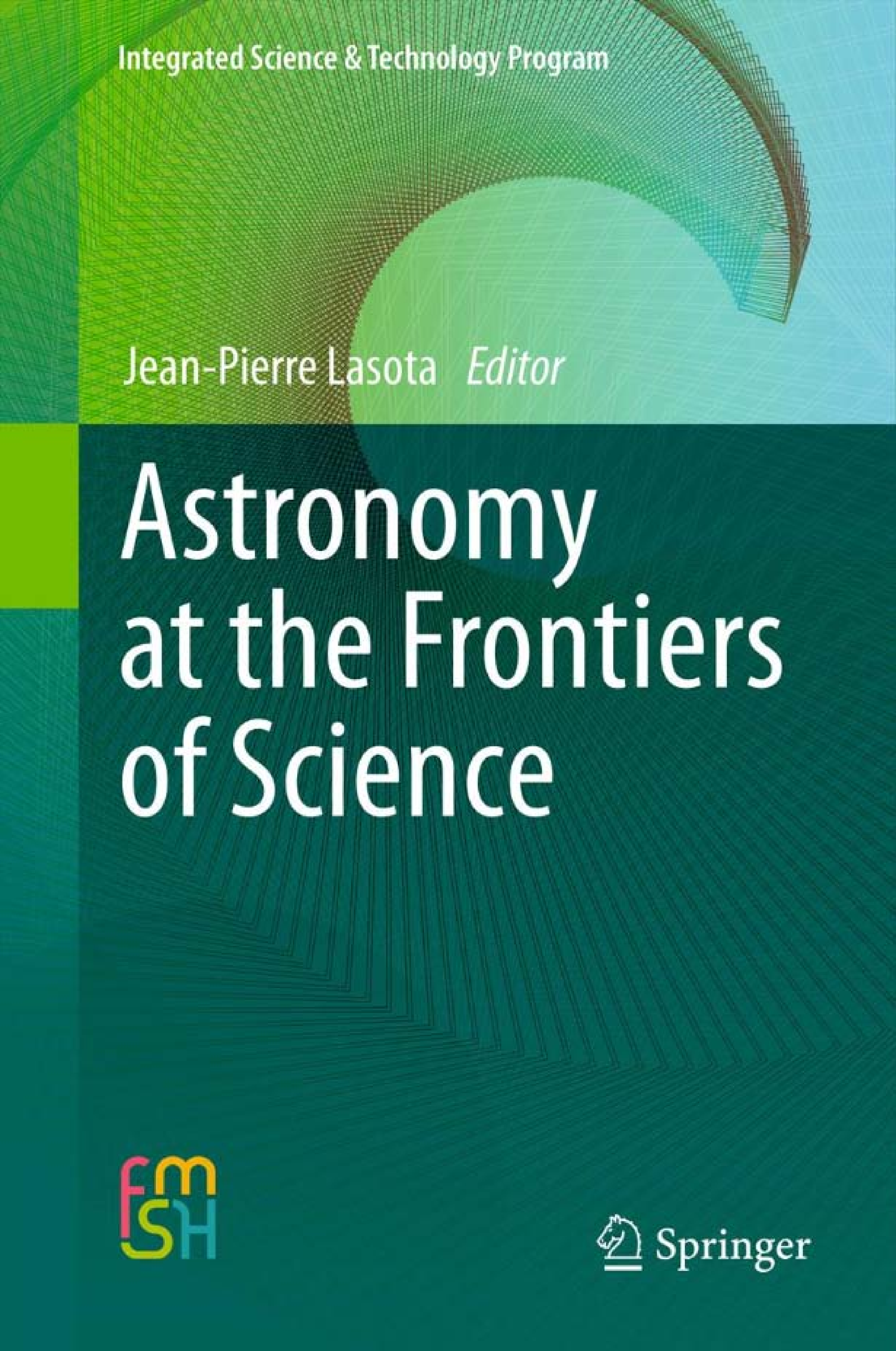


Integrated Science & Technology Program

Jean-Pierre Lasota *Editor*

Astronomy at the Frontiers of Science

Springer

Astronomy at the Frontiers of Science

INTEGRATED SCIENCE & TECHNOLOGY PROGRAM

Volume 1

Aims and Scope

A hub for a large number of national and international scientific and academic networks, the Fondation Maison des Sciences de l'Homme/FMSH (http://www.msh-paris.fr/) helps forge relations between institutions, researchers, and research teams worldwide in order to facilitate research across disciplines. No one discipline is given priority and there are no territorial limitations. The FMSH is committed to its role as a place for experimenting with and fine-tuning ideas, new theories, and types of organization. Founded upon the conviction that interdisciplinarity is essential for understanding the complex questions facing society in a global setting in order to make sense of diverse cultures, the FMSH's mission is to act as "a facilitating, interdisciplinary, and international incubator."

Integrated Science & Technology Program (ISTP) Book Series

Today, the FMSH continues to occupy a unique place in research and to provide a uniquely stimulating environment for intellectual meetings and exchanges. The creation of the *Integrated Science & Technology Programs* (ISTP) facilitates this interdisciplinary exchange and pursues three main goals:

- to highlight the success of integrated approaches in science
- to examine national and international policy with regard to research in science and technology
- to address the cultural, economic, political and social impact of scientific and technological development

The ISTP Book Series includes volumes of selected invited papers and is intended for researchers, practitioners, undergraduate and graduate students in all disciplines who are dealing with complex problems and looking for cutting-edge research as well as methodological tools to set up truly transversal science and technology projects.

Themes

The work generated by exchanges and cooperation between researchers from all countries has helped to delineate research themes that often touch upon the major questions facing contemporary societies. For the Foundation, this means taking a risk on "unconventional" projects that correspond to its mission of "incubator." Through its broad-reaching programs, which address a wide array of cultures and civilizations while maintaining a clear focus on historical depth, international comparisons, and interdisciplinarity, the FMSH assists interdisciplinary projects that seek to improve our understanding of key problems linked to the reshaping of contemporary societies. Examples of themes include:

- Environmental Issues and Sustainable Development
- Earth Sciences and Space Sciences
- Science, Technology and Society
- Cognitive Sciences and Technologies
- Information Technology and Communication
- Health and Life Sciences
- Economics and Finance

Springer is pleased to collaborate with the Fondation Maison des Sciences de l'Homme in publishing this Book Series.

For further volumes:
http://www.springer.com/series/8815

Jean-Pierre Lasota
Editor

Astronomy at the Frontiers of Science

Springer

Editor
Jean-Pierre Lasota
Institut d'Astrophysique de Paris
Boulevard Arago 98bis
75014 Paris
France
lasota@iap.fr

ISSN 2211-2375 e-ISSN 2211-2383
ISBN 978-94-007-1657-5 e-ISBN 978-94-007-1658-2
DOI 10.1007/978-94-007-1658-2
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011933945

Cover design: deblik

Printed on acid-free paper

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

Preface

Astronomy is certainly the oldest science and that of astronomer probably the oldest profession. This second assertion is notoriously debatable, but one can safely assume that in a primitive civilized society the (remunerated) shaman or priest had to be an astronomer to be credible. For a long time astronomy played a very important role in society (and parts of the present book describe some aspects of this role); it is only quite recently that astronomy has been relegated to the status of a more or less ordinary scientific research activity, so that today being an astronomer is just being a scientist like any other. Remnants of past glories survive in countries such as England and Scotland which still have an Astronomer Royal, but these are exceptions.

However, astronomy still has a special place among the other sciences and in society at large. It is by nature an interdisciplinary activity: it incorporates mathematics, various branches of physics, chemistry and biology. In building their instruments astronomers use, and often develop, the latest technology and in order to treat and understand their data, they must use the fastest computers and most refined software.

Astronomy is easy to advertise and popularize: not only does it provide (or try to provide) answers to fundamental questions about the origin and fate of the Universe, as well as dealing with objects such as black holes; astronomers also produce beautiful images and movies which are fascinating to see even without understanding what they represent. Despite its relative decline in importance astronomy is still a special science, and with the expansion of global means of communication it is, in a sense, more special than ever. An astronomer is also somebody who is perceived as possessing special and fascinating knowledge. Quite often at dinners, if the food is good, I claim to be an accountant just to be able to eat and not answer questions about black hole interiors.[1]

---

[1] For some reason this seems to be the most fascinating subject for amateurs.

The present book describes the various relations and interactions of astronomy with other branches of physics and other sciences, as well as the relations between astronomy and technology, industry, politics and philosophy (and superstition), describes modern instruments and discusses what it means to be an astronomer.

Part I, *Astronomy and Physics*, discusses the place of astronomy among various branches of (mostly high-energy) physics. The main characteristic of astronomy is that it is a science of observation, while physics is usually a science of experiment. Guillaume Dubus, rather provocatively, considers whether, in spite of this, the Universe can be considered a laboratory. He shows how in some cases it must be so regarded, but also convincingly warns against the misuse of the laboratory "paradigm" in astronomy. His objections to "physics without limits" find a resonant echo in the essay by George Ellis later in the book. In the following chapter, Paolo Mazzali deals with the main implication of astronomy's observational nature: the necessity of finding standards of distance without the possibility of testing them in situ. A French research minister used to say that astronomers cannot be serious scientists because they are not even sure what is the value of the Hubble constant.[2] Chapter 2 well describes the seriousness of astronomers, who, in finding standard candles, have to struggle with the complex nature of objects and processes in the Universe.

The next three chapters are devoted to what is nowadays called astroparticle physics. This label is somewhat justified in the cases described in Chaps. 3 and 4 and much less so in the case of gravitational-wave astronomy, presented in Chap. 5. But all three chapters deal with what is really a novelty in astronomy: observations of the Universe through non-electromagnetic messengers such as cosmic-rays, neutrinos and gravitational radiation. Cosmic rays have been observed for a long time but the hope that they can be used in astronomy is recent. As is discussed in detail (and from different perspectives) by both Eli Waxman and Günter Sigl, enthusiastic reports that this is already the case were at best premature. The sources of cosmic rays are still unknown. Neutrinos from the Sun and from a supernova have already been observed (their role in the "Universe as a lab" is discussed by Dubus) but here the question is of high energies that have yet to be observed. Waxman explains why we should expect them to be emitted in quantities observable from Earth but for the moment detectors have been observing light emitted by plankton and bacteria – a very nice example of multi-disciplinary research, but not related to astronomy. Sigl also deals with the problems of testing theories in particle physics, which nicely complements the discussion by Dubus. It is interesting to observe the differences of approach to high-energy astrophysics between Dubus, Waxman and Sigl. As for gravitational waves, they have yet to be observed. The detectors are there, the signals too, but for the moment they are produced by trucks on Louisiana roads or earthquakes in Indonesia. We know that the sources are there, it is simply that events which would produce waves detectable on Earth are rather rare. Alessandra

---

[2] The minister being a geophysicist, it would have been too easy (and also unwise) to reply with a remark about the poor predictability of volcano eruptions.

Buonanno aptly describes all the problems, theoretical as well as observational and instrumental, of (future) gravitational astronomy. In less than 5 years the first gravitational-wave signal should hit (so to speak) the (by then improved) detectors.

Part II, *Astronomy in Society* begins with an article by David Aubin about the rise and (relative) fall of the role of observatories as the central places of knowledge creation and storage. The period covered is from the seventeenth century until today, and Aubin shows that for a long time observatories were more than just astronomical; they also involved experimental activity. It is interesting to note that in France, the CNRS National Institute for the Sciences of the Universe (INSU) created Science of the Universe Observatories (French acronym OSU), which regroup within university structures what Aubin calls observatory sciences. It remains to be seen if these OSUs will play, even in part, the role which Observatories played in past times.

Modern astronomy projects are mostly complex technological, technical and managerial enterprises. Modern astronomical instruments employ cutting edge, sometimes unique, technology. There are therefore very strong links and interactions between astronomy laboratories and industry. Astronomy can inspire engineers, but quite often astronomers find the fulfilment of their desires in industry products and inventions. This complex relationship is well explained by two seasoned practitioners, James Lequeux and Laurent Vigroux. They also provide useful examples of what is meant in astronomy by expensive.

Whatever the details, building and running telescopes, launching and using space observatories and probes cost substantial amounts of money. And except in the United States of America, it is public money only. The imagination and curiosity of astronomers are unlimited and the same is true of the variety of practical means of satisfying these needs. This is characteristic of astronomy and in this it differs from the other “big science”, particle physics. It makes no sense to build two accelerators of the same type at the same time, but building several big telescopes does. Or at least it is not absurd a priori. Hence the need, in astronomy, to plan, select, prioritize etc. In her learned and entertaining essay Virginia Trimble describes how this was done in the USA, whose astronomy community was the first to be involved in such exercises. Of special interest are the tables comparing the dreams with their fulfilments. I also appreciate her sobering comments on the attitude of her colleagues towards international collaborations.

Johannes Andersen describes the remarkable effort of unifying European astronomy, which mainly means unifying its planning and general strategy. I am not sure that 50 years ago European astronomy was really a backwater (incidentally it was interesting to read, in Chap. 8, that the old Canadian-French Hawaii Telescope even recently was still producing more publications than some of the US giants), but progress in the last half-century has been tremendous. Andersen tells the story of Astronet, a structure through which European astronomy has been able to define a road-map for its future development. A remarkable achievement, thanks to the help of the European Union (a good example of the role it should play) and the effort and determination of a small group of people, several of whom I have the privilege to count as collaborators and friends.

One of them is Fabienne Casoli, who wrote the chapter devoted to space astronomy – probably the most complex and difficult aspect of astronomical research. The combination of technological and quality requirements, budget constraints, competition with other branches of space research and, last but not least, political (local and international) considerations, makes the planning and realization of space-astronomy projects complicated and often frustrating. All this Fabienne Casoli describes expertly. She ends her chapter with the words: "One must be confident that space astronomy will continue developing and producing fascinating results about the Universe". This is understandable when one is in charge of astrophysics missions in a space agency. Let us hope the Universe will live up to this expectation.

Part III is called: *The Tools of Observation and the Profession of Astronomer*. I decided to start it with two complementary approaches to contemporary astronomy: the use of very-small and very-large telescopes. Neither would be possible without recent advances in detection and data processing technology but they are fundamentally different in almost every other respect. In spite of this Udalski would know perfectly well how to use a 42 m telescope, and Charles would have no difficulty in using a robotic telescope (in fact he got one at SAAO). This is the beauty and strength of astronomy. It is a fundamentally interdisciplinary science but a translator is almost never needed when two astronomers talk. Andrzej Udalski recounts the very successful realization of the ideas of Bohdan Paczyński, who advocated using small telescopes for great science. Small telescopes are now used to discover extrasolar planets and GRB afterglows, to observe variable stars and in many other fields. OGLE IV regularly observes one billion celestial objects. In such a case it is not the telescope that costs most, it is the instruments and data processing. In the case of large and very-large telescopes everything is expensive and complicated. Phil Charles describes in detail the challenges of planning such telescopes and the difficulties encountered when building and commissioning the existing ones. The story he tells is based on his own experience with SALT – the 10 m -class telescope in South Africa. But this telescope is also special: "It is an icon for driving science and technology education in a developing nation." This is one of the natural roles of astronomy, but unfortunately we have not been able to include its description in this book.

Extremely large telescopes (for the moment the idea of building 100 m telescopes has been dropped and the largest telescope planned will have a diameter of "only" 42 m) imply extremely large problems. Eric Ruch devotes his chapter to the "heart" of a telescope: its mirrors. After a very interesting review of the history and progress of mirror technology he discusses the challenges offered by the extremely large telescope. As Phil Charles explains the mirrors of such telescopes must be segmented, composed from many smaller mirrors. The challenge is best summarized by a sentence from Ruch's article that I cannot resist quoting: "… the challenge of the optical industry will be to produce more mirrors for these telescopes in a period of 5 years than in the whole history of astronomy and much more accurately polished than they were in the past". He explains in detail how this will be done.

What will we do with all these data when even Udalski's small telescope provides more than 30 Tb a year? Who will analyze these extremely large amounts of information? Of course analyzing observations is only one part of the question. The main goal of scientific research is not analyzing data but understanding them and their implications for our knowledge of the nature of the Universe and its components. Nowadays, because of this deluge of data, astronomy departments prefer to hire people who are skilful with computers and pay less attention to their education in physics. As a result, even at the best universities it is possible to come across professors for whom the laws of physics are a mystery (I don't mean that they do not have a sufficient grasp of the supposed "theory of everything", aka M-theory, but simply that they are not familiar with the law of energy conservation). This is a very worrying aspect of modern scientific research. In his chapter Mark Allen addresses another, not totally unrelated, question about the ever-growing amount of data: how to store, organize and distribute it. Astronomers have found an original solution that is well adapted to the existing worldwide information net: the Virtual Observatory. Through this, by accessing a global database, any scientist can perform multi-wavelength observations of objects or systems he is interested in. Allen describe show the VO works and provides examples of scientific results obtained through such virtual observations.

Generally, astronomers are very good at sharing their knowledge and results with the general public. This task is made easier by the esthetic qualities of astronomical images (of course these are only a part, and often not the most important, of observational results, but spectra, say, are much less sexy than images, at least for the layman). But if it is easy to attract and impress, it is more difficult (and more important) to educate and explain. Since antiquity the main role in astronomy education has been played by planetaria, which nowadays present celestial shows on a dome-shaped projection screen. Often such planetaria are part of science museums. Mike Shara is a well-known practising astronomer and curator of the American Museum of Natural history in New York, and in his entertaining article he describes how a modern museum/planetarium operates. Of course some aspects of the way this planetarium operates are uncommon (not many planetaria employ actors such as Harrison Ford) but one gets the general idea of the possible scope of education through astronomy.

Finally, in the last article in this part of the book, Bernard Fort describes what it is to be an astronomer. The reader should be warned that he is not a typical representative of the profession. First, he is deeply worried by some fundamental epistemological questions. Most astronomers are similar to the majority of physicists, who, according to Einstein, differ from clergymen in that, while these are interested in the general laws of nature, the physicists, very often, are not. Second, he made a fundamental discovery, which is not the destiny of most of us. I should also mention that I disagree with his view that scientific theories are cultural products. But his narration of his technologically driven research adventure is really fascinating. And when he writes: "Only the confrontation between theoretical models and observations make it possible to perfect our knowledge of the history of the Universe. This confrontation cannot end before any observation is coherently interpreted by a single physical model", one is reassured that he has forgotten all about his "cultural construct" fancy.

The last part of the book *Astronomy at the Frontiers of Knowledge* contains four chapters. The first concerns an association astronomers are not very happy about. For a very long time astronomers were also astrologers; quite often (depending on the country and the religious and political system) their astronomical skills were just supposed to provide tools for divination in the service of the ruler. Unfortunately, in our post-modern times, the relative decline of the societal role of astronomy seems to have had no influence on the ubiquity of astrology. The main difference with ancient times is that nowadays no astronomer is involved in the horoscope business. On the contrary, many astronomers try to debunk this pseudo-science. Astrologers usually reply that astronomers do not know their methods, which they allege to be rigorous and sound. To the best of my knowledge, Marek Abramowicz is the only professional astronomer who answered this challenge by learning the methods of astrology and becoming a (non-practicing) astrologer. When he unmasks astrology, he knows exactly what he is writing about. But his chapter is interesting also because of its historical and methodological perspective.

Cosmology is a very particular branch of science because of the uniqueness of its object. This disturbing characteristic and it consequences are addressed by George Ellis, who critically analyzes some particulars of contemporary cosmology. He calmly questions and rebuts various assertions made by (too) many theoretical physicists for whom anything goes in the Universe. As I have mentioned before, such lack of restraint also worries Guillaume Dubus, but most astronomers just follow the bandwagon (or try to jump on it) without thinking. I greatly value Ellis's skepticism regarding the primitive reductionism that seems to be lately in vogue.

Thérèse Encrenaz provides us with a fascinating account of the exploration of the bodies of our Solar system. Planetology is a branch of astronomy (some would like it to be considered as a sort of "external" geophysics, usually to justify the unification of astronomy with particle physics, but this is nefarious idea) but it is the only branch where astronomers can study objects in situ: where one can get directly under the surface of a celestial body. Of course, the beauty of astronomy is that learning about the Sun's interior does not entail (pace Auguste Comte) having to go there, but still the prospect of sending drilling rovers to the surface of Mars and Europa (as described by Fabienne Casoli) has a sort of fantasy-like quality.

Of course the aim of all this drilling is to find water, and who says water, says life (or rather a possibility of life). Hence the existence of astrobiology. Although, according to Virginia Trimble, it is "said by some to be a subject with no subject matter to study", Muriel Gargaud & Stéphane Tirard provide a convincing rebuttal to such mischievous assertions. But their article is mainly about working in an interdisciplinary subject. Some readers will be surprised by the amount of space they devote to the problem of the common tongue to be used in this field. They should not be. Many examples show that it is an important, sometimes crucial, problem. Günter Sigl also mentions this difficulty. Not surprisingly, since one of the well-known mishaps of research on ultra-high-energy cosmic rays was the misunderstanding by particle physicists of what is meant in astronomy by a catalogue. Another example:

years ago, high-energy astrophysicists from one of the French space laboratories asked my advice about observing nuclear lines in the spectrum of an X-ray Nova. I told them not to try. Being γ-ray astronomers, they did not realize that X-ray astronomers misused the term "Nova" (which normally designates the light emitted by a thermonuclear explosion on the surface of a white-dwarf, whereas the so-called X-ray Nova results from an accretion-disc eruption, and my friends would be looking in vain for nuclear lines there). So finally a translator is sometimes needed even in astronomy.

The present book provides a fairly broad, although not complete, picture of the relation of astronomy to the other sciences and its place in society in general.

The book is the first of the Integrated Sciences & Technology Programs (ISTP) Book Series of the Fondation Maison des Sciences de l'Homme/FMSH[3] published by Springer.

I am grateful to Angela Procoli from the FMSH and Jean-Marie Hameury, director for Astronomy and Astrophysics at the INSU/CNRS for entrusting me with the task of editing this volume. I also thank Jean-Marie Hameury for his critical and careful reading of the manuscripts of this book.

Institut d'Astrophysique de Paris, UMR 7095 CNRS, Jean-Pierre Lasota
UPMC, Univ Paris 06, 98bis Boulevard Arago,
75014, Paris, France

Astronomical Observatory, Jagiellonian University,
ul. Orla 171, 30–244 Kraków, Poland

---

[3] http://www.msh-paris.fr/.

Contents

Part III The Tools of Observation and the Profession of Astronomer

Part IV Astronomy at the Frontiers of Knowledge

Contributors

Marek Artur Abramowicz After graduation in 1968 at Wroclaw University he worked two years in its Institute of Mathematics. In 1970 he moved to the Institute of Astrophysics in Warsaw, and since then never worked as a mathematician. In 1974 he received his Ph.D. at Warsaw University in theoretical physics. The same year he started his long odyssey: Stanford, Austin, Oxford, Trieste (SISSA and ICTP), Copenhagen (Nordita), and finally Gothenburg, where from 1994 he is a full professor of astrophysics and chair. Since 2007 he is also a visiting professor at the Copernicus Center in Warsaw. His interests include theory of black holes, accretion disks, and the nature of inertial forces. He supervised about twenty doctorates, including those of Piero Madau, Omer Blaes and Jufu Lu.

Mark G. Allen obtained his Ph.D. in Astronomy and Astrophysics from the Australian National University in 1998. After a postdoctoral position at the Space Telescope Science Institute he joined the Observatoire de Strasbourg, and is currently a researcher in the Centre National de la Recherche Scientifique (CNRS). His astronomical interests include active galactic nuclei and making world wide astronomical data interoperable though the development of 'Virtual Observatories'.

Johannes Andersen is Professor of Astronomy at the University of Copenhagen, Denmark, and currently Director of the Nordic Optical Telescope, La Palma, Spain. He has also held visiting appointments for five years in France, Canada and the USA, and observed extensively at ESO in Chile. His main scientific interests are the structure and evolution of stars and the Milky Way Galaxy, astronomical instrumentation, and international cooperation, with some 360 publications in these fields. He has served as Chair of the ESO Scientific and Technical Committee, as General Secretary of the International Astronomical Union, on the Bureau of COSPAR, and as Chair of the Board of the ERA-NET ASTRONET, and currently chairs the OPTICON Telescope Directors' Forum and a number of other committees.

David Aubin is Professor of History of Sciences at the Université Pierre et Marie Curie (UPMC) in Paris and a member of the Institut de mathématiques de Jussieu. The author of several articles on the history of mathematics and of the observatory sciences, he is the co-editor with Charlotte Bigg and H. Otto Sibum of *The Heavens on Earth: Observatories and Astronomy in Nineteenth-Century Science and Culture*. He has also published a collection of essays on the ninetieth-century transits of Venus in the *Cahiers François Viète* (Nantes, 2006).

Alessandra Buonanno is a theoretical physicist working in gravitational-wave physics and cosmology. Her research focuses in particular upon the dynamics and gravitational-wave emission from coalescing binary black holes, the interface between analytical and numerical relativity, and post-Newtonian theory. She has conducted research to design advanced optical configurations of laser interferometer gravitational-wave detectors operating at and below the standard quantum limit. She is also interested in studying physical sources of gravitational-wave emission in the early Universe. Alessandra Buonanno received her *Laurea* degree and her Ph.D. in Physics from the University of Pisa. She held postdoctoral positions at CERN, at the Institut des Hautes Études Scientifiques near Paris, and at the California Institute of Technology. She was also a staff researcher at the France's Centre Nationale de la Recherche Scientifique. Today she is physics professor at the University of Maryland. She has been a Richard C. Tolman Fellow, an Alfred P. Sloan Fellow, and she has recently become a Fellow of the International Society on General Relativity and Gravitation.

Fabienne Casoli is in charge of the space science and exploration program at CNES, the French Space Agency. As an astrophysicist, she has studied galaxy evolution and star formation. She has been director of the Institut d'Astrophysique Spatiale in Orsay near Paris and President of the French Astronomical Society SF2A. She has been involved in the beginnings of ASTRONET, a network of European funding agencies for astronomy, funded by the European Union seventh Framework Program (FP7).

Phil Charles has a Ph.D. in X-ray Astronomy, obtained at the Mullard Space Science Laboratory of University College London in 1976, which led to his involvement in NASA's HEAO-1 X-ray observatory through the Space Sciences Laboratory of the University of California, Berkeley. There he expanded his research activities to include optical studies of X-ray sources with large ground-based telescopes, and multi-wavelength studies of high energy sources have been a feature throughout his career. In 1980, Charles was appointed to a University Lectureship in Astronomy at Oxford University and became a Fellow of St Hugh's College, where he continued his work in coordinated optical and X-ray observations of accreting, compact X-ray sources involving white dwarfs, neutron stars and black holes. This led to him becoming Principal Investigator for ISIS, the principal spectrograph on the 4.2 m William Herschel Telescope on La Palma, and included a spell of 5 years on La Palma as Head of the Astronomy Group, and Deputy Officer-in-Charge with the Royal Greenwich Observatory. From 1994–1999 Charles was Head of Astrophysics at Oxford. In 2000, Charles joined the University of Southampton as Professor and Head of Astronomy in the School of Physics & Astronomy, and was instrumental in seeing Southampton join the multi-national collaboration that constructed SALT, the Southern African Large Telescope. In 2004, Charles took leave-of-absence from Southampton to become Director of the South African Astronomical Observatory following the sudden passing of its previous Director, Bob Stobie. SAAO is now responsible for operating SALT on behalf of the SALT partnership. Charles also holds an Honorary Professorship at the University of Cape Town, and has fostered close links between SAAO and UCT. He is involved in a number of international scientific collaborations between South Africa, Spain, India and the UK.

Guillaume Dubus is a high-energy astrophysicist with the CNRS at the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG). His main research interest is physical processes around compact objects, especially in binary systems with a black hole or neutron star. His current focus is understanding high energy gamma-ray emission from those systems. Guillaume Dubus has been involved since 2003 with the HESS collaboration, which operates a gamma-ray observatory in Namibia. He is also involved with the Fermi/LAT collaboration, a space-based gamma-ray detector launched in 2008. Prior to moving to Grenoble, Guillaume Dubus held his CNRS position at the Laboratoire Leprince-Ringuet, a particle physics lab at the École Polytechnique.

George F.R. Ellis, FRS, is Professor Emeritus of Applied Mathematics at the University of Cape Town. He has worked extensively on general relativity theory and relativistic cosmology. Published books include *The Large Scale Structure of Space Time* with Stephen Hawking, *Dynamical Systems and Cosmology* with John Wainwright, and *On the Density of Matter in the Universe* with Peter Coles. Recently he has been engaging with issues in the philosophy of cosmology. He is Past President of the International Society of General Relativity and Gravitation and is presently co-Editor in Chief of the Journal of General Relativity and Gravitation.

Thérèse Encrenaz is senior scientist at CNRS (France). She works at LESIA (Laboratory for Space Research and Instrumentation in Astrophysics) at the Observatoire de Paris. Her field of expertise is the study of planetary atmospheres, mostly by remote sensing techniques (infrared and millimeter spectroscopy) and, more generally, the study of the origin and evolution of solar-system objects. Her publication lists includes about 200 articles in refereed journals and about ten lecture and/or popular books. Thérèse Encrenaz has been using many large ground-based telescopes and has been involved in many planetary space missions, including Galileo toward Jupiter, Mars Express, Venus Express and Rosetta. She has been a Mission Scientist of the European ISO (Infrared Space Observatory) mission. She has been heading the Space Research Department of Paris Observatory between 1992 and 2002, and she has been the vice-president of the Scientific Council of Paris Observatory between 2007 and 2011.

Bernard Fort started to work in astronomy at the Paris Observatory on the solar corona. After 1976, he was among the first few astronomers who promoted CCD techniques after guessing they will provide the best astronomical detectors for extragalactic astronomy. With this objective in mind, in 1982 he moved to Toulouse where with the financial support of CNRS he developed a CCD laboratory dedicated to deep sky observations with the Canada France Hawaii telescope. With his team he developed the first multi – object spectrographs used at CFHT and ESO and discovered giant gravitational arcs in cluster of galaxies. His enthusiasm has allowed his team and students to play a leading role in a new field of cosmology: the mapping of dark matter in cluster of galaxies and cosmic shear observations. For his pioneering observational work, Bernard Fort received several astronomical prices, including the Janssen gold medal of astronomy from the French Sciences Academy, and he was also elected to the Royal Astronomical Society. He was Director of the Institut d'Astrophysique de Paris from 1998 to 2005, where he is still associated as CNRS emeritus research director.

