invisible world, which was a place of visions, dreams, ancestors, and spirits, not electromagnetic waves (Lattas, 2000).

This is more than just a quirky anthropological moment; it leads us to the heart of something special about technology in human societies. In fact, the Kaliai's preoccupations are mirrored in a pattern that existed at the birth of these particular technologies in the West. It's worthwhile to recall that the realms of physics and what is now called parapsychology were not particularly separate in the late 1800s. Sir Oliver Lodge, for example, a physicist and radio pioneer, was an active member of the Society for Psychical Research, with a strong interest in communications with the dead; this was not an eccentric tangent to his work in electromagnetism, but deeply embedded within it (Raia, 2007). Moreover, the general public of the mid-1800s "often found it hard to distinguish between telegraphy and spiritualism" (Noakes, 1999), since the new technology of telegraphy promised communication with those residing in the next *county* in much the same way that mediums promised communication with spirits residing in the next *world*. Some found telegraphy too implausible to really believe, and so the Electric Telegraph Company had to aggressively market "the idea of communicating with invisible people via the electric fluid" (Noakes, 1999). But at the same time, some telegraph engineers themselves thought that communications with the dead were just as real as telegraphy, and worthy of similar exploration (Noakes, 1999). Likewise, in the early days of radio, some supposed that radios could be designed to allow the reception of voices of the dead (Sterne, 2003, p. 189). And in the 1800s, in the early days of sound recording, it was very important to people that these technologies allowed "the possibility of preserving the voice beyond the death of the speaker" (Sterne, 2003, p. 287). In a sense, there are ghosts in these machines.

We already know, of course, that in technologically modern societies today, most people have minimal or no understanding of how their telephone or radio works; the devices might as well be magic. It is also easy to appreciate that in any human society, new technologies are often used to reconfigure social relationships. But the facts mentioned above push us further than those basic realizations; indeed, they suggest that the development of radio technology on Earth wasn't just about pure scientific discovery, or even just about practical concerns such as economics, military might, and the concerns of empire. It also involved very human questions about mortality and spiritual existence. Cultural details such as these are not often mentioned in general histories of these technologies; rather, they are often deemed irrelevant, and the story is recast as a linear, logical progression, and the technology itself as simply rational and utilitarian.

## 25.5 Technology – Not Just Our Tools

Heidegger noted that when we ask "What is technology?", the usual answers are that it is a human activity, and it is a means to an end (Heidegger, 1977). Above, I addressed some human dimensions of technological practice – what *we* do with *it* – so now I should address what *it* does with *us*. As philosophers and students of technology have come to understand, technology has agency. Sometimes it does our bidding – we use it. But it also sometimes tells us what to do, what to see, and even what to believe.

Many thinkers have wrestled with how best to characterize the dance between humanity and our technology: in particular, who leads? The answers lie along a spectrum. Some have suggested that we are technology's masters, while others have suggested that we humans are little more than the "sex organs of machines".[7] The most nuanced answers recognize that neither human beings, nor our things, always have the upper hand. Hans Jonas said it well: "The relation of means to ends is not unilinear but circular… new technologies may suggest, create, even impose new ends, never before conceived, simply by offering their feasibility." (Jonas, 2004 (1979), p. 19). Similarly, Langdon Winner proposed that technologies are forms of life: for example, we can turn off the television set in the literal sense, by pressing the "power" button, but television as a social phenomenon cannot be turned off (Winner, 1986).

Indeed, technology does have its own imperatives sometimes. Ursula Franklin gives the example of radar traps, designed to ensure that motorists obey speed limits. Not long after these were introduced, radar detectors or "fuzz-busters" were developed. Next came a device which police could use to locate fuzz-busters… And so it goes on, with the initial objective – speed reduction – not really being achieved (Franklin, 1990, p. 57). The obvious analogy is the arms race phenomenon, with which we are all familiar. But this also recalls Edward Tenner's *Why Things Bite Back*, a study of the unintended and sometimes counterproductive effects of our technology. He gives the examples of early car power door locks, which increased drivers' sense of safety… but simultaneously greatly increased their risk of being

7 Marshall McLuhan's phrase, quoted in Andrew Feenburg 2004(1992), p. 211.

locked out of the car and exposed to unsafe situations. Similarly, car alarm systems malfunction so frequently that in urban areas the sound of an alarm going off initiates little more than neighbours' irritation. And home security systems have such a high false alarm rate that they effectively divert police officers from doing real law enforcement. These "revenge effects", as Tenner calls them, happen because new technologies "react with real people in real situations in ways we could not foresee." (Tenner, 1997, p. 11). Pulling ideas like these together in his critical synthesis *The Technological Bluff*, Jacques Ellul offered this summary of technology's unpredictability and influence:

> "First, all technical progress has its price.
>
> Second, at each stage it raises more and greater problems than it solves.
>
> Third, its harmful effects are inseparable from its beneficial effects.
>
> Fourth, it has a great number of unforeseen effects." (Ellul, 1990)

Indeed, it is obvious that modern technology has transformed our social world profoundly; for example, it is easy to see that the speed at which telegraphs, radios, telephones, and satellites allow us to exchange information has fundamentally changed global society. Compared to all human societies before about 1800, which had to rely on traveling humans and horses to move messages, our world is very different (Franklin, 1990, p. 43). Even more than this, however, television and the internet deeply affect our apprehension of the world in which we live; humanity's "relationship with the real" is now being changed at the most basic levels (Augé, 1999, p. 2).

But the story is much older and far deeper than that. In terms of the full sweep of humanity's history, it has been argued that our technology has been instrumental to our actual physical evolution. From the taming of fire to the development of clothing to the invention of baby slings or corrective lenses, our things have enabled physical changes in our species. In this sense, it may be said that "things evolved us", rather than the other way around.[8] Or, as Andy Clark and others have put it, we have always been human-technology symbionts: cyborgs (Clark, 2003).

## 25.6 What, Then, Does it Mean to have Radio Technology?

Having explored some historical and philosophical perspectives about technology, and some of the intriguing workings of technology in human societies, it is time to return to the main question: If the technology of an

[8] Timothy Taylor, pers comm

extraterrestrial civilization is detected via SETI, what could we assume about that civilization?

But instead of making a list of hypotheticals which cannot yet be empirically evaluated, I propose to probe what we can say about *ourselves*, as the bearers of radio transmission and detection technology. (I've chosen this as my example because, although there are several kinds of SETI search underway, the dominant search technology has been the radiotelescope.) The stories of the development of radio wave detectors and receivers, and the history of radioastronomy, have been told many times and many ways, so I will not attempt a comprehensive retelling here. Rather, I want to highlight several themes: the diversity of early radio wave detectors; the role of contingency and the military in radar and radioastronomy, specifically the case of Arecibo; and the agency of the technology itself.

### 25.6.1 Diversity of early radio technology: not a simple path

In tracing the history of any modern technology, it is tempting to track it backward, one precursor at a time, in a straight line to its earliest antecedent. This makes a convincing and satisfying story. However, played forwards, technological history doesn't work that way. At any given time, most complex technologies exist in diverse forms, and it is not always possible to predict which ones will survive and which will perish[9].

(Nor do most histories fairly acknowledge the contributions of all the individuals involved in a technology's development. So it is that books are still being written to "set the record straight" about who really "invented radio", teasing apart the long list of characters involved, from Hertz to Marconi to Tesla to Popov to Bose. Naturally, national pride occasionally plays a role in these debates.)

In the case of early radio wave detectors, Phillips describes a host of variations, including some lesser-known ones (Phillips, 1980, Chapter 9). Electrostatic detectors had a receiver with fine wires or needles which, when connected to an aerial, moved slightly due to electrostatic attraction when a signal was present. Gas flame detectors had a sensitive flame, which became disturbed by minor fluctuations in the gas flow, which in turn were induced by the presence of a signal. This produced enough variation in flame size and sound to convey Morse signals to an audience. Another variant made Morse code audible in another way, by using the signal to "open and close a valve in an air pipe", which in turn activated a whistle (Phillips, 1980, p. 191). There were also systems using fluid jet relays and mechanical relays.

But there were also others, rather more surprising, including a method for receiving Morse code by mouth. Presented in 1921 in New York, this

---

[9] Stephen Jay Gould's book *Wonderful Life* makes the parallel argument for biological evolution.

idea used two silver electrodes positioned in the mouth. When an amplified audio-frequency signal was picked up by the electrodes, it produced a rather nasty taste. Using this method, the lucky subject could read Morse signals at a rate of five to ten words per minute (Phillips, 1980, p. 198). Its drawbacks, however, included the unpleasant taste, induced toothache, and effects upon the user's eyes.

Even better, Phillips describes the "Physiological Detector" of Lefeuvre in 1916, based on Galvani's experiment which demonstrated that electrical impulses would make the muscles of a dead frog's leg twitch. An unfortunate frog was mounted to a board, and the sciatic nerve was connected to the aerial circuit; a signal made the leg convulse. In turn, the leg was connected to a pointer moving along a rotating drum, recording the signal. Drawbacks included the onset of rigor mortis, which could be avoided by using a pithed frog instead of dead one, but of course even this wouldn't last long.

Best (or worst) of all, was the brain coherer of Collins, published in 1902. Using a standard spark transmitter and receiver, he added a brain into the apparatus, "in an attempt to find out whether it would act like a coherer" (Phillips, 1980, p. 202). The dead cat brain didn't work very well, but the anaesthetized cat worked rather better. Ultimately, Collins was pleased to find that a fresh human brain (provenance unspecified) worked very well indeed, even picking up signals from an approaching thunderstorm.