Muriel Gargaud is CNRS research scientist at the Laboratoire d'Astrophysique de Bordeaux. She studies the origins and evolution of life on Earth and its possible distribution elsewhere in the Universe. She is vice-president of the Société Française d'Exobiologie and has just produced, as chief editor, an encyclopaedia of astrobiology of 2,000 entries (Springer, 2011) to which more than 300 scientists have contributed.

Jean-Pierre Lasota is CNRS Emeritus Research Director at the Institut d'Astrophysique de Paris and Professor of physics at the Astronomical Observatory of the Jagellonian University in Cracow. He has worked mainly on problems of relativistic and high-energy astrophysics, structure and evolution of close binary systems, in particular on accretion onto black-holes. For many years he was Director of the Department of Relativistic Astrophysics and Cosmology at the Paris Observatory. During the construction of the gravitational wave detector *Virgo* he served on its steering committee and later was member of the EGO Scientific and Technical Advisory Committee. Since 2002 has been scientific advisor for astroparticle physics at the French National Institute for the Sciences of the Universe (INSU). He published recently *La Science des Trous Noirs* (Odile Jacob 2010).

James Lequeux started in radioastronomy in 1954 as a student. After a long military service, he observed the structure of continuum radiosources with an interferometer in Nançay and obtained a Ph.D. in 1962. Then he worked on the construction of the large Nançay radiotelescope, and founded the first infrared astronomy group in Meudon in 1966. He was deeply involved in the genesis of the IRAM project. He directed the Marseilles Observatory from 1983 to 1988. Later, he was also involved in the science with the Infrared Space Observatory (ISO) as an associate scientist. He was during 15 years one of the two Editors-in-chief of the European journal *Astronomy & Astrophysics*. After a career in various fields of astrophysics, mainly interstellar matter and evolution of galaxies, his post-retirement interests turned to the history of astronomy. He published several books and a large number of papers in this field.

Paolo A. Mazzali was born in La Spezia, Italy. He was educated in the Italian school system, and spent most of his youth playing football (soccer). After making the probably incorrect decision that a career in science was more exciting than one in football, he went to the University of Manchester, UK, where he received an Honours B.Sc. in Physics in 1984. From there he moved to UCLA, where he obtained a Ph.D. in Astronomy under the guidance of Mirek Plavec in 1989. His topic of research was radiation transport in stellar winds, which took him to Munich to work first with Rolf Kudritzki's group at the LMU and then with Leon Lucy at ESO as a postdoc. There he became interested in applying radiation transport to Supernovae, in particular using Montecarlo techniques. His next mistake was to imagine that a scientific career in Italy would be a good thing, so he accepted a research Astronomer position at Trieste Astronomical Observatory. He spent several years there working mostly on Supernovae, his main interest being trying to extract quantitative information from the data, the quality of which was becoming better and the quantity larger and larger. He spent significant time in Japan working with Ken Nomoto's group, and was involved in the discovery and interpretation of the first GRB/SN, 1998bw. His other interest, Type Ia Supernovae, finally took him back to Munich, where he helped to coordinate an EU-RTN led by Wolfgang Hillebrandt. He was eventually promoted to an Associate professorship in Trieste, and later transferred to Padova Observatory. He has continued to work in the area of Supernova research, and lately he has been seconded to the Scuola Normale Superiore, in Pisa, a decision motivated by the impression that life in a university in Italy would be better than in a research institute. This was the third mistake, and there are certainly more to come. Still, he enjoys doing science, when he can.

Eric Ruch has graduated in optical engineering from the Institute of Optics in Paris. He has joined REOSC in 1985, has worked in lens design and holds three patents. He has been Project manager for several space and astronomy projects, such as the Infrared Space Observatory (ISO) for the European Space Agency and the HELIOS I and II project for the CNES. Since 2006, he is in charge of business development for space and astronomy activities of the REOSC department in Sagem. Eric Ruch is also teaching optical systems and optical design at the Institute of Optics Graduate School in Paris since 1992.

Michael Shara is Curator in the Department of Astrophysics at the American Museum of Natural history. He produces space-shows seem by millions every year. He also leads a team of research astrophysicists that studies the structure and evolution of novae and supernovae; collisions between stars and the remnant descendants of those collisions; and the populations of stars inhabiting star clusters and galaxies. Prior to joining the Museum, he was with the Space Telescope Science Institute at Johns Hopkins for 17 years where he was responsible for the peer review committees for the Hubble Space Telescope. He is also adjunct professor of Astronomy at Columbia University. Dr. Shara frequently observes with the Hubble Space Telescope and other large ground-based telescopes. In his spare time he is a gung-ho scuba diver enjoying the company of sharks in the South Pacific.

Günter Sigl is Professor at the University of Hamburg since 2007. He obtained his Ph.D from the Ludwig-Maximilians University in Munich in 1993. In the years 1993–1996 he was a "Feodor Lynen fellow" (awarded by the Alexander von Humboldt foundation) at the University of Chicago where he was later a research scientist until 1999. From 1999–2005 he was a French Centre National de la Recherche (CNRS) Chargé de Recherche at the Institut d'Astrophysique de Paris (IAP) and in 2005–2007 Directeur de Recherche at the Astroparticules et Cosmologie (APC) institute. He specialises in astroparticle physics theory and phenomenology. He is in particular interested in ultra-high energy cosmic rays, their propagation and the development of the general numerical code CRPropa, multi-messenger simulations including photons and neutrinos, dark matter indirect signatures from radio to gamma-rays, neutrino astrophysics, neutrino oscillations in supernovae and the early Universe, astrophysical and cosmological constraints on photon mixing with light axion-like particles or hidden photons, and in the origin and evolution of primordial magnetic fields. He is also an associated member of the Pierre Auger Observatory.

Stéphane Tirard is Professor of epistemology and history of science and director of the François Viète Center of Epistemology and History of Science at the University of Nantes. In particular he studies the history of theories of the origins and evolution of life on Earth, and the topic of limits of life. He recently published *Histoire de la vie latente, Paris*, Vuibert-Adapt, 2010.

Virginia Trimble is a native Californian and graduate of Hollywood High School, UCLA (BA, physics & astronomy, 1964), and Caltech (Ph.D. astronomy, 1968). She holds an honorary MA from Cambridge University and a Doctora Honoris Causa from the University of Valencia. Trimble is currently Professor of Physics & Astronomy at the University of California, Irvine, and an advisory staff member of Las Cumbres Observatory Global Telescope Network. From 1973 to 2003 she spent half of each academic year at the University of Maryland, where her late husband, Joseph Weber, inventor and builder of the first detectors for gravitational radiation, was tenured. Her current research interests include the structure and evolution of stars, galaxies, and the universe, and of the communities of scientists who study them, this last generally being called history of science and scientometrics. She has served as a vice president or council member of the American Astronomical Society, the American Physical Society, the International Astronomical Union, Sigma Xi, and a few other professional organizations. She is currently a member of Council and of its executive committee of the American Association for the Advancement of Science and vice-chair (to become chair in 2013) of the California-Nevada section of the APS.

Andrzej Udalski, Professor of Astronomy at Warsaw University Observatory is the leader of the Optical Gravitational Lensing Experiment (OGLE) – the longest microlensing survey operating continuously since 1992 that for the last 18 years has been uninterruptedly bringing new top rank astrophysical discoveries.

Laurent Vigroux is the President of the ESO Council since 2009. Before, he was the French scientific delegate to the ESO Council since 2002 and he is member of the ALMA Board since 2006. He is the director of the Institut d'Astrophysique de Paris since 2005, after having lead the Service d'Astrophysique of the French Atomic Energy Commission (CEA) in Saclay for 9 years. He was a member of the ASTRONET (an European network of astronomy funding agencies) working group in charge of the definition of the European road map for large facilities in astronomy, and co-lead the panel concerned with facilities in the UV, visible, infrared and radio domain. He was involved in large project for ground based telescopes and for space observatories: PI of MEGACAM, a wide field imaging camera for the Canada-France-Hawaii telescope, instrument scientist of the infrared camera ISOCAM on board of ISO, a satellite launched by ESA in 1995, and he is co-PI of SPIRE, one of the three instruments on board of the Herschel Space Observatory launched by ESA in 2009. Beyond design and operation of instruments, most of his scientific activities are related to the observations of galaxies and the modeling of their evolution.

Eli Waxman was born in Petach-Tikva, Israel, in 1965. He has gained his higher education at the Hebrew University of Jerusalem, within the framework of the IDF Talpiot Program. In 1994 he has completed his Ph.D. thesis "Self-similar solutions to Euler's equations" under the supervision of Professors D. Shvarts, NRCN, and G. Rakavy of the Hebrew University of Jerusalem. From 1994 to 1998 he was a long-term member at the Institute for Advanced Study in Princeton, NJ. In 1998 Eli Waxman returned to Israel and joined the Physics Faculty of the Weizmann Institute of Science, where he serves presently as professor of physics and head of the Particle Physics & Astrophysics department. His research field is theoretical astrophysics, with focus on high-energy neutrino and cosmic-ray astrophysics, relativistic astrophysics, supernovae and gamma-ray bursts. Eli Waxman is married to Vered, and he is the father to Iddo (19) and Chen (20).

Part I
Astronomy and Physics

Chapter 1
The Universe as a Laboratory for High-Energy Physics

Guillaume Dubus

Abstract Physics is validated through careful experimental work and its progress is punctuated by great experiments: Newton decomposing light with prisms, Thomson's discovery of the electron, Michelson's experiment on the speed of light through ether etc. Direct experimentation, whether ground-based or space-based, remains the method of choice. Yet, high-energy physics, the study of the fundamental constituents of matter and their interactions, has moved to the point where it can address conditions that cannot be tested by direct experimentation. Can the distant Universe then be used as a laboratory? How have astronomical observations tested and expanded our knowledge of high-energy physics? Is this affecting the way astrophysics is done? These are the questions addressed in this contribution.

Keywords Astroparticle physics • Relativistic processes • Gamma-ray burst: general • Cosmology

1 Can the Universe Be Used as a Laboratory for Physics?

Using the Universe as a *laboratory* for physics may appear as wishful thinking, if not entirely preposterous. Laboratories are visualised as ordered spaces, controlled environments in which scientists with white coats design and carry out experiments, experiments that are analysed and refined until all of their parameters are understood, all of their uncertainties subdued. The outcome is an experimental protocol leading to results that can be repeated and verified by others. In contrast, "the Universe" conjures up images of something inaccessible, beyond our reach and

---

G. Dubus (✉)
UJF-Grenoble 1/CNRS-INSU, Institut de Planétologie et d'Astrophysique
de Grenoble (IPAG) UMR 5274, Grenoble, F-38041, France
e-mail: Guillaume.Dubus@obs.ujf-grenoble.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science
& Technology Program 1, DOI 10.1007/978-94-007-1658-2_1,

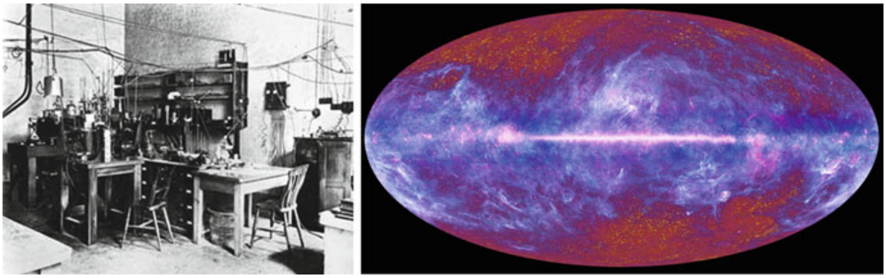


Fig. 1.1 Famous laboratories of physics: Rutherford's lab at the Cavendish in Cambridge *circa* 1920 and the microwave sky, showing foreground emission from the Milky Way and the Cosmic Microwave Background (CMB), as seen by ESA's *Planck* observatory *circa* 2010 (Picture credits: Courtesy AIP Emilio Segrè Visual Archive and ESA Planck LFI, HFI consortia [6])

our control, of something unintelligible of which we are only a passive spectator (Fig. 1.1). The two views would seem irreconcilable and this probably stems from the deeply rooted preconception that the world populated by mankind and the heavens are distinct spheres governed by different rules.

1.1 Gravity: The Historical Showcase

Yet, one of the deepest foundations of science is that the laws derived on Earth should apply equally well anywhere else in the Universe. Indeed, the beginning of modern science is usually traced back to the discovery of the laws of gravitation and planetary motion. For the first time, laws divined on Earth are seen to apply up to the achievable accuracy to phenomena outside our realm. Confidence in the measurements can be increased by independent, repeated or simultaneous, observations. We have a clear experimental protocol to test a theory using space as our laboratory. Moreover, Newton's law unifies various phenomena under the same umbrella: from the fall of the apple to the movement of the Moon, tides, the shape of planets, the evolution of their orbits and spins can all be calculated to provide predictions amenable to tests via observations. The theory succeeds because it organises a large set of facts and because it proposes new observables.

Not only is the Universe accessible to the human mind but space provides a vast playground to test theories in the absence of other effects that can plague measurements or on scales impossible to realise on Earth. This is both enviable and delicate: we elaborate hypotheses as to the pertinent physics at work and improve the apparatus with which we observe but we have no control on the experimental setup. This can limit the precision to which a value can be derived. For instance, measuring the value of the gravitational constant – one of the least-well constrained fundamental constants – can be done only by careful direct experimentation [15].

This has not prevented astronomical observations from verifying predictions of general relativity, such as gravitational lensing, that are inaccessible to direct experimentation.

In verifying our knowledge of physics, we seek to match predictions from established theories with observations in novel environments. *Expanding* our knowledge of physics boils down to the search for disagreement. A subtle issue is then to decide if the mismatch represents a true deviation from known physics or a simply a deficiency in the observation or the interpretation. The explanation by general relativity of the advance of perihelion of Mercury, that differed significantly from the expectations from Newton's theory of gravity, was all the more compelling that the observational issues and several interpretations based on classical physics had been carefully considered and discarded. Even then, the decisive observation was that of the deviation of starlight near the Sun, an observation that stemmed from a distinctive prediction of general relativity that could not be accounted for by any classical interpretation.

1.2 Laboratories in High Energy Physics

Newton's law of gravitation symbolises the first steps of a program that continues to this day, the endeavour to render intelligible the world around us through science. Modern high energy physics is a consecration of this vision with the explicit goal of achieving a *theory of everything* that would explain the fundamental constituents of our Universe and the basic laws governing their interactions. Progress is made through the extensive use of induction and falsification, a logical sequence that starts with the casting of hypotheses, continues by conceiving and designing apparatus to infirm these, by analysing their results, abandoning the dead branches of ideas not borne out by experiments and that concludes with the recast of new hypothesis as one progresses in the tree of knowledge. The story of the discovery of the neutrino epitomises this concerted effort balancing theory and experimentation [11]. Today, the *Large Hadron Collider* (LHC) at CERN (Fig. 1.2), the quintessential modern-day laboratory in high energy physics, involves thousands of scientists from around the world organised around its four detectors. The size and cost of the machine leave little place for hit and miss. It is designed to make specific measurements that will test quantitative theoretical predictions, most prominently to find evidence for the Higgs boson, a particle thought to be at the origin of mass. In 1964, J. R. Platt wrote of high energy physics that

> the theorists in this field take pride in trying to predict new properties or new particles explicitly enough so that if they are not found the theories will fall [25].

Platt was arguing that the astounding string of successes achieved by high energy physics compared to other branches of knowledge was due to the systematic use of strong inference. The LHC is undoubtedly a crowning achievement of this method.

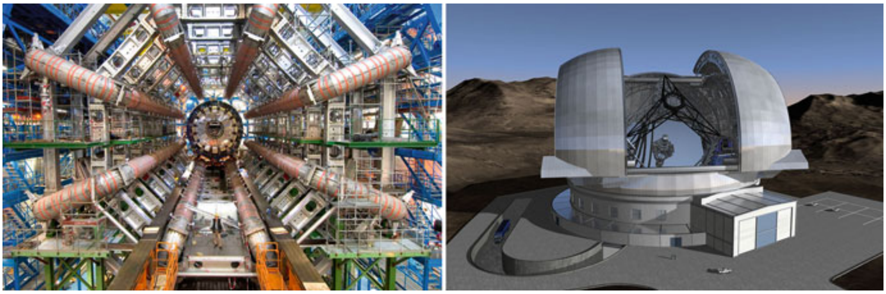


Fig. 1.2 Flagships of high-energy physics and astrophysics: the ATLAS detector at CERN's *Large Hadron Collider* (LHC) and ESO's planned *European Extremely Large Telescope* (E-ELT) (Credits: ATLAS experiment at CERN and ESO, [29])

1.3 Laboratories in Astrophysics

Astrophysics also has its string of successes in the last century fueled by rapid technological advances that have vastly expanded the number of observables (low fluxes, wide fields, fast timing, multi-wavelength, etc., [35]) and by a liberal application of inductive reasoning constantly challenged by these new observations. What hypotheses explain the widest set of observations? Are they supported by new observations, anticipated or not?

A major point of intersection is cosmology where astronomical observations have been used to infer that the dominant constituents in the Universe are dark matter and dark energy. The evidence is all the more compelling that it comes from different independent sets of observations (for dark matter: big bang nucleosynthesis, the rotation curves of galaxies, confinement of hot gas in clusters etc). Hypotheses concerning their nature are formulated and then tested through the usual means of high energy physics (e.g. search for dark matter particles at the LHC or with sensitive detectors in underground laboratories) or by using astronomical observations (e.g. gamma-rays emitted when dark matter particles decay [5]). The method matters, observation replacing experimentation, not the means. Observations are a perfectly legitimate way of testing hypotheses and, in this sense, the Universe is indeed a laboratory for high energy physics.

Observatories, on the ground or in space, in ever greater sizes and with ever more sensitive detectors, mustering ever greater resources and investments, have become the focal points of an astrophysical community organised and structured increasingly like the high energy physics community. The *European Extremely Large Telescope* (E-ELT, Fig. 1.2), the quintessential modern-day laboratory in astrophysics and the future flagship of ESO, an organisation modelled on CERN, is representative of this evolution. One of its main objectives is the study of dark energy. This convergence of high energy physics and astrophysics has not gone unnoticed and the last section will come back to this.

2 How Have Astronomical Observations Tested and Expanded Our Knowledge of High-Energy Physics?

The theory of gravity and its tests using observations of the Universe has been mentioned. This section provides other examples of how data on astrophysical phenomena have been used to test and expand frontier knowledge in high energy physics.

2.1 High Energy Physics with the Sun

The link between high energy physics and astrophysics is, perhaps, most visible in cosmology. Yet, there is no need to look far to find several examples illustrating fruitful exchanges between astronomical observations and fundamental discoveries in high-energy physics. Our Sun has provided, and continues to provide, a useful laboratory from which major results have emerged.

2.1.1 The Discovery of Helium

Helium is the second most abundant chemical element in the Universe but it was discovered only in 1868 when Janssen and Lockyer noticed a strong line in solar spectra that corresponded to no known element. Wollaston and Fraunhofer had discovered absorption lines in spectra of the Sun in the early 1800s. Bunsen and Kirchhoff had established in the late 1850s that spectral lines in hot gases allow its elements to be identified and had found cesium in this way [9]. The interpretative framework for the observations of new lines in the Sun's spectrum was set. Yet, the attribution of the unidentified lines in the solar chromosphere to a new element was met with skepticism. It took 30 years before helium could be successfully isolated on Earth by Ramsay and others. This demonstrated that the constituents of the Universe can be determined remotely, provided the laws of physics are universal. There is a kinship between these observations and current work that shows baryonic matter accounts for less than 10% of the matter content of the Universe.

2.1.2 Nucleosynthesis

The source of the Sun's energy was a major puzzle until progress in nuclear physics made it possible to establish that this is provided by the fusion of hydrogen into helium in the core. The application to astrophysical objects also led to new discoveries for nuclear physics. Fusion opened up the possibility that elements up to iron could be manufactured by the stars, heavier elements being obtained by neutron capture. Whereas the paths involved in the fusion of hydrogen into helium were described by Hans Bethe in 1939, it was not possible to go beyond and produce significant quantities of carbon from lighter elements in stars given the nuclear reaction rates known at the time. Hoyle conjectured in 1953 that synthesising carbon required the existence of an as-yet unknown resonance at 7.68 MeV in an excited nuclei of ^{12}C, a hypothesis that was quickly confirmed by direct experimentation [19]. Nowadays, nucleosynthesis intimately connects nuclear physics and astrophysics. The application of high energy physics theory to the big bang explains the abundances of light elements measured today. Our current understanding of the origin of everything we manipulate in daily life is entirely derived from the combination of high energy physics theory and astronomical observations. The theory and measurements are so delicately intertwined that it is possible to set upper limits on the density of exotic particles in the early universe because of the observable effects they would have on nucleosynthesis, thereby testing models for dark matter [17].

2.1.3 The Standard Solar Model and Neutrino Oscillations

Knowledge of the nuclear reaction rates yield the energy input rate in the core of the Sun from which its structure may be derived using radiative transfer and hydrodynamics. Conditions vary with mass or composition, giving predictions of the radius, luminosity or colours for different stars that continue to be investigated in ever greater details by stellar astrophysicists. Using these stellar structures, astrophysicists can calculate how stars oscillate in response to perturbations. The observation of these oscillations in the Sun, helioseismology, brings exquisite constraints on the internal structure of our star: the sound speed in the Sun's interior derived from these measurements matches theory to within 0.1%.

Nuclear reactions in the core of the Sun produce neutrinos that can escape freely from the core (the dominant reaction in the Sun is the $p-p$ nuclear fusion chain $4p \rightarrow^4 \mathrm{He} + 2e^+ + 2\nu_e + 25$ MeV which occurs at temperatures $\approx 10^7$ K). In the 1960s, high energy physicists started programs to detect these solar neutrinos, which they did except there was a dearth of detections compared to predictions (Fig. 1.3). Inaccuracies in nuclear reaction rates and problems with the experimental setups were successively ruled out as the missing solar neutrino problem became acute [11]. Helioseismology then ruled out that the problem was due to inadequate astrophysical knowledge, leaving only neutrino oscillations as the solution [4]. Neutrinos propagating through vacuum or matter have a mixed probability of appearing as one of three flavours (ν_e, ν_μ, ν_τ; there are independent constraints on the number of neutrino families – three – including from big bang nucleosynthesis), which requires that neutrinos have a non-zero mass. Neutrinos of one type produced in the Sun's core are missed when they appear as another flavour to which the detector is not sensitive. This was confirmed in the past decade using neutrinos created when cosmic rays hit the atmosphere and with neutrinos produced in nuclear

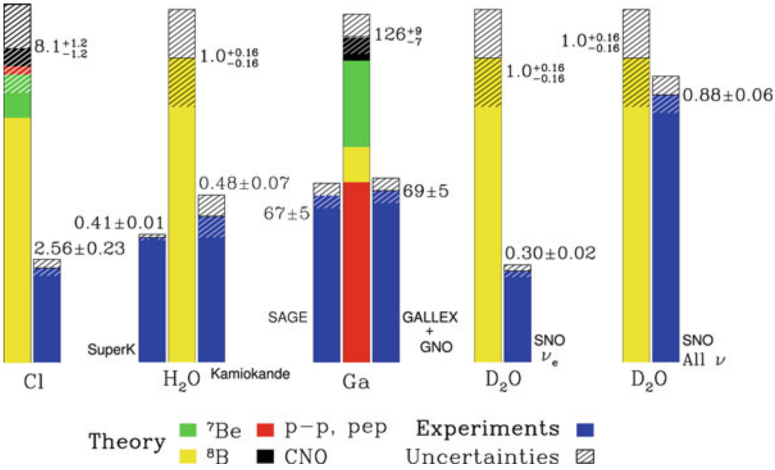


Fig. 1.3 Fusion in the Sun's core results in neutrino emission. This graph compares predicted and observed neutrino fluxes (in 2005) for several experiments (plot credit: J. Bahcall [3]). Each set of bars corresponds to a detection technique (Cl, water, etc). For each technique, the detected neutrino rate from various experiments is compared to the expected rate using the standard solar model and weak interaction model. The contribution to the neutrino rate from each nuclear fusion process ongoing in the Sun ($p - p$, ^{8}Be, etc) is detailed in the theoretical bar plot. The uncertainties in the expected and detected rates are also shown. Some of the detection techniques clearly led to large disagreements between expected and detected neutrino rates. The 40 year long effort to understand whether the discrepancies revealed problems with nuclear, solar or neutrino physics led to the discovery of neutrino oscillations. Solar neutrinos oscillate between flavours, not all of which are detectable by the experiments (only the *Sudbury Neutrino Observatory*, SNO, was sensitive to all neutrino types and has an observed flux matching predictions)

reactors. Neutrino oscillations are not part of the standard model of particle physics and measuring precisely how neutrino flavours mix is the focus of much activity.

Today, the same methodology is being used to constrain the properties of some dark matter particles using the Sun. Some (e.g. neutralinos) can be captured by the Sun's gravitational field, concentrate in its core and annihilate. Others (e.g. axions) can be created in the core and carry energy away from it. Constraints can be derived from the observable consequences on stellar models (including the solar neutrino flux !) or from the search on Earth for a flux of such particles from the Sun [5, 24].

2.2 *High Energy Astrophysics*

High energy astrophysics exemplifies the successful use of the Universe as a laboratory. The first *deliberate* attempts to constrain fundamental theories of high-energy physics from astrophysics can probably be traced back to the early 1960s and the beginnings of X-ray astronomy. This was all summed-up by Rees in 1974:

> The traditional kind of astrophysicist is, in a sense, an "applied" physicist, who computes models for stars and galaxies based on relatively well-understood properties of atoms and nuclei, Newtonian gravity, and other branches of classical physics. But recently radio and X-ray observations have revealed some fascinating cosmic objects and phenomena where the inferred energies, densities, and gravitation field strengths are so extreme that we cannot be confident that we know the relevant physics. The physical assumptions themselves, and not merely the astrophysical models, are then vulnerable to observational test; and the astrophysicist can feel that he has a symbiotic rather than a parasitic relationship with his physicist colleagues [27].

High energy astrophysics has since then sought to test and push theories to their limits or even beyond. Here are a few examples.

2.2.1 Neutron Stars

The detection of steady, rapid radio pulsations from an astrophysical source by A. Hewish and J. Bell in 1967 can only be explained by the rotation of an extremely dense object. A normal star or even a white dwarf would be disrupted by centrifugal forces if forced to rotate on periods shorter than 1s. Stellar oscillations would not be expected to gradually slow down, as observed with pulsar periods. Gold and Pacini independently recognised in 1968 that magnetised neutron stars (pulsars) were the solution. White dwarfs were observationally known at the time but neutron stars, more compact objects supported by neutron degeneracy pressure and nuclear interactions instead of electron degeneracy pressure as in white dwarfs, were known only to theorists interested in highly condensed states of matter.

The existence of objects that squeeze a solar mass of material to super-nuclear densities offers the opportunity to constrain the behaviour of matter under the most extreme conditions [21]. Various hypotheses for the equation of state of matter at these densities can be distinguished through the measurement of the mass and radius of neutron stars. The accretion of material onto the neutron stars can lead to crustal heating and runaway nuclear fusion, which are used to constrain processes in nuclear physics (neutrino cooling, capture processes).

Many neutron stars are also inferred to possess huge magnetic field B resulting probably from the amplification of the star's field during collapse (conservation of magnetic flux $\propto BR^2$). Magnetars harbour fields of several 10^{15} G when the strongest man-made magnetic fields only reach 10^6 G. This is well above the critical field for which the Compton wavelength of an electron becomes equal to the radius of its gyration around magnetic field lines

$$B_{\rm crit} = \frac{(m_e c^2)^2}{e c \hbar} \approx 4 \times 10^{13}\ \mathrm{G}. \tag{1.1}$$

Quantum effects cannot be neglected at such extreme field intensities offering new prospects to test QED, the modern theory of electromagnetism [16]. For example, vacuum birefringence (never experimentally verified) means that photons travelling along or perpendicular to magnetic field lines will propagate differently.

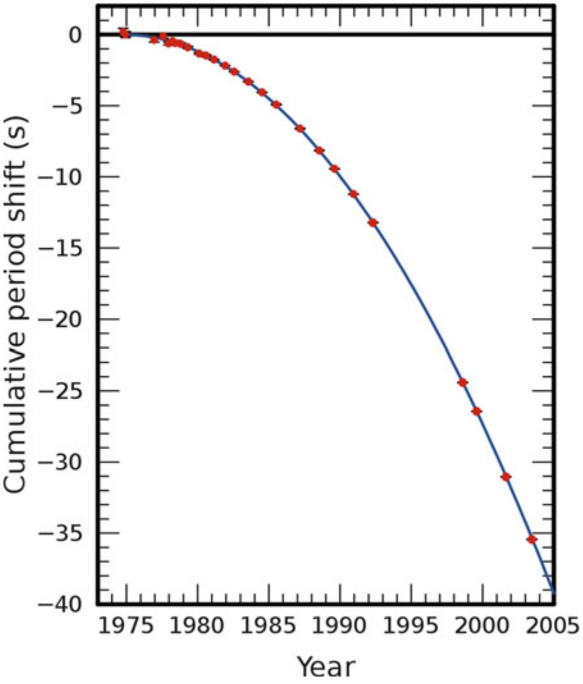


Fig. 1.4 Precise timing of radio pulsars allows many tests of general relativity. Here, the measured orbital decay of pulsar PSR B1913+16 is compared to the expected decay (solid line) due to gravitational wave emission over a timespan of 30 years [33]. The observation of this decay by Hulse and Taylor in 1978, using a fraction of the dataset shown here, constituted the first (indirect) proof for gravitational waves, a prediction of general relativity (Plot credit: Wikipedia [33])

Neutron stars may also provide ways to constrain the coupling between photons and the hypothetical axion (the probability for a conversion of photon to axion is $\propto (BL)^2$ where L is the path length).

The most emblematic use of neutron stars to test physics has been the determination by Hulse and Taylor in 1974 of the rate at which the 8 h orbital period of the binary pulsar PSR B1913+16 decreased (Fig. 1.4). The binary parameters can be determined extremely precisely using the doppler shift of the 59 ms pulse as the neutron star moves in its orbit. Hulse and Taylor showed that the time of periastron passage gradually decreased. General relativity predicts that binary motion in tight orbits will generate gravitational waves carrying away orbital energy and angular momentum. The theoretical calculations match the observations so precisely that they are now used to constrain alternate theories of gravity. The discovery in 2003 of the binary system PSR J0737-3039 where pulses from each of the two neutron stars are detected brought even more possibilities to test general relativity. Pulsars make extremely accurate clocks. A daring proposal is to use very precise timing of an array of millisecond pulsars (old neutron stars with very stable pulsations) spread across the sky to search for slight deviations due to the passage of low-frequency gravitational waves. Big bang theory predicts a relic background of such gravitational waves.

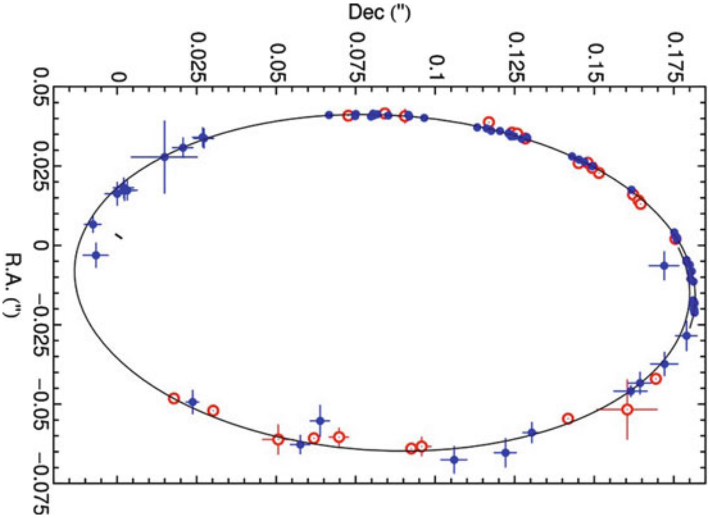


Fig. 1.5 Stellar and gas dynamics have revealed the presence of black holes in the Universe. The current best evidence is the 16-year long Keplerian orbit of a star (S2) in the centre of our Galaxy (At right, from [14], figure copyright 2009 reprinted with permission from the AAS). S2 passes within 18 light-hours from the derived center-of-mass (shown by a small line close to the origin). Only a 4×10^6 $M_\odot$ black hole can explain such a large mass enclosed by such a tight orbit

2.2.2 Black Holes

Black holes are one of the strongest links between fundamental theory and astrophysical observations. They are a clear prediction of general relativity, entirely and fully described by their mass, spin and charge. X-ray observations showed the existence of very compact objects in tight orbit around normal stars and with masses well above the maximum mass ($\approx 3M_\odot$) above which no known physical process can prevent a neutron star from collapsing onto itself. Only black holes fit the bill.

However, the best evidence for a black hole now comes from the observation of the movement of stars in our Galactic Centre. The orbit of the closest star approaches within 100 AU ($\sim$the size of our Solar System) of an object with a mass of 4×10^6 $M_\odot$, Sgr A* (Fig. 1.5). This mass and the density of matter it implies rule out every known alternative but a black hole [13].

Observations clearly favour the existence of black holes. For all practical purposes their presence in the hearts of galaxies and in some binaries is certain. *Proving* their existence is an extremely difficult task, underlining some of the difficulties that can arise when using the Universe as a lab. Even the stringiest constraints on the minimum density of matter enclosed by the stars at our Galaxy's centre will not prove that Sgr A* is a black hole rather instead of some exotic object not yet thought of. Proving an object is black hole requires finding evidence for its defining characteristic: the horizon beyond which light is trapped. Indirect evidence for horizons was inferred from the brighter X-ray emission from neutron stars compared to black holes, which is attributed to energy released at the surface of neutron stars but that disappears behind the horizon in black holes. High-resolution imaging of the region around Sgr A* at mm or infrared wavelength may lead to observing the black-hole's silhouette within the next decade [26] but the ultimate proof can be brought only by observations of merging black-holes (see Chap. 5).

Astrophysicists are also busy trying to find ways to measure the spin of black holes by the reddening it causes on emission or the space drag it imposes on accreting material. Measuring these properties through the observation of X-ray spectral lines and/or quasi-periodic oscillations is a major goal of the future *International X-ray Observatory*. Such measurements can lead to tests of general relativity in the strong field regime (when the curvature GM/R^3c^2 is high [26]). Mention should also be made of Hawking radiation from black holes, a prediction combining quantum mechanics and relativity, which is therefore at the frontiers of theoretical knowledge. However, Hawking temperatures of astrophysical black-holes are much lower than that of the CMB so instead of emitting they absorb radiation. Radiation from hypothetical primordial mini black-holes has still to be observed.

2.2.3 Cosmic-ray Physics

There is every second, in a surface of a square meter, a proton or nucleus with an energy greater than 100 GeV impacting the Earth's atmosphere. The cosmic origin of these particles has been known since 1912 when Victor Hess showed that this ionising flux increases with altitude. Many ground-based or space-based particle detectors have measured the flux, composition, energy and arrival direction of cosmic rays. Their observed energies reach several 10^{20} eV. The collision of such a particle with a proton at rest in the atmosphere yields more than 10^{14} eV in the centre-of-mass frame, one order-of-magnitude above the energies reached with the LHC. The discovery of the positron (antimatter) by Anderson in 1932 (Fig. 1.6), of the muon (1936), the pion (1947) and other particles were made using observations of cosmic rays. Accelerators, with controlled injections and collisions, became the tool of choice after World War II. Observations of ultra-high energy cosmic rays (UHECR, $>10^{18}$ eV, [23]) still push the limits of particle interaction models derived from accelerator data.

One hundred years after their cosmic origin was established, we still do not understand where cosmic rays come from (see also Sigl, Waxman). In fact,

> at first [cosmic rays] were utilised mainly as a convenient source of energetic particles for particle physicists during the pre-accelerator days. Only in the early 50s was their astrophysical significance fully realized [35].

Cosmic rays are charged particles so their trajectories are scrambled by propagation and diffusion on Galactic magnetic fields. Up to 10^{15}–10^{18} eV, cosmic rays probably get their energy from Fermi acceleration in the supernova remnants of our Galaxy. Accelerating particles to greater energies puts enormous requirements on the magnetic field and size of the astrophysical source (gamma-ray bursts are

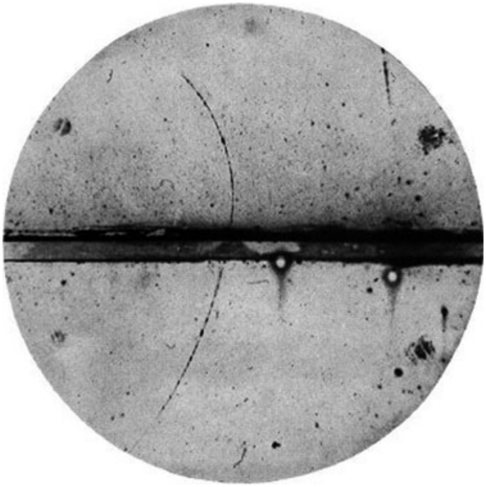


Fig. 1.6 Earth is bathed by a continuous flux of particles with energies greater than what can be achieved in man-made accelerators. The study of cosmic rays has led to the discovery of several fundamental particles, starting with the positron. This is C. D. Anderson's picture of a 63 MeV positron of cosmic origin going through his cloud chamber from his discovery article (Figure copyright 1933 reprinted with permission of the APS from [2])

thought to be the most likely sources of UHECR). Because of this, UHECR have been suggested to be the product of the decay of exotic particles or topological defects. UHECR are not confined by Galactic magnetic fields and can have an extragalactic origin. If UHECR are protons, then they have enough energy to create e^-e^+ pairs and pions by interacting with photons from the 2.7 K cosmic microwave background. There should be an observable diminution in the flux of UHECR due to this energy loss above $\approx 5 \times 10^{19}$ eV (this is called the Greisen-Zatsepin-Kuzmin or GZK cutoff). The characteristic energy-loss length implies that protons with energies $>3 \times 10^{20}$ eV come from within $\approx$30 Mpc from us. The idea that UHECR hinted at new physics was entertained when the AGASA reported results inconsistent with a GZK cutoff. For instance, this could be due to a violation of Lorentz invariance (required by special relativity and that implies, for instance, the conservation of $E^2 - p^2c^4$ in any frame). The *Auger* collaboration operates a gigantic detector array in Argentina built largely for the purpose of settling this question. They have accumulated in the recent years a dataset superseding all others. The *Auger* dataset shows the expected GZK cutoff and also an anisotropy in the arrival directions of UHECR, firmly pointing to astrophysical sources.

Cosmic rays at lower energies are also being investigated for signatures of frontier physics. Reports of an excess of electrons and positrons with energies around 100 GeV and of an excess in the e^+/e^- ratio compared to the standard astrophysical model were interpreted as the contribution from the decay of dark matter particles. This has not been entirely corroborated by other measurements and our current knowledge of astrophysical sources and e^-e^+ propagation in the Galaxy are still too uncertain to rule out a conventional explanation [22]. The *Alpha Magnetic Spectrometer* (AMS), due for launch on the last space shuttle mission, will provide high quality measurements of the cosmic-ray spectrum at these energies as well as search for antimatter helium, which is not expected to occur in known astrophysical sources and, if detected, would require a revision of the role of antimatter in the evolution of the Universe.

2.2.4 Multi-messenger Astronomy

The detection of an anisotropy in UHECR arrival directions opens up the prospect of identifying the sources using images reconstructed from the cosmic-ray arrival directions. *Multi-messenger* astronomy using cosmic ray, neutrino and gravitational wave detectors brings new sources of information on the Universe complementing photon astronomy, exactly like radio, IR, X-ray and gamma-ray astronomy complement visible light. It is too early to tell exactly how observations by these instruments will challenge physics but there is no doubt that they will be used for this purpose.

The first (and only) astrophysical image of the sky in neutrinos shows the Sun [18]. The detection of an excess of neutrinos detected in coincidence with the collapse of supernova SN 1987A vindicated the standard supernova scenario but also triggered efforts towards building a neutrino detector capable of identifying other astrophysical sources. Neutrino emission must occur in the sources of cosmic rays since interactions with high-energy protons produce pions that decay into particles including high-energy neutrinos [12] (see also Waxman). The most advanced project is *ICECUBE* at the South Pole.

The *Virgo* and *LIGO* collaborations search for gravitational waves from phenomena involving masses of order of the mass of the Sun (e.g. binary neutron star coalescence). They use km-sized laser interferometers to measure the slight deviation in path length (smaller than the size of a nucleus) caused by the passage of a gravitational wave (see also Buonanno chapter). The planned upgrades will make binary mergers observable within 100 Mpc. The merger rate in this volume is $\gtrsim$1/year and this should lead to the first direct detection of gravitational waves. This would be a tremendous intellectual and technological achievement [30]. A space mission, *LISA*, is also proposed. With arms of millions of km, the interferometer should be sensitive to the gravitational waves from merging massive black holes throughout the observable universe. The waveform detected during mergers provides an unrivalled means of seeing how the theory of gravity works at its extreme. The exact distance to the event can be deduced by comparison to theoretical waveforms so that, if an electromagnetic counterpart and a redshift are found, this will give a new, precise and independent way to calibrate the extragalactic distance scale.

3 Is This Affecting the Way Astrophysics Is Done?

Although the first use of the Universe as a laboratory is arguably the comparison of the movement of planets with the predictions of Newton's law of gravity, it is only in the last hundred years or so that astronomical observations have been increasingly used for insight and tests of physical theories. This has led to successes, some of which have been recapped above, and ambitious proposals to test the very foundations of physics. It has also led to pitfalls and has somewhat affected the way astrophysics is done.

3.1 Convergence

The equations of general relativity can be used to describe the evolution of the Universe as a whole and this introduced a significant qualitative change to the way astrophysics is perceived. From an effort to understand the workings of objects and phenomena in the sky, astrophysics becomes a path to fundamental insights into the nature of the world around us. Any initial skepticism that pertinent calculations or observations can be made on the Universe as a whole were blown away by the discovery of its expansion and the cosmic microwave background. Seemingly far-fetched hypotheses like inflation are actually being verified by precise measurements of the perturbations left on the CMB.

Cosmology has become such a fertile meeting ground between high energy physics and astrophysics that even the most basic tenets of physics are now thought to be within the realm of experimentation, including the universality of the laws of physics. For example, we can test whether the fundamental constants governing the laws of physics changed with time [32]. There are claims that the ratio of the frequencies of spectral lines changes with redshift, implying that the fine structure constant (the constant involved in the calculation of energy levels in atoms and molecules) had a different value in the early Universe. Even more ambitious ideas are that cosmological observations can test the Copernican principle [31] or constrain the existence of other universes, some of which may be governed by entirely different laws of physics [28]. How confident we have become in the use of the Universe as a laboratory (see Ellis chapter)!

Nowadays, the Universe as a laboratory has become a pillar in the justification of the development and funding of astrophysics. Understanding *the extremes* or the *physics* of the Universe stands alongside the quest for the origins and the search for life in the top questions of both the 2007 European *ASTRONET* report (see Andersen) and the 2010 US Decadal Survey (see Trimble chapter). Such is the perceived symbiosis that one could read in a *Science magazine* special issue on particle astrophysics

> researchers have begun explorations at the boundaries between particle physics, astrophysics, and astronomy [...] It's likely that in the next 10 years, one of these efforts will lead to a major discovery [8].

There is ground for optimism but this should not blind us to some difficulties discussed below.

3.2 Pitfalls

With the increasing pace of research in physics, the pressure from funding agencies, are we sometimes going too far in wanting to identify new phenomena with new physics? The detection of very high energy gamma rays from the vicinity of the

Galactic Centre or of an excess in the positron fraction in the composition of cosmic rays were promptly interpreted as signatures of dark matter although explanations are readily found that involve no new physics or astrophysics (respectively: standard electromagnetic emission from the vicinity of the central black hole or a pulsar wind nebula, injection of positrons by nearby pulsars). The temptation to put forward ground-breaking hypotheses from experimental data is neither new nor condemnable in itself. After all, eminent physicists like Niels Bohr were prepared to abandon energy conservation to interpret β decay before Pauli hypothesised the existence of the neutrino. However, Bohr and Pauli were faced with a phenomenon that could not be satisfyingly explained by any theory at the time (unlike the examples above) and Pauli's conjecture actually led to verifiable consequences (the particle had to have such and such property that could be observed in such and such a way [11]).

Recently, the *Fermi Gamma-ray Space Telescope* observed an 31 GeV photon emitted in a distant ($z = 0.9$) gamma-ray burst (GRB 090510). Gamma-ray bursts are thought to be produced when a massive star collapses or a binary star merges to form a black hole. This photon, which had the highest energy ever observed in a GRB, arrived 0.8 s after the start of the event as measured with lower energy photons. The lag was used to place a lower limit on the energy scale at which Lorentz invariance may be broken. More prosaically, the question is whether light propagates at the same speed in vacuum regardless of its energy. Some theories of quantum gravity (theories thus going beyond standard physics) propose that this is not the case. A delay would arise in the arrival time of photons of different energies emitted at the same time. This delay can be written as

$$\Delta t \propto \frac{1}{H_0}\frac{\Delta E}{E_{\mathrm{QG}}}. \tag{1.2}$$

where E_{QG} is the energy scale at which this effect appears and H_0 is the Hubble parameter ($\approx$70 km s^{-1} Mpc^{-1}). Assuming the delay due to quantum gravity is less than the observed $\approx$1s delay between the start of the burst and the detection of the 31 GeV photon sets a lower limit on E_{QG} slightly greater than the Planck energy scale $E_p = (\hbar c^5/G)^{1/2} \approx 10^{28}$ eV, as can easily be derived from the above equation [1].

Is there much to be derived from this exercise? Some articles in the press hailed this as a test of Einstein's theory of relativity: it isn't since c is implicitly assumed to be constant when using the observed delay as an upper limit on Δt. The lower limit on E_{QG} excludes some theories of quantum gravity, a theory of which is required in the search for a theory of everything but which is not required at all to explain GRBs. In fact, delayed high energy emission in a GRB is much more likely to reflect the astrophysics of black hole formation than some fundamental property of our Universe. The observation of a delay is not a major puzzle in itself. Therefore, given our limited understanding of the astrophysics of the source, it is unlikely that observations of delayed emission will lead to the robust detection of some trick in the speed of light or that great insights into a quantum theory of gravity will be gained from these constraints.

Not all astrophysical results have fundamental consequences. In fact, few do and it would be a mistake to analyze them and judge their worth from the unique vantage point of high-energy physics [34]. Physics at the frontiers should also be no excuse for physics without limits. Anything goes in the Universe, who's there to check anyway? Astrophysics relies on a wide body of evidence continuously tested for consistency. Astronomical phenomena can rarely be studied in isolation so that assuming non-standard physics (e.g. a new particle) is never entirely without consequences on other subfields (e.g. stellar evolution). The relevant use of the Universe as a laboratory for high-energy physics, especially when it comes to finding evidence for new physics, requires well-identified astrophysics.

3.3 Divergence

Differences will and should remain between astrophysics and high energy physics. A recent CERN press release stated that

> as soon as they have "re-discovered" the known Standard Model particles, a necessary precursor to looking for new physics, the LHC experiments will start the systematic search for the Higgs boson [7].

Whereas new particle accelerators redo measurements previously made before moving into new territory, astronomical observations are not all guaranteed to yield the same results because of changing conditions in the astrophysical source unbeknownst to us. New telescopes do check their results against previous measurements (if only for calibration purposes) but all astronomical observations are essentially unique with an importance for future work that cannot be assessed *a priori*. There is little hierarchy in the archival value of astrophysical data: observations taken in the eighteenth century can be as important as data taken yesterday with cutting-edge instrumentation (e.g. historical records that date supernovae remnants seen today).

The phenomena that can be observed, or are actually observed, are not decided by our understanding of physics and so care must be taken that we do not narrow our perspectives by focusing on specific measurements, leaving opportunities for the unexpected to be identified [10]. Accurate measurements in cosmology involve the processing of huge amounts of observational data into a few numbers like the acceleration of the expansion rate of the Universe with redshift. These same data might be used for many other studies, some we can imagine and others we cannot yet. Indeed,

> our celestial science seems to be primarily instrument-driven, guided by unanticipated discoveries with unique telescopes and novel detection equipment. With our current knowledge, we can be certain that the observed universe is just a modest fraction of what remains to be discovered [20].

Contingency and serendipity play major roles in the observation of the Universe and this should not be forgotten when we use it as a laboratory [34].

Acknowledgements I acknowledge support from the European Commission under contract ERC-StG-200911.

References

1. A.A. Abdo et al. (Fermi collaboration), Nature **462**, 311 (2009)
2. C.D. Anderson, Phys. Rev. **43**, 491 (1933), http://prola.aps.org/abstract/PR/v43/i6/p491_1. Cited 20 Aug 2010
3. J.N. Bahcall, A.M. Serenelli, S. Basu, Astrophys. J. Lett. **621**, L85 (2005). http://www.sns.ias.edu/~jnb/SNviewgraphs/snviewgraphs.html. Cited 20 Aug 2010
4. J.N. Bahcall, M.H. Pinsonneault, S. Basu, J. Christensen-Dalsgaard, Phys. Rev. Lett. **78**, 171 (1997)
5. G. Bertone, D. Hooper, J. Silk, Phys. Rep. **405**, 279 (2005)
6. Cavendish lab picture, British Information Services, 30 Rockfeller Plaza, New York, http://photos.aip.org ; http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=47333. Cited 2 Sept 2010
7. CERN press release 19/3/2010, http://public.web.cern.ch/press/pressreleases/Releases2010/PR05.10E.html. Cited 27 Aug 2010
8. A. Cho, Science **315**, 56 (2007)
9. J. Emsley, *Nature's Building Blocks: An A-Z Guide to the Elements*, 2nd edn. (Oxford University Press, New York, 2002)
10. A.C. Fabian, in *Serendipity*, ed. by M. de Rond, I. Morley (Cambridge University Press, Cambridge, 2010)
11. A. Franklin, *Are There Really Neutrinos?* 2nd edn. (Westview Press, Boulder, 2003)
12. T.K. Gaisser, F. Halzen, T. Stanev, Phys. Rep. **258**, 173 (1995)
13. R. Genzel, F. Eisenhauer, S. Gillessen, Rev. Mod. Phys. **82**, 3121 (2010)
14. S. Gillessen et al., Astrophys. J. Lett. **707**, L114 (2009)
15. G.T. Gillies, Rep. Prog. Phys. **60**, 151 (1997)
16. A.K. Harding, D. Lai, Rep. Prog. Phys. **69**, 2631 (2006)
17. K. Jedamzik, M. Pospelov, New J. Phys. **11**, 105028 (2009)
18. M. Koshiba, *Nobel Lecture* http://nobelprize.org/nobel_prizes/physics/laureates/2002/koshiba-lecture.pdf (2002)
19. H. Kragh, Arch. Hist. Exact. Sci. **64**, 721 (2010). doi:10.1007/s00407-010-0068-8
20. K.R. Lang, Science **327**, 5961 (2010)
21. J.M. Lattimer, M. Prakash, Science **304**, 536 (2004)
22. J. Lavalle, in *Invisible Universe*, ed. by J.-M. Alimi, A. Füzfa. AIP Conference Proceedings, vol. 1241 (American Institute of Physics, Melville, 2010), p. 398
23. M. Lemoine, G. Sigl (eds.), *Physics and Astrophysics of UHECRs*, Lecture Notes in Physics, vol 576 (Springer, Berlin, 2001)
24. K. Nakamura et al., Particle Data Group, J. Phys. G **37**, 075021 (2010)
25. J.R. Platt, Science **146**, 347 (1964)
26. D. Psaltis, Living Rev Relativ **11**, 9 (2008). http://www.livingreviews.org/lrr-2008-9. Cited 25 Aug 2010
27. M. Rees, PNAS **72**, 4685 (1974)
28. M. Tegmark, Found. Phys. **38**, 101 (2008)
29. The ATLAS Experiment at CERN, http://www.atlas.ch and ESO http://www.eso.org/public/images/e-elt-1_2008/. Cited 2 Sept 2010
30. K.S. Thorne, in *Particle and Nuclear Astrophysics and Cosmology in the Next Millenium*, ed. by E.W. Kolb, R.D. Peccei (World Scientific, Singapore, 1995)
31. J.-P. Uzan, in *Dark Energy, Observational and Theoretical Spproaches*, ed. by P. Ruiz-Lapuente (Cambridge University Press, Cambridge, 2010)

32. J.-P. Uzan, Rev. Mod. Phys. **75**, 403 (2003)
33. J.M. Weisberg, J.H. Taylor, in *Binary Radio Pulsars*, ed. by F.A. Rasio, I.H. Stairs. ASP Conference Series, vol. 328, (2005), p. 25. http://en.wikipedia.org/wiki/File:PSR_B1913+16_period_shift_graph.svg. Cited 2 Sept 2010
34. S.D. White, Rep. Prog. Phys. **70**, 883 (2007)
35. L. Woltjer, *Europe's Quest for the Universe* (EDP Sciences, Les Ulis, 2006)

Chapter 2
Standard Candles in Astronomy

Paolo A. Mazzali

Abstract One of the basic missions of Astronomy is to measure distances in the cosmos. This is usually done using the method of *standard candles*, which requires identifying astronomical objects or phenomena with a repeatable luminosity, and to measure that luminosity. Objects suitable as standard candles range from stars to supernovae, but also properties of the light of galaxies and the distribution of galaxies in clusters are useful standard candles. More luminous objects can be used to measure larger distances, looking back into the evolution of the Universe. We review here some of the history of determining astronomical distances, and discuss some of the most recent applications and results.

Keywords (Cosmology:) distance scale • Stars:variables: cepheids • (Stars:) supernovae: general

1 Introduction

Whenever we try to measure a distance we need a meter stick, a unit to which distances can be related. One of the main tasks of astronomy is to measure the distances to celestial objects, and ultimately the size of the Universe. This presents the problem that the sizes that confront us are extremely large. In fact, the Universe is about as large as the size reached by light that has been travelling since the beginning of the Universe itself, which occurred about 14 billion years ago. Astronomers have to be very inventive about measuring distance.

We know that the Sun is about 150 million kilometers away from the Earth. This is the distance light travels in about 8 min. However, the nearest star other

---

P.A. Mazzali (✉)
INAF–Osservatorio di Padova, Vicolo dell'Osservatorio, 5, 35122 Padova, Italy

Max-Planck-Institut für Astrophysik, 85748 Garching, Germany
e-mail: paolo.mazzali@oapd.inaf.it

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_2,

than the Sun, α Centauri, is already more than 4 light years away! The Sun and the Earth sit at the periphery of a large spiral galaxy, the Milky Way, which contains hundreds of billions of stars. Already determining distances within the Milky Way is a formidable task.

For stars that are sufficiently close that the Earth's motion around the Sun causes a change in the perspective from which they are seen projected against the more distant, apparently "fixed" stars, we can measure an angular motion relative to this apparently fixed background. This motion, called parallax, allows a very precise geometrical determination of the distance if we know our measuring unit, which in this case is the size of the Earth's orbit. However, even with the most advanced space-based instruments such as Gaia, which is to be launched in 2012, this can be done for distances of up to about 30,000 light years, which is still well within our Galaxy. Beyond this, there is essentially no way to measure distance directly.

2 Standard Candles

Therefore, we need to come up with indirect methods to determine distance. Astronomy looks at light from the Cosmos. Since we know that the speed of light is a universal constant, it is natural to try to use light posts as meter sticks: if we know the luminosity (i.e. the intrinsic energy output in the form of electromagnetic radiation) of a particular light source, we can measure its distance by simply comparing the "observed" brightness to the intrinsic luminosity.

Suppose we have a 100 W light bulb. We can measure the intensity of the light by measuring the flux crossing a unit surface (e.g. 1 cm^2), equivalent to the size of a CCD detector in a small digital camera, at a distance of 1 m. This may be used as our reference. If we then move further from the source of light, we will see it becomes dimmer. This dimming is a function of the square of the distance, since light travels in all directions, and it spreads over a spherical surface. So, for a given source luminosity L, the flux F observed at a distance d is given by $F = L/4\pi d^2$. Using the measured flux to determine a distance based on a known Luminosity is the concept of a Standard Candle.

However, nothing is easy in astronomy. Since we do not know the distance to astronomical sources, we can only measure the flux at Earth. Astronomical sources, e.g. stars, have different luminosities, as can be seen from their different colours and spectra. Only for the Sun and for those stars which are so close that the method of parallax can be applied can we measure both the flux and the distance independently, and hence derive the luminosity. For the Sun very precise measurements can be made, but the Sun is not a very luminous star,[1] and twins of the Sun cannot be observed to very large distances. In other galaxies, a Sun would hardly be distinguishable from the general stellar glare.

---

[1] Here we distinguish between the term "luminous" which refers to the effective light output of a star, and the terms "bright" vs. "faint", which are relative statements that refer to the observed flux and hence depend on the distance to the source.

There are, however, much more luminous stars than the Sun. An O-type star, with a temperature of more than 30,000 K (vs. 5,800 K for the Sun), is about one million times more luminous. This alone tells us that we may be able to see an O star about 1,000 times further away. If we can see with our naked eye a star like the Sun to a distance of a few light years, we should be able to see an O star to distances of a few thousand light years, which is still within the Milky Way.

Distances in astronomy are actually measured not in light years but rather in parsecs. This distance is based on a parallax, and it corresponds to the distance at which a source would appear to move against the sky background by $2''$ over a time of 6 months, as the Earth moves from one side of the Sun to the other while it moves along its orbit. This distance corresponds to $\sim 3 \times 10^{13}$ km, or 3.26 light years. Geometrically, it is the length of the side of the imaginary right-handed triangle which has the distance from the Earth to the Sun as the short side while the angle opposite to it is $1''$.

3 The Universe Close to Us

With our naked eye we can even see other galaxies. The two closest ones are known as the Large and Small Magellanic Clouds (LMC, SMC), and are visible in the southern sky. Magellan called them clouds because they appear like diffuse patches of light. They are small satellites of our own Galaxy, located at distances of $\sim$50 kpc (LMC) and $\sim$60 kpc (SMC). The LMC hosts about ten billion stars, while the SMC is smaller, containing only several 100 million, and is therefore less luminous and dimmer as seen from the Earth. We can see these galaxies because their total light output is several 100 million times larger than that of the Sun, or several 100 times brighter than an O star. Still, they are sufficiently far away that we cannot distinguish individual stars with our naked eye.

A nearby galaxy which can be seen from the northern hemisphere is Andromeda. This is a spiral galaxy similar to ours, although a bit bigger. It contains about 10^{12} stars, and it is located almost 800 kpc (i.e. more than 2.5 million light years) away. Although Andromeda is somewhat brighter than the Milky Way, it is so distant that it looks very faint, and it is just about the furthest celestial object visible with the naked eye, with the exception of some GRB optical counterparts (see below). And this is only the nearest large galaxy. As seen with our naked eye, the Universe is a rather small place.

Of course we do not simply rely on our eyes to explore the Universe. Our telescopes have reached sizes and capabilities that were simply unthinkable only 50 years ago. Using the most powerful telescopes presently available, such as the European Southern Observatory's (ESO) 8 m Very Large Telescopes (VLT), or the Hubble Space Telescope (HST), astronomers can probe the Universe at optical wavelengths to unprecedented depths. With these instruments it is possible to see objects with apparent magnitudes as large as $\sim$26. In the units in which the brightness of astronomical sources is measured,[2] this means being able to see things that are $\sim 10^{21}$ times fainter than our Sun in daytime, or about one billion times fainter than the brightest stars at night. This makes it possible to see very bright stars not only in galaxies that belong to the Local Group, like Andromeda, but also, using HST, in galaxies that belong to the neighbouring large cluster of galaxies, known as the Virgo cluster because it is located in the direction of the constellation Virgo. The Virgo cluster contains some 1,500 galaxies. The core of the Virgo cluster is located about 25 times as far away from us as Andromeda, and light from it takes more than 60 million years to reach us.

Determining the distance to these galaxies is important for a number of reasons. First, we want to know the size, and hence the density of the local Universe. Second, galaxies are bright and can be seen to larger distances, so that one may want to compare the apparent and intrinsic magnitudes of galaxies further away than Virgo to those in Virgo and compute relative distances. Third, galaxies host very bright events which can be used as lighthouses, supernovae. If we know the distance to nearby galaxies accurately through independent methods, and are therefore able to calibrate the luminosity of any Supernova in those galaxies, we can use supernovae to measure larger distances.

Galaxies beyond the Local Group move away from us, showing the first signs of the expansion of the Universe. If we can measure distances to these galaxies and correlate that with their motion away from us, we can map the expansion of the Universe. The expansion is revealed by the increasing redshift of the spectra of galaxies as we look at fainter and fainter galaxies, which are further away. This observation led Edwin Hubble [33] to hypothesize that the Universe is expanding. The Hubble constant (H_0) measures the rate of the expansion of at least the local Universe. It has the units of km s^{-1} Mpc^{-1} and it relates the expansion velocity to a distance, the so-called Hubble law.[3] Clearly, if the Universe is expanding, all things must have been much closer together in the past. This is how the theory of the Big Bang basically started.