This short tour of early 20th-century eccentricities in radio wave detection underscores the diversity of technological forms which once existed, but never achieved widespread use. But it is easy enough to imagine that if social conditions were different, or our bodies were different, surprising forms of radio wave detector could have become dominant, in turn founding quite different trajectories for radio technology as a whole.

### 25.6.2 Contingency and the military connection in radar and radioastronomy's development

The case of Arecibo provides a useful exploration of contingency in radioastronomy's development, and the entanglement of radioastronomy with military history.

Radioastronomy itself is arguably the progeny of ionospheric physics, which as a whole was constantly supported by commercial labs such as Bell Telephone Laboratories (Jansky's home base), and by the military, because of its applications to communications. Most postwar radioastronomers had backgrounds in ionosphere studies and wartime radar (Gillmor, 1986).

The 1950s were, of course, a time of great tension between the Soviet bloc and the West, and this drove the development of radar and radioastronomy facilities. High-power radar systems were used to monitor Soviet ICBM tests in the mid-50s, and 85-foot dishes were in use in the US Air Force's

Ballistic Missile Early Warning radar system in 1958. Many sites detected *Sputnik I*'s beeps in orbit in 1957, but the Mk1 (Lovell Telescope) at Jodrell Bank actually detected the third stage of the carrier rocket, and Millstone in Massachusetts was the first to detect the satellite by active radar (Stone and Banner, 2000). These kinds of accomplishments assured funding for the technology's development, and the systems could then also be used for studies of the ionosphere, moon, and other planets.[10]

It was within this context that Arecibo was proposed by William E. Gordon and colleagues in 1958 (Gordon, 1958), primarily as a radar system for studying the Earth's ionosphere. Initially, Arecibo was funded by the US Advanced Research Projects Agency, and administered by the US Air Force, and managed by Cornell. It wasn't until 1971 that the National Science Foundation took over as the funder, and the National Astronomy and Ionosphere Center was created to manage the facility.[11] Of course, the NSF itself had been established in 1950 partly because WWII had emphasized that being a leader in the development of new technologies was essential to America's welfare: The NSF's official purpose was "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes."[12]

Arecibo was intended to study the ionosphere, particularly to gain information about the potential effects of atomic weapons in relation to long-distance communications.[13] Arecibo was also designed to be capable of planetary radar and passive radioastronomy, but these were secondary to Gordon's main interest in communications and scatter propagation – the idea was to make it a multi-purpose facility, more attractive for funders (Gillmor, 1986, p. 127). The dish's size was predicated on the assumption that the ionosphere studies would have to use incoherent backscatter; however, interestingly, it was discovered thereafter that the scatter is in fact coherent, which means that the dish could have been much smaller than its actual 300 meters… even a then-standard 85-foot dish would have done.[14]

At about the same time, the US National Security Agency and Naval Research Laboratory attempted, and then abandoned, a project to build a 600-foot dish at Sugar Grove, in West Virginia, for the purposes of eavesdropping, via moon reflections, on Soviet radar signals and radio communications (Bamford, 1982). Had the huge, and hugely expensive, Sugar Grove project been proposed just a couple of years earlier, and had it succeeded, Arecibo might well have never come to exist in its present form.

10 Stone and Banner 2000, and www.jb.man.ac.uk/public/story/mk1.html

11 www.research.cornell.edu/VPR/CWC172-04/pdfs/NAIC_timeline.pdf

12 www.nsf.gov/about/history/

13 Jim Condon, personal communication

14 Jim Condon, personal communication

The point here is simply this: Arecibo proved to be an extraordinary instrument for pure research in radioastronomy, but it could easily have been otherwise. It was designed and built at least partly for military purposes, the exceptional receiving area and extremely powerful transmitter were actually overkill for its intended purpose, and other contemporary developments might easily have eliminated Arecibo.

But it worked out the way it did, and these contingencies led to Arecibo being one of the preeminent telescopes for SETI observations, but also the transmitter for the first major Active SETI message of 1974, still the strongest ever sent.[15]

Did *any* of this have to be so? For example, did radar and radioastronomy have to be associated with the military machine, or cotemporal with the development of weapons of mass destruction? Perhaps on Earth, yes. But there is nothing inherent in the technology itself that made it so: that is, if there hadn't been extensive warfare during the 20th century, radio technologies might have been developed for other remote sensing purposes. After all, radar has a plethora of uses in environmental observation, and there are many good reasons to study the troposphere and ionosphere that have nothing to do with the activities of enemy nations (Armand and Polyakov, 2005).

### 25.6.3 The agency of the technology: effective but not determining

It's commonplace in science that many technologies end up being used for purposes which were neither imagined nor imaginable when they were first designed and built. While this might be overstating the case for SETI, it's notable that radio SETI itself was initially an outgrowth of the instruments already available. For example, Arecibo was of course built for purposes other than SETI, although it has been extensively used by SETI, has been involved in much of the public awareness about SETI, through media and SETI@home, and although SETI has been invoked as a reason to keep the facility open (Grossman, 2007). As another example, Green Bank's first telescope, the Howard Tatel 85 Foot Telescope/Tatel 1, was used in Project Ozma shortly after it was built, but it was built for other reasons (Lockman et al., 2007).

However, we might ask: within that historical context, could the Tatel 1 possibly have been designed, funded, and built *for* SETI searching? It seems improbable. But it's also salient that many other telescopes just like it have never been used for SETI. The point here is that the telescope's existence made SETI possible, but it didn't entail it. It was necessary to the birth of modern SETI, but not sufficient.

---

[15] www.seti.org/site/pp.asp?c=ktJ2J9MMIsE&b=17907

What else was needed? Of course, the Ozma receiver had to be built, and the physics worked out by Frank Drake, like the contemporary speculations of Cocconi and Morrison, was also necessary (Cocconi and Morrison, 1959) But even that was not enough. The other key ingredient was the *idea* of alien beings dwelling among the stars, and that is a very old idea indeed, a cultural constant for millennia. In fact, it has even been suggested that the tendency to anthropomorphize our environment, including the skies, is a deeply ingrained quirk of the human psyche, rooted deep in our evolutionary history (Guthrie, 1995). Without this tendency and this cultural history, would anyone ever have thought to use a radiotelescope to look for extraterrestrials?

### 25.6.4 A brief summary regarding the path to radio technology on Earth…

The presently dominant forms of radio technology on Earth need not have developed exactly as they did. Their histories are full of contingencies, twists and turns, and conditions which were necessary but not sufficient for subsequent stages to develop. Cultural factors were constantly in play.

Given this knowledge, there are many thought experiments one might perform to explore further. For example, what if these technologies had been discovered within a different social context, for example in a society where religious specialists controlled technology? What if they had been discovered in a different economy, without the free market which confers an economic advantage to rapid communications? What if they had been discovered in a peaceful world? Could they still have developed to the point of being able to receive or transmit signals across interstellar distances? Perhaps so.

In short, if we ask what we could assume about a society which SETI could detect, there are few simple answers. Substantive assumptions about a civilization's structure or history are difficult to justify, for the *necessary* cultural correlates of radio technology seem to be few.

## 25.7 In Closing

I return to H.G. Wells, and *War of the World*s, to a crucial, oft-cited moment when a terrified man asks, "What are these Martians?"

The narrator replies, "What are we?" (Wells, 2003 (1989), p. 96).

To think about technology on other worlds, we must go beyond the simple superficial analogue of Earth, by resisting the easy assumptions in our tales about our own technological progress. Ironically, one way to achieve that critical distance is by going deeper into our own human world. When we do, we see that technology does not follow simple paths, that its development is influenced by contingency as well as necessity, that it is intertwined with culture, and that it is embroiled in history. We and our things are bound together in an intricate dance. Being technological is no simple matter.

If humanity ever does encounter, whether from a distance or up close, technology from another world, we will urgently need answers to crucial questions: not just "what is it?", but also "what are they?" The answer to the first question can only be a fraction of the answer to the second. But if we rigorously examine the relationship between technology and its makers on Earth, we might better comprehend the relationship between these questions. And then, we may better know what, and whom, we seek.

**Acknowledgements**

My thanks go, as ever, to the scientists in the SETI community for their congeniality, and also to my colleagues in the International Academy of Astronautics SETI Permanent Study Group for their kind invitation to first present these thoughts in Hyderabad, India as the Billingham Cutting Edge Lecture 2007. I am particularly honored because John Billingham's work in SETI has been a great inspiration to me. For their patient answers to my questions about the history of Arecibo, I am indebted to Jim Condon and Frank Drake. The astronomers of NRAO Green Bank have been teaching me about radioastronomy for several years now, and I am grateful to them. For research and travel and moral support, inspiration, and encouragement, I owe thanks, variously, to Allen Tough (the generous sponsor of the BCEL), Mark Knackstedt, York University's Faculty of Arts, and the Social Sciences and Humanities Research Council of Canada. Despite all this assistance, errors are surely mine.

## References

Armand, N.A. and V.M. Polyakov. 2005. *Radio Propagation and Remote Sensing of the Environment.* London: CRC Press.

Augé, Marc. 1999. *The War of Dreams*. London: Pluto Press.

Bamford, James. 1982. *The Puzzle Palace.* Boston: Houghton Mifflin.

Basalla, George. 1988. *The Evolution of Technology*. Cambridge University Press.

Bogdanov, Alexander. 1984 (1908). *Red Star: The First Bolshevik Utopia*. Edited by L.R. Graham and R. Stites, translated by C. Rougle. Bloomington: Indiana University Press.

Clark, Andy. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence.* Oxford University Press.

Cocconi, Giuseppe and Philip Morrison. 1959. Searching for Interstellar Communications. *Nature* **184: 4690**, 844–846.

Crossley, Robert. 2004. H.G. Wells, Visionary Telescopes and the 'Matter of Mars'. *Philological Quarterly* **83:1**.