---

[2] Magnitudes are logarithmic units of flux, and a larger magnitude implies a fainter object. This somewhat confusing system stems from the early days of astronomy, before the telescope was even invented, and reflects the way the human eye perceives light. The brightest stars were assigned to "first magnitude", somewhat fainter stars "second magnitude", etc. A difference of 5 magnitudes is equivalent to a difference of 100 times in flux. Objects visible with the naked eye have apparent magnitudes between ~ 5 (the faintest ones) and ~ -1 (a bright star like Sirius). The Sun in daytime has apparent magnitude -26.7, the full Moon at night -12.6.

[3] The unit of the Hubble "constant" is of course the inverse of time, but astronomers use km s^{-1} Mpc^{-1} so that when it is multiplied by the distance in Megaparsecs, the resulting speed of recession is in km/s (Hubble's law).

4 Stars as Standard Candles

As we have seen, for distances up to the Virgo cluster it is possible to use stars as distance indicators. Stars can act as standard candles if we know exactly how luminous they are. This method works as follows. Suppose that for a nearby, hot, bright star we can measure the distance via the parallax method. We need to identify some property of the star which can be observed and recognised at large distances. One such property is the spectral type: the spectrum of a star depends primarily on the star's temperature, which is the result of the luminosity and the size of the star through $T^4 \propto L/R_*^2$. Stars of different luminosity and size are therefore characterised by different spectra. In order to determine a star's spectral type, one obviously needs to obtain a good spectrum. Spectroscopy cannot be performed at the same distance as photometry, as in this case light is dispersed along wavelength and the flux at each wavelength is much smaller than the entire stellar flux, but it is possible to obtain spectra with a good signal-to-noise ratio of at least the brightest stars in galaxies as far as the Virgo Cluster. The stars that are best suited for this exercise are blue, A and B-type supergiants (SG), which have absolute magnitudes up to -9.5, and are therefore more than $\sim 10^5$ times as bright as the Sun.[4] Large, 8 m telescopes can perform spectroscopy for objects with apparent magnitude down to $\sim$23, so these stars would be observable in the Virgo Cluster (which has a distance modulus $\mu \approx 31$ mag[5]). Work in this direction has been spearheaded by R.-P. Kudritzki, among others, who showed that it is possible to determine distances fairly accurately in this way [40]: the stellar spectrum must be modelled so that parameters such as stellar temperature and gravity can be determined. These in turn relate to a star's luminosity.

5 Indirect Standard Candles

5.1 The Tip of the Red Giant Branch

Apart from massive stars, few other astrophysical objects can be regarded as useful standard candles. Therefore there has been a search for properties which indirectly indicate a known luminosity. Among these indirect methods is the luminosity of the Tip of the Red Giant branch in the Hertzsprung-Russell diagramme (HRD).[6] The

---

[4] The absolute magnitude of a celestial object is the magnitude it would have if it was observed from a distance of 1 pc.

[5] This is difference between absolute and apparent magnitude, and is related to the distance in parsecs as $\mu = m - M = -5 + 5 \log D(pc)$.

[6] The HRD displays stars according to their spectral type (i.e. temperature) and their luminosity. During the course of their evolution, stars move in the HRD following well-known "evolutionary tracks". Initially, stars sit on a line known as the "Main Sequence" After exhausting their core hydrogen, stars expand and become cooler, moving to the top right of the HRD, where "Red Giants" (RG) are located. The most massive stars become Red supergiants (RSG).

brightest end of the so-called RG branch varies in luminosity depending on the age of the stellar population which is observed, but its luminosity can be predicted based on the distribution of stars in the HRD.

5.2 The Planetary Nebulae Luminosity Function

Another method is the Planetary Nebulae Luminosity Function (PNLF). At the end of the RG phase, low-mass stars eject their outer envelopes. These are illuminated by the star beneath and shine like a nebula. Early observers could not resolve this structure and thought it was made of planets, hence its name. Planetary Nebulae (PN) have a standard distribution of luminosities which can be calibrated locally and used to measure distances. These measurements can be very effective, but they are affected by uncertainties and cannot be made at distances larger than the direct stellar methods.

6 "Standardizable" Candles

6.1 Cepheid Stars

More accurate methods require more complicated procedures. One very useful "indirect" standard candle is the relation between the pulsation period of a class of variable stars, the so-called Cepheids (named after the prototype, δ Cephei), and their luminosity. Cepheid variables are fairly massive ($\sim$10 solar masses) but relatively cool stars (spectral type F8-K2, with temperatures of 6–4,000 K). They are $\sim 10^4$ times as luminous as the Sun ($M \sim -5$), and although they are not the brightest stars they can be seen to reasonably large distances.

Cepheids pulsate because their surface properties oscillate between two different ionization stages. In the lower ionization stage, the star is cool and it has less opacity since helium is singly ionized. This tends to make the star contract, leading to an increase in surface temperature, and consequently in ionization. At the end of this contraction phase helium is doubly ionized. This leads to an increase of the opacity, and the following increased absorption of radiation causes the star to expand and cool again. The ionization degree now decreases, and the cycle completes. When the star has a smaller radius it is less luminous. This cycle repeats very precisely, and there is a tight relationship whereby brighter Cepheids have longer periods. If the period can be measured, the luminosity can be derived and compared to the observed flux to determine the distance. Measuring the variability of a light source is a relatively simple exercise, especially with instruments such as HST, which are not affected by atmospheric dispersion and can resolve stars in crowded fields. Therefore, it is possible to obtain photometry and study the light curves of Cepheids as far out as the Virgo Cluster.

Using the great resolving power of HST two independent groups, one lead by Wendy Freedman and the other by Alan Sandage and Gustav Tammann observed Cepheids in galaxies in the Virgo Cluster and derived their distances. Their results do not quite agree. Freedman's group obtained a larger value of the Hubble constant (72 km s^{-1} Mpc^{-1}, [25]) than the Sandage/Tammann group, who obtained a value closer to 62 km s^{-1} Mpc^{-1}[71]. Several uncertainties may have affected their results, including a dependence of the Period-Luminosity relation on the metal content of the stars.[7] Presently, the larger value of H_0 is favoured because it yields an age of the Universe more in line with the accepted "concordance Cosmology", as well as with measurements of gravitational lensing [70] and measurements of the Cosmic Microwave Background radiation (CMB).

6.2 Galaxy-Based Measurements

Despite the uncertainties, we have been able to measure distances in the local Universe with a precision better than 10%. To go further requires using brighter observables. Galaxies can be seen to large distances, however their precise luminosity is not known and it is difficult to make them a standard. Yet they have other properties which can be useful. Since they are made of individual stars, it is possible to predict that the light emitted from their surface will vary in intensity with a specific pattern, which reflects the number of stars in a given unit surface. This point-to-point fluctuation of a galaxy's surface brightness correlates with the total galaxy luminosity, and may therefore be used to measure distance. This method, called Surface Brightness Fluctuation (SBF) requires accurate imagery of the galaxies, and may be affected by uncertainties due to metallicity, age, morphology, dust properties. Tonry et al. [74] measured distances to galaxies located some 50 Mpc from us, i.e. $\sim$160 Mly. This is beyond the Virgo Cluster, in the so-called Hubble flow: at these distances the local motion of galaxies caused by gravitational interaction with their local neighbours becomes small compared to overall expansion velocity and can therefore be neglected. This is the best region to sample in order to measure the "present" value of H_0. The typical error for these measurements is 0.13 magnitudes, i.e. 13% in flux, 6.5% in distance.

Another relation involving galaxies as secondary standard candles is the Tully-Fisher (TF) Relation and variations thereof. The original TF relation [75] calibrates the luminosity of a galaxy based on the rotational velocity of its disc. Since the rotational velocity depends on the mass of the galaxy, and mass in turn correlates with luminosity, the relatively easily observable rotation curve (i.e. how rapidly parts

---

[7] In astronomy, "metals" include all elements heavier than helium.

of the galaxy located at different radial distances from the nucleus of the galaxy revolve around the center of the galaxy itself) can be transformed into a luminosity. When combined with the observed brightness, this then yields a distance. This relation applies to spiral galaxies. For elliptical galaxies an equivalent relation, the $\mathrm{D_n} - \sigma$ relation [18], compares the velocity dispersion within an elliptical galaxy and the galaxy's angular size. This method seems to give good relative distances, and it could be used to distances larger that those which can be probed with the TF relation for spiral galaxies, because elliptical galaxies are brighter. The problem is that it is difficult to calibrate these distances because elliptical galaxies do not host cepheid stars.

7 Supernovae as Distance Indicators

7.1 Supernovae as Standard Candles

The description above suggests that in order to measure distances well into the past, at redshifts of $z = v/c \sim 1$ (which corresponds to looking back at the Universe when it was about half its present age), a standard or standardizable candle is required that is bright and can be measured just with photometry.

One of the brightest phenomena in the Universe is a supernova (SN), the explosion that marks the end of the life cycle of different types of stars. The name derives from the Latin "novus", i.e. new. "Stella Nova" was the term used by astronomers in the Renaissance to define celestial objects that would suddenly appear in the sky, "new stars". It was actually first used by Tycho Brahe in his book "Stella Nova" as an attribute to what we now know was a Supernova, SN 1572, from the year in which it was observed. Tycho's SN is still visible as a remnant (SNR) in the constellation Cassiopea. Chinese and Japanese astronomers had previously observed SNe, which they called "guest Stars" [14].

Later, however, the term Nova was used by astronomers for phenomena which are more common than SNe but intrinsically different from a physical point of view, representing the sudden increase in luminosity of a white dwarf accreting mass from a companion. A white dwarf is a compact, hot object, which is formed at the end of the life of a star of less than about eight solar masses. White dwarfs slowly cool down, turning into dark and dead bodies. Interaction with a binary companion can change this uneventful destiny. The accreted material is unstable to nuclear burning, and when it ignites it can be ejected, with the attendant display of optical light. Only in the early 1900's it was realised that some of these bursts are actually much more luminous. The term supernova was introduced by Fred Zwicky in 1926. In 1934, Baade and Zwicky [4, 5] recognised the difference between "Novae" and "Supernovae". The latter stood out by being much brighter, so that they could be seen easily in external galaxies. In fact, SNe can be about as bright as the entire galaxy in which they occur. Baade and Zwicky based these suggestions on very few events, including the historical Tycho SN, and a SN which occurred in Andromeda in 1885 (SN 1885A: the modern naming system adds sequential letters to the year of discovery: when the alphabet runs out double letters are used: aa, ab, etc.).

Based on their very meagre database, Baade and Zwicky already suggested that SNe have a characteristic luminosity, which is comparable to that of a galaxy. In 1938 Baade [3] published a mean value of the absolute luminosity of SNe, and suggested that they could be used as standard candles. (In practice the estimate was incorrect, because he did not know the exact distance to Andromeda, which he thought to be much shorter than it really is.)

But those were early days. Only shortly thereafter Minkowski [51] recognised that there are at least two groups of SNe, based on their spectra. Some (e.g.SNe 1940B and 1941A) show hydrogen lines, while others do not (e.g.SNe 1937C, 1937D). He called the latter Type I, and the former Type II.

This was the beginning of a journey of discovery. The classification of SNe is now rather complex (see [19]), but we know that fundamentally SNe come in two flavours. One is the core collapse of massive stars, which leaves behind a very compact object, only a few kilometres in radius. It can be made of neutrons (a neutron star) as was originally suggested by Baade and Zwicky in 1934 [4] or be a black hole, depending on how much matter falls back on the compact object initially formed [26]. This group of SNe includes H-rich SNe II, which are produced by stars that collapse when they still have their H-envelope, but also certain subtypes of SNe I (SNe Ib, Ic) and SNe IIb (see [19]). These various subtypes are characterised respectively by the presence of little H but a lot of He (SNe IIb), no H but He (SNe Ib), and no H or He but significant Oxygen (SNe Ic). Since H and He are found in massive stars, and a predominance of O is typical of the cores of massive stars, all these SN types are thought to arise from the collapse of a massive stellar core. These SNe have a wide range of properties (luminosity, mass, energy), and cannot as such be used as standard candles.

This leaves us with the other flavour of SNe, Type Ia. These SNe show no H, no He, no O, but have strong Silicon and Sulphur lines, which is not compatible with massive stellar cores. They also are characterised by a large luminosity, and are on average much more luminous than core collapse SNe. In the 1960s and 1970s the subtypes of SNe I had not yet been discovered, but a suggestion was made that SNe I could be used as standard candles [39]. In 1973 Whelan and Iben [78] first proposed a physical mechanism for SNe I: the explosion of a degenerate carbon-oxygen (CO) white dwarf accreting material from a companion. This scenario is still the favourite one, despite uncertainties. It is in principle similar to the Nova scenario, but the accretion rate must be higher, so that a shell of accreting material (composed mostly of H) can stably burn and grow on the surface of a CO WD which has mass close to the Chandrasekhar limit – the largest possible mass an electron degenerate star can have, $\approx$1.38 solar masses. At that point, the temperature in the innermost regions of the WD exceeds 10^9 K, and thermonuclear reactions start which burn carbon and oxygen into heavier elements, producing energy and unbinding the star.

7.2 SNe Ia and Dark Energy

Having established the principle that SNe Ia could be used as distance indicators, there remains the problem of discovering them at large distances. This was not doable with the telescopes and detector technology of the 1960s and 1970s. The first pioneering attempt was performed by a Danish-led team, which managed to discover a SN at redshift $z \approx 0.3$ using the 1.5 m Danish Telescope at the ESO observatory on La Silla, Chile [53]. Obviously, accurate measurements were not possible, but this remains an outstanding achievement for the time. In the 1990s, with the advent of new, very powerful 8 m-class telescopes (VLT, Keck, Gemini, Subaru), of adaptive optics and of improved CCD detectors, the prospects for detecting SNe Ia at high redshift became much brighter.

Saul Perlmutter of UC Berkeley advocated the feasibility of the project, expecting to measure the deceleration of the Universe's expansion as it was slowed down by gravity. This was the prediction for a Universe in which the density of matter was just sufficient to stop ultimately its expansion, a favourite model for Big Bang cosmology at the time. In cosmology's terms, this means that the total density of the Universe $\Omega_{tot} = 1$.[8] Since visible matter clearly is insufficient to "close" the Universe, providing only $\sim$4.5% of the closure density, other forms of matter were thought to dominate. This would be in the form of "dark matter", non-visible matter which is detected for example in the halos of galaxies through its influence on the orbital properties of visible objects. Indeed [57] discovered a SN (SN 1992bi) at $z = 0.458$, and argued that SNe at high redshift could be used to measure cosmological distances. The technique they proposed was to use SNe Ia as actual standard candles, assuming a "typical" peak absolute magnitude (-18.86 magnitudes). This approach was soon to be revolutionised, however, adding confidence to our use of SNe Ia as distance indicators.

One of the difficulties involved with calibrating the absolute magnitude at peak of SNe Ia so that they could be used as standard candles is the dispersion caused by observing SNe in nearby galaxies for which the distance and reddening are uncertain [12]. Available data, covering a few dozen SNe discovered mostly serendipitously, seemed to indicate that indeed SNe Ia were all alike, and assumptions were made about their intrinsic colour, which was used to correct for any difference, attributing it to reddening.

However, in 1991 this state of affairs had begun to change. During that year two SNe were discovered that did not conform with the norm. One, SN 1991T [20, 61] was very luminous ($\sim$0.5 mag more than the norm), while the other, SN 1991bg [21, 42] was very dim, almost 2 mag less luminous than the "typical" SN Ia luminosity. Not only that, but both SNe also exhibited spectra that were distinctly different from those of typical SNe Ia, although both the light curve properties and

---

[8] The parameter Ω measures the density ρ in the Universe in units of the critical density ρ_c: $\Omega \equiv \rho/\rho_c = 8\pi G\rho/3H^2$.

the elements seen in the spectra indicated that these events were indeed SNe Ia. The spectral differences could be explained with the presence of more highly ionised species for SN 1991T and less highly ionised ones for SN 1991bg, in agreement with their different luminosities [45, 46]. This, which seemed to deal a blow to the possible use of SNe Ia as standard candles, was instead the beginning of the era of Supernova Cosmology, as well as a major step in our understanding of these explosions.

In a visionary and pioneering work, Mark Phillips [60] redefined the concept of SNe Ia as standard candles. He suggested that SNe Ia are not standard candles per se, but come with a range of luminosities. While it would be very difficult to measure directly the exact luminosity, this is fortunately correlated with a purely observational quantity: the shape of the light curve, i.e. the way in which a SN Ia becomes bright and subsequently fades. In particular, brighter SNe have broader light curves. In the terms he used, the luminosity decline in the 15 days following maximum light is smaller ($\sim$1 magnitude) for the most luminous SNe like SN 1991T and larger ($\sim$2 magnitudes) for the least luminous ones like SN 1991bg. He called this quantity Δm_{15}, and showed that the correlation holds in different wavebands but it is steepest, and hence easiest to measure and distinguish, in the B band, which is where most of the SN light is emitted [$M_B = -21.73 + 2.70\ \Delta m_{15}(B)$, with an uncertainty of about 0.3 mag]. He used only 9 SNe, all located in nearby galaxies for which the distance had been estimated using the TF or the SBF methods. All of these issues (small sample, uncertain distances and reddenings) affected his error estimate, but nevertheless the road was paved for the use of SNe Ia as distance indicators, not as direct standard candles, but rather as “Standardizable Candles”. The luminosity of a SN is in fact not determined by a direct measurement of its peak magnitude, but rather by following the evolution of its light curve. The shape of the light curve indicates the SN luminosity and this in turn, when compared to the observed magnitude taking into account any reddening, for example via its influence on the colour, gives a distance. A distance thus derived is known as a “Luminosity Distance”. Finally, a measurement of the redshift of the SN or its host galaxy is necessary to derive the relation between distance and recession velocity, the Hubble Law.

Inspired by this work, more and more accurate measurements of SN light curves were taken. Adam Riess, William Press and Robert Kirshner (Harvard) developed an automated package that solves for the shape of the SN light curve and gives a more accurate estimate of the SN peak magnitude than just measuring a magnitude difference over 15 days [64]. Saul Perlmutter, on the other hand, defined the concept of “stretch”: regardless of their actual breadth, all SNe Ia light curves can be stretched in time and warped back to a “typical” light curve. A SN more luminous than the average has a broader light curve than the “typical” SN Ia light curve, and this is parametrised by a stretch factor greater than 1. The opposite holds for underluminous SNe.

It was with the help of these tools that two teams set out to discover SNe Ia at very large distances in order to map the expansion of the Universe. Perlmutter’s Supernova Cosmology Project and the High-z Supernova Team led by Brian Schmidt at Mt. Stromlo Observatory, Canberra, both searched for SNe Ia in distant fields using 4–8 m telescopes. Light curves were obtained for the SN that were discovered, so that the luminosity of these SNe could be determined using the relation with light curve shape that holds for less distant SNe. A first tentative measurement of the cosmological parameters was presented by Perlmutter et al. [58]. Soon thereafter the two groups published two papers which agreed on the totally unexpected result that the Universe is not decelerating under its own gravity but rather is accelerating in its expansion, as if under the influence of some unknown force [59, 65]. If the Universe were decelerating, distant SNe would be closer than what their redshift would suggest when projected on to a Hubble law of constant expansion velocity. The light from the SN was emitted billions of years before it was detected, when the relative velocity was larger, but because of deceleration the distance that light would need to travel to get to us would actually be less than what one would estimate at the time of emission, when the redshift was imparted. The SN would thus be seen at "too bright" for its redshift. Instead, the opposite was found. Since distant SNe look fainter than expected, their light had to travel a larger distance than what the redshift indicates.

These results caused a paradigm shift in astrophysics and cosmology. In our present picture of the Universe some unknown form of energy, associated with the vacuum, is the prevailing force. This "Dark Energy" plays the role of Einstein's cosmological constant, Λ. The original Λ was introduced in order to stop the Universe from collapsing under its own gravity, which was a natural consequence of the theory of General Relativity. At the time (1916) Einstein did not know about the motion of galaxies, which Hubble only announced in 1929. Actually, the discovery of the expansion of the Universe made Einstein retract his idea. The form of Λ which seems to be active, on the other hand, accelerates the expansion, at least in the latter part of the Universe's life, i.e. from $z \sim 0.5$ onwards. Studies of the exact nature of Dark Energy are ongoing (see Chap. 18).

7.3 From SN Ia Cosmology to Concordance Cosmology

Meanwhile other studies, such as the analysis of the power-spectrum of the anisotropies of the CMB, the 2-degree Field redshift survey (2dF), and the properties of Baryonic Acoustic Oscillations (BAO) have confirmed that the total density of the Universe Ω_{tot} has a value consistent with 1, and that only a small part of this is due to baryons or Dark Matter. 2dF was a spectroscopic survey of more than 200,000 galaxies to distances of up to 400 Mpc over an area of ∼1,500 square degrees of the sky. This allowed an accurate measurement of the way galaxies cluster, which is related to the total mass density of the Universe. The power spectrum of the spacial distribution of galaxies yields an estimate of the matter density of the Universe $\Omega_m \approx 0.28 \pm 0.04$ [56].

Measures of the power spectrum of the CMB, which was first mapped by the Cobe satellite [8], showed that the Universe has the anisotropies predicted by Big Bang inflationary theory. Later, more accurate measurements obtained with

WMAP [68] confirmed that the Universe is “flat”, i.e. $\Omega_{tot} \simeq 1$ (first peak of power spectrum), that $\Omega_m \approx 0.26$ (height of the peaks), and that $\Omega_b \approx 0.04$ (position of second peak).

These experiments in themselves measure the matter content of the Universe. Type Ia SNe can also measure Ω_{tot} and the evolution of its components over time, although not very accurately: they rather measure the quantity $\Omega_m/2 - \Omega_\Lambda (= q_0)$. The combination of these results supports the present “Concordance cosmology”, where $\mathrm{H}_0 = 72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{tot} \approx 1.00$, $\Omega_m \approx 0.27$, of which $\Omega_b \approx 0.04$, and $\Omega_\Lambda \approx 0.73$ [38].

7.4 Reliability of SNe Ia as Standardizable Candles

So, SNe Ia have played a major role as standard candles, and will continue to do so with the new proposed satellites which should collect large datasets (WFIRST, Euclid), and the blind surveys which will discover SNe Ia from the ground (Pan-Starrs, the Palomar Transient Factory, Skymapper, LSST in the future). Given the revolutionary aspect of the result, the question arises how good are SNe Ia as standardizable candles. Obviously, a large amount of work is being done in order to understand how SNe Ia explode, why they have a large range of properties which are however tightly related (e.g.luminosity and light curve). Even the nature of the progenitors of SNe Ia is an open question.

A lot of observational effort has gone into perfecting the art of standardizing SNe Ia using increasingly large samples and more and more sophisticated analysis techniques. The accumulation of data helps reducing statistical errors on the estimate of the cosmological parameters, but it cannot eliminate systematic uncertainties on the validity of the standard candle. Among these uncertainties are the uniqueness of the luminosity – light-curve shape relation, the possible dependence of SN properties on environment, the evolution of SN Ia properties with age and/or redshift, contamination by other SN types, intervening absorption.

The first question, uniqueness, addresses the problem of dispersion of SN Ia properties: namely, is it possible that SNe Ia that have the same intrinsic peak luminosity have different light curve shape, or, alternatively, can SNe Ia that have the same light-curve shape have really different luminosities, and if so, by how much? This question ultimately addresses the physics of SNe Ia, as well as the nature of their progenitors. These areas are the subject of much work. We could go back to the very beginning and ask the fundamental question, why do SNe Ia behave the way they do.

Naively, a bright SN with a broad light curve could be thought to have released simply more mass. We now understand, after an initial suggestion in [15], that the source of light for SNe (not just SNe Ia) is the radioactive decay of ^{56}Ni. This isotope is produced by a series of α-captures, i.e. the successive addition of He nuclei, starting from lighter elements such as C and O (α-chain). ^{56}Ni has an equal number of protons and neutrons, and is the last isotope along the α-chain whose production actually yields energy. ^{56}Ni is one of the most tightly bound nuclei. Building on ^{56}Ni to produce something heavier actually requires energy, as does breaking ^{56}Ni apart into lighter isotopes. Therefore, when ^{56}Ni has been produced, as much nuclear energy as possible has been extracted. Burning $\sim$1 solar mass of C + O to ^{56}Ni produces more than 10^{51} ergs of energy, which is easily more than what is required to unbind a White Dwarf and eject its material at high velocity. ^{56}Ni is unstable. It decays via electron capture to ^{56}Co with a half-life of about 6 days. ^{56}Co is also unstable, and it decays to stable ^{56}Fe with a half-life of about 77 days. Both decays produce γ-rays and positrons, which deposit their energy in the SN ejecta. This energy is ultimately converted into the optical radiation that makes SNe bright. ^{56}Fe is actually the most stable nucleus: it has the smallest mass per nucleon, with a total mass of 55.94 amu.

The uniformity of the behaviour of SNe Ia was initially taken as an indication that the progenitors are all similar. A degenerate C + O White Dwarf (WD) is the best candidate: this type of star, the leftover of stars of masses of 3–5 solar masses, will explode if it happens to approach the Chandrasekhar limit. This can happen if a WD, which typically has a mass of less than $1M_{\odot}$, accretes hydrogen from a companion in a binary system. This scenario guarantees repeatability, is not limited to very young stellar populations like core-collapse SNe, and is one that can be physically modelled. Progress has in fact been made in this direction [32], although details are not yet clear.

Why then do SNe Ia show a range of luminosities? This probably depends on how the explosion takes place. A CO WD ignites when it reaches a mass close to the Chandrasekhar limit. This is actually a coincidence: thermonuclear reactions start because the temperature in the WD reaches very high values ($\sim 10^9$ K), and this happens at masses just below the highest limit above which the WD would simply collapse under its own gravity. Evidently, after ignition starts, something must happen that is not the same in all WDs, otherwise the outcome would always be the same. The reason for diversity probably lies in the way burning occurs. If a Chandrasekhar-mass CO WD ignites and burns explosively (supersonically, i.e. through a detonation), it would turn entirely into ^{56}Ni. This would actually be a perfect standard candle, but we know this is not what happens in reality because we see the presence of elements other than ^{56}Ni and its decay products, ^{56}Co and ^{56}Fe, in the spectra of SNe Ia. Alternatively, if burning proceeds subsonically, at least initially, the WD would begin to expand as nuclear reactions generate energy. This leads to a decrease of the density. Burning at low density does not go all the way to ^{56}Ni, but stops somewhere along the way, resulting in an increased production of "Intermediate Mass Elements" such as Silicon and Sulphur, which are indeed observed in the spectra of SNe Ia [13]. Models where burning is only subsonic, however, do not seem to produce much ^{56}Ni. A way to produce SNe Ia with different ^{56}Ni content is a hybrid mechanism called a "delayed detonation" (DD, [36]). In this scenario burning starts subsonically at high density, incinerating the inner parts of the WD to Nuclear Statistical Equilibrium (NSE) and producing some ^{56}Ni and other Fe-group isotopes while the WD begins to expand. However, at some point a transition occurs to supersonic burning, which consumes the rest of the star within a second or so. Since the detonation occurs at lower densities because the WD has somewhat expanded, the nucleosynthetic products can be a mixture of ^{56}Ni and IME, with a general stratification where the heavier elements, which are produced at higher densities, lie deeper. This configuration can give rise to the variety of observed SNe. Burning to IME only generates almost the same kinetic energy as burning to NSE, so the star would still explode. How can the explosion then be tuned to reproduce the observed range of SN Ia luminosities, corresponding to ^{56}Ni masses of between $\sim$0.1 and $\sim$1 $M_{\odot}$ [16]? One possibility is that while all explosions start as deflagrations, the transition to a detonation occurs at different times, leading to different outcomes: an early transition results in more burning at high density, resulting in production of more ^{56}Ni and a brighter SN, while a later transition would mean more WD expansion and smaller ^{56}Ni production. A similar outcome would be obtained if the deflagration could have different strengths: a stronger deflagration again expands the WD more before the detonation sets in, and vice versa. Since the physical reason for this is actually not yet understood, this remains only a possibility. Analysis of SN spectra reveals that the combined amount of material burned to NSE and IME is actually the same in all SNe Ia, lending credence to delayed detonation as the underlying explosion mechanism of the bulk of SNe Ia [49].

Would these events be precise copies of one another so that a single parameter, the mass of ^{56}Ni, can describe them all? Most likely not, although they would be similar. The details of the explosion are complex, and must depend on the exact properties of the progenitor, which are not known in detail. For example, a change in the metal content of the WD should lead to somewhat different nucleosynthesis, in particular to different ratios of stable Fe-group isotopes and ^{56}Ni [73]. This may affect the luminosity-light curve shape relation [35, 48]. Furthermore, the explosion is an intrinsically 3D phenomenon, and SNe Ia may look different from different viewing points. This may cause a dispersion of $\sim$5 % [35]. If there is any evolution in the properties of the progenitors, which may be more metal rich at the larger redshifts where elliptical galaxies formed, and should be on average younger at higher redshift, SNe Ia may not obey exactly the same luminosity-light curve shape relation. If this was not taken into account, the estimated distance may be incorrect. Extensive work seems to rule out major observable differences between nearby and distant SNe Ia [10, 23], but the high-redshift data are obviously not as good as local data, so there may be room for uncertainty as differences may be subtle. Differences in fact seem to exist between SNe originating in different types of galaxies [69]. If these changes can be quantified and are predictable they may be corrected for, as long as the host galaxy type is known.

The large range of properties of SNe Ia may ultimately be difficult to explain within a single scenario. A recently revived alternative possibility is that some fraction of the SNe Ia come from the merging of two white dwarfs whose combined mass exceeds the Chandrasekhar mass [55]. This is called the double-degenerate scenario. Also, within the single-degenerate scenario, the donor star may not yield hydrogen to the white dwarf companion, but rather helium. In this case the white dwarf may explode before reaching the Chandrasekhar mass [22, 43].

These alternative channels are potentially quite interesting, and require further investigation.