Denning, Kathryn. 2005. 'L' on Earth. IAC-05-A4.2.03, Paper for session "SETI II: Interdisciplinary Aspects", International Astronautical Congress, Fukuoka, Japan, October. International Astronautical Congress online archive. **Or** see K. Denning, in press. 'L' on Earth. In Douglas Vakoch and Albert Harrison (Eds) *Culture in the Cosmos: Extraterrestrial Life and Society,* Berghahn Press.

Denning, Kathryn. 2010(2006). Ten Thousand Revolutions. *Acta Astronautica* doi:10.1016/j.actaastro.2009.11.019 Originally presented at International Astronautical Congress, Valencia, Spain, 2006.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton & Company.

Dick, Steven. 1998, *Life on Other Worlds: The 20th Century Extraterrestrial Life Debate.* Cambridge University Press.

Ellul, Jacques. 1990. *The Technological Bluff.* Trans: Geoffrey Bromley. Grand Rapids, Michigan: William B. Eerdmans Publishing.

Feenburg, Andrew. 2004 (1992). Democratic Rationalization: Technology, Power, and Freedom. Reprinted in David M. Kaplan (Ed.) *Readings in the Philosophy of Technology.* Oxford: Rowan and Littlefield.

Fitting, Peter. 2000. Estranged Invaders: The War of the Worlds. In Patrick Parrinder (Ed.) *Learning from Other Worlds: Estrangement, Cognition and the Politics of Science Fiction and Utopia*. Liverpool: Liverpool University Press, 127–145.

Franklin, Ursula. 1990. *The Real World of Technology*. Toronto: CBC Enterprises.

Gerrie, James. 2005. Techno-Eschatology, in *Techne: Research in Philosophy and Technology* **9: 2**. http://scholar.lib.vt.edu/ejournals/SPT/v9n2/gerrie_te.html

Gillmor, C. Stewart. 1986. Federal Funding and Knowledge Growth in Ionospheric Physics, 1945–81. *Social Studies of Science* **16**,105 –33.

Gordon, William E. 1958. Incoherent Scattering of Radio Waves by Free Electrons with Applications to Space Exploration by Radar. *Proceedings of the IRE* November 1958, 1824–1829.

Gould, Stephen Jay. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*. W.W. Norton.

Grossman, Lisa. 2007. Profs to Defend Arecibo Funding in DC. *Cornell Daily Sun*, 30 August 2007. http://cornellsun.com/node/23811

Guthke, Karl S. 1990. *The Last Frontier: Imagining Other Worlds, from the Copernican Revolution to Modern Science Fiction*. Ithaca: Cornell University Press.

Guthrie, Stewart. 1995. *Faces in the Clouds: A New Theory of Religion*. Oxford University Press.

Heidegger, M. 1977. *The Question Concerning Technology and Other Essays.* New York: Harper & Row.

Heilbroner, Robert. 1967, Do Machines Make History? *Technology and Culture* **8**, 335–45.

Jonas, Hans. 2004 (1979). Toward a Philosophy of Technology. Reprinted in David M. Kaplan (Ed.) *Readings in the Philosophy of Technology.* Oxford: Rowman and Littlefield.

Kurzweil, Ray. 2005. *The Singularity is Near.* New York: Viking.

Lattas, Andrew. 2000. Telephones, Cameras and Technology in West New Britain Cargo Cults. *Oceania* **70: 4**, 325–344.

Lockman, Felix J., Ghigo, F. and Balser, S.. 2007. *But it was Fun: The First Forty Years of Radio Astronomy at Green Bank*. National Radio Astronomy Observatory.

Lowell, Percival. 1895. *Mars*. Boston: Houghton, Mifflin and Co.

Lowell, Percival. 1906. *Mars and Its Canals.* New York: MacMillan.

Marx, Karl. 1959 (1850). *The Communist Manifesto*. In Max Eastman (Ed.) *Capital, the Communist Manifesto and Other Writings.* New York: The Modern Library.

Michaud, Michael 2007. *Contact with Alien Civilizations: Our Hopes and Fears about Encountering Extraterrestrials.* New York: Copernicus Books.

Molotch, Harvey. 2005. *Where Stuff Comes From*. New York: Routledge.

Noakes, Richard. 1999. Telegraphy is an Occult Art: Cromwell Fleetwood Varley and the Diffusion of Electricity to the Other World. *British Journal for the History of Science* **32**, 421–459.

Phillips, Vivian J. 1980. *Early Radio Wave Detectors.* London: Peter Peregrinus Ltd.

Raia, Courtney Grean. 2007. From Ether Theory to Ether Theology: Oliver Lodge and the Physics of Immortality. *Journal of the History of the Behavioral Sciences*, **43:1**, 19–43.

Sheehan, William. 1988. *Planets and Perception: Telescopic Views and Interpretations*, 1609–1909. Tucson: University of Arizona Press.

Sterne, Jonathan. 2003. *The Audible Past: Cultural Origins of Sound Reproduction.* Duke University Press.

Stone, Melvin L. and Gerald P. Banner. 2000. Radars for the Detection and Tracking of Ballistic Missiles, Satellites, and Planets. *Lincoln Laboratory Journal* 12:2, 217–244.

Tenner, Edward. 1997. *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. New York: Vintage Books.

Trigger, Bruce. 2003. *Understanding Early Civilizations*. Cambridge University Press.

Wells, H.G. 2003 (1898). *The War of the Worlds*. Edited by Martin Danahay. Peterborough: Broadview Literary Texts.

Winner, Langdon. 1986. Technologies as Forms of Life. In *The Whale and the Reactor: A Search for Limits in an Age of Technology*. Chicago: University of Chicago Press.

Wright, Ronald. 2004. *A Short History of Progress*. Toronto: Anansi Press.

# 26

# After Contact – Then What?

Albert A. Harrison,
*University of California, Davis*

Few topics ignite the imagination as do the prospects of encountering extraterrestrial life – and what this may mean for individuals, societies, and cultures. Until recently speculation fell largely within the realms of philosophy, science fiction, and UFO studies. By 1960, however, the theoretical feasibility of interstellar transmissions coupled with Frank Drake's initial empirical search, Project Ozma, established a need to put such speculation on a firmer footing. Drake's work had gained the attention of Donald Norman, a psychologist who was developing a report on the peaceful uses of outer space for the US Congress. Whereas most of this report dealt with topics such as communications satellites, remote sensing, and human space exploration, portions dwelled on the possible implications of the discovery:

> "The consequences for attitudes and values are unpredictable, but would vary profoundly in different cultures and between groups within complex societies… Whether or not earth would be inspired to an all-out space effort by such a discovery is moot: societies sure of their own place in the universe have disintegrated when confronted by a superior society, and others have survived even though changed. Clearly, the better we can come to understand the factors involved in responding to such crises the better prepared we may be." (Committee on Science and Astronautics, 1961: 48).

The physical and biological scientists in charge of the search understood that their efforts could have enormous consequences for society, and encouraged

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection,
DOI 10.1007/978-3-642-13196-7_26, 

colleagues from fields as diverse as anthropology, history, psychology, sociology, and theology to explore intellectual and emotional responses to the discovery. Fifty years later, such research falls under such rubrics as "the cultural aspects of SETI" and "the societal implications of astrobiology" and is an important part of the overall SETI effort (Billingham et al., 1999; Harrison, 1997; Michaud, 2007; Tough, 2000). In addition to forecasting immediate and long-term consequences for humanity, SETI social scientists study public attitudes and support, signal decryption and interpretation, potential behavioral characteristics of extraterrestrial organisms and societies, and even the possible course of interstellar affairs (Harrison et al., 2000).

Effective planning is a daunting task (Harrison, 2008). Those who approach this task should try to keep an open mind, understand when they are making assumptions, when they are dealing with facts, and when they are faced with genuine uncertainties. As much as possible, they should keep their biases under control. They should avoid rejecting possibilities on emotional rather than rational bases, over-relying on simplifying assumptions, and seeking information that confirms their expectations while denying evidence to the contrary. Planning should be more than an academic exercise. It is critical to involve leaders who may have to make key decisions and, if that is impossible, then their close advisors. At least, influential people should be made aware of the issues that are under discussion. The best of plans will prove useless if presidents, generals, governors, spymasters, bishops, chief executive officers and other powerful people are caught by surprise. And, if governmental or military commissions are developing their own plans behind closed doors, plans that are developed in the open marketplace of ideas may offer useful alternatives.

## 26.1 Evidence-based Predictions

The effects of discovery will depend on the nature of the ETI (for example, peaceful or warlike), the type of the discovery (a message from a distant galaxy or the crash of a flying saucer in New Mexico), the trajectory of the discovery (a dawning realization or a sudden insight), the political climate (during peace or war) and much else. To simplify the discussion – and in the spirit of this celebratory volume – we will limit our discussion to the discovery of a microwave or laser transmission from outside our Solar System. This is the standard SETI detection scenario – a "dial tone at a distance" – or, as Seth Shostak has recently described it, a bottle that once contained a message found washed up on a beach (Shostak, 2009). Given the state of the art, once the dial tone is discovered more powerful instruments will have to be deployed to detect any information superimposed upon it. If successful, this in turn would be followed by a long, difficult, and perhaps unsuccessful struggle to interpret the message. We will avoid assumptions that, piled on top of one another, offer encouragement that the message will be immediately understandable, or that we will be able to interact in any meaningful sense with a civilization that is many light years away.

Although the long-term effects of contact may be profound and eventually infiltrate every sphere of human existence, there are strong justifications for focusing on the immediate effects. By definition, these are the very first results that we will have to manage! Second, we can proceed with only a few assumptions to understand initial reactions to a dial tone at a distance. (For example, we do not have to assume faster than light communication.) Finally, society is continually changing and long-term effects (such as they may be) will impact a society that is likely to be very different from that of today. Consider a member of the Committee on Science and Astronautics in 1960 trying to envision the results of a discovery on society in 2010. He or she would have no knowledge of home computers, virtual reality, i-pods, the dissolution of the Soviet Union, the war against terror, and other scientific, technological and social changes that shape life today.

As we anticipate human reaction to "contact" we should favor evidence over myth. But where can we find evidence? One source, as Steven J. Dick has pointed out, is historical episodes when radically different cultures have encountered one another, either physically, or through the exchange of information (Dick, 1996). These include the arrival of Europeans in the Americas, the relentless westward expansion of white Americans into Indian territories, the British in India, the Europeans in Africa, and much else.

Consider the plight of Emperor Moctezuma (also known as Montezuma) and his Mesoamerican peoples, the Aztecs (Weir, 2005). By the 16th century the Spaniards had established a base in Cuba and had completed two expeditions into what was then known as Mexica, sometimes establishing friendly relations and other times entering into armed conflict. In 1519, Hernando Cortez – a civilized and literate man who came from an upper crust family and who had studied law at the University of Salamanca – led a third expedition that would eventually take him to the heart of the Aztec Empire. Cortez had cannons, horses, dogs and other technology unknown to the Aztecs. He and his 400 soldiers also had an advantage that they did not know about. Aztec priests had predicted the return of the god Quetzalcoatl from across the Eastern Sea (Atlantic) that very year, and the "fair haired, bearded men with their strange animals and fierce weaponry that moved over the ground so easily" fit Quetzalcoatl's description (Weir, 2005, p.58). Thus, Moctezuma accepted Cortez's arrival as the second coming, immediately pledged allegiance to King Charles V of Spain and turned over his empire to the Spaniards. In the beginning Cortez and Moctezuma got along famously.

But after Cortez left on a brief trip, his more skeptical and less pacifistic deputy took the opportunity to butcher Aztec royalty during a Fiesta. This led to insurrection and battle. Moctezuma was murdered by his own people and in the end a sophisticated empire was ruined by an ancient belief and a remarkable resemblance between Quetzalcoatl and Cortez.

The ongoing history of exploration, encounters, subjugation, colonization and independence include both encouraging effects (such as the spread of literacy and new technology) and discouraging effects (political domination, exploitation, forced religious conversions). These are difficult enough to unravel with the benefit of hindsight, never mind (as we might prefer) armed only with foresight. Perhaps more useful for our purposes are episodes when significant numbers of people believed that extraterrestrial life had been discovered (Harrison, 1997). These include a brief period in the mid 1830s when readers of the *New York Sun* read that a powerful telescope revealed "bat men" on the moon; a somewhat later but lengthier period when astronomers including Percival Lowell claimed that Martians had dug an elaborate series of canals to distribute the remaining water on their drought-stricken Mars; Orson Welles' famous 1938 radio broadcast of H.G. Wells' story of an invasion from Mars; an on-going avalanche of claims that UFOs are extraterrestrial spaceships; and in the 1960s discoveries of quasars and pulsars whose naturally-caused microwave emissions initially gave the appearance that they were under intelligent control. In most cases people responded with curiosity, interest, and even delight. Widespread tales of panic following Orson Welles' broadcast were exaggerated and sensationalistic and ignored the fact that some of the actions interpreted as "panic" constituted sensible self-defensive and socially responsible behaviors given the listeners' (false) understanding of the situation. Responses to UFO reports are mixed, and include fear, inspiration, and derision. Accurate or not, UFO reports have prompted many otherwise disinterested people to ponder the possibilities and implications of extraterrestrial life. In my view the UFO "debate" has profound implications for people's reactions to a bona fide discovery and is one more battleground in our transformation from citizens of the world to citizens of the cosmos (Harrison, 2007).

Because they bear close resemblance to real scientific discoveries two historical prototypes – the Mars rock and the EQ Pegasi (EQ Peg) hoax – are of particular interest. The Mars rock (also known as ALH84001) is one of many meteorites that were ejected from Mars eons ago that landed on Earth (Dick and Strick, 2005). Studying such rocks for signs of organic activity on Mars millennia ago is difficult, in part because this requires highly skilled investigators using state-of-the-art equipment and painstaking techniques to rule out the possibility that any organic matter present may be due to contamination after arriving on Earth. Careful study by NASA scientists and contractors revealed several strands of evidence suggestive of biotic activity, including an incredibly tiny structure that bore a remarkable resemblance to a worm. In August 1996, NASA staged a press conference days before the publication of their peer-refereed scientific report in the highly prestigious journal *Science.* While it is considered bad form to hold a press conference prior to the publication, this was necessitated because the findings slipped out prior to publication. (One culprit was the girlfriend of a science advisor to the President who stole proofs from the advisor to sell to the tabloids.) Citing the familiar mantra that "extraordinary claims require extraordinary evidence" several influential scientists heaped criticism on the report. Today, the debate over whether or not the rock shows evidence of biological activity on Mars continues, and the balance of evidence seems to swing back and forth. The important consideration for us is that despite skepticism on the part of many scientists the discovery caught the public's fancy and led to increased governmental funding for astrobiology (Dick and Strick, 2005).

In late 1998, Art Bell, whose "Coast to Coast" talk radio show often discussed UFOs, announced that an anonymous amateur astronomer had intercepted an extraterrestrial transmission (Shostak, 1999, 2009). The signal came from the direction of the constellation EQ Pegasi, one of 88 constellations and one that struck Shostak as an unlikely home for ET. On October 29, the story was picked up by the BBC and gained international attention. Scientists suspected that the detection was a hoax, and attempts to confirm the discovery failed. The story persisted (dwindling in importance) for several days before it was refuted. Commenting on this, Shostak expressed vexation and gratitude, the latter because it "added a modicum of real experience to the endless theorizing about what would happen in the case of a signal 'hit'" (Shostak. 2009: 245–246). In his analysis the claims were met by a mix of uncertainty, confusion, and aggressive interest by the media.

As revealed in John Schuessler's summary of 60 years of opinion polls as presented in newspapers (Schuessler, 2000; 2004) at least half of the people surveyed believe in extraterrestrial life. A 1966 Gallup Poll found that 34% of the respondents believed that there are people "like us" who live on other planets in the universe (*Des Moines Register*, May 8, 1966). Men and women were equally likely to support the idea of extraterrestrial life, and support was stronger among college and high school graduates than among people who had only a grade school education. (Later polls also found that more highly educated people are more open to the idea of extraterrestrial life.) A similar poll seven years later found that the percentage of people believing that people "somewhat like us" living on other planets had risen to 46% (*The Houston Post*, December 6, 1973). Five years later, the percentage rose to 65% (*St. Louis Post Democrat*, June 7, 1978). From then on, the results become repetitious. From the late 1970s through the first decade of our present century somewhere between 50 and 60 percent of survey respondents believe in life "out there," a figure that varies somewhat depending on the wording of the question, whether or not people are allowed to state "don't know," the specific population surveyed, and the survey sponsor (those sponsored by tabloids typically offer higher figures). Most recently, 56% of the respondents to a 2008 Scripps Howard Poll thought that ETI was likely, and 31% thought Earth had already been visited by extraterrestrials (Hargrove and Stemple, 2008).

Other surveys suggest that people feel prepared for the discovery, and expect to react positively or to be unaffected (National Institute for Discovery Science, 1999). Very few people expect to have their lives disrupted, but they are less confident that everyone else will remain on an even keel. Is it possible that the Committee on Science and Astronautics (1961) rested on a general tendency to see oneself as prepared but doubt the abilities of others? A shrink might describe this as a projection of personal anxieties and insecurities onto other people. Certainly the Committee on Science and Astronautics report has a somewhat paternalistic tone, assuming that some information might have to be withheld to avoid upsetting the people.

## 26.2 Initial Response

Conditions have changed since the first microwave search. Today, we live in an era when cosmic evolution, modern astronomy, space exploration, and all facets of astrobiology (including SETI) encourage redefining ourselves as citizens of the universe. Weird physics, modern technology and science fiction make anything seem possible. A majority of people "believe in" extraterrestrial intelligence and a significant percentage believe that ET has visited Earth. Most of those surveyed feel ready to cope with the discovery, and humanity has survived convincing claims (including false claims of SETI detections) that ET had been discovered. Under such conditions, it is tempting to expect rather mild responses to a real discovery. Compared to finding bug-eyed monsters, flying saucers, alien abductors and counterfeit humans working at the next desk, intercepting a dial tone at a distance is pretty mild stuff.

When taking the Rorschach Inkblot Test people make up stories about meaningless inkblots and in the process reveal their personalities. It has long been noted that ET offers people another wonderful opportunity to project their hopes and fears onto a blank screen. Powerful myths – of gods, demons, and magical beings – will play a role, as well as countless mental images from science, science fiction, and documentaries that sometimes stray from the facts. This, coupled with personal experience, values, and peer pressures, creates expectations which in turn will shape people's views of

ET. One of the most established findings in social psychology is that, faced with uncertainty, people seek information from others, searching for bits and pieces of information to get a clear picture, put their emotions in perspective, and (often) to achieve consistency with group expectations (Harrison and Johnson, 2000). Typically, people are less influenced by pronouncements of authorities and media accounts than by the more immediate views of their family and friends – people who are similar to themselves. The tendency to validate personal views by comparing them with other people may be particularly pronounced in this case, given widespread expectations that governmental authorities will withhold the truth.