Other correlations between SN Ia luminosity and various observable properties have been found, confirming that SNe Ia are indeed good standardizable candles. Among these are the $B-V$ colour of SNe Ia 12 days after maximum [77], or the width of the nebular emission lines of Fe [47], which is a direct consequence of ^{56}Ni being produced in different amounts. It is however a fact that SNe Ia in spiral and elliptical galaxies have on average different luminosities, the former showing a larger spread of properties, extending to the most luminous sub-types, which are not found in the latter group, which on the other hand contains basically all the dimmest events [30]. If all these effects could be corrected for empirically, and better still understood, the prospects would be very bright for SNe Ia as standard candles.

Other issues such as contamination by other SN types may be relevant if for example Type Ic SNe (core-collapse events without the outer H and He layers of the star, see [19] for a review of SN types) could rival in luminosity with SNe Ia, but we know that this is a rare occurrence. Besides, SNe Ic can usually be sorted out on the basis of their colour [62]. Dust may also be a cause of concern, as it would dim the SN flux. However, normal dust would also redden the flux, and this can be rather easily spotted. Only completely gray dust may fool us [1]. Such dust may be located in the otherwise empty space between galaxies, but there is evidence that its amount is small enough that it should not cause worries ([41] and references therein).

8 Further Afield

The main limitation in pushing SN Ia observations to very large redshifts may come from the fact that it will always be very difficult to see SNe Ia at redshifts much beyond $z = 2$. On the one hand, their faintness makes it hard to discover SNe Ia at those distances, while on the other, as one looks back in time to the earliest part of the life of the Universe the number of SNe Ia is expected to decrease. Therefore some standard candle which is active in the young Universe and is very bright would ultimately be desirable. Presently, there are two candidates: Baryonic Acoustic Oscillations (BAO) and Gamma-ray Bursts (GRB).

8.1 Baryonic Acoustic Oscillations

BAOs provide not a standard candle but rather a “standard ruler”. The characteristic length scale of the distribution of matter in the Universe was established at the time when H in the Universe first recombined and consequently matter and radiation decoupled, at an age of $\approx$ half a million years (redshift $\sim$1,100). This scale has been expanding with the Universe itself, and it can be measured at different redshifts, providing a “standard ruler” which can be used to map the expansion history of the Universe in detail, as soon as technology allows. So far, the Sloan Digital Sky Survey (SDSS) has sampled the Universe out to $z \sim 0.35$ [17], and a team in Australia (WiggleZ) is making measurements at $z \sim 0.7$ [9]. The experiment consists of measuring the probability that a galaxy is found within a certain distance of another one. The expectations are for a large correlation for galaxies at small separation, as matter tends to cluster because of gravity, and a low correlation at large separation. The BAO signal is a bump in the correlation function, at a comoving separation equal to the sound horizon. Today, the sound horizon is $\sim$150 Mpc [17].

8.2 Gamma-Ray Bursts

The idea of using GRBs as distance indicators has been around since the optical afterglow of the first GRB was discovered, a result obtained by the Dutch-Italian satellite BeppoSax [76], and it was clear that GRBs are extra-galactic sources [50]. The optical counterparts of GRBs are very bright (reaching apparent magnitude $V = 5$, [63]), and are detectable out to the highest redshifts [66, 72].

GRBs are very brief but very powerful flashes of radiation (see Chaps. 1 and 3), lasting up to a few minutes, mostly in the γ-rays and X-rays. They are divided observationally into two groups, “long-soft” and “short-hard”, based on their duration (longer or shorter than about 2 s) and their spectral properties (long GRBs tend to have softer spectra). Long GRBs are associated with the collapse of massive stars, also a BeppoSax result [27, 34, 44], while short GRBs are probably caused by the merging of two compact objects, e.g.two neutron stars [54]. The optical afterglow of a GRB can be as bright as mag -22 for a few hours, and therefore GRBs can be seen out to redshifts of $\sim$8, where they reach an apparent magnitude of $\sim$5, [63]. But can GRBs be used as standard candles? Certainly not directly, as their intrinsic luminosity is not constant. However, they may possibly be used as standardizable candles, like SNe Ia. A relation has been found between the peak of the spectral distribution of a GRB (E_{pk}) and the isotropic emitted energy E_{iso}, which assumes that GRBs radiate isotropically [2]. However, the energy content of GRB would be unrealistically large if they radiated isotropically, while this is not a concern if they are beamed events. E_{pk} can be measured from a spectrum of the GRB’s γ- and X-ray emission, while deriving E_{iso} obviously requires assumptions regarding the distance. Further studies have shown that if the beaming is corrected for, and E_{iso} is transformed into E_{tot}, the real emitted energy, an even tighter correlation with E_{pk} is obtained [29]. This suggests that long GRBs may indeed be standard energy reservoirs [24]. In this case, long GRBs may be used as standard candles as long as an independent measurement of the redshift is available (e.g.from the emission lines of the host galaxy). These results are heavily disputed, and they may indeed be the artificial outcome of the limited range of energies detected by γ- and X-ray satellites [52]. Also, at low redshift several examples have been found of weak GRBs, which do not obey the correlations. While this may not be a source of uncertainties at high redshift, where these events are too faint to be seen, great care should be exercised when using GRBs for cosmology, because they cannot be calibrated at low redshift. Our understanding of GRBs also needs to be greatly improved before we can use them with confidence as distance indicators.

9 Other SN Standard Candles

We have discussed SNe Ia as standard candles, but neglected all classes of core-collapse SNe. These are in fact so varied in their behaviour, as are the masses and the state of the stars from which they originate, that it is essentially impossible to use them as standard candles. There may be some exceptions, however.

9.1 SNe IIP

The first is the behaviour of SNe IIP in the plateau phase. In this phase, which lasts $\sim$100 days, the light curve of a SN IIP is dominated by hydrogen recombination and is rather constant in luminosity. The recombination wave moves inwards in the SN ejecta, which are in turn expanding. It is therefore possible to apply a variation of the Baade-Wesselink (BW) method to determine the change in the angular radius of the emitting photosphere, and hence the distance to the SN. The original BW method of estimating distance relies on stellar pulsations. A pulsating star changes radius and luminosity with a regular cycle (e.g.cepheids, see above). This leads to a change in the temperature of the star, which can in turn be determined from observations of its photosphere through spectra: if the photosphere is expanding/contracting, this is shown by a Doppler shift (blue/red) of absorption lines, which are formed at or above the photosphere. Luminosity, radius and temperature are linked through the relation $L = 4\pi R^2 \sigma_B T^4$, where σ_B is the Stefan-Boltzmann constant. If we can determine T and L independently from spectroscopy, we can estimate R. In addition, if a velocity measurement through the Doppler shift of the lines tells us how rapidly the photosphere expands/contracts, we can measure both the ratio of the minimum and maximum radii and their difference ($\Delta R = vt$, where t is the pulsation half-period). Hence we can solve for the actual radii. The luminosity and the variation in angular size can also be obtained from atmospheric models.

In the plateau phase of SNe IIP, the inward motion of the photosphere can be measured from the change in blueshift of the absorption lines over time (the blueshift decreases as the SN expands). The method of Expanding Photospheres (EPM) uses this to measure distances to SNe IIP. If we know the time of explosion and the instantaneous position of the photosphere, we can obtain the radius as $R = vt$, where t is the time elapsed since the explosion. With an independent determination of the temperature T, the luminosity L can be determined with the formula above. We can then relate L to the observed flux F through $F = L/4\pi D^2$, where D is the distance. However, an accurate determination of the temperature is not easy. Therefore, it is easier to obtain the distance following the expansion of the photosphere: two measurements of v at two different times yield both the ratio of the radii at the two times and their difference (through geometry and $\Delta R = v(t_2 - t_1)$). We can then plot the evolution of the radius and extrapolate it back to t_0, the time of explosion [37]. This method can yield accurate distances out to what is allowed by the relatively low luminosity of the plateau phase of SNe IIP, i.e. out to $z \sim 0.05$ [67].

A potentially even more accurate method is actually to determine the properties of SNe IIP through spectral modelling, thus avoiding the use of photometry and of assumptions about the properties of the radiation field. Presently, this method (SEAM: Spectral-fitting Expanding Atmospheres Method, [7]) and EPM do not yield consistent results, and this requires further investigation.

9.2 Pair Instability SNe

Another possible type of SN which may be used as a standardizable candle is the Pair Instability SN (PISN). This is the explosion of an extremely massive star (initial masses $>140 M_\odot$). The core of these stars can become so hot that electron-positron pairs are produced through the annihilation of energetic γ-rays. The loss of γ-rays can significantly reduce the outwards pressure of radiation which normally balances these massive stars against their own gravitational pull and keeps them from collapsing. This starts contraction of the star, with a consequent further heating of the core. If sufficiently high temperatures are reached in the core, thermonuclear reactions can start like in SNe Ia, burning oxygen explosively. The star may go through a phase of pulsations, where pair-production induces contraction which is then overcome by the energy produced by thermonuclear reactions, but eventually a final explosion disrupts the star, leaving no remnant, again like in a SN Ia [6, 11, 79]. The potentially useful feature of these SNe is that their explosion energy, and therefore also the nucleosynthesis, depend essentially only on the mass of the star. Additionally, and most usefully, the explosion should produce very large amounts of ^{56}Ni ($\sim$3–7 $M_\odot$, [31]), and therefore the ensuing SNe should be much brighter than SNe Ia, reaching peak magnitudes ~ -22 to -23, and should be observable to much larger distances. Another useful aspect of PISN is that very massive stars are much more likely to form and to survive as massive objects in the metal-poor environments of the early Universe when stars first formed [31]. Theory predicts that more massive stars produce more energetic explosions and brighter SNe [31]. The first probable detection of a PISN is very recent [28]. Like SNe Ia, more massive PISNe are expected to have broader light curves. This field is still in its infancy, but if a unique relation between luminosity and light curve shape could be established for PISNe, as it was for SNe Ia, then PISNe could be very useful distance indicators, and could be used to sample the epoch of formation of the first stars, at redshifts of $\sim$10, when the Universe was only about half a billion years old.

10 Conclusions

Astronomy has gone a long way from gazing at the motion of objects in the sky to mapping the cosmos in which we live. A combination of physical rigour and astronomical ingenuity has given birth to astrophysics, the modern version of astronomy. Our understanding of the processes that govern the birth, life and death of stars and of structures like galaxies has been instrumental in mapping the Universe. The accuracy of this mapping is undoubtedly going to increase in the forthcoming future, taking advantage of new space and ground-based instruments. As always in science, new discoveries set the stage for new questions, so surely we will see more exciting developments in our efforts to understand the Universe.

Acknowledgements The author is deeply grateful to Brian Schmidt and Elena Pian for critically reading earlier versions of this manuscript.

References

1. A.N. Aguirre, Astrophys. J. **512**, L19 (1999)
2. L. Amati et al., Astron. Astrophys. **390**, 81 (2002)
3. W. Baade, Astrophys. J. **88**, 285 (1938)
4. W. Baade, F. Zwicky, Phys. Rev. **46**, 76 (1934)
5. W. Baade, F. Zwicky, Proc. Natl. Acad. Sci. **20**, 259 (1934)
6. Z. Barkat, G. Rakavy, N. Sack, Phys. Rev. Lett. **18**, 379 (1967)
7. E. Baron, P.E. Nugent, D. Branch, P.H. Hauschildt, Astrophys. J. **616**, L91 (2004)
8. C.L. Bennett et al., Astrophys. J. **464**, L1 (1996)
9. C. Blake et al., Mon. Not. R. Astron. Soc. **406**, 803 (2010)
10. S. Blondin et al., Astron. J. **131**, 1648 (2006)
11. J.R. Bond, W.D. Arnett, B.J. Carr, Astrophys. J. **280**, 825 (1984)
12. D. Branch, D.L. Miller, Astrophys. J. **405**, L5 (1993)
13. D. Branch, J.B. Doggett, K. Nomoto, F.-K. Thielemann, Astrophys. J. **294**, 619 (1985)
14. D.H. Clark, F.R. Stephenson, Historical Supernovae (Pergamon Press, Oxford/New York, 1977)
15. S.A. Colgate, C. McKee, Astrophys. J. **157**, 623 (1969)
16. G. Contardo, B. Leibundgut, W.D. Vacca, Astron. Astrophys. **359**, 876 (2000)
17. D.J. Eisenstein et al., Astrophys. J. **633**, 560 (2005)
18. S.M. Faber, R.E. Jackson, Astrophys. J. **204**, 668 (1976)
19. A.V. Filippenko, Annu. Rev. Astron. Astrophys. **35**, 309 (1997)
20. A.V. Filippenko et al., Astrophys. J. **384**, L15 (1992)
21. A.V. Filippenko et al., Astron. J. **104**, 1543 (1992)
22. M. Fink, F.K. Röpke, W. Hillebrandt, I.R. Seitenzahl, S.A. Sim, M. Kromer, Annu. Rev. Astron. Astrophys. **514**, A53 (2010)
23. R.J. Foley et al., Astrophys. J. **684**, 68 (2008)
24. D.A. Frail et al., Astrophys. J. **562**, L55 (2001)
25. W.L. Freedman et al., Astrophys. J. **553**, 47 (2001)
26. C.L. Fryer, Astrophys. J. **522**, 413 (1999)
27. T.J. Galama et al., Nature **395**, 670 (1998)
28. A. Gal-Yam, et al., Nature **462**, 624 (2009)

29. G. Ghirlanda, G. Ghisellini, D. Lazzati, Astrophys. J. **616**, 331 (2004)
30. M. Hamuy, S.C. Trager, P.A. Pinto, M.M. Phillips, R.A. Schommer, V. Ivanov, N.B. Suntzeff, Astron. J. **120**, 1479 (2000)
31. A. Heger, S.E. Woosley, Astrophys. J. **567**, 532 (2002)
32. W. Hillebrandt, J.C. Niemeyer, Annu. Rev. Astron. Astrophys. **38**, 191 (2000)
33. E. Hubble, Proc. Natl. Acad. Sci. **15**, 168 (1929)
34. K. Iwamoto et al., Nature **395**, 672 (1998)
35. D. Kasen, F.K. Röpke, S.E. Woosley, Nature **460**, 869 (2009)
36. A.M. Khokhlov, Astron. Astrophys. **245**, 114 (1991)
37. R.P. Kirshner, J. Kwan, Astrophys. J. **193**, 27 (1974)
38. E. Komatsu et al., Astrophys. J. Suppl. S. **180**, 330 (2009)
39. C.T. Kowal, Astron. J. **73**, 1021 (1968)
40. R.-P. Kudritzki, Astron. Nachr. **331**, 459 (2010)
41. B. Leibundgut, Annu. Rev. Astron. Astrophys. **39**, 67 (2001)
42. B. Leibundgut et al., Astron. J. **105**, 301 (1993)
43. E. Livne, D. Arnett, Astrophys. J. **452**, 62 (1995)
44. A.I. MacFadyen, S.E. Woosley, Astrophys. J. **524**, 262 (1999)
45. P.A. Mazzali, I.J. Danziger, M. Turatto, Astron. Astrophys. **297**, 509 (1995)
46. P.A. Mazzali, N. Chugai, M. Turatto, L.B. Lucy, I.J. Danziger, E. Cappellaro, M. della Valle, S. Benetti, Mon. Not. R. Astron. Soc. **284**, 151 (1997)
47. P.A. Mazzali, E. Cappellaro, I.J. Danziger, M. Turatto, S. Benetti, Astrophys. J. **499**, L49 (1998)
48. P.A. Mazzali, P. Podsiadlowski, Mon. Not. R. Astron. Soc. **369**, L19 (2006)
49. P.A. Mazzali, F.K. Röpke, S. Benetti, W. Hillebrandt, Science **315**, 825 (2007)
50. M.R. Metzger, S.G. Djorgovski, S.R. Kulkarni, C.C. Steidel, K.L. Adelberger, D.A. Frail, E. Costa, F. Frontera, Nature **387**, 878 (1997)
51. R. Minkowski, Publ. Astron. Soc. Pac. **53**, 224 (1941)
52. E. Nakar, T. Piran, Mon. Not. R. Astron. Soc. **360**, L73 (2005)
53. H.U. Norgaard-Nielsen, L. Hansen, H.E. Jorgensen, A. Aragon Salamanca, R.S. Ellis, Nature **339**, 523 (1989)
54. B. Paczynski, Astrophys. J. **308**, L43 (1986)
55. R. Pakmor, M. Kromer, F.K. Röpke, S.A. Sim, A.J. Ruiter, W. Hillebrandt, Nature **463**, 61 (2010)
56. W.J. Percival et al., Mon. Not. R. Astron. Soc. **327**, 1297 (2001)
57. S. Perlmutter et al., Astrophys. J. **440**, L41 (1995)
58. S. Perlmutter et al., Astrophys. J. **483**, 565 (1997)
59. S. Perlmutter et al., Astrophys. J. **517**, 565 (1999)
60. M.M. Phillips, Astrophys. J. **413**, L105 (1993)
61. M.M. Phillips, L.A. Wells, N.B. Suntzeff, M. Hamuy, B. Leibundgut, R.P. Kirshner, C.B. Foltz, Astron. J. **103**, 1632 (1992)
62. D. Poznanski, A. Gal-Yam, D. Maoz, A.V. Filippenko, D.C. Leonard, T. Matheson, Publ. Astron. Soc. Pac. **114**, 833 (2002)
63. J.L. Racusin et al., Nature **455**, 183 (2008)
64. A.G. Riess, W.H. Press, R.P. Kirshner, Astrophys. J. **438**, L17 (1995)
65. A.G. Riess, et al., Astron. J. **116**, 1009 (1998)
66. R. Salvaterra et al., Nature **461**, 1258 (2009)
67. B.P. Schmidt, R.P. Kirshner, R.G. Eastman, M.M. Phillips, N.B. Suntzeff, M. Hamuy, J. Maza, R. Aviles, Astrophys. J. **432**, 42 (1994)
68. D.N. Spergel et al., Astrophys. J. Suppl. S. **148**, 175 (2003)
69. M. Sullivan et al., Mon. Not. R. Astron. Soc. **406**, 782 (2010)
70. S.H. Suyu, P.J. Marshall, M.W. Auger, S. Hilbert, R.D. Blandford, L.V.E. Koopmans, C.D. Fassnacht, T. Treu, Astrophys. J. **711**, 201 (2010)
71. G.A. Tammann, A. Sandage, B. Reindl, Astron. Astrophys. Rev. **15**, 289 (2008)
72. N.R. Tanvir et al., Nature **461**, 1254 (2009)

73. F.X. Timmes, E.F. Brown, J.W. Truran, Astrophys. J. **590**, L83 (2003)
74. J.L. Tonry, A. Dressler, J.P. Blakeslee, E.A. Ajhar, A.B. Fletcher, G.A. Luppino, M.R. Metzger, C.B. Moore, Astrophys. J. **546**, 681 (2001)
75. R.B. Tully, J.R. Fisher, Astron. Astrophys. Rev. **54**, 661 (1977)
76. J. van Paradijs et al., Nature **386**, 686 (1997)
77. X. Wang, L. Wang, X. Zhou, Y.-Q. Lou, Z. Li, Astrophys. J. **620**, L87 (2005)
78. J. Whelan, I. Iben Jr., Astrophys. J. **186**, 1007 (1973)
79. S.E. Woosley, S. Blinnikov, A. Heger, Nature **450**, 390 (2007)

Chapter 3
High Energy Cosmic Ray and Neutrino Astronomy

Eli Waxman

Abstract Cosmic-rays with energies exceeding 10^{19} eV are referred to as Ultra High Energy Cosmic Rays (UHECRs). The sources of these particles and their acceleration mechanism are unknown, and for many years have been the issue of much debate. The first part of this review describes the main constraints, that are implied by UHECR observations on the properties of candidate UHECR sources, the candidate sources, and the related main open questions.

In order to address the challenges of identifying the UHECR sources and of probing the physical mechanisms driving them, a "multi-messenger" approach will most likely be required, combining electromagnetic, cosmic-ray and neutrino observations. The second part of the review is devoted to a discussion of high energy neutrino astronomy. It is shown that detectors, which are currently under construction, are expected to reach the effective mass required for the detection of high energy extra-Galactic neutrino sources, and may therefore play a key role in the near future in resolving the main open questions. The detection of high energy neutrinos from extra-Galactic sources will not only provide constraints on the identity and underlying physics of UHECR sources, but may furthermore provide information on fundamental neutrino properties.

Keywords Acceleration of particles • Cosmic rays • Neutrinos • Gamma-Ray burst: general • Relativistic processes

---

E. Waxman (✉)
Particle Physics & Astrophysics Department, Weizmann Institute of Science, Rehovot, Israel
e-mail: eli.waxman@weizmann.ac.il

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_3,

1 Introduction

Cosmic-rays (CRs) with energies exceeding $\sim 10^{19}$ eV are referred to as Ultra High Energy Cosmic Rays (UHECRs). The sources of these particles, which are probably extra-Galactic, and their acceleration mechanism are unknown, and for many years have been the issue of much debate (e.g. [29, 32, 55, 79] and references therein). The first part of this chapter, Sect. 2, describes the main constraints that are implied by UHECR observations on the properties of candidate UHECR sources. The constraints derived under the assumption that UHECRs are protons, which is supported by most observations but questioned by some (see Sects. 2.1–2.3 and 4), are summarized in Sect. 2.6. In Sect. 2.7 it is shown that GRBs are the only known type of sources that satisfy these constraints. Testable predictions for the spectrum and arrival direction distribution of UHECRs, made by the GRB model of UHECR production, are also described.

The challenges of identifying the UHECR sources, and of probing the physical mechanisms driving them, may be met with the help of high energy neutrino detectors [16, 52, 56, 99]. This is discussed in Sect. 3. In Sect. 3.1 it is shown that detectors, which are currently under construction, are expected to reach the effective mass required for the detection of high energy extra-Galactic neutrino sources (see Figs. 3.8 and 3.9), and may therefore play a key role in the near future in resolving the main open questions. GZK and GRB neutrinos are discussed in Sects. 3.2 and 3.3 respectively. In Sect. 3.4 we point out that the detection of high energy neutrinos from extra-Galactic sources will not only provide constraints on the identity and underlying physics of UHECR sources, but may furthermore provide information on fundamental neutrino properties.

The main open questions associated with the production of UHECRs are summarized in Sect. 4. It is argued that a "multi-messenger" approach, combining electromagnetic, cosmic-ray and neutrino data, would be required in order to provide answers to these questions.

2 What We (Don't) Know About the Sources of UHECRs

The origin of CRs of all energies is still unknown (see [23, 34, 79] for reviews). The cosmic ray properties change qualitatively as a function of particle energy, as illustrated in Fig. 3.1. The spectrum steepens around $\sim 5 \times 10^{15}$ eV (the "knee") and flattens around 5×10^{18} eV (the "ankle"). Below $\sim 10^{15}$ eV, the cosmic rays are thought to originate from Galactic supernovae. However, this hypothesis has not yet been confirmed (e.g. [37] and references therein). The composition is dominated by protons at the lowest energies, and the fraction of heavy nuclei increases with energy. The proton fraction at $\sim 10^{15}$ eV is reduced to $\sim 15\%$ [31, 36]. At yet higher energies, there is evidence that the fraction of light nuclei increases, and that the cosmic-ray flux above 5×10^{18} eV is again dominated by protons [33, 41, 45].

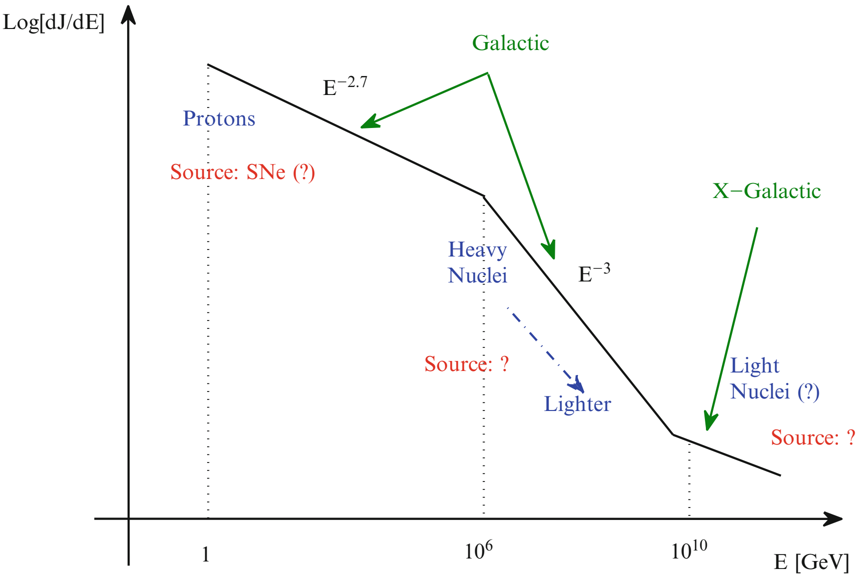


Fig. 3.1 A schematic description of the differential CR spectrum, dJ/dE, with some comments on what we know (or don't) about the composition and origin of the CRs

The composition change and the flattening of the spectrum around 10^{19} eV suggest that the flux above and below this energy is dominated by different sources. At energies of $E_{19} \equiv E/10^{19}\text{eV} \sim 1$ the Larmor radius of CRs in the Galactic magnetic field is

$$R_L = \frac{E}{ZeB} \approx 3B_{-5.5}^{-1}E_{19}Z^{-1}\text{kpc}, \tag{3.1}$$

where $B = 10^{0.5}B_{-5.5}\mu\text{G}$ is the value of the Galactic magnetic field and Z is the CR charge. Since the Galactic magnetic field cannot confine protons above 10^{19} eV, it is believed that the nearly isotropic cosmic ray flux at $E > 5 \times 10^{18}$ eV originates from extra-Galactic (XG) sources. In what follows we focus on this XG component.

2.1 Composition

At low energy, $<$10 TeV, CR particles are detected by space or balloon born detectors, which provide a direct measurement of the primary CR composition. At higher energies, the flux is too low to be detectable by space/balloon born detectors, and CRs are detected indirectly through the "air-showers", the large number of lower energy particles, they produce as they propagate and lose energy in the atmosphere. The low flux at the highest energies,

$$J(>10^{20}\text{eV}) \approx 1/100\,\text{km}^2\text{year}\,2\pi\,\text{sr}, \tag{3.2}$$

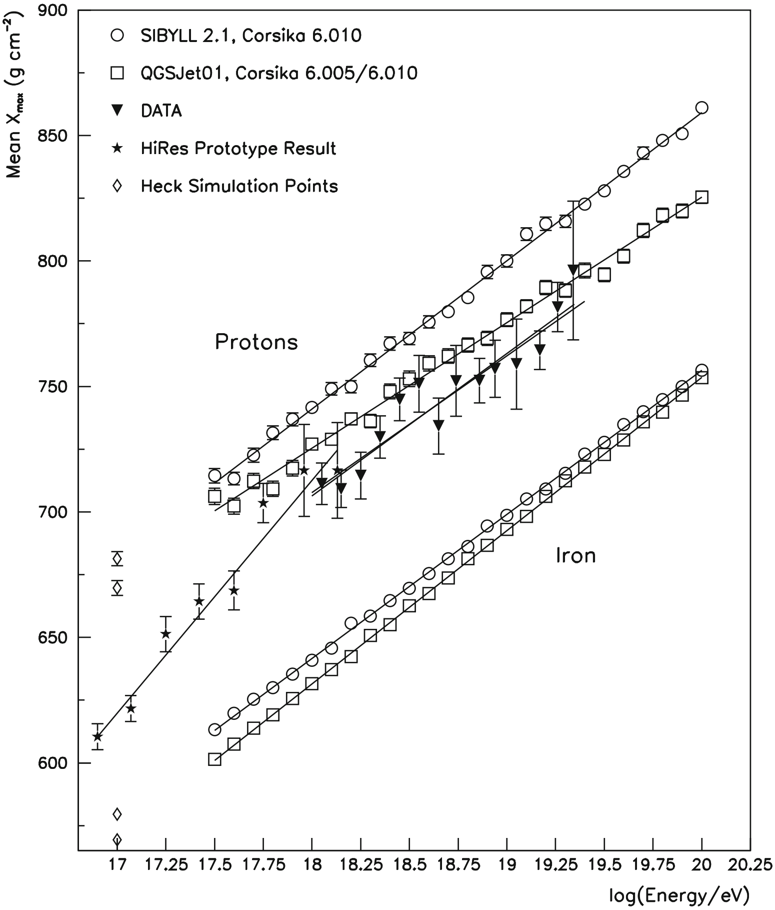


Fig. 3.2 Average depth of shower maximum as function of energy: Measurements by the HiRes detector compared to predictions for proton and iron primaries based on various model extrapolations of the *pp* cross section (Adapted from Abbasi et al. [3])

requires detectors with effective area of many 100's km^2. The primary composition is constrained at high energies mainly by the average and variance of $X_{\max}$, the depth in the atmosphere at which the shower contains the largest number of high energy particles, obtained for showers of fixed energy (fluctuations in individual shower development are large, leading to fluctuations in the depth of maximum which are not small compared to the dependence on the primary mass). X_{max} is larger for higher energy particles. Since a high energy heavy nucleus behaves roughly as a group of independent lower energy nucleons, X_{max} and its variance are larger at fixed energy for lighter nuclei.

Figure 3.2 presents the main evidence for the transition to lighter nuclei at higher energy: $X_{\max}$ grows with energy faster than model predictions for fixed composition,

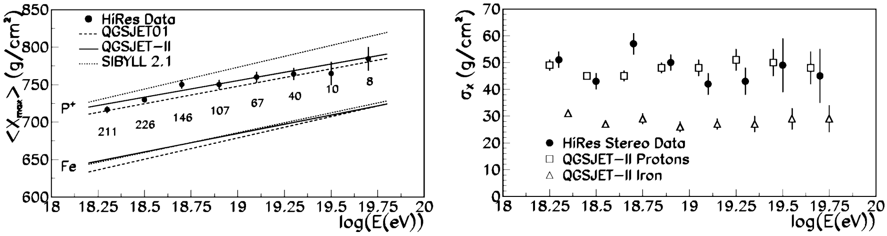


Fig. 3.3 Average and standard deviation of $X_{\max}$ at the highest energies: Measurements by the HiRes detector compared to model predictions (Reprinted figures with permission from Abbasi et al. [2]; http://link.aps.org/abstract/PRL/v104/p161101. Copyright (2010) by the American Physical Society)

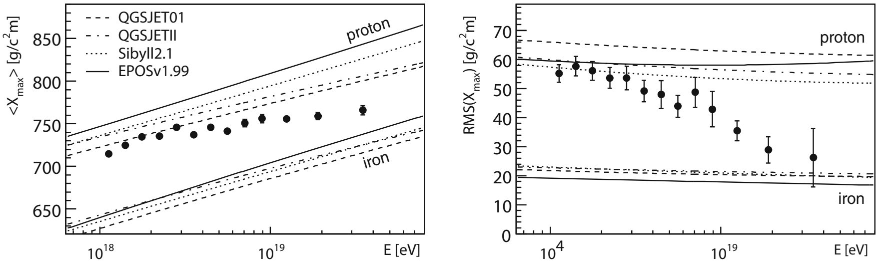


Fig. 3.4 Same as Fig. 3.3, for the PAO data (Reprinted figures with permission from Abraham et al. [8]; http://link.aps.org/abstract/PRL/v104/p091101. Copyright (2010) by the American Physical Society)

becoming consistent with pure proton composition at $\sim 10^{18}$ eV. At the highest observed energies, there is some discrepancy between the results reported by the HiRes observatory and by the Pierre Auger Observatory (PAO). While the HiRes observatory reports the average and variance of $X_{\max}$ to be consistent with a pure proton composition all the way up to $10^{19.7}$ eV (Fig. 3.3), the PAO reports $X_{\max}$ and σ_X evolution which suggests a transition back to heavier nuclei at the highest energies (Fig. 3.4).