Potential implications for science, religion and politics are likely to be of lesser concern than the immediate consequences for individuals and the people that they care about. People who are intolerant of ambiguity will be in for a rough time. People who are fearful and insecure, and those whose belief systems are stressed, may be challenged. But many of them will alleviate their stress by denying the reality of the discovery or rationalizing it within their belief system. Do we really expect people who are able to dismiss a mountain of evidence in support of Darwinian evolution to be upset by the discovery of a Martian fossil or a signal from a distant galaxy? People who already endorse the idea of extraterrestrial intelligence (especially those who believe that ET has already visited Earth) are likely to react with "I knew it all along," "I told you so," and "What is it that you are keeping from me?" Conventional and highly practical people may be oblivious to the philosophical significance of the discovery, and those who live in poverty, are ravaged with illness, or engaged in genocide or war may not care.

## 26.3 Sectors

Culture will play an important part and also age, gender, family and friends. Reactions will vary across different segments or sectors of society: Government, Science, Religion, Business, and the Media (Harrison, 2007). Within each sector we find people who have interests in common (even though they may squabble with one another from time to time) and who, with varying degrees of motivation and skill, try to help society survive.

### 26.3.1 Government

Government plans for all sorts of far-fetched contingencies, including future wars with current allies. It is not clear that, apart from attempting to defuse fears of UFO invasion in bygone decades (Hoyt, 2000), how much thought government has given to the discovery of ETI. Claims of secret government panels addressing the extraterrestrial origins of UFOs are hard to assess precisely because they are secret. From time to time, people affiliated with government (including the military and intelligence communities) openly express interest, but this may be a personal interest and "not reflect the official policies or views of the United States Government."

After he left the presidency, comments by Bill Clinton revealed that interest extends to Government's highest levels (Wozniak, 2005). After expressing support for astrobiology, Clinton confessed that during his years in the White House he used his position to learn more about UFOs. He rapidly concluded that the crash landing at Roswell was an illusion, but based on the beliefs of many people in his administration sent an emissary to uncover the truth about Area 51 in Nevada. The emissary returned, reporting that there were no extraterrestrial artifacts there, only "boring" defense research projects. If there were UFO secrets, Clinton remarked, they were "concealed from me too."

Roberto Pinotti expects contact will create a "crisis of authority" and that Earth's super powers would quickly be reduced to the status of "Andorra, Monaco, and the Republic of San Marino" (Pinotti, 1990, p. 163). He adds "today's establishment worldwide would have everything to lose from any form of contact with ETI, since it would be the first victim in a frontal collision between different civilizations" (1990, p. 164). Alexander Wendt and Raymond Duvall (2008) argue that the mere prospect of extraterrestrial intelligence threatens terrestrial political systems. Government is highly anthropocentric and not receptive to the idea of ETI. In the old days, Nature and God helped rule, but today, nature is "assumed to lack the cognitive capacity and/or subjectivity to be sovereign, and while God might have ultimate sovereignty, even most religious fundamentalists grant that it is not exercised directly in the temporal world" (Wendt and Duvall, 2008, p. 1). All terrestrial governments rest on an assumption of people governing people. The prospect of extraterrestrial sovereignty poses a "metaphysical threat" which leads to denial and refusal to look closely for extraterrestrial life.

Seth Shostak and Donald E. Tarter offer opposing views of Government's reaction to a SETI detection. Shostak is convinced that it is a matter for scientists, and the discovery of that dial tone at a distance will be so remote and irrelevant to politics and everyday life that the entire matter will be left to scientists (Shostak, 2009). The search is "safe" because we will not reveal our presence to them, and there is no easy way for them to get here from there. Donald Tarter takes the contrary position that scientists are deluding themselves if they think that Government would ignore the discovery (Tarter, 2000). As long as SETI remains little more than an exercise the government can afford to treat the activity with benign neglect. Following an actual discovery, no national leader would risk leaving the matter entirely within the hands of the scientific community. To gain control, authorities will debrief scientists, demand that they sign security oaths, send government agents to monitor signals and hire government analysts try to make sense of the situation. Any hint of danger in the signal itself, or any public unrest, will strengthen government regulation. Decide for yourself if the same governments that spend billions on intelligence gathering, monitor phone calls and computers, fight cyberterrorism and make old ladies take off their shoes before getting on an airplane will be able to ignore evidence of a technologically advanced extraterrestrial civilization.

Government's low credibility in the eyes of the public will make it difficult to mount an effective response. Public trust in the US government began to slide in the 1950s and has reached crisis proportions today (Harrison and Thomas, 1996). Skepticism about Government candor cuts across many topics including presidential assassinations, UFOs, soldiers missing in action, AIDS, and countless other topics. Influential people, not just the uneducated, poor, and disenfranchised, expect that Government will suppress evidence of extraterrestrial life. In a 1999 Roper Poll, less than 10% of all respondents thought that government would immediately apprise the public of the findings (National Institute for Discovery Science, 1999). But Shostak notes that cover-ups would be very difficult in the case of SETI. "The nature of SETI precludes cover-ups. There is no way to hide the evidence of a true extraterrestrial signal: it is present in the sky, and anyone with an antenna can check it out for themselves. It cannot be secreted away by nefarious groups and hidden from view" (Shostak, 1999).

Government would do well to have a strong central planning effort to deal with "detection day." The principal goal should be to have a workable plan in place. This would establish jurisdictions, assign roles, and above all make provision for a coordinated but flexible response. To minimize public perceptions of dishonesty Government should avoid premature statements (confident pronouncements before the facts are in) and coordinate claims made by different spokespersons so that they are consistent with one another. Government could enlist the help of damage control experts (such as those that tell businesses what to do following recall of an unsafe product) and third party public relations experts whose goals are to inform and persuade.

### 26.3.2 Science

Science would want to learn as much as possible about nonhuman intelligence, irrespective of whether it's a Mars rock, a dial tone, or a spaceship. Each of the other sectors would rely heavily on science for assessing the situation and for working on ways to communicate. Some scientists have suggested (with great humility and modesty) that since they would come closest to matching ET's IQ they would make the best communicators and good-will ambassadors (Harrison, 2007). The public does not necessarily disagree. When asked with whom they would choose to make first contact on Earth scientists came in first (29%), the military and a private organization were tied for second and third (20%), and Government came in fourth (14%) only slightly ahead of religious leaders (11%).

Following discovery Science expects to become an instant winner. The finding is regularly touted as one of the greatest discoveries of all time. It would constitute a capstone for two intellectual revolutions: the Copernican revolution (that displaced us from the central to a very peripheral position in a universe that is 12–15 billion light years in diameter) and the Darwinian revolution (that displaced us from the pinnacle of creation by finding continuities with other species). In Science's view, the discovery would prove that life's mysteries can be solved by applying empirical methods to study the material universe of matter and energy. Some people construe this as a triumph of science over religion (Davies, 2000).

Science will win financially. Following the discovery of the Mars Rock the Clinton Administration poured more money into astrobiology, and one can only wonder if this finding also prompted the President to send an emissary to Area 51. Since the implications of discovering a technologically advanced civilization are even more profound, the discovery is likely to trigger a tremendous infusion of research funds. The search will accelerate in the hopes of finding additional civilizations, and researchers who formerly sat on the sidelines will jump into the fray. Expect a plethora of instant SETI experts (who had never before given serious thought to the matter) in many fields, not only in the sciences.

### 26.3.3 Religion

Science's expectation that proof positive of extraterrestrial life will sound a death knell for Religion is not shared by Religion. Leading theologians are not intimidated and many would view the discovery as greater testimony to God's power and creativity (Peters, 2009). Michael Ashkenazi conducted interviews with 21 theologians, 17 of whom believe that extraterrestrial intelligence exists, and found Buddhism and Hinduism among the religions that already accept the idea of extraterrestrial life or will not be particularly troubled by confirmation of its existence (Ashkenazi, 1992). He was more concerned by Christianity, Judaism, and Islam, which rest on the conviction that man is created in God's image. Ashkenazi expects that Catholicism and other hierarchically organized Christian religions will interpret the event and make policy and then church members will fall in line. Protestant sects are varied and autonomous, and are likely to react in different ways, with fundamentalist and dogmatic sects reacting badly.

Victoria Alexander surveyed a national sample of clergy including 563 Protestant ministers, 396 Roman Catholic priests, and 41 rabbis (Alexander, 1998). Most of the clergy who returned their questionnaires did not believe that their faith or that of their congregation would be compromised. More than three quarters of the respondents strongly disagreed (29%) or disagreed (48%) with the statement that "Official confirmation of the discovery of an advanced, technologically superior extraterrestrial civilization would have severe negative effects on the country's moral, social, and religious foundations." Would the discovery cause their congregations to question their fundamental beliefs regarding the origin of life? Eight out of ten ministers, priests and rabbis thought not. Only 15 percent of the respondents thought that the discovery would constitute a religious threat.

Alexander's findings were corroborated and extended by Ted Peters' recent survey of 1325 adherents to the world's established religious traditions (Roman Catholicism, mainline Protestantism, evangelical Protestantism, Orthodox Christianity, Mormonism, Judaism, and Buddhism) and nonbelievers (Peters, 2009). He found that 83 to 94 percent of the respondents (depending on religious affiliation) strongly disagreed with the statement "Official confirmation of the discovery of intelligent beings living on another planet would so undercut my beliefs that my beliefs would face a crisis." They were slightly less confident that their religious tradition would face a conflict, but still 67 percent of the Catholics, 73 percent of the evangelical Protestants and 99 percent of the mainline Protestants did not think that the discovery would cause a crisis within their religions. Intriguingly, respondents could see how such a discovery could cause a crisis within other people's religions. It was respondents from the nonreligious group – the agnostics and atheists – who were the most certain (69%) that the discovery would trigger a religious crisis. Where are the polls supporting claims that confirmation of extraterrestrial intelligence would shatter religious beliefs? This is particularly important given Peters' finding that nonreligious people have exaggerated concerns about religious people's reactions. Will some religious people be troubled by the discovery of ETI? Probably, but the same will be true for some scientists, some politicians, some newspaper reporters, some business people, and some psychologists!

Better we should think of religion as a useful ally in framing overall response to the discovery of nonhuman intelligence. Theologians and religious leaders could interpret new information with respect to present-day religious beliefs and values and communicate this to the public. Religion could resist pressures to demonize the ETI or, alternatively, treat them as saviors. Mainstream religion would work against the emergence and proliferation of dangerous cults. As anthropologist Jim Funaro pointed out, we should not "underrate religion's survivability and its usefulness as an adaptive tool. The discovery may stimulate a worldwide resurgence in religious activity. Religion may have an advantage over Science as we attempt to adapt to strong and widespread emotional impact… Religion has already had considerable experience dealing with ETs… Religion can answer questions that science cannot… Religion provides a built-in, self-activating mechanism for responding to widespread societal stress. In the actual event of encountering extraterrestrial life, some of the needs of humanity as a whole may require the kind of nonscientific solutions provided by religion. Given the number of unknowns in the contact equation, we should not ignore the potential value of any of our adaptive resources" (Harrison and Connell, 2001, pp. 29–31).

### 26.3.4 Business

Typically, the introduction of uncertainty drives the stock market down and gold prices up. A change of corporate leadership, fluctuations in unemployment, the latest product from a major manufacturer, the slightest hint of an adjustment of interest rates – almost anything, it seems, can send the stock market soaring or skidding.

A quick check of the Dow Jones Industrial Average (DJIA) for August 1996 (when NASA held a Mars rock press conference) shows that the DJIA ended the month up 22 points at 5616. From October 29, 1998 when the media picked up the EQ Peg hoax through the end of November that same year, the DJIA rose over 600 points, closing at 9116. Of course, nobody knows what would happen following a real detection that remained in the news for several days, but performance during these "dry runs" was far from disheartening.

Over time, particularly if there were growing consensus as to what ET "is like," Business would find new merchandizing opportunities. Obvious merchandise includes sweat shirts, souvenirs, video games, candy bars, jewelry, CD albums, and theme parks. (Perhaps we should include underground shelters, emergency rations, and shotguns.) Business may explore possibilities for terrestrial – extraterrestrial partnerships, perhaps licensing terrestrial rights for extraterrestrial inventions.

### 26.3.5 Media

Media's primary purpose is to attract and retain a large audience for the benefit of its sponsors. In the wake of a discovery Media may be torn between providing a full and accurate account of the news, sensationalizing, and dribbling out bits and pieces ("More after this commercial break!"). The elite media, such as *The New York Times*, the *Washington Post* and national public radio and television are strong allies of science and strive for accuracy (Goode, 2000). Tabloids and their radio equivalents are less inhibited when it comes to placing entertainment over education.

Reporters are trained to seek out the dramatic. Often the unusual, weird, bizarre and strange are given free reign, sometimes to the detriment of the truth. But even reporters striving for accuracy may encounter difficulties as they try to break complex stories into a dozen or so simple declarative sentences to audiences with limited attention spans and intelligence.

Two of the factors that will affect media coverage are time pressures and the news hole. People are interested in fast breaking developments so it is important to get the news out in a hurry. This means that stories are launched following tips and preliminary announcements. Often stories reach the public before they are verified and without the benefit of elaborating detail. More than once newspaper readers have read front page predictions of large meteorites striking Earth at some later date (but not late enough to alleviate the worries of youngsters who have a long life expectancy) only to find back page retractions following more refined calculations.

The news hole refers to the size of the story – the number of column inches that an editor wants to devote to the story, or the number of minutes (or hours) that a radio producer must fill (Fischer, 2009). The news hole may be too small or too large for a story. If it is too small, the reporter may not be able to explain the context or include necessary supporting detail. If it is too large, the reporter may have to spin an elaborate yarn or represent the views of other (typically unnamed) reporters as "experts." In the case of discovering ET there may be very little in the way of facts or "hard" news, so the choices of filling the news hole include embellishing the facts or reporting "soft" news," that is, human interest stories. These include the experiences and views of scientists, religious leaders, politicians, celebrities, and everyday folk plucked off the street. Compared to hard news soft news is highly subjective and in some cases (such as reporting on disasters) more likely to rest on supposition and perpetuate myths (Fischer, 2008).

The Rio Scale was developed to help Media gauge the significance of a claimed discovery (Almar and Tarter, in press). "Level of Importance" depends on the class of phenomena (for example, an omni-directional beacon as compared to a specific message aimed at Earth), discovery type (a radiotelescope search or reinterpretation of ancient documents), and distance (a neighboring planet or a distant galaxy) as well as the credibility of the person who claims the discovery (ranging from "obviously fake or fraudulent" to "absolutely reliable, without any doubt.") As revised, the Rio Scale is intended to set an appropriate level of alertness on a 10-point linear scale to help scientists and reporters understand an event's importance. The scale has been applied to a number of science fiction scenarios, false alarms, and hoaxes (Shostak and Almar, 2006). For example, the EQ Peg hoax began with an importance value of 1-2 which increased to 3-4 following mistaken claims of verification and becoming a BBC news story but then dropped to 0 when leading observatories failed to confirm the discovery and other evidence of a hoax was found (Shostak and Almar, 2006).

Educational outreach and working with Media has an important correlate of the SETI effort. As Carol Oliver points out, the public learns about science from many different sources: television and newspapers to be sure, but also from textbooks and college courses, art, literature, music, coworkers, family, friends and, of course, the Internet. Communications experts no longer think in terms of “an” audience, but in terms of multiple audiences, each with its characteristic belief systems, motives, levels of intelligence and attentiveness, and so forth (Oliver, 2003). Different campaigns are needed to inform different audiences just as politicians would use different messages, spokespersons and outlets to recruit votes from students and retired people.

## 26.4 Conclusion

The discovery of a dial tone at a distance may indeed be one of the greatest scientific discoveries of all time, and of immense philosophical interest, but historical prototypes and survey results suggest only modest effects on most people’s everyday lives. Fifty years of technological and social changes, coupled with 50 years of well-publicized searches, have prepared the citizens of modern western societies for the discovery in ways that could not have been envisioned by futurists in 1960. Over the years, scientists involved in the greater field of astrobiology (the study of the origin, distribution and fate of life in the universe), along with SETI researchers, have assembled more and more evidence consistent with the hypothesis of life out there. These include the discovery that life’s chemical building blocks are widely distributed, over 400 hundred extrasolar planets, and strange creatures known as extremophiles that live under conditions that were once considered lethal (Darling, 2001). Given what we know today, as compared to what we knew in 1960, the discovery of extraterrestrial intelligence, while not necessarily anticlimactic, should come as less of a surprise.

Any difficulties associated with our transition to a post-contact era are more likely to be of our own rather than ET’s making. ET will be a blank slate, most likely unaware of our existence. The difficulties that we do experience will flow from our expectations and biases, our desire for closure in the absence of firm information, and perhaps bumbling on the part of agencies and organizations that have failed to give the matter advance thought.

Of course, all of this is based on the standard SETI detection scenario. Although a radio telescope search makes this seem the most likely, there are other ways that the discovery could come about. These include finding emissions from fusion, antimatter, and other advanced propulsion systems; infrared emissions, and unexpected cosmic rays. More possibilities include finding evidence of extraterrestrial artifacts or inventions on Earth; inspecting DNA for encoded messages; quantum communication, and finding ET on the Internet. Or, perhaps there will be a breakthrough in UFO studies (Dick, 1996). Perhaps somewhere within the vast UFO files lies an authentic encounter, or as President Clinton laughingly suggested, one or more invisible government bureaucrats are zealously guarding knowledge of an extraterrestrial presence (Wozniak, 2005).

While such possibilities strike many scientists as essentially impossible, Nicholas Taleb points out that many of the powerful forces that shape society are "outliers" or qualify as "black swans" in the sense that they were inconceivable before they were discovered, just as black swans were inconceivable to Englishmen before they were found in Australia (Taleb, 2008). Black swans – which include home computers, the Internet, and Google – are "outliers" in the sense that nothing in the past could point convincingly to their existence. Because of black swans, change proceeds by fits, starts, and jumps, rather than orderly trends. Taleb adds that since we seek order and coherence, we develop after the fact explanations that make black swans seem logical and predictable, perhaps forgetting that we are viewing events through the rearview mirror rather than windshield of a car. We need to think broadly to avoid the "Titanic effect" which occurs when we are so certain that an event (such as the sinking of the Titanic) is unlikely that we give the matter no further thought.

Adaptation to the post-contact world may involve several phases (Harrison et al., 2000). The sheer discovery will force us to assimilate the knowledge that we are not alone in the Universe. This, by itself, will affect our philosophy, our science worldview, our religion, and our culture. If we can discern and interpret a message, we may gain scientific, technical or other information that will provide answers to both philosophical and practical questions. And later still, if we communicate with and interact with the extraterrestrial culture, we may develop a long-term dialogue. Maybe a thousand years from now, first contact will be part of ancient history and we will be interacting with many different extraterrestrial civilizations (Harrison and Dick, 2000). By that time, we may draw on experience rather than imagination, and find that "they" are neither all good nor all bad, but are complex creatures with unique patterns of strengths and weaknesses.

## References

Alexander, V. 1998. The Alexander UFO Religious Crisis Survey. httpwww.com/kelleher/articles/Alexander/survey_religion.html. Accessed April 14, 2006.

Almar, I. and Tarter, J. (in press). The Discovery of ETI as a High Consequence, Low Probability Event. *Acta Astronautica*.

Ashkenazi, M. 1992. Not the Sons of Adam: Religious Responses to ETI. *Space Policy* **8**, 341–349.