The origin of this discrepancy is not yet understood. However, a few comments are in place. It was noted in [107] that the analysis of the PAO data, presented in [8], is not self consistent: according to this analysis, σ_X measured at the highest energy implies an Fe fraction $>90\%$, while the measured value of $\langle X_{\max} \rangle$ implies an Fe fraction $<60\%$. This inconsistency may reflect some experimental problem, but may also reflect a modification of the hadronic interaction cross section which is not accounted for in the models used for shower calculations. It should be emphasized that the theoretical $X_{\max}$ calculations depend on extrapolation of hadronic models to energies well beyond those currently tested in accelerators. The theoretical and experimental uncertainties in the extrapolation of the pp cross-section to

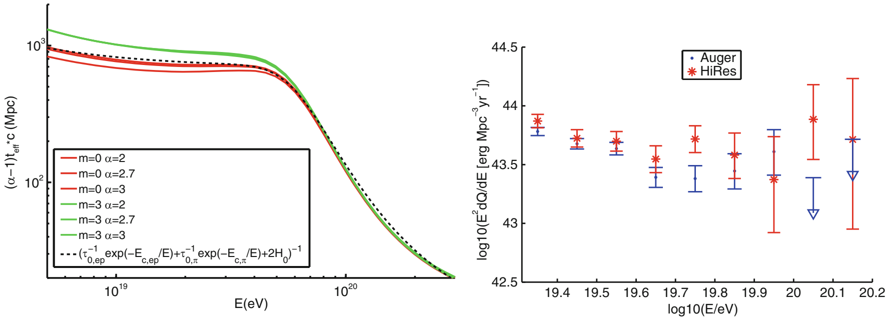


Fig. 3.5 *Left*: A comparison of direct numerical calculations of the effective CR life time (*solid lines*) with the analytic approximation of [60] using $\{E_{c,ep} = 9.1 \times 10^{18}$ eV, $\tau_{0,ep} = 0.5 \times 10^9$ year, $E_{c,\pi} = 3.5 \times 10^{20}$ eV, $\tau_{0,\pi} = 1.4 \times 10^7$ year$\}$ (*dashed line*), for CR generation following $d\dot{n}/dE(E,z) \propto (1+z)^m E^{-\alpha}$. *Right*: The local ($z = 0$) energy generation rate as measured by Auger [35] and Hires [4] assuming that the CRs are purely protons, for $\alpha - 1 = 1$ (For different values of α, the spectrum should be multiplied by an energy independent factor $(\alpha - 1)$; $Q \equiv \dot{n}$). Statistical and systematic errors in the experimental determination of event energies lead to ∼50% uncertainty in the flux at the highest energies. The absolute energy scales of the Auger and Hires data where not altered in this figure (Adapted from Katz et al. [60])

center-of-mass energies ≥ 100 TeV are a possible source of biases in shower reconstruction (e.g. [89]). It is therefore difficult to draw a firm conclusion regarding primary composition at the highest energies based on current shower measurements.

2.2 Generation Rate and Spectrum

Let us assume first that the UHECRs are protons of extra-Galactic origin. As they propagate, high-energy protons lose energy as a result of the cosmological redshift and as a result of production of pions and $e+e-$ pairs in interactions with cosmic microwave background (CMB) photons. The local intensity of UHECRs may be written as

$$\frac{dJ(E)}{dE} = \frac{c}{4\pi} \frac{d\dot{n}_0(E)}{dE} t_{\text{eff.}}(E), \tag{3.3}$$

where $d\dot{n}_0(E)/dE$ is the local ($z = 0$) proton production rate (per unit volume and proton energy) and $t_{\text{eff.}}$ is the effective energy loss time of the proton (this equation is, in fact, a definition of $t_{\text{eff.}}$). The left panel of Fig. 3.5 shows $t_{\text{eff.}}$ for proton generation following $d\dot{n}/dE(E,z) \propto (1+z)^m E^{-\alpha}$. The rapid decrease in the effective life time, or propagation distance $ct_{\text{eff.}}$, above $\sim 6 \times 10^{19}$ eV, commonly termed the "Greisen-Zatsepin-Kuzmin (GZK) suppression" [47, 109], is due to photo-production of pions by the interaction of protons with CMB photons (The proton threshold energy for pion production on $\sim 10^{-3}$ eV CMB photons is $\sim 10^{20}$ eV). Since proton propagation is limited at high energies to distances $\ll c/H_0$, e.g. to $\sim$100 Mpc at 10^{20} eV, the dependence of $t_{\rm eff.}$ on redshift evolution (m) is not strong.

Using Eq. 3.3 and the measured UHECR intensity, it is straightforward to infer the local production rate of UHECRs. The right panel of Fig. 3.5 shows that the energy generation rate above $10^{19.5}$ eV is roughly constant per logarithmic CR energy interval, $\alpha \approx 2$ and

$$E^2 \frac{d\dot{n}_0(E)}{dE} \approx 10^{43.5} \text{erg/Mpc}^3\text{year}. \tag{3.4}$$

In other words, the observed CR spectrum is consistent with a generation spectrum $d\dot{n}/dE \propto E^{-2}$ modified by the GZK suppression. Since both observations and models for particle acceleration in collisionless shocks, which are believed to be the main sources of high energy particles in many astrophysical systems, typically imply $\alpha \approx 2$ (see [34, 98] for reviews of particle acceleration in non-relativistic and relativistic shocks respectively), this supports the validity of the assumption that UHECRs are protons produced by extra-Galactic objects.

The following point should, however, be made here. Heavy nuclei lose energy by interaction with CMB and IR photons, that leads to spallation. Since the effective life time of such nuclei is not very different from that of protons, the consistency of the observed spectrum with a model of extra-Galactic sources of protons with generation spectrum of $d\dot{n}/dE \propto E^{-2}$ can not be considered as a conclusive evidence for the UHECRs being protons.

One of the important open questions is at what energy the transition from Galactic to extra-Galactic (XG) sources takes place. A simple model with $E^2 d\dot{n}_0(E)/dE = 5 \times 10^{43} \text{erg/Mpc}^3\text{yr}$ and a transition from Galactic to XG sources at 10^{19} eV is consistent with observations [60]. In such a model, the Galactic flux is comparable to the XG one at 10^{19} eV, and negligible at $>10^{19.5}$ eV. Other models, however, have been proposed, in which the Galactic-XG transition occurs well below 10^{19} eV (e.g. [29] and references therein). Such models are motivated mainly by the argument that they allow one to explain the $\sim 5 \times 10^{18}$ eV spectral feature by pair production (in proton interactions with the CMB). The transition energy in such models is therefore well below 5×10^{18} eV. As explained in [60], a Galactic-XG transition at $\sim 10^{18}$ eV requires fine tuning of the Galactic and XG contributions (to produce the smooth power-law observed), and is disfavored by the data: it requires that Auger systematically underestimates the energy of the events by 40% (well above the stated uncertainty) and it requires $d\dot{n}_{p,\rm XG}/d\varepsilon \propto \varepsilon^{-2.7}$, which is inconsistent with the $>10^{19}$ eV data.

Finally, one notes that if the generation spectrum of XG CRs extends over many decades below 10^{19} eV, the total XG CR energy production rate, $Q_{\rm XG}^{z=0}$, might exceed significantly the UHECR production rate, $Q_{10^{19}\rm eV}^{z=0} \equiv (E^2 d\dot{n}_0/dE)_{>10^{19}\rm eV}$, given by Eq. 3.4. For the $d\dot{n}/dE \propto 1/E^2$ spectrum inferred from observations, the “bolometric correction” will be $Q_{\rm XG}/Q_{10^{19}\rm eV} \approx \ln(10^{20}\text{eV}/E_{\rm min}) \sim 10$, where $E_{\rm min} \ll 10^{19}$ eV is the low energy to which the spectrum extends.

2.3 Anisotropy: Source and Composition Clues

The propagation of UHECRs is limited at the highest energies to distances $\sim$100 Mpc. The galaxy distribution is not homogeneous over such a distance scale. Thus, if the distribution of UHECR sources is correlated with that of galaxies, one expects an anisotropy in the UHECR arrival direction distribution reflecting the inhomogeneity of the galaxy distribution [104]. Figure 3.6 shows the integrated galaxy density out to 75 Mpc and the predicted anisotropy of the UHECR intensity. Also shown are the (angular) positions of the 27 Auger events with energy exceeding 5.7×10^{19} eV. The distribution of these events is inconsistent with isotropy at a 98% confidence level for a source density $n_s = 10^{-4}\text{Mpc}^{-3}$, corresponding to the lowest allowed source density (see Sect. 2.4), and at a 99% confidence level for $n_s = 10^{-2}\text{Mpc}^{-3}$, corresponding to the density of galaxies. The angular distribution of CR arrival directions is consistent with a UHECR source distribution that follows the galaxy distribution (for detailed discussions see [54, 59, 61, 88]). This provides some support to the association of the sources with known extra-Galactic astrophysical objects.[1] The more recent PAO analysis (valid for $n_s \to \infty$) of a larger number of events, 58 above 5.5×10^{19} eV, yields inconsistency with isotropy at a 99% confidence level [7].

UHECRs may suffer significant deflections as they cross dense large scale structures, such as galaxy clusters and large scale galaxy filaments, in which the energy density of the plasma is large enough to support strong magnetic fields. Such deflections may distort the anisotropy pattern expected based on the galaxy distribution. An estimate of the expected deflection may be obtained assuming that all large scale structures support a magnetic field with energy density comprising a fraction ε_B of the plasma thermal energy density. The deflection expected in this case for a propagation distance d is (see [59, 62] for a detailed derivation)

$$\theta \approx 0.3^\circ \frac{L}{1\ \text{Mpc}} \left(\frac{f}{0.1} \frac{d}{100\ \text{Mpc}} \frac{\lambda}{10\ \text{kpc}} \right)^{1/2} \left(\frac{\varepsilon_B}{0.01} \right)^{1/2} \left(\frac{E/Z}{10^{20}\ \text{eV}} \right)^{-1} . \quad (3.5)$$

Here, Z is the particle charge, f is the fraction of the volume filled by filaments of diameter L, and λ is the field coherence length. The deflections are not expected therefore to distort significantly the anisotropy map.

The anisotropy signal provides also a test of the primary UHECR composition. If one records an anisotropy signal produced by heavy nuclei of charge Z above an energy E_{thr}, one should record an even stronger (possibly much stronger) anisotropy

---

[1] The evidence in the PAO data for a clustering of events in the region around Cen A has triggered much discussion (see discussion and references in [69]). However, it is difficult to quantify the level of significance of the evidence for clustering, since it is based on an a posteriori analysis, as noted in [7]. Moreover, one must keep in mind that Cen A lies in front of one of the largest concentrations of matter in the local ($d \sim 50$Mpc), Universe, $\{l = -51^\circ, b = 19^\circ\}$, so that an excess of events from that direction does not necessarily imply that Cen A is the source.

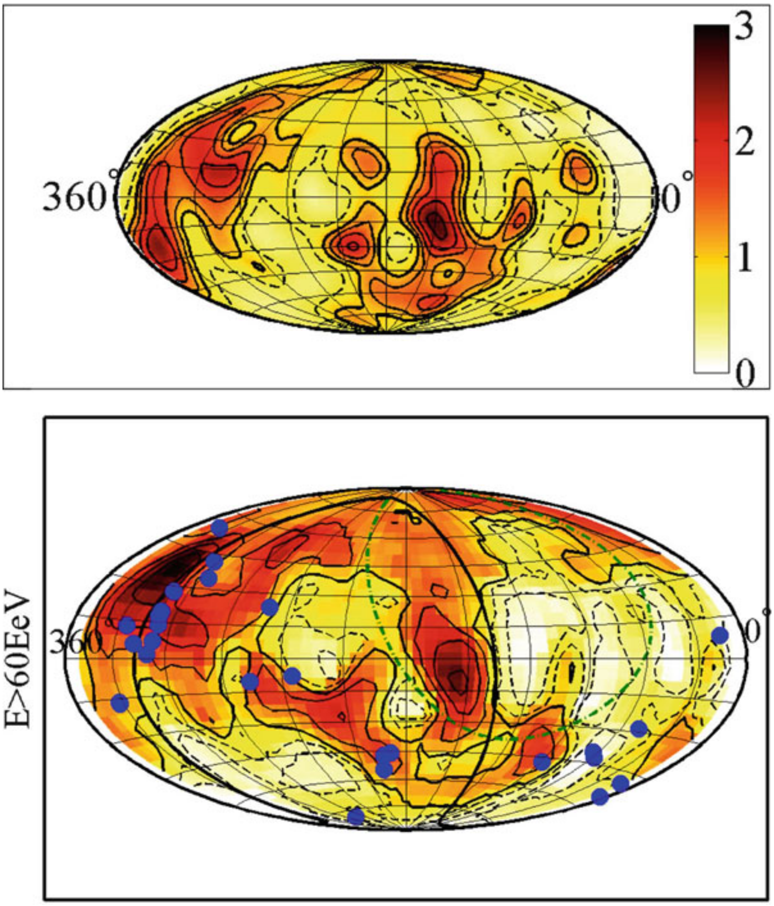


Fig. 3.6 *Top*: The integrated galaxy density out to a distance of 75 Mpc, normalized to the mean integrated density. The contours are logarithmic, ranging from 0.5 to 4 with three contours per density doubling. Dashed curves represent under-density. *Bottom*: The positions of the 27 Auger events with energy exceeding 5.7×10^{19} eV [5], overlaid on the UHECR intensity map, $J(\hat{\Omega})$, predicted in a model in which the UHECR source distribution follows the galaxy density distribution (with a bias $b[\delta] = 1 + \delta$ for $\delta > 0$, $b = 0$ otherwise, where δ is the fractional galaxy over density). The coordinates are Galactic and J is normalized to its all sky average. The contours denote $J/\bar{J} = (0.7, 0.9, 1, 1.1, 1.3, 1.5)$, with dashed lines representing under-density. The thick solid line denotes the super-galactic plane. The dashed-dotted green line marks the boundary of Auger's coverage (corresponding to a zenith angle of 60°) (Adapted from Kashti et al. [59])

at energies $>E_{\rm thr}/Z$ due to the proton component that is expected to be associated with the sources of the heavy nuclei. This is due to the fact that particles of similar rigidity E/Z propagate in a similar manner in the inter-galactic magnetic field and based on the plausible assumptions that (i) a source accelerating particles of charge Z to energy E will accelerate protons to energy E/Z, and (ii) there are at least as many protons accelerated as there are heavy nuclei. The anisotropy signal is expected to be stronger at lower energy since the signal increases as the number of particles produced by the source, $E^{-\alpha+1}$, while the background increases as the square-root of the number of all observed CRs, $\sim E^{-(2.7-1)/2}$ (see [69] for a detailed discussion). Thus, if the PAO $> 5.7 \times 10^{19}$ eV anisotropy signal is real, the lack of detection of stronger anisotropy at lower energy disfavors a heavy nuclei composition at $\sim 6 \times 10^{19}$ eV.

2.4 Source Density

The arrival directions of the 27 PAO events and ∼30 HiRes events above 6×10^{19} eV show no evidence for "repeaters", i.e. multiple events that may be associated with a single source given the small deflection angles expected. The lack of repeaters implies that the number of sources contributing to the flux, N_s, should satisfy $N_s > N^2$, where N is the number of events (for identical sources each producing on average N/N_s events and $N^2/N_s \ll 1$, the probability for repeaters is $\sim N^2/N_s$). This suggests that there should be more than $\sim 10^{3.5}$ independent sources contributing to the (all sky) flux (note that HiReS and PAO observed the northern and southern hemispheres respectively). For protons, the effective propagation distance is ∼200 Mpc, see Fig. 3.5, implying a lower limit on the source density of

$$n_s > 10^{-4}\ \mathrm{Mpc}^{-3} \tag{3.6}$$

(for a more detailed analysis see [39, 104]). For comparison, the density of galaxies is roughly $10^{-2}\mathrm{Mpc}^{-3}$.

2.5 Source Constraints: Minimum Power and Speed

The essence of the challenge of accelerating particles to $>10^{19}$ eV can be understood using the following simple arguments ([99], for a more detailed derivation see [100]). Consider an astrophysical source driving a flow of magnetized plasma, with characteristic magnetic field strength B and velocity v. Imagine now a conducting wire encircling the source at radius R, as illustrated in Fig. 3.7. The potential generated by the moving plasma is given by the time derivative of the magnetic flux Φ and is therefore given by $V \approx \beta BR$ where $\beta = v/c$. A proton which is allowed to be accelerated by this potential drop would reach energy $E \sim \beta eBR$. The situation is somewhat more complicated in the case of a relativistic outflow, with $\Gamma \equiv (1-\beta^2)^{-1/2} \gg 1$. In this case, the proton is allowed to be accelerated only over a fraction of the radius R, comparable to R/Γ. To see this, one must realize that as the plasma expands, its magnetic field decreases, so the time available for

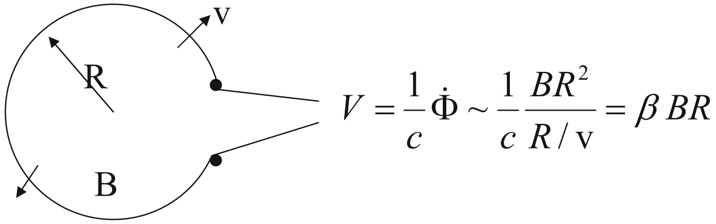


Fig. 3.7 Potential drop generated by an unsteady outflow of magnetized plasma

acceleration corresponds to the time of expansion from R to, say, $2R$. In the observer frame this time is R/c, while in the plasma rest frame it is $R/\Gamma c$. Thus, a proton moving with the magnetized plasma can be accelerated over a transverse distance $\sim R/\Gamma$. This sets a lower limit to the product of the magnetic field and source size, which is required to allow acceleration to E, $BR > \Gamma E/e\beta$. This constraint also sets a lower limit to the rate L at which energy should be generated by the source. The magnetic field carries with it an energy density $B^2/8\pi$, and the flow therefore carries with it an energy flux $> vB^2/8\pi$ (some energy is carried also as plasma kinetic energy), which implies $L > vR^2B^2$ and therefore

$$L > \frac{\Gamma^2}{\beta}\left(\frac{E}{e}\right)^2 c = 10^{45.5}\frac{\Gamma^2}{\beta}\left(\frac{E}{10^{20}\ \mathrm{eV}}\right)^2 \mathrm{erg/s}. \tag{3.7}$$

Another constraint on the source results from the requirement that the acceleration is not suppressed by synchrotron emission of the accelerated particle. Let us consider a relativistic source. Denoting by B' the magnetic field in the plasma rest frame, the acceleration time of a proton is $t'_{\mathrm{acc}} > E'/eB'c$ where $E' = E/\Gamma$ is the proton energy in the plasma frame. The synchrotron loss time, on the other hand, is given by $t'_{\mathrm{syn}} \approx (m_p/m_e)^2(6\pi E'/\sigma_T c\gamma'^2 B'^2)$ where $\gamma' = E'/m_pc^2$. Requiring $t'_{\mathrm{acc}} < t'_{\mathrm{syn}}$ sets an upper limit on B' (which depends on E and Γ). Requiring this upper limit to be larger than the lower limit derived in the previous paragraph, $B'R > E/e$, sets a lower limit to Γ (which depends on R and E). Relating the source radius R to an observed variability time (of the radiation emitted by the source) through $R = 2\Gamma^2 c\delta t$, the lower limit is [100]

$$\Gamma > 10^2\left(\frac{E}{10^{20}\ \mathrm{eV}}\right)^{3/4}\left(\frac{\delta t}{10\ \mathrm{ms}}\right)^{-1/4}. \tag{3.8}$$

This implies that the sources must be relativistic, unless their characteristic variability time exceeds $\approx 10^6$ s.

2.6 Summary of Source Constraints

The evidence for a transition to a light composition, consistent with protons, at few $\times 10^{18}$ eV (Sect. 2.1), the consistency of the spectrum above $\sim 10^{19}$ eV with a $d\dot{n}/dE \propto 1/E^2$ generation spectrum modified by the GZK suppression (Sect. 2.2), and the hints for a light composition from the anisotropy signal (Sect. 2.3), suggest that the UHECRs are protons produced by extra-Galactic sources. If this is indeed the case, the discussion of the preceding sections implies that their sources must satisfy several constraints:

- The sources should produce protons with a local ($z = 0$) rate and spectrum (averaged over space and time) $E^2 d\dot{n}_0/dE \approx 10^{43.5}\mathrm{erg/Mpc^3yr}$;
- The density of sources (contributing to the flux at $\sim 5 \times 10^{19}$ eV) should satisfy $n_s > 10^{-4}\mathrm{Mpc^{-3}}$;
- The power output of the individual sources should satisfy $L > 10^{45.5}\Gamma^2\beta^{-1}\mathrm{erg/s}$;
- The Lorentz factor of the flow driven by the source must satisfy $\Gamma > 10^2$ $(\delta t/10\mathrm{ms})^{-1/4}$ where δt is the characteristic source variability time.

No sources that satisfy the constraint $L > 10^{46}\mathrm{erg/s}$ are known to lie within a $\sim$100 Mpc distance. One may argue, of course, that there are "dark sources", i.e. sources that produce such power output (and UHECRs) but do not produce much radiation and are hence not known. One can not rule out the existence of such sources. On the other hand, we do not have direct evidence for their existence either. Putting aside such a caveat, the lack of known sources of sufficient luminosity suggests that the sources are transient. The transient duration T must be shorter than the time delay between the arrival of photons and protons from the source. The protons are delayed due to magnetic field deflection by $\Delta t \sim \theta^2 d/c$, where θ is estimated in Eq. 3.5. This yields

$$\Delta t(E,d) \sim 10^4 \left(\frac{d}{100\,\mathrm{Mpc}}\right)^2 \left(\frac{E}{10^{20}\,\mathrm{eV}}\right)^{-2} \mathrm{year}. \tag{3.9}$$

Due to the random energy loss of the protons during their propagation, and due to the possibility of multiple paths between source and observer, the arrival of protons of energy E is delayed and spread over a similar time $\Delta t(E,d)$. For $T < \Delta t$, the effective number density of sources contributing to the flux at energy E is $\sim \dot{n}_s \Delta t[E, d_{\mathrm{eff}}(E)]$, where $\dot{n}_s$ is the transient rate (per unit volume) and $d_{\mathrm{eff}}(E) \sim ct_{\mathrm{eff}}(E)$.

2.7 "Suspects", Predictions

Only two types of sources are known to satisfy the above minimum power requirement: active galactic nuclei (AGN) – the brightest known steady sources, and gamma-ray bursts (GRBs) – the brightest known transient sources.[2] Both AGN (e.g. [83]) and GRBs [73, 90, 100] have therefore been suggested to be UHECR sources. The absence of AGN with $L > 10^{46}$ erg s^{-1} within the GZK horizon had motivated the suggestion [44] that UHECRs may be produced by a new, yet undetected, class of short duration AGN flares resulting from the tidal disruption of stars or accretion disk instabilities. The existence of tidal disruption flares is likely. However, they are yet to be detected and whether their properties are consistent with the constraints derived above is yet to be determined (see also [105]).

Let us consider then the GRB transients. First, consider the minimum power and minimum speed constraints that should be satisfied by individual sources: Eqs. 3.7 and 3.8. For GRBs, the (luminosity function averaged) peak luminosity is $L_\gamma \approx 10^{52}$erg/s ([51, 92], note that [51] gives $L_{50-300\text{ keV}}$ which is ≈ 0.1 of $L_{0.1-10\text{ MeV}}$ given in [92]), and typical values of Γ and δt are $\Gamma \simeq 10^{2.5}$ and $\delta t \sim 10$ ms [70, 71, 82, 93]. Thus, both constraints are satisfied. It is worth noting that $\Gamma > 10^2$ is inferred for GRBs based on the photon spectrum (in order to avoid large pair production optical depth), i.e. based on arguments which are different than those leading to the $\Gamma > 10^2$ constraint of Eq. 3.8.

Next, let us consider the global constraints on the rate, Eq. 3.6, and average energy production rate, Eq. 3.4, of the sources. The local, $z = 0$, GRB rate is $\dot{n}_s^{z=0} \sim 10^{-9}\text{Mpc}^{-3}\text{yr}^{-1}$ (assuming $\dot{n}_s$ evolves rapidly with redshift, following the star formation rate, i.e $\dot{n}_s^{z=0} \ll \dot{n}_s^{z=1.5}$ [51, 92]), implying, using Eq. 3.9, $n_s(E) \sim \dot{n}_s^{z=0}\Delta t[E, d_{\text{eff}}(E)] \sim 10^{-4}(d_{\text{eff}}/200\text{ Mpc})^2(E/0.5\times 10^{20}\text{ eV})^{-2}\text{Mpc}^{-3}$, consistent with Eq. 3.6. The local, $z = 0$, GRB energy production rate in ∼1 MeV photons is given by $\dot{n}_s^{z=0}L_\gamma\Delta t$, where Δt is the effective duration (the average ratio of the fluence to the peak luminosity) corrected for redshift (the observed duration is $1+z$ larger than the duration at the source), $\Delta t \approx 10\text{s}/(1+z) \sim 4$ s (using $z = 1.5$ as a characteristic redshift). This yields $E_\gamma \equiv L_\gamma\Delta t \approx 10^{52.5}$erg and $Q^{z=0}_{\text{MeV,GRB}} \equiv \dot{n}_s^{z=0}E_\gamma \approx 10^{43.5}\text{erg/Mpc}^3\text{yr}$, similar to the required UHECR energy production rate given in Eq. 3.4, $Q^{z=0}_{10^{19}\text{eV}} \equiv (E^2 d\dot{n}_0/dE)_{>10^{19}\text{eV}} \approx 10^{43.5}\text{erg/Mpc}^3\text{yr}$ (for a more detailed discussion see [67, 95, 96]; for additional energy production by “low-luminosity GRBs” and “heavy baryon loading GRBs” see [77] and [106] respectively, and references therein).

As noted at the end of Sect. 2.2, if the generation spectrum of XG CRs extends over many decades below 10^{19} eV, the total XG CR energy production rate, $Q^{z=0}_{\text{XG}}$, may exceed significantly the UHECR production rate, $Q_{\text{XG}}/Q_{10^{19}\text{eV}} \sim 10$. Estimating the ratio of $Q^{z=0}_{\text{XG}}$ to the total photon energy production by GRBs, $Q^{z=0}_{\gamma,\text{GRB}}$, as $Q^{z=0}_{\text{XG}}/Q^{z=0}_{\gamma,\text{GRB}} = Q^{z=0}_{\text{XG}}/Q^{z=0}_{\text{MeV,GRB}} \sim 10$ is, however, quite uncertain. This is due to uncertainties in the redshift evolution of the GRB rate and luminosity function,

---

[2] It was recognized early on ([55] and references therein) that while highly magnetized neutron stars may also satisfy the minimum power requirement, it is difficult to utilize the potential drop in their electro-magnetic winds for proton acceleration to ultra-high energy (see, however, [18]).

in the "bolometric correction" for the CR production rate, and in the bolometric correction, $Q^{z=0}_{\gamma,\mathrm{GRB}}/Q^{z=0}_{\mathrm{MeV,GRB}} > 1$, that should also be applied to the photons.

If GRBs are the sources of UHECRs, then some interesting predictions can be made regarding the spectrum and angular distribution of events at the highest energies [97]. Due to the rapid decrease of d_{eff} with energy, the total number of sources contributing to the flux, $\sim(4\pi/3)d^3_{\mathrm{eff}}\dot{n}_s\Delta t[E,d_{\mathrm{eff}}(E)]$ drops rapidly with energy. This implies that, for $\dot{n}_s \sim 10^{-9}$ Mpc^{-3}yr^{-1} and adopting the estimate of Eq. 3.5 for the deflection angle, only a few sources contribute to the flux above $\sim 3\times10^{20}$ eV. Moreover, the spectrum of these sources should be rather narrow, $\Delta E/E \sim 1$, since the energy dependent time delay $\Delta t(E,d)$ implies that higher (lower) energy particles arrived (will arrive) in the past (future). Testing this prediction, which requires a large number of events detected above $\sim 3\times10^{20}$ eV, may require large exposure, exceeding even that of PAO, which may be provided by space born detectors [81, 87].

3 High Energy Neutrino Astronomy

UHECR sources are likely to be sources of high energy neutrinos. The interaction of high energy protons (nucleons) with radiation or gas, either at or far from the source, leads to production of charged pions, via $p\gamma$ and $pp(n)$ interactions, which decay to produce neutrinos (e.g. $p+\gamma \to n+\pi^+$, $\pi^+ \to \mu^+ + \nu_\mu \to e^+ + \nu_\mu + \bar{\nu}_\mu + \nu_e$). In Sect. 3.1 we estimate the minimum detector size, which is required to detect such neutrinos. In Sect. 3.2 we comment on the importance of the detection of "GZK neutrinos". The prospects for detection of GRB neutrinos, and the possible implications of such detection for the study of GRBs, are discussed in Sect. 3.3. Prospects for the study of fundamental neutrino properties using high energy GRB neutrinos are discussed in Sect. 3.4. For most of the discussion of this section, we adopt the assumption that UHECRs are protons.