Billingham, J., Heyns, R., Milne, D., Doyle, S., Klein, M., Heilbron, J., Ashkenazi, M., Michaud, M.A.G., Lutz, J. and Shostak S. 1999. *Societal Implications of the Detection of an Extraterrestrial Civilization.* SETI Institute Press, Mountain View, CA.

Committee on Science and Astronautics. 1961. Proposed Studies on the Implications of Peaceful Space Activities for Human affairs. U.S. House of 3 Representatives, Eighty-Seventh Congress, First Session, 24 March.

Darling, D. 2001. *Life Everywhere: The Maverick Science of Astrobiology*. Basic Books, New York.

Davies, P. 2000. Transformations in Spirituality and Religion. In A. Tough (Ed.) *When SETI Succeeds: The Impact of High Information Contact.* Foundation for the Future, Bellevue, WA, 51–52.

Dick, S.J. 1996. *The Biological Universe: The Twentieth Century Extraterrestrial Life Debate and the Limits of Science*. Cambridge University Press, Cambridge.

Dick, S.J. and Strick, J.E. 2005. *The Living Universe: NASA and the Development of Astrobiology*. Rutgers University Press, New Brunswick, NJ

Fischer, H.W. 2008. *Response to Disaster: Fact Versus Fiction and its Perpetuation, the Sociology of Disaster*. University Press of America, Lanham, MD.

Goode, E. 2000. *Paranormal Beliefs*. Waveland Press, Prospect Heights, IL

Hargrove, T. and Stemple, G.H. 2008. Scripps UFO Poll Findings. http://www.scrippsnews.come/node 37461 accessed September 19, 2009.

Harrison, A.A. 1997. *After Contact: The Human Response to Extraterrestrial Life*. Plenum, NY.

Harrison, A.A. 2000. The relative stability of belligerent and peaceful societies: implications for SETI. *Acta Astronautica*, **46**: 707–712.

Harrison, A.A. 2003. Confirmation of ETI: initial organizational response. *Acta Astronautica*, **53**: 229–236.

Harrison, A.A. 2007. *Starstruck: Cosmic Visions in Science, Religion, and Folklore*. Berghahn, New York.

Harrison, A.A. 2008. Planning for Contact: Fantasy Documents or Guidelines for Action? IAC-08-A4.2.6.

Harrison, A.A. 2009. The future of SETI. *Futures*, **41**: 554–561.

Harrison, A.A, 2010. The ETI myth: idolatrous fantasy or plausible inference? *Theology and Science* **8**, 51–67.

Harrison, A.A., Billingham, J., Dick, S. J., Finney, B., Michaud, M.A.G., Tarter, D.E., Tough, A. and Vakoch, D.A. 2000. The Role of the Social Sciences in SETI. In A. Tough (Ed.) *When SETI Succeeds: The Impact of High Information Contact*. Foundation for the Future, Bellevue, WA, 71–80.

Harrison, A.A. and Connell, C. (Eds) 2001. Workshop on the Societal Implications of Astrobiology Final Report. NASA, Ames Research Center, 29–31. http:www/http://astrobiology.arc.nasa.gov/workshops/societal/societal_report.pdf accessed 22 June 2009

Harrison, A.A. and Dick, S.J. 2000. Contact: Long-term Implications for Humanity. In A. Tough (Ed.) *When SETI Succeeds: The Impact of High Information Contact*. Foundation for the Future, Bellevue, WA, 7–31.

Harrison, A.A. and Johnson, J.T. 2000 ETI: Our First Impressions. *Acta Astronautica* **46**, 713–718.

Harrison, A.A. and Thomas, J.M. 1996. The Kennedy Assassination, Unidentified Flying Objects and Other Conspiracies: Psychological and Organizational Factors in the Perception of Cover-up. *Systems Research and Behavioral Science* **14**, 113–128.

Hoyt, D. 2000. UFOCRITIQUE: UFOs, Social Intelligence, and the Condon Report. Master's thesis, Virginia Polytechnic Institute and State University.

Michaud, M.A.G. 2007. *Contact with Alien Civilizations*. Copernicus, New York.

National Institute of Discovery Science. 1999. Roper Poll Data – Survey Data and Methodology. http://www.anv.net/nids/RoperPollData.shtml accessed June 13, 1999.

Oliver, C. 2003. SETI and the Media. Thesis submitted to the Management Committee for the Master of Sciences Communication, Central Queensland University.

Peters, T. 2009. Astrotheology and the ETI Myth. *Theology and Science* **7**, 3–20.

Pinotti, R. 1990. ETI, SETI, and Today's Public. *Space Policy* **6**, 161–167.

Schuessler, J. 2000. Public Opinion Surveys and Unidentified Flying Objects: 50+ years of Sampling Public Opinion. Unpublished paper, Mutual UFO Network, Morrison, CO.

Schuessler, J. 2004. Public Opinion Polls and UFOs. Unpublished paper, Mutual UFO Network, Morrison, CO.

Shostak S (1997) Media reactions to a SETI Success. Acta Astronautica, 41: 623-627.

Shostak, S. 1999. EQ Peg. A Detection? http.//www.seti.org/eqpeg.html. Accessed April 19, 1999.

Shostak, S. 2009. *Confessions of an Alien Hunter: A Scientist's Search for Extraterrestrial Intelligence*. Washington, National Geographic.

Shostak, S. and Almar, I. 2006. The Rio Scale Applied to Fictional SETI Detections. IAA-02-IAA.9.1.6.06

Taleb, N.N. 2008. *The Black Swan: The Impact of the Highly Improbable*. Random House, NY.

Tarter, D.E. 2000. Security Considerations in Signal Detection. *Acta Astronautica* **46**, 725–728.

Tough, A. 2000. *When SETI Succeeds: The Implications of High Information Contact*. Foundation for the Future, Bellevue, WA.

Weir, S. 2005. *Encyclopedia Idiotica: History's Worst Decisions and the People who Made Them*. Barron's Educational Series, Hauppauge, NY.

Wendt, A. and Duvall, R. 2008. Sovereignty and the UFO. *Political Theory* **36:4**, 607–633.

Wozniak, L. 2005. Clinton's Worldview: Part Two. http://www.finacenasia.com/articles/4E718926–9027–7E17–4B7EF9778FE76A1C.cfm. Accessed September 14, 2005.

# Epilogue

## Hungarians as Martians: The Truth Behind the Legend

**Philip Morrison,**
*Professor of Physics and Astronomy, MIT (deceased)*

### Editor's Introduction

Much has been written in these pages about the Fermi Paradox, first articulated at Los Alamos in around 1950. It has been reported that one of the Hungarian physicists in attendance on that eventful morning replied to Enrico Fermi's query "Where are they?" with the response, "We are right here, and we call ourselves Hungarians." Thus was born the legend that Hungarians are in fact Martians – or so I had thought. The quip has been widely quoted in the literature, by John McFee and others, and I cited it in late 1997 in one of my "Ask Dr. SETI" columns in the pages of *SearchLites*, the quarterly newsletter of the SETI League, Inc.

Shortly thereafter, there arrived at SETI League headquarters a wonderfully decorative envelope (see Figure 1) from Prof. Philip Morrison, SETI pioneer and co-author of the first scientific SETI article (also much cited in the present volume). I was privileged to have Phil Morrison as a mentor, and am proud to say he was my friend. He and his wife, Phylis, were ardent SETI League members, and I always enjoyed receiving their encouraging letters. This one provided a new perspective on the Martian legend, which (as you will see) Los Alamos alumnus Phil Morrison asked

H. Paul Shuch, *Searching for Extraterrestrial Intelligence*, The Frontiers Collection, DOI 10.1007/978-3-642-13196-7, © Springer-Verlag Berlin Heidelberg 2011

me to keep under wraps. I respected his wishes throughout his life, but now choose to enter his recollection into the historical record.

Here, then, is Phil's letter in its entirety. Enjoy!

13 January, 1998

Dear Paul,

I hope you are on good terms with Dr. SETI. I have a rather delicate matter to send to him via you. I do not want this to be a matter of discussion now, but to stay in the archives and minds of the League until sometime when it is again relevant.

The story he cites (winter 1998 issue of *SearchLites*) of Fermi and Szilard is simply a folk tale, a delightful one perhaps, that grew in postwar Los Alamos, if McPhee is to be trusted. I know this because I am indeed the originator of the theory of Martian origin of the Hungarians. Of course the talents and energies of famous examples were taken as evidence; there, folk lore and history concur. But my reasoning was far different and I believe more cogent. Why would Martians come to the Danube? Is it nicer there than on advanced Mars? Nor would they be fearful of the *barbaroi.*

No, the answer is clearer. The Martians simply were planning, at least on a contingent basis, the eventual need to occupy Earth. Such an expedition is extraordinarily difficult beyond all history. It is naive – this was a wartime story with the tone of that era – to suppose that the first Martians on Earth would be the combatants of the forces of conquest. Even just across the channel to Normandy an invasion was not like that. The Allies knew a great deal about Europe before the landing, and had strong covert support already in place. The earliest Martians to come to Earth were indeed sent as the first intelligence assets. They would plan for a safe base, a large number on staff, and a long lead time to learn all about this planet. A few months or years would not do; you need a millennium or two, and a nation with a strange language provides the safest long-term cover. Their unconcealed intelligence, beauty (recall the Gabors!), and energy are clearly beyond earthly level. (The gypsies are a false note; that people are surely emigrant refugees from Rajasthan in northwest India, whose own language is close to Rajastani. They reached Romania before they came to Hungary, and indeed were found over all Europe west to Spain and Britain.)

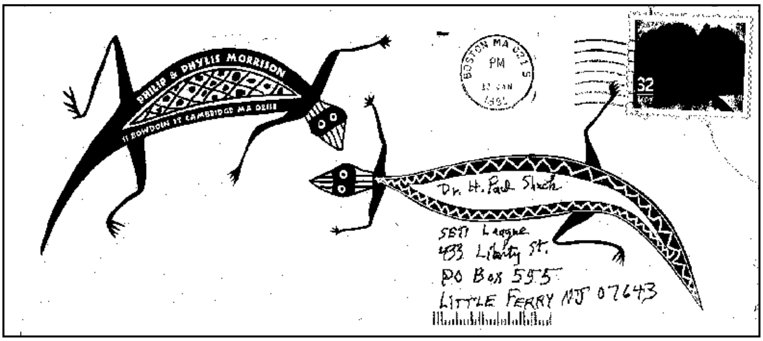
Why strong Hungarian interest in nuclear weapons? Easy; they were finally organizing to divide earthkind in a way that would weaken us profoundly. Szilard began to propose the A-bomb a few years before fission was known. The discovery of fission showed that it was necessary to set the bait. The very threat of a long nuclear war would make us simpletons much easier targets. Their superiority in such simple issues as weapons was not at risk. Get going on high strategy; the time has arrived! They did, and it almost worked, nor is the last word said.

I made up and told my tale widely at Los Alamos in 1945 or maybe 46, first probably to Stan Ulam, long before the McPhee contacts, and indeed before Szilard ever came to Los Alamos, if he ever got there. (I am not certain whether or not he came postwar either; possibly he did.)

My high point in this long-elaborated spoof was telling the great Hungarian aerodynamicist Theodor von Karman, who enjoyed it greatly. "I do not deny", he said, at Cornell some years postwar. This is documented, if not dated, by my aerodynamical friend and associate of Karman, William R Sears, writing in *Physics Today* in 1986, and in his 1994 autobiography called *A Twentieth Century Life*, publisher Parabolic Press, PO Box 3032, Stanford Ca 94309. You would like Sears' book a lot.

I am pleased enough with this funny story not to lose it to local rumors recorded by a writer who wasn't there at the time. It is a delight to see just how fiction has slowly turned into slightly implausible folklore. There is of course nothing important about the credit; that is why I do not want you to print my version in rejoinder. Please, no controversies! But I thought you'd enjoy having the truth discreetly on hand, just in case.
with best wishes and a happy New Orbit to all of you,

Phil M.



# Afterword

## Looking Forward

Professor Allen Tough,
*University of Toronto (Emeritus)*

In recent years, scientists and the general public have realized that intelligent life may well be found throughout the Universe. It is extremely unlikely that we are the only civilization in our Galaxy. It may even contain dozens or hundreds of civilizations scattered among its 400 thousand million stars. If we receive a richly detailed message from one of these civilizations or engage in a lively dialogue, the effects on our civilization could be pervasive and profound.

Contact with intelligent life from somewhere else in our Galaxy will probably occur sometime in humanity's future. It might take the form of a richly detailed radio or laser message from the distant civilization, for instance, or a super-intelligent probe that reaches our planet. We humans have been practicing observational SETI science for a mere 50 years, with no definitive results. But it is not unreasonable to expect that contact might well occur within the next 50 years.

Few events in the entire sweep of human history would be as significant and far-reaching, affecting our deepest beliefs about the nature of the Universe, our place in it, and what lies ahead for human civilization. Seeking contact and preparing for successful interaction should be two of the top priorities on our civilization's current agenda.

Such contact will surely be an extraordinary event in all of human history. Over the next thousand years, several significant events will no doubt have a powerful positive impact on human society. But making contact with another civilization will likely be the event with the highest positive impact of all.

Currently, only a few hundred scientists, social scientists, artists, engineers, and technicians around the world are involved in the pursuit of such contact: the search for extraterrestrial intelligence (SETI). If they succeed, their quest will have an immeasurable impact on human culture, science, philosophy, and society.

Any other civilizations in our Galaxy are probably much older than human civilization. Two factors support this assumption. First, the vast majority of stars in our Galaxy are much older than our Sun, many of them millions of years older. It follows, then, that any civilizations on planets revolving around those stars likely arose much earlier than did our own civilization. Second, it seems quite possible that some civilizations survive for a million years or even longer. If the civilizations in our Galaxy range in age from a few thousand years up to a million years old, then we are one of the youngest: by most definitions, human civilization is not much more than 10,000 years old.

Because other civilizations in our galaxy are thousands of years older than human civilization, they have probably advanced in certain ways beyond our present level of development. Some civilizations presumably fail to survive once they discover nuclear weapons or other means of causing their own extinction, but surely others learn to cope successfully with such problems and then survive for a very long time. Some of them may be hundreds of thousand, or even millions of years more advanced than we are. Imagine what we might learn from them!

Though SETI has not yet succeeded in detecting any repeatable evidence, the range of strategies and the intensity of the efforts are growing rapidly, making success all the more likely in the next half-century. More than one strategy may succeed, of course, so that by the year 3000 we may well be engaged in dialogue with several different civilizations (or other forms of intelligence) that originated in various parts of our Milky Way galaxy.

Because of our society's focus on the immediate present and on the very short-term future, it is difficult to switch into a long-term perspective. As a result, most oral and printed discussions of contact focus on the immediate and short-term effects. The short-term effects of SETI success are likely to be chaotic, frenzied, unsettling – perhaps marked by resistance and conflict, by extreme media reactions, and by political maneuvering or even warfare (military or covert). If handled well, presumably most of these short-term effects will fade within a few years. Our discussion here will focus on the potential effects on human civilization several decades or centuries after contact occurs.

When SETI succeeds, two types of contact are possible:

1. One possibility is simply evidence that another advanced intelligence exists somewhere in the Universe, with little information about its characteristics and no dialogue. One example is evidence of a Dyson sphere or some other major astroengineering project many light-years away, with no additional information about its creators. Another example is a radio message that arrives from many light-years away but is not successfully decoded even after many years of effort.

2. The second possibility is contact that yields a rich storehouse of knowledge about the alien intelligence and its history, technology, science, values, social organization, and so on. This could occur through an encyclopedic radio or optical message that we manage to decode. Because of recent progress in nanotechnology, artificial intelligence, and space exploration, we now realize that close-up contact with a small but super-smart probe is at least as likely a scenario. In fact, by monitoring our telecommunications, the probe will likely have learned our languages and be able to communicate with us quite effectively: no decoding necessary!

We might well receive practical information and advice that helps our human civilization to survive and flourish. Possible examples include technology, transportation, a new form of energy, a new way of producing food or nourishing ourselves, a feasible solution to population growth, more effective governance and social organization, fresh views on values and ethics, and inspiration to shift direction dramatically in order to achieve a reasonably positive future. The message might also bring home to people the importance of eliminating warfare or at least eliminating weapons of extraordinary destruction. Viewing ourselves from an extraterrestrial perspective might be very useful in reducing our emphasis on differences and divisions among humans, and instead seeing ourselves as one human family.

We might gain new insights and knowledge about deep major questions that go far beyond ordinary practical day-to-day matters. Topics in an encyclopedia-like message or close-up dialogue could include astrophysics, the origin and evolution of the Universe, religious questions, the meaning and purpose of life, and answers to philosophical questions. We might receive detailed information about the other civilization (which might be deeply alien to us) and about its philosophies and beliefs. Similar information could be provided about several other civilizations throughout our Galaxy, too. We might even receive a body of knowledge accumulated over the past billion years through contributions by dozens of alien civilizations throughout the Galaxy.

What sorts of consequences will contact have for our religious ideas and institutions? Some religions may be deeply shaken by contact, or at least need to re-examine their set of beliefs. It seems clear, however, that humanity's religions have already flourished over many centuries despite a variety of scientific discoveries that conflict with religious views. And several religions have already incorporated the idea of extraterrestrial life. Although some religious leaders may denounce an extraterrestrial dialogue, most will surely embrace it as further evidence of God's infinite greatness.

Richly detailed information from an alien civilization might transform our view of ourselves and our place in the Universe, even our ultimate destination. We might gain a much deeper sense of ourselves as part of intelligent life and evolving culture throughout the universe – or at least part of a galactic family of civilizations. We might develop a deeper sense of meaning and connectedness to a Universe filled with biology and intelligence. A new cosmotheology or global/cosmic ethic might arise, or a powerful secular movement of altruistic service to the Universe and its long-term flourishing.

We might eventually play a role in some grand galactic project in art, science, philosophy, or philanthropy. Such projects might aim to solve fundamental mysteries of the Universe, help other civilizations develop and flourish, or spread harmonious life throughout their region of the universe.

If we incorporate alien knowledge and advice into our human society, we may experience severe disruption, at least for a short time. We might suffer from enormous culture shock, temporarily feel inferior, or lose confidence in our own culture. Will our science or philosophy "lose its nerve" when faced with far superior knowledge, and permanently retreat into trivia or resistance rather than embracing the new? Massive and rapid change could occur in the sciences if alien science is deeply different, in business and industry if we learn about new processes and products, in the legal system if we move toward cosmic or universal laws, and in the armed forces and their suppliers if we eliminate the threat of war. Probably all of this should be regarded as simply the major cost we have to pay for incorporating new knowledge and possibilities.

During the past five decades, the scientific search for extraterrestrial intelligence has become quite mainstream within science. Several strategies have already been implemented and more are being considered. Public interest is high. It now seems quite possible that our first contact with another civilization will occur within the five decades to come. This first contact will, in turn, lead on to redoubled efforts using a variety of strategies to achieve contact with additional civilizations. Of all the positive events that occur during the next thousand years, this surely will have the most profound and pervasive impact on human civilization. It truly will be an extraordinary event.

# Index