3.1 Neutrino Flux Upper Bound, Detector Size, Detectors' Status

The energy production rate, Eq. 3.4, sets an upper bound to the neutrino intensity produced by sources which, as GRBs and AGN jets, are for high-energy nucleons optically thin to $p\gamma$ and $pp(n)$ interactions. For sources of this type, the energy generation rate of neutrinos can not exceed the energy generation rate implied by assuming that all the energy injected as high-energy protons is converted to pions (via $p\gamma$ and $pp(n)$ interactions). Using Eq. 3.4, the resulting upper bound ($\nu_\mu + \bar{\nu}_\mu$, neglecting mixing) is [25, 102]

$$E^2_\nu \Phi_\nu < \frac{1}{4}\xi_Z t_H \frac{c}{4\pi} E^2 \frac{d\dot{n}_0}{dE} \approx 10^{-8}\xi_Z \left(\frac{E^2 d\dot{n}_0/dE}{10^{44}\mathrm{erg/Mpc^3yr}}\right) \mathrm{GeV\,cm^{-2}s^{-1}sr^{-1}}. \tag{3.10}$$

Here t_H is the Hubble time and the $1/4$ factor is due to the fact that charged and neutral pions (which decay to photons) are produced with similar probability, and that muon neutrinos carry roughly half the energy of the decaying pion. In the derivation of Eq. 3.10 we have neglected the redshift energy loss of neutrinos produced at cosmic time $t < t_H$, and implicitly assumed that the cosmic-ray generation rate per unit (comoving) volume is independent of cosmic time. The quantity ξ_Z in Eq. 3.10 has been introduced to describe corrections due to redshift evolution and energy loss. For source evolution following the star-formation rate evolution, $\propto (1+z)^3$, $\xi_z \approx 5$.

The upper bound is compared in Fig. 3.8 with the current experimental limits and with the expected sensitivity of planned neutrino telescopes. The figure indicates that km-scale (i.e. giga-ton-scale) neutrino telescopes are needed to detect the expected extra-Galactic flux in the energy range of ~1 TeV to ~1 PeV, and that much larger effective volume is required to detect the flux at higher energy. The Baikal, AMANDA, and ANTARES optical Cerenkov telescopes have proven that the construction of km-scale neutrino detectors is feasible, and the IceCube detector, the construction of which is well underway, is expected to reach its designed target effective mass of ~1 Gton in 2011.

3.2 GZK Neutrinos

As discussed in Sect. 2.2, protons of energy exceeding the threshold for pion production in interaction with CMB photons, $\sim 5 \times 10^{19}$ eV, lose most of their energy over a time short compared to the age of the universe. If UHECRs are indeed protons of extra-Galactic origin, their energy loss should produce a neutrino intensity similar to the upper bound given by Eq. 3.10. Since most of the pions are produced in interactions with photons of energy corresponding to the Δ-resonance, each of the resulting neutrinos carry approximately 5% of the proton energy. The neutrino background is therefore close to the bound above $\sim 5 \times 10^{18}$ eV, where neutrinos are produced by $\sim 10^{20}$ eV protons. The intensity at lower energies is lower, since protons of lower energy do not lose all their energy over the age of the universe (The GZK intensity in Fig. 3.8 decreases at the highest energies since it was assumed that the maximum energy of protons produced by UHECR sources is 10^{21} eV). The results of detailed calculations of the expected GZK neutrino intensity [43] are in agreement with the qualitative analysis presented above.

The detection of GZK neutrinos will be a milestone in neutrino astronomy. Most important, it will allow one to test the hypothesis that the UHECRs are protons (possibly somewhat heavier nuclei) of extra-Galactic origin (e.g. [75] and references therein). Moreover, measurements of the flux and spectrum would constrain the redshift evolution of the sources. Finally, detection of ultra-high energy neutrinos may allow one to test for modifications of the neutrino interaction cross section due to new physics effects at high (100 TeV) energies [15, 27, 65].

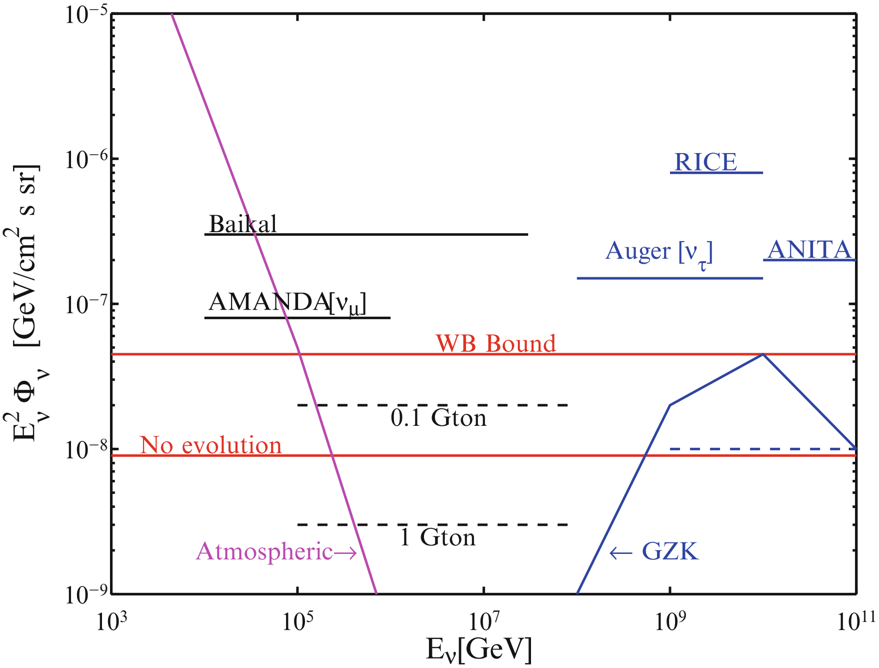


Fig. 3.8 The upper bound imposed by UHECR observations on the extra-Galactic high energy muon neutrino ($\nu_\mu + \bar{\nu}_\mu$) intensity [25,102] (red lower-curve: no evolution of the energy production rate, red upper curve (WB): assuming evolution following star formation rate), compared with the atmospheric muon-neutrino background and with several experimental upper bounds (various solid lines). The theoretical bound does not include the effect of neutrino oscillations. Such oscillations are expected to change the $\nu_e : \nu_\mu : \nu_\tau$ flavor ratio from $1:2:0$ to $1:1:1$ (e.g. [68]), leading to an upper bound which is $\approx 1/2$ that shown in the figure for each flavor. Shown are the muon and all flavor upper bounds of the optical Cerenkov observatories AMANDA [9, 11] and BAIKAL [22], the all flavor upper bounds of the coherent Cerenkov radio detectors RICE [63] and ANITA [46], and the ν_τ upper bound of the PAO [6]. The curve labelled "GZK" shows the muon neutrino intensity (not corrected for oscillations) expected from UHECR proton interactions with micro-wave background photons [30]. Black dashed curves show the expected sensitivity (for few years operation) of 0.1 Gton (ANTARES, http://antares.in2p3.fr/) and 1 Gton (IceCube, http://icecube.wisc.edu/; Km3Net, http://www.km3net.org/home.php) optical Cerenkov detectors. The blue dashed curve is the expected sensitivity of detectors of few 100 Gton (few 100 km^3) effective mass (volume), that may be achieved with proposed radio detectors [12, 26, 27, 66] or with proposed (optical) extensions of IceCube [53]. For a detailed discussion of the current experimental status see [16, 56]

3.3 Neutrinos from GRBs

GRB gamma-rays are believed to be produced within a relativistic expanding wind, a so called "fireball", driven by rapid mass accretion onto a newly formed stellar-mass black hole. It is commonly assumed that electrons are accelerated to high energy in collisionless shocks taking place within the expanding wind, and that synchrotron emission from these shock accelerated electrons produces the observed γ-rays (see [70, 71, 82, 93] for reviews). If protons are present in the wind, as assumed in the fireball model, they would also be accelerated to high energy in the region where electrons are accelerated. If protons are indeed accelerated, then high energy neutrino emission is also expected.

3.3.1 100 TeV Fireball Neutrinos

Protons accelerated in the region where MeV gamma-rays are produced will interact with these photons to produce pions provided that their energy exceeds the threshold for pion production,

$$E_\gamma E \approx 0.2\Gamma^2\,\mathrm{GeV}^2\,. \tag{3.11}$$

Here, E_γ is the observed photon energy. The Γ^2 factor appears since the protons and photon energies in the plasma rest frame (where the particle distributions are roughly isotropic) are smaller than the observed energy by the Lorentz factor Γ of the outflow. For $\Gamma \approx 10^{2.5}$ and $E_\gamma = 1$ MeV, proton energies $\sim 10^{16}$ eV are required to produce pions. Since neutrinos produced by pion decay typically carry 5% of the proton energy, production of $\sim 10^{14}$ eV neutrinos is expected [101].

The fraction of energy lost by protons to pions, f_π, is $f_\pi \approx 0.2$ [48, 101]. Assuming that GRBs generate the observed UHECRs, the expected GRB muon and anti-muon neutrino flux may be estimated using Eq. 3.10 [101, 102],

$$E_\nu^2\Phi_\nu \approx 10^{-8}\frac{f_\pi}{0.2}\left(\frac{E^2 d\dot{n}_0/dE}{10^{44}\mathrm{erg/Mpc^3yr}}\right)\mathrm{GeV\,cm^{-2}s^{-1}sr^{-1}}. \tag{3.12}$$

This neutrino spectrum extends to $\sim 10^{16}$ eV, and is suppressed at higher energy due to energy loss of pions and muons [86, 101, 102] (for the contribution of Kaon decay at high energy see [19]). Equation 3.12 implies a detection rate of ~ 10 neutrino-induced muon events per year (over 4π sr) in a 1 Gton (1 cubic-km) detector [13, 28, 50, 76, 101]. The upper limit on the GRB neutrino emission provided by the AMANDA (~ 0.05 Gton) detector approaches the flux predicted by Eq. 3.12, see Fig. 3.9, and the 1 Gton IceCube detector, which will be completed at the beginning of 2011, will reach a sensitivity that may allow one to test this model's predictions [1].

Since GRB neutrino events are correlated both in time and in direction with gamma-rays, their detection is practically background free. The main background is due to atmospheric neutrinos, which produce neutrino-induced muons, travelling in a direction lying within a cone of opening angle $\Delta\theta$ around some direction, at a rate

$$J^A_{\nu\to\mu} \simeq 4\times 10^{-3}\left(\frac{\Delta\theta}{0.5^o}\right)^2\left(\frac{E}{100\ \mathrm{TeV}}\right)^{-\beta}\ \mathrm{km^{-2}year^{-1}}, \tag{3.13}$$

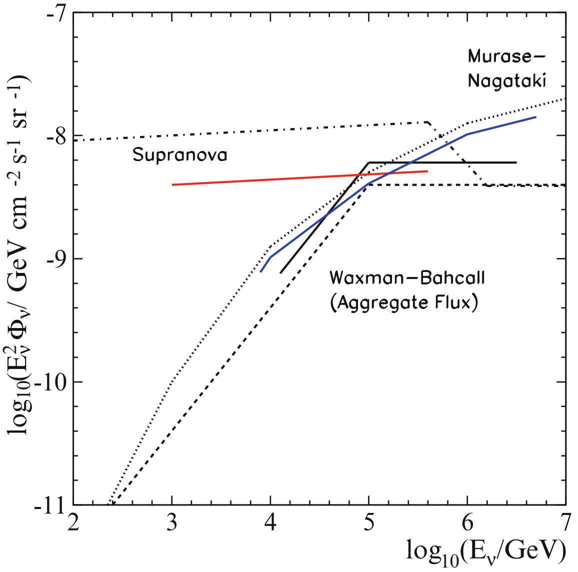


Fig. 3.9 AMANDA flux upper limits (*solid lines*, 90% confidence) for muon neutrino energy spectra predicted by the models of [76, 101] for the ∼100 TeV internal shock fireball neutrinos (Sect. 3.3.1), and for the muon neutrino energy spectrum predicted by Razzaque et al. [84] for the precursor supernova ("supranova") model (Sect. 3.3.2). The upper bounds are compared with the fluxes predicted by the models ([94, 101]– *thick dotted line*, [76]– *thin dotted line*, [84]– *dot-dashed line*) (Adapted from Achterberg et al. [10])

with $\beta = 1.7$ for $E < 100$ TeV and $\beta = 2.5$ for $E > 100$ TeV. At high energies, the neutrino induced muon propagates at nearly the same direction as the incoming neutrino, and km-scale neutrino telescopes will be able to determine the incoming neutrino direction to better than $\sim 0.5^o$. For a known source direction, therefore, the neutrino search is practically background free.

3.3.2 TeV Neutrinos

The 100 TeV neutrinos discussed in the previous sub-section are produced in the same region where GRB γ-rays are produced and should therefore accompany the 10 to 100 s γ-ray emission phase (note, however, that it was pointed out in [78] that if the late, $\sim 10^4$ s, X-ray/UV flares are produced by late internal shocks within the fireball, the emission of 100 TeV neutrinos may be extended to accompany these flares). Their production is a generic prediction of the fireball model: it is a direct consequence of the assumptions that energy is carried from the underlying engine as kinetic energy of protons and that γ-rays are produced by synchrotron emission of shock accelerated particles. Neutrinos may be produced also in other stages of fireball evolution, at energies different than 100 TeV. The production of these neutrinos is dependent on additional model assumptions. We discuss below some examples of $\sim$1 TeV neutrino emission predictions, that depend on the properties of the GRB progenitor. For a discussion of $\sim 10^{18}$ eV neutrino emission during the afterglow phase see [40, 74, 91, 103] and the reviews [70, 71, 94].

The most widely discussed progenitor scenarios for long-duration GRBs involve core collapse of massive stars. In these "collapsar" models, a relativistic jet breaks through the stellar envelope to produce a GRB. For extended or slowly rotating stars, the jet may be unable to break through the envelope. Both penetrating (GRB producing) and "choked" jets can produce a burst of $\sim$10 TeV neutrinos by interaction of accelerated protons with jet photons, while the jet propagates in the envelope [17, 72, 85] (it was pointed out in [17] that neutrino production by kaon decay may dominate over the pion decay contribution, extending the neutrino spectrum to $\sim$20 TeV). The estimated event rates may exceed $\sim 10^2$ events per yr in a km-scale detector, depending on the ratio of non-visible to visible fireballs. A clear detection of non-visible GRBs with neutrinos may be difficult due to the low energy resolution for muon-neutrino events, unless the associated supernova photons are detected.

In the two-step "supranova" model, interaction of the GRB blast wave with the supernova shell can lead to detectable neutrino emission, either through nuclear collisions with the dense supernova shell or through interaction with the intense supernova and backscattered radiation field [42, 49, 84]. As indicated by Fig. 3.9, the upper limits provided by AMANADA on the muon neutrino flux suggest that "supranova"s do not accompany most GRBs.

3.4 Neutrino Physics Prospects

In addition to testing the GRB model for UHECR production and to providing a new handle on the physics of GRB sources, detection of high energy GRB neutrinos may provide information on fundamental neutrino properties [101].

Detection of neutrinos from GRBs could be used to test the simultaneity of neutrino and photon arrival to an accuracy of $\sim$1 s. It is important to emphasize here that since the background level of neutrino telescopes is very low, see Eq. 3.13, the detection of a single neutrino from the direction of a GRB on a time scale of months after the burst would imply an association of the neutrino with the burst and will therefore establish a time of flight delay measurement. Such a measurement will allow one to test for violations of Lorentz invariance (as expected due to quantum gravity effects) [14, 38, 57, 101]), and to test the weak equivalence principle, according to which photons and neutrinos should suffer the same time delay as they pass through a gravitational potential. With 1 s accuracy, a burst at 1 Gpc would reveal a fractional difference in (photon and neutrino) speed of 10^{-17}, and a fractional difference in gravitational time delay of order 10^{-6} (considering the Galactic potential alone). Previous applications of these ideas to supernova 1987A (see [24] for review), yielded much weaker upper limits: of order 10^{-8} and 10^{-2} respectively. Note that at the high neutrino energies under discussion deviations of the propagation speed from that of light due to the finite mass of the neutrino lead to negligible time delay even from propagation over cosmological distances (less than $\sim 10^{-10}$ s at 100 TeV).

High energy neutrinos are expected to be produced in GRBs by the decay of charged pions, which lead to the production of neutrinos with flavor ratio $\Phi_{\nu_e} : \Phi_{\nu_\mu} : \Phi_{\nu_\tau} = 1 : 2 : 0$ (here Φ_{ν_l} stands for the combined flux of ν_l and $\bar{\nu}_l$). Neutrino oscillations then lead to an observed flux ratio on Earth of $\Phi_{\nu_e} : \Phi_{\nu_\mu} : \Phi_{\nu_\tau} = 1 : 1 : 1$ [21, 68] (see, however [58]). Up-going τ's, rather than μ's, would be a distinctive signature of such oscillations. It has furthermore been pointed out that flavor measurements of astrophysical neutrinos may help determining the mixing parameters and mass hierarchy [108], and may possibly enable one to probe new physics [20, 68].

4 Outlook: Open Questions and Multi-messenger Astronomy

The validity of the constraints imposed on the properties of candidate UHECR sources, as summarized in Sect. 2.6, depends on the validity of the inference that the highest energy particles are protons, and on the validity of the assumption that the particles are accelerated by some electromagnetic process, for which the constraints derived in Sect. 2.5 are valid. The inference that the highest energy particles are protons is supported by the HiRes and PAO UHECR spectrum, by the properties of air showers as measured by HiRes, and by the anisotropy hints. However, the shower properties reported by PAO appear to be inconsistent with a pure proton composition at the highest energy (and possibly also with a heavy nuclei composition, see Sect. 2.1). Given this, and the fact that the pp cross section at the high energies under discussion is not well known, the possibility that the highest energy particles are heavy nuclei can not yet be excluded. If the particles are indeed heavy nuclei of charge Z, the minimum power requirement, Eq. 3.7, would be reduced by a factor Z^2, and could possibly be satisfied by local steady sources like AGN (e.g. [80]).

Thus, although we have strong arguments suggesting that UHECR sources are protons produced by transient XG sources, and that the sources should satisfy the constraints given in Sect. 2.6, which point towards GRBs being the likely sources, we are still missing a direct proof of the validity of these conclusions. The open questions that require conclusive answers are:

- *Composition.* Is the composition indeed dominated by protons, or is there a transition back to heavier nuclei at the highest energies? What is the cross section for pp interaction at high, >100 TeV, energy?
- *Galactic- XG transition.* At what energy does the flux become dominated by XG sources?
- *Sources.* Are the sources indeed transient? If so, are the sources GRBs or other transients?

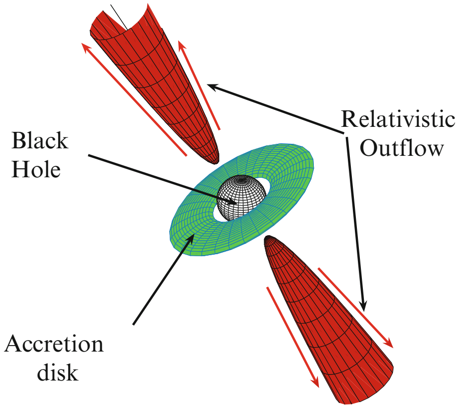


Fig. 3.10 GRBs and AGN are believed to be powered by black holes. The accretion of mass onto the black hole, through an accretion disk, releases large amounts of gravitational energy. If the black hole is rotating rapidly, another energy source becomes available: The rotational energy may be released by slowing the black hole down through interaction with the accretion disk. The energy released drives a jet-like relativistic outflow. The observed radiation is produced as part of the energy carried by the jets is converted, at large distance from the central black hole, to electromagnetic radiation

- *Acceleration.* Are UHECRs accelerated, as suspected, in collisionless (relativistic) shocks? A theory of such shocks based on basic principles is not yet available (e.g. [98] and references therein).

In addition to the open questions listed above, the physics of the candidate UHECR sources is also not well understood. As we have shown, UHECR sources are required to produce very large power and are likely to be driving relativistic outflows, see Eqs. 3.7 and 3.8. These requirements suggest that the sources are powered by the accretion of mass onto black holes, as believed to be the case for GRBs and AGN. GRBs are most likely powered by the accretion of a fraction of a Solar mass on a $\sim$1 s time scale onto a newly born Solar mass black hole [70, 71, 82, 93]. Recent observations strongly suggest that the formation of the black hole is associated with the collapse of the core of a very massive star. AGN are believed to be powered by accretion of mass at a rate of $\sim$1 Solar mass per year onto massive, million to billion Solar mass, black holes residing at the centers of distant galaxies [64]. As illustrated in Fig. 3.10, the gravitational energy released by the accretion of mass onto the black hole is assumed in both cases to drive a relativistic jet, which travels at nearly the speed of light and produces the observed radiation at a large distance away from the central black hole. The models describing the physics responsible for powering these objects, though successful in explaining most observations, are largely phenomenological: the mechanism by which the gravitational energy release is harnessed to drive jets, the mechanism of jet collimation and acceleration, and the process of particle acceleration (and radiation generation), are not understood from basic principles. In particular, the answer to the question of whether the jet energy outflow is predominantly electromagnetic or kinetic, which has major implications to our understanding of the mechanism by which the jets are formed, is not known despite many years of photon observations.

These open questions are unlikely to be answered by UHECR observatories alone. For example, given the uncertainties in the high energy *pp* cross section, it is not clear that studying shower properties would determine the primary composition. The composition could be constrained by an energy dependent anisotropy study (see Sect. 2.3). However, the conclusions of such an analysis would depend on some assumptions regarding the sources [69]. In addition, UHECR observatories are unlikely to identify the sources. Although they may provide a conclusive evidence for the correlation between the distribution of UHECR sources and that of matter in the local universe, and possibly discriminate between steady and transient sources (which may require large exposure that can be provided only by space-born detectors, see Sect. 2.7), this would not determine which type of objects the sources are. It should be emphasized that electromagnetic observations are equally unlikely to resolve the open questions: despite many years of observations we are still lacking direct evidence for acceleration of nuclei in any astrophysical object, and fundamental questions related to the physics of the sources (e.g. the content of relativistic jets) remain unanswered.

Thus, resolving the UHECR puzzles would require a "multi-messenger" approach, combining data from UHECR, γ-ray and neutrino detectors. Neutrino astronomy is likely to play an important role in this context: detection of GZK neutrinos (see Sect. 3.2), combined with accurate measurements of the UHECR flux and spectrum, may allow us to determine the UHECR composition (and constrain the UHE *pp* and neutrino interaction cross sections); detection of high energy neutrino emission from electromagnetically identified sources may allow us to identify the UHECR sources; neutrino observations will provide new constraints on the physics driving the sources, which can not be obtained using electromagnetic observations, since they can escape from regions which are opaque to electromagnetic radiation (see Sect. 3.3 for examples related to GRBs).

Finally, it should be realized that if the UHECR sources are steady, identifying the sources by directly detecting their neutrino emission is highly improbable, due to the fact that the effective area of a $1\,\mathrm{km}^2$ neutrino detector is $\approx 3\times10^{-4}\,\mathrm{km}^2$ at 10^3 TeV, $\approx 10^{-7}$ of the area of $>10^{19}$ eV CR detectors (hence, neutrinos will not be detected unless the neutrino luminosity of the sources exceeds their UHECR luminosity by a factor $>10^3$). In this case, identifying the sources will require a theoretical analysis combining electromagnetic, CR and neutrino data.

References

1. R. Abbasi, Y. Abdou, T. Abu-Zayyad, J. Adams, J.A. Aguilar, M. Ahlers, K. Andeen, J. Auffenberg, X. Bai, M. Baker, et al., Astrophys. J. **710**, 346 (2010). doi:10.1088/0004-637X/710/1/346

2. R.U. Abbasi, T. Abu-Zayyad, M. Al-Seady, M. Allen, J.F. Amman, R.J. Anderson, G. Archbold, K. Belov, J.W. Belz, D.R. Bergman, S.A. Blake, O.A. Brusova, G.W. Burt, C. Cannon, Z. Cao, W. Deng, Y. Fedorova, C.B. Finley, R.C. Gray, W.F. Hanlon, C.M. Hoffman, M.H. Holzscheiter, G. Hughes, P. Hüntemeyer, B.F. Jones, C.C.H. Jui, K. Kim, M.A. Kirn, E.C. Loh, J. Liu, J.P. Lundquist, M.M. Maestas, N. Manago, L.J. Marek, K. Martens, J.A.J. Matthews, J.N. Matthews, S.A. Moore, A. O'Neill, C.A. Painter, L. Perera, K. Reil, R. Riehle, M. Roberts, D. Rodriguez, N. Sasaki, S.R. Schnetzer, L.M. Scott, G. Sinnis, J.D. Smith, P. Sokolsky, C. Song, R.W. Springer, B.T. Stokes, S. Stratton, S.B. Thomas, J.R. Thomas, G.B. Thomson, D. Tupa, A. Zech, X. Zhang, Phys. Rev. Lett. **104**(16), 161101 (2010). doi:10.1103/PhysRevLett.104.161101
3. R.U. Abbasi et al. (HiRes Collaboration), Astrophys. J. **622**, 910 (2005). doi:10.1086/427931
4. R.U. Abbasi et al. (HiRes Collaboration), Phys. Rev. Lett. **100**(10), 101101 (2008). doi:10.1103/PhysRevLett.100.101101
5. J. Abraham et al. (Pierre Auger Collaboration), Astropart. Phys. **29**, 188 (2008). doi:10.1016/j.astropartphys.2008.01.002
6. J. Abraham, P. Abreu, M. Aglietta, C. Aguirre, E.J. Ahn, D. Allard, I. Allekotte, J. Allen, P. Allison, J. Alvarez-Muñiz, et al., Phys. Rev. D **79**(10), 102001 (2009). doi:10.1103/PhysRevD.79.102001
7. J. Abraham et al. (Pierre Auger Collaboration), arXiv:0906.2347 (2009)
8. J. Abraham, P. Abreu, M. Aglietta, E.J. Ahn, D. Allard, I. Allekotte, J. Allen, J. Alvarez-Muñiz, M. Ambrosio, L. Anchordoqui, et al., Phys. Rev. Lett. **104**(9), 091101 (2010). doi:10.1103/PhysRevLett.104.091101
9. A. Achterberg, M. Ackermann, J. Adams, J. Ahrens, K. Andeen, J. Auffenberg, X. Bai, B. Baret, S.W. Barwick, R. Bay, et al., Phys. Rev. D **76**(4), 042008 (2007). doi:10.1103/PhysRevD.76.042008
10. A. Achterberg, M. Ackermann, J. Adams, J. Ahrens, K. Andeen, J. Auffenberg, J.N. Bahcall, X. Bai, B. Baret, S.W. Barwick, et al., Astrophys. J. **674**, 357 (2008). doi:10.1086/524920
11. A. Achterberg, M. Ackermann, J. Adams, J. Ahrens, K. Andeen, J. Auffenberg, X. Bai, B. Baret, S.W. Barwick, R. Bay, et al., Phys. Rev. D **77**(8), 089904 (2008). doi:10.1103/PhysRevD.77.089904
12. P. Allison, J. Beatty, P. Chen, A. Connolly, M. Duvernois, P. Gorham, F. Halzen, K. Hanson, K. Hoffman, A. Karle, J. Kelley, H. Landsman, J. Learned, C. Miki, R. Morse, R. Nichol, C. Rott, L. Ruckman, D. Seckel, G. Varner, D. Williams, Nucl. Instrum. Methods Phys. Res. A **604**, 64 (2009). doi:10.1016/j.nima.2009.03.031
13. J. Alvarez-Muñiz, F. Halzen, D.W. Hooper, Phys. Rev. D **62**(9), 093015 (2000). doi:10.1103/PhysRevD.62.093015
14. G. Amelino-Camelia, J. Ellis, N.E. Mavromatos, D.V. Nanopoulos, S. Sarkar, Nature **393**, 763 (1998). doi:10.1038/31647
15. L.A. Anchordoqui, J.L. Feng, H. Goldberg, Phys. Rev. Lett. **96**(2), 021101 (2006). doi:10.1103/PhysRevLett.96.021101
16. L.A. Anchordoqui, T. Montaruli, Annu. Rev. Nucl. Part. Sci. **60**, 129 (2010)
17. S. Ando, J.F. Beacom, Phys. Rev. Lett. **95**(6), 061103 (2005). doi:10.1103/PhysRevLett.95.061103
18. J. Arons, Astrophys. J. **589**, 871 (2003). doi:10.1086/374776
19. K. Asano, S. Nagataki, Astrophys. J. **640**, L9 (2006). doi:10.1086/503291
20. H. Athar, M. Jeżabek, O. Yasuda, Phys. Rev. D **62**(10), 103007 (2000). doi:10.1103/PhysRevD.62.103007
21. H. Athar, C.S. Kim, J. Lee, Mod. Phys. Lett. A **21**, 1049 (2006). doi:10.1142/S021773230602038X
22. A.V. Avrorin, V.M. Aynutdinov, V.A. Balkanov, I.A. Belolaptikov, D.Y. Bogorodsky, N.M. Budnev, R. Wischnewski, O.N. Gaponenko, K.V. Golubkov, O.A. Gres, T.I. Gres, O.G. Grishin, I.A. Danilchenko, Z. Dzhilkibaev, G.V. Domogatsky, A.A. Doroshenko, A.N. D'Yachok, V.A. Zhukov, A.M. Klabukov, A.I. Klimov, K.V. Konishchev, A.A. Kochanov,

A.P. Koshechkin, L.A. Kuzmichev, V.F. Kulepov, D.A. Kuleshov, E. Middell, M.B. Milenin, R.R. Mirgazov, S.P. Mikheev, E.A. Osipova, A.I. Panfilov, L.V. Pan'kov, G.L. Pan'kov, D.P. Petukhov, E.N. Pliskovsky, V.A. Poleshchuk, E.G. Popova, P.G. Pokhil, V.V. Prosin, M.I. Rozanov, V.Y. Rubtsov, O.V. Suvorova, B.A. Tarashchansky, S.V. Fialkovsky, B.A. Shaibonov, A.A. Sheifler, A.V. Shirokov, C. Spiering, I.V. Yashin, Astron. Lett. **35**, 651 (2009). doi:10.1134/S1063773709100016

23. W.I. Axford, Astrophys. J. Suppl. **90**, 937 (1994). doi:10.1086/191928
24. J.N. Bahcall, *Neutrino Astrophysics* (Cambridge University Press, Cambridge, 1989)
25. J. Bahcall, E. Waxman, Phys. Rev. D **64**(2), 023002 (2001). doi:10.1103/PhysRevD.64.023002
26. S.W. Barwick, J. Phys. Conf. Ser. **60**, 276 (2007). doi:10.1088/1742-6596/60/1/060
27. S.W. Barwick, Nucl. Instrum. Methods Phys. Res. A **602**, 279 (2009). doi:10.1016/j.nima.2008.12.039
28. J.K. Becker, M. Stamatikos, F. Halzen, W. Rhode, Astropart. Phys. **25**, 118 (2006). doi:10.1016/j.astropartphys.2005.12.006
29. V. Berezinsky, Adv. Space Res. **41**, 2071 (2008). doi:10.1016/j.asr.2007.02.065
30. V.S. Beresinsky, G.T. Zatsepin, Phys. Lett. B **28**, 423 (1969). doi:10.1016/0370-2693(69)90341-4
31. K. Bernlöhr, W. Hofmann, G. Leffers, V. Matheis, M. Panter, R. Zink, Astropart. Phys. **8**, 253 (1998). doi:10.1016/S0927-6505(98)00002-4
32. P. Bhattacharjee, Phys. Rep. **327**, 109 (2000). doi:10.1016/S0370-1573(99)00101-5
33. D.J. Bird, S.C. Corbato, H.Y. Dai, B.R. Dawson, J.W. Elbert, B.L. Emerson, K.D. Green, M.A. Huang, D.B. Kieda, M. Luo, S. Ko, C.G. Larsen, E.C. Loh, M.H. Salamon, J.D. Smith, P. Sokolsky, P. Sommers, J.K.K. Tang, S.B. Thomas, Astrophys. J. **424**, 491 (1994). doi:10.1086/173906
34. R. Blandford, D. Eichler, Phys. Rep. **154**, 1 (1987). doi:10.1016/0370-1573(87)90134-7
35. J. Bluemer, for the Pierre Auger Collaboration, arXiv:0807.4871 (2008)
36. T.H. Burnett, S. Dake, J.H. Derrickson, W.F. Fountain, M. Fuki, J.C. Gregory, T. Hayashi, R. Holynski, J. Iwai, W.V. Jones, A. Jurak, J.J. Lord, O. Miyamura, H. Oda, T. Ogata, T.A. Parnell, F.E. Roberts, S. Strausz, T. Tabuki, Y. Takahashi, T. Tominaga, J.W. Watts, J.P. Wefel, B. Wilczynska, H. Wilczynski, R.J. Wilkes, W. Wolter, B. Wosiek, The JACEE collaboration. Astrophys. J. l **349**, L25 (1990). doi:10.1086/185642
37. Y. Butt, Nature **460**, 701 (2009). doi:10.1038/nature08127
38. S. Coleman, S.L. Glashow, Phys. Rev. D **59**(11), 116008 (1999). doi:10.1103/PhysRevD.59.116008
39. A. Cuoco, S. Hannestad, T. Haugbølle, M. Kachelrieß, P.D. Serpico, Astrophys. J. **702**, 825 (2009). doi:10.1088/0004-637X/702/2/825
40. Z.G. Dai, T. Lu, Astrophys. J. **551**, 249 (2001). doi:10.1086/320056
41. B.R. Dawson, R. Meyhandan, K.M. Simpson, Astropart. Phys. **9**, 331 (1998). doi:10.1016/S0927-6505(98)00031-0
42. C.D. Dermer, A. Atoyan, Phys. Rev. Lett. **91**(7), 071102 (2003). doi:10.1103/PhysRevLett.91.071102
43. R. Engel, D. Seckel, T. Stanev, Phys. Rev. D **64**(9), 093010 (2001). doi:10.1103/PhysRevD.64.093010
44. G.R. Farrar, A. Gruzinov, Astrophys. J. **693**, 329 (2009). doi:10.1088/0004-637X/693/1/329
45. T.K. Gaisser, T. Stanev, S. Tilav, S.C. Corbato, H.Y. Dai, B.R. Dawson, J.W. Elbert, B. Emerson, D.B. Kieda, M. Luo, S. Ko, C. Larsen, E.C. Loh, M.H. Salamon, J.D. Smith, P. Sokolsky, P. Sommers, J. Tang, S.B. Thomas, D.J. Bird, Phys. Rev. D **47**, 1919 (1993). doi:10.1103/PhysRevD.47.1919
46. P.W. Gorham, P. Allison, S.W. Barwick, J.J. Beatty, D.Z. Besson, W.R. Binns, C. Chen, P. Chen, J.M. Clem, A. Connolly, P.F. Dowkontt, M.A. Duvernois, R.C. Field, D. Goldstein, A. Goodhue, C. Hast, C.L. Hebert, S. Hoover, M.H. Israel, J. Kowalski, J.G. Learned, K.M. Liewer, J.T. Link, E. Lusczek, S. Matsuno, B.C. Mercurio, C. Miki, P. Miočinović, J. Nam, C.J. Naudet, J. Ng, R.J. Nichol, K. Palladino, K. Reil, A. Romero-Wolf, M. Rosen,

L. Ruckman, D. Saltzberg, D. Seckel, G.S. Varner, D. Walz, Y. Wang, F. Wu, Phys. Rev. Lett. **103**(5), 051103 (2009). doi:10.1103/PhysRevLett.103.051103
47. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966). doi:10.1103/PhysRevLett.16.748
48. D. Guetta, M. Spada, E. Waxman, Astrophys. J. **559**, 101 (2001). doi:10.1086/322481
49. D. Guetta, J. Granot, Phys. Rev. Lett. **90**(20), 201103 (2003). doi:10.1103/PhysRevLett.90.201103
50. D. Guetta, D. Hooper, J. Alvarez-MuñIz, F. Halzen, E. Reuveni, Astropart. Phys. **20**, 429 (2004). doi:10.1016/S0927-6505(03)00211-1
51. D. Guetta, T. Piran, E. Waxman, Astrophys. J. **619**, 412 (2005). doi:10.1086/423125
52. F. Halzen, D. Hooper, Rep. Prog. Phys. **65**, 1025 (2002). doi:10.1088/0034-4885/65/7/201
53. F. Halzen, D. Hooper, J. Cosmol. Astropart. Phys. **1**, 2 (2004). doi:10.1088/1475-7516/2004/01/002
54. D. Harari, S. Mollerach, E. Roulet, Mon. Not. R. Astron. Soc. **394**, 916 (2009). doi:10.1111/j.1365-2966.2008.14327.x
55. A.M. Hillas, Ann. Rev. Astron. Astrophys. **22**, 425 (1984). doi:10.1146/annurev.aa.22.090184.002233
56. K.D. Hoffman, New J. Phys. **11**(5), 055006 (2009). doi:10.1088/1367-2630/11/5/055006
57. U. Jacob, T. Piran, Nat. Phys. **3**, 87 (2007). doi:10.1038/nphys506
58. T. Kashti, E. Waxman, Phys. Rev. Lett. **95**(18), 181101 (2005). doi:10.1103/PhysRevLett.95.181101
59. T. Kashti, E. Waxman, J. Cosmol. Astropart. Phys. **5**, 6 (2008). doi:10.1088/1475-7516/2008/05/006
60. B. Katz, R. Budnik, E. Waxman, J. Cosmol. Astropart. Phys. **3**, 20 (2009). doi:10.1088/1475-7516/2009/03/020
61. H.B.J. Koers, P. Tinyakov, J. Cosmol. Astropart. Phys. **4**, 3 (2009). doi:10.1088/1475-7516/2009/04/003
62. K. Kotera, M. Lemoine, Phys. Rev. D **77**(12), 123003 (2008). doi:10.1103/PhysRevD.77.123003
63. I. Kravchenko, C. Cooley, S. Hussain, D. Seckel, P. Wahrlich, J. Adams, S. Churchwell, P. Harris, S. Seunarine, A. Bean, D. Besson, S. Graham, S. Holt, D. Marfatia, D. McKay, J. Meyers, J. Ralston, R. Schiel, H. Swift, J. Ledford, K. Ratzlaff, Phys. Rev. D **73**(8), 082002 (2006). doi:10.1103/PhysRevD.73.082002
64. J.H. Krolik, *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment* (Princeton University Press, Princeton, 1998)
65. A. Kusenko, T.J. Weiler, Phys. Rev. Lett. **88**(16), 161101 (2002). doi:10.1103/PhysRevLett.88.161101
66. H. Landsman, L. Ruckman, G.S. Varner, in *International Cosmic Ray Conference*, vol. 4, ed. by R. Caballero et al. (Universidad Nacional Autnoma de Mexico, Mexico, 2008), pp. 827–830
67. T. Le, C.D. Dermer, Astrophys. J. **661**, 394 (2007). doi:10.1086/513460
68. J.G. Learned, S. Pakvasa, Astropart. Phys. **3**, 267 (1995). doi:10.1016/0927-6505(94)00043-3
69. M. Lemoine, E. Waxman, J. Cosmol. Astropart. P. **11**, 9 (2009). doi:10.1088/1475-7516/2009/11/009
70. P. Mészáros, Ann. Rev. Astron. Astrophys. **40**, 137 (2002). doi:10.1146/annurev.astro.40.060401.093821
71. P. Mészáros, Rep. Prog. Phys. **69**, 2259 (2006). doi:10.1088/0034-4885/69/8/R01
72. P. Mészáros, E. Waxman, Phys. Rev. Lett. **87**(17), 171102 (2001). doi:10.1103/PhysRevLett.87.171102
73. M. Milgrom, V. Usov, Astrophys. J **449**, L37+ (1995). doi:10.1086/309633
74. K. Murase, Phys. Rev. D **76**(12), 123001 (2007). doi:10.1103/PhysRevD.76.123001
75. K. Murase, J.F. Beacom, Phys. Rev. D **81**(12), 123001 (2010). doi:10.1103/PhysRevD.81.123001
76. K. Murase, S. Nagataki, Phys. Rev. D **73**(6), 063002 (2006). doi:10.1103/PhysRevD.73.063002

77. K. Murase, K. Ioka, S. Nagataki, T. Nakamura, Phys. Rev. D **78**(2), 023005 (2008). doi:10.1103/PhysRevD.78.023005
78. K. Murase, S. Nagataki, Phys. Rev. Lett. **97**(5), 051101 (2006). doi:10.1103/PhysRevLett.97.051101
79. M. Nagano, A.A. Watson, Rev. Mod. Phys. **72**, 689 (2000). doi:10.1103/RevModPhys.72.689
80. A. Pe'Er, K. Murase, P. Mészáros, Phys. Rev. D **80**(12), 123018 (2009). doi:10.1103/PhysRevD.80.123018
81. A. Petrolini, Nucl. Instrum. Methods Phys. Res. A **588**, 201 (2008). doi:10.1016/j.nima.2008.01.040
82. T. Piran, Rev. Mod. Phys. **76**, 1143 (2004). doi:10.1103/RevModPhys.76.1143
83. J.P. Rachen, P.L. Biermann, Astron. Astrophys. **272**, 161 (1993)
84. S. Razzaque, P. Mészáros, E. Waxman, Phys. Rev. Lett. **90**(24), 241103 (2003). doi:10.1103/PhysRevLett.90.241103
85. S. Razzaque, P. Mészáros, E. Waxman, Phys. Rev. D **69**(2), 023001 (2004). doi:10.1103/PhysRevD.69.023001
86. J.P. Rachen, P. Mészáros, Phys. Rev. D **58**(12), 123005 (1998). doi:10.1103/PhysRevD.58.123005
87. Y. Takahashi, the JEM-EUSO Collaboration, New J. Phys. **11**(6), 065009 (2009). doi:10.1088/1367-2630/11/6/065009
88. H. Takami, T. Nishimichi, K. Yahata, K. Sato, J. Cosmol. Astropart. Phys. **6**, 31 (2009). doi:10.1088/1475-7516/2009/06/031
89. R. Ulrich, R. Engel, S. Müller, F. Schüssler, M. Unger, Nucl. Phys. B Proc. Suppl. **196**, 335 (2009). doi:10.1016/j.nuclphysbps.2009.09.064
90. M. Vietri, Astrophys. J. **453**, 883 (1995). doi:10.1086/176448
91. M. Vietri, Phys. Rev. Lett. **80**, 3690 (1998). doi:10.1103/PhysRevLett.80.3690
92. D. Wanderman, T. Piran, Mon. Not. R. Astron. Soc. **406**, 1944 (2010). doi:10.1111/j.1365-2966.2010.16787.x
93. E. Waxman, in *Supernovae and Gamma-Ray Bursters*, ed. by K. Weiler. Lecture Notes in Physics, vol. 598 (Springer, Berlin, 2003), pp. 393–418
94. E. Waxman, Nucl. Phys. B Proc. Suppl. **118**, 353 (2003)
95. E. Waxman, Astrophys. J. **606**, 988 (2004). doi:10.1086/383116
96. E. Waxman, ArXiv:1010.5007 (2010)
97. E. Waxman, J. Miralda-Escude, Astrophys. J. **472**, L89+ (1996). doi:10.1086/310367
98. E. Waxman, Plasma Phys. Contr. F. **48**, B137 (2006). doi:10.1088/0741-3335/48/12B/S14
99. E. Waxman, Phys. Scripta Vol T **121**, 147 (2005). doi:10.1088/0031-8949/2005/T121/022
100. E. Waxman, Phys. Rev. Lett. **75**, 386 (1995). doi:10.1103/PhysRevLett.75.386
101. E. Waxman, J. Bahcall, Phys. Rev. Lett. **78**, 2292 (1997). doi:10.1103/PhysRevLett.78.2292
102. E. Waxman, J. Bahcall, Phys. Rev. D **59**(2), 023002 (1999). doi:10.1103/PhysRevD.59.023002
103. E. Waxman, J.N. Bahcall, Astrophys. J. **541**, 707 (2000). doi:10.1086/309462
104. E. Waxman, K.B. Fisher, T. Piran, Astrophys. J. **483**, 1 (1997). doi:10.1086/304205
105. E. Waxman, A. Loeb, J. Cosmol. Astropart. P. **8**, 26 (2009). doi:10.1088/1475-7516/2009/08/026
106. S.D. Wick, C.D. Dermer, A. Atoyan, Nucl. Phys. B Proc. Suppl. **134**, 81 (2004). doi:10.1016/j.nuclphysbps.2004.08.013
107. G. Wilk, Z. Wlodarczyk, ArXiv e-prints (2010)
108. W. Winter, Phys. Rev. D **74**(3), 033015 (2006). doi:10.1103/PhysRevD.74.033015
109. G.T. Zatsepin, V.A. Kuz'min, Sov. J. Exp. Theor. Phy. Lett. **4**, 78 (1966)

Chapter 4
Interdisciplinary Aspects of High-Energy Astrophysics

Günter Sigl

Abstract Modern astrophysics, especially at GeV energy scales and above is a typical example where several disciplines meet: the location and distribution of the sources is the domain of astronomy. At distances corresponding to significant redshift cosmological aspects such as the expansion history come into play. Finally, the emission mechanisms and subsequent propagation of produced high energy particles is at least partly the domain of particle physics, in particular if new phenomena beyond the Standard Model are probed that require baselines and/or energies unattained in the laboratory. In this contribution we focus on three examples: highest energy cosmic rays, tests of the Lorentz symmetry and the search for new light photon-like states in the spectra of active galaxies.

Keywords Acceleration of particles • Astroparticle physics • Cosmic rays • Relativistic processes • Elementary particles

1 Introduction

High energy astrophysics is nowadays a very interdisciplinary research field which either uses input from or provides new output to other fields including astronomy, cosmology, particle physics and even philosophy and (astro)biology. Examples of where this becomes especially obvious include the use of active galactic nuclei to probe the formation of structure at very high redshift of order ten, high energy cosmic rays as probes for the annihilation or decay of dark matter and the use of "standard candles" (see Chap. 2) such as exploding white dwarfs and (more recently) gamma-ray bursts to probe the expansion history of the Universe.

G. Sigl (✉)
II. Institut für theoretische Physik, Universität Hamburg, Luruper Chaussee 149,
D-22761 Hamburg, Germany
e-mail: guenter.sigl@desy.de

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_4,

A particular problem that sometimes occurs at these intersections arises from the different languages spoken by the different communities. In general, however, a lot of progress has been made in that respect. This is the case in particular in astroparticle physics, a still young but meanwhile well established research discipline in its own right. This can be seen not least from the fact that funding agencies in most countries have developed programs and instruments aiming specifically at this field.

The present paper can naturally cover at most a tiny fraction of interesting examples for such interfaces between neighboring research fields. We specifically focus on three topics at the interface between astronomy, high energy astrophysics and particle physics: first, ultra-high energy cosmic rays, traditionally understood as particles with energies above 10^{18} eV, have been observed with energies up to a few times 10^{20} eV, which is a macroscopic energy of about 50 J, presumably of just one elementary particle. Therefore, very likely, the sources of these ultra-energetic particles have to be exceptionally powerful and visible in other wavelengths and channels. The search of these sources has thus a strong relation to astronomy.

Second, the macroscopic energies of these particles make them natural test beams for particle physics at energies that cannot be achieved in the laboratory in the foreseeable future. In particular, tiny violations of fundamental symmetries of Nature, such as the Lorentz symmetry, may become magnified at large energies. We are still lacking a description of gravity that is consistent with quantum mechanics and the way gravity unifies with the electromagnetic, weak and strong interactions may only manifest itself at energies approaching the Planck scale. In this case, high energy astrophysics may be an indispensable tool for the phenomenology of quantum gravity.

Finally, at the opposite, low energy end, new physics may also exist in the form of very light particles that may morph into photons and vice versa. The strongest constraints on such possibilities that are often motivated by models of fundamental physics such as string theory and loop quantum gravity often come from astrophysical and cosmological observations which offer the largest baselines and the highest energies.

2 Astronomy with the Highest Energy Particles of Nature?

The research field of ultra-high energy cosmic rays started in 1938 when Pierre Auger proved the existence of extensive air showers (EAS) caused by primary particles with energies above 10^{15} eV by simultaneously observing the arrival of secondary particles in Geiger counters many meters apart [21]. Since that time, ultra-high energy cosmic rays (UHECRs) have challenged the imagination of physicists and astrophysicists alike. The first cosmic ray with energy above 10^{20} eV was discovered by John Lindsley in 1963 at the Volcano Ranch Observatory [57]. The record holder is probably still the famous "Fly's Eye event" of $\simeq 3 \times 10^{20}$ eV [22] and quickly, scientists were looking for astronomical sources [32]. Around the same time, the Akeno Giant Air Shower Array (AGASA) caused excitement because it observed an UHECR spectrum continuing seemingly as a power law around 10^{20} eV [46]. This was contrary to expectations because the famous Greisen-Zatsepin-Kuzmin (GZK) effect [44] predicts that nucleons loose their energy within about 20 Mpc above a threshold of $\simeq 6 \times 10^{19}$ eV [81] due to pion production on the cosmic microwave background which is a relic of the early Universe. As long as we do not live in a strong over-density of UHECR sources, this would predict a strong suppression of the UHECR flux above that threshold, often somewhat misleadingly called the "GZK cutoff". Meanwhile, a flux suppression consistent with the GZK effect has been observed by the more recent High Resolution Fly's Eye [2] and Pierre Auger [6] instruments and it is likely that the AGASA spectrum was due to an overestimate of the UHECR energies.

These more recent, higher statistics data, however, raised other, no less interesting questions: For the first time, the Pierre Auger Observatory which observes the Southern hemisphere from Argentina has accumulated enough statistics at the highest energies to see signs of anisotropy: a significant correlation with the 12th edition of the Véron-Cetty and Véron catalog of nearby AGNs was observed for events with energies above 56 EeV [5]. This is very suggestive because it is also the energy scale above which the GZK effect limits the range of primary cosmic rays to $\sim$50 Mpc. This does not necessarily mean that these objects represent the sources, but it suggests that the real UHECR sources follow an anisotropic distribution that is similar to nearby AGNs. This may not be surprising if the sources are astrophysical accelerators which follow the local large scale structure. Unfortunately, with accumulation of more data, these correlations have weakened [9]. The fraction of events above 55 EeV correlating with the Véron-Cetty and Véron Catalog has came down from $69^{+11}_{-13}\%$ to $38^{+7}_{-6}\%$ compared to 21% expected for isotropy. If one divides the sky distribution into a component correlating, for example, with the 2MASS redshift survey and an isotropic component, this corresponds to a relatively large isotropic fraction of 60–70% [9]. Still, an excess of correlations is seen with 2MASS redshift survey at 95% confidence level. On the other hand, in the Northern hemisphere, the HiRes experiment has not seen any correlations [1].

The nature and location of UHECR sources is thus still an open question in which general theoretical considerations play a significant role. Accelerating particles of charge eZ to an energy $E_{\rm max}$ requires an induction $\mathscr{E} \gtrsim E_{\rm max}/(eZ)$. With $Z_0 \simeq 100\,\Omega$ the vacuum impedance, this requires dissipation of a minimal power of [23, 58]

$$L_{\rm min} \simeq \frac{\mathscr{E}^2}{Z_0} \simeq 10^{45}\, Z^{-2} \left(\frac{E_{\rm max}}{10^{20}\,{\rm eV}}\right)^2 {\rm erg\,s^{-1}}\,. \tag{4.1}$$

When expressing the square of the product of the magnetic field in an accelerator with its size in terms of a luminosity, this condition can be expressed in terms of the Hillas-criterium [47] which states that the gyro-radius of a charged particle at the maximal acceleration energy must fit within the accelerator. Equation 4.1 suggests that the power requirements are considerably relaxed for heavier nuclei which is easy to understand because an estimate solely based on motion of charged particles in magnetic fields can only depend on their rigidity E/Z. However, the

Hillas criterion and Eq. 4.1 are necessary but in general not sufficient since they do not take into account energy loss processes within the source. Extensions of the conditions on UHECR sources that include energy-loss processes have recently been discussed in Ref. [71]. An interesting argument linking UHECR sources to their luminosity at radio frequencies has been put forward by Hardcastle [45]. He concludes that if UHECRs are predominantly protons, then very few sources should contribute to the observed flux. These sources should be easy to identify in the radio and their UHECR spectrum should cut off steeply at the observed highest energies. In contrast, if the composition is heavy at the highest energies then many radio galaxies could contribute to the UHECR flux but due to the much stronger deflection only the nearby radio galaxy Centaurus A may be identifiable.

In fact, the Pierre Auger data reveal a clustering of super-GZK events towards the direction of Centaurus A (NGC 5128) [9, 65], whereas other directions on the sky with an overdensity of potential UHECR accelerators such as the Virgo cluster containing the prominent radio galaxy M87 show an apparent deficit in such events [42]. This is somewhat surprising since, although Cen A is the closest radio galaxy and the third-strongest radio source in the sky, it is a relatively weak elliptical radio galaxy (see, e.g., [72]), making it difficult to reach the required UHECR energies. However, one should note that the UHECR events observed towards Cen A could at least partly originate from sources within the Centaurus galaxy cluster which is located just behind Cen A and is itself part of the Hydra-Centaurus supercluster. In any case, due to its closeness, Cen A has been observed in many channels. For example, its lobes have been detected in 200 MeV gamma-rays by Fermi LAT [36], and its core was observed by Fermi LAT [35]. These observations and its potential role as a major local UHECR accelerator has lead to many multi-messenger model building efforts for Cen A [52,72]. As an example, in Ref. [52] it was pointed out that proton acceleration in the jet of Cen A is hard to reconcile with Cen A observations in TeV gamma-rays by HESS [11] if gamma-rays are produced by proton-proton interactions. Instead, $\mathrm{p}-\gamma$ interactions in the core are consistent with these observations.

We note in passing that another potential UHECR source are gamma-ray bursts (GRBs) (see, e.g., [30]). Although GRBs individually have more than adequate power to achieve the required maximal acceleration energies, they may be disfavored in terms of local power density compared to an UHECR origin in AGNs and radio galaxies.

Another interesting new question concerns the chemical composition of highest energy cosmic rays: the depth in the atmosphere where particle density in the giant air showers observed by the Pierre Auger Observatory is maximal, and in particular the fluctuations of the depth of shower maximum from event to event, when compared with air shower simulations, point towards a heavy composition for energies $10^{19}\,\mathrm{eV} \lesssim E \lesssim 4\times10^{19}\,\mathrm{eV}$. At higher energies statistics is insufficient to determine the variance of the depth of shower maximum [8]. On the other hand, HiRes observations are consistent with a light composition above $\simeq 1.6\times10^{18}\,\mathrm{eV}$ and up to $\simeq 5\times10^{19}\,\mathrm{eV}$ above which statistics is insufficient to determine composition [3]. This could indicate that statistics is still too limited to draw firm conclusions or that the Northern and Southern hemispheres are significantly different in terms of UHECR composition. In addition, there are significant uncertainties in hadronic cross sections, multiplicities and inelasticities that can influence predicted air shower shapes and none of the existing hadronic interaction models consistently describes the shower depth and muon data of the Pierre Auger experiment [84, 85]. Note that the center of mass energy for a UHECR interacting in the atmosphere reaches a PeV $= 10^{15}$ eV, which is still a factor of a few hundred higher than the highest energies reached in the laboratory, at the Large Hadron Collider (LHC) at CERN. It is therefore not excluded that the true chemical composition is light on both hemispheres and the UHECR data teaches us something fundamental about hadronic interactions at energies unattainable in the laboratory.

The question of chemical composition is linked to other observables such as the UHECR spectrum. Unfortunately, the current statistics is still insufficient to gain significant information on the chemical composition from the observed spectrum. The flux suppression observed above $\simeq 4 \times 10^{19}$ eV is qualitatively consistent with either proton or nuclei heavier than carbon up to iron nuclei [16, 18, 19]. In the latter case, the main energy loss process responsible for the "cut-off" is photo-disintegration on the CMB and infrared backgrounds. It should be noted, however, that the observed flux suppression could also be due to the intrinsic maximal acceleration energies attained in the sources, although it would possibly be somewhat of a coincidence that this energy should be close to the GZK energy.

The UHECR chemical composition can in principle also be tested independently with the flux of secondary cosmogenic neutrinos [12,15,18,55] and photons [40,49]: These secondaries are essentially produced by pion production on the constituent nucleons of a nucleus with a given atomic number A. Therefore, if the maximal acceleration energy $E_{\max}$ is not much larger than 10^{21} eV then for mass numbers A approaching iron group nuclei, the energy of the constituent nucleons will be below the GZK threshold for pion production on the CMB and secondary gamma-ray and neutrino production can only occur by interactions with the infrared background, with a rate suppressed by the relative target photon number density which is a factor of a few hundred. As a result, the cosmogenic neutrino and photon fluxes depend strongly on injection spectrum, maximal acceleration energy and chemical composition, but it may not always be easy to break the resulting degeneracies.

Finally, the question of chemical composition of UHECRs is strongly linked with the question of deflection angles in cosmic magnetic fields. In a field with rms strength B and coherence length l_c the rms deflection angle of a cosmic ray of energy E and charge Ze traveling a distance d is given by Waxman and Miralda-Escude [86]

$$\begin{aligned}\theta(E,d) &\simeq \frac{(2dl_c/9)^{1/2}}{r_g}\\ &\simeq 0.8^\circ Z\left(\frac{E}{10^{20}\,\mathrm{eV}}\right)^{-1}\left(\frac{d}{10\,\mathrm{Mpc}}\right)^{1/2}\left(\frac{l_c}{1\,\mathrm{Mpc}}\right)^{1/2}\left(\frac{B}{10^{-9}\,\mathrm{G}}\right),\end{aligned} \tag{4.2}$$

where $r_L = E/(ZeB)$ is the Larmor radius. For an order of magnitude estimate for the deflection angles in the Galactic magnetic field we use $l_c \sim 100\,\mathrm{pc}$, $d \sim 10\,\mathrm{kpc}$, $B \sim 3\,\mu\mathrm{G}$ which gives $\theta(E) \sim 1^\circ Z(10^{20}\,\mathrm{eV}/E)$. Thus, protons around the GZK cut-off, $E \sim 60\,\mathrm{EeV}$, will be deflected by a few degrees or less, whereas iron nuclei can be deflected by several dozens of degrees. This immediately raises the issue that the Galactic magnetic fields are likely to destroy any possible correlation with the local large scale structure in case of a heavy composition. Detailed numerical simulations demonstrate that the relatively large deflections of a heavy composition can considerably distort the images of individual sources and even of the local large scale structure as a whole [41].

Large scale extra-galactic magnetic fields (EGMF) are much less well known than Galactic magnetic fields [56]. One reason is that one of the major detection methods for the EGMF, the Faraday rotation of the polarization of radio emission from a distant source which is a measure of the line of sight integral of the plasma density times the parallel magnetic field component, is only sensitive to fields at a given location stronger than $\sim 0.1\,\mu\mathrm{G}$. Fields below that strength require much higher statistics data than currently available, but still have a strong effect on UHECR deflection, as obvious from Eq. 4.2. As a statistical average over the sky, an all pervading EGMF is constrained to be $\lesssim 3 \times 10^{-7}\,(l_c/\mathrm{Mpc})^{1/2}\,\mathrm{G}$ [24]. Assuming an EGMF whose flux is frozen and follows the large scale structure gives the more stringent limit $B \lesssim 10^{-9} - 10^{-8}\,\mathrm{G}$, but the fields in the sheets and filaments can in this case be up to a micro Gauss. This is also the scale which is routinely observed in galaxy clusters which are the largest virialized structures in the Universe. Beyond galaxy cluster scales at best only hints exist on the EGMF properties, for example in the Hercules and Perseus-Pisces superclusters [88]. It is expected, however, that in the future large scale radio telescopes such as Lofar and SKA will improve observational information on the EGMF in the large scale structure dramatically. We note in this context that the EGMF in the voids is expected to be very weak and uncontaminated by astrophysical processes. This makes voids excellent probes of relic seed magnetic fields from the early Universe [43]. It is exciting that the non-observation at GeV energies by Fermi of certain distant blazars that were seen at TeV energies by HESS suggests a *lower limit* $E \gtrsim 3 \times 10^{-16}\,\mathrm{G}$ on the EGMF in the voids [66]. This is because the TeV gamma-rays seen by HESS would initiate electromagnetic cascades that should be detectable by Fermi unless an EGMF of that strength deflects these cascades into a diffuse halo around the source whose flux is then below the Fermi sensitivity. However, void fields at that level are not relevant for UHECR propagation.

As long as better observational information on the EGMF is not available, one way of proceeding is to build models of the EGMF using large scale structure simulations. Two major techniques for doing this are a magnetohydrodynamic version of a constrained smooth particle hydrodynamics code [31] and Eulerian grid-based hydro+n-body codes [78]. The magnetic fields are followed passively and are seeded either uniformly or around cosmic shocks through the Biermann battery mechanism. The normalisation is then constrained by the largest fields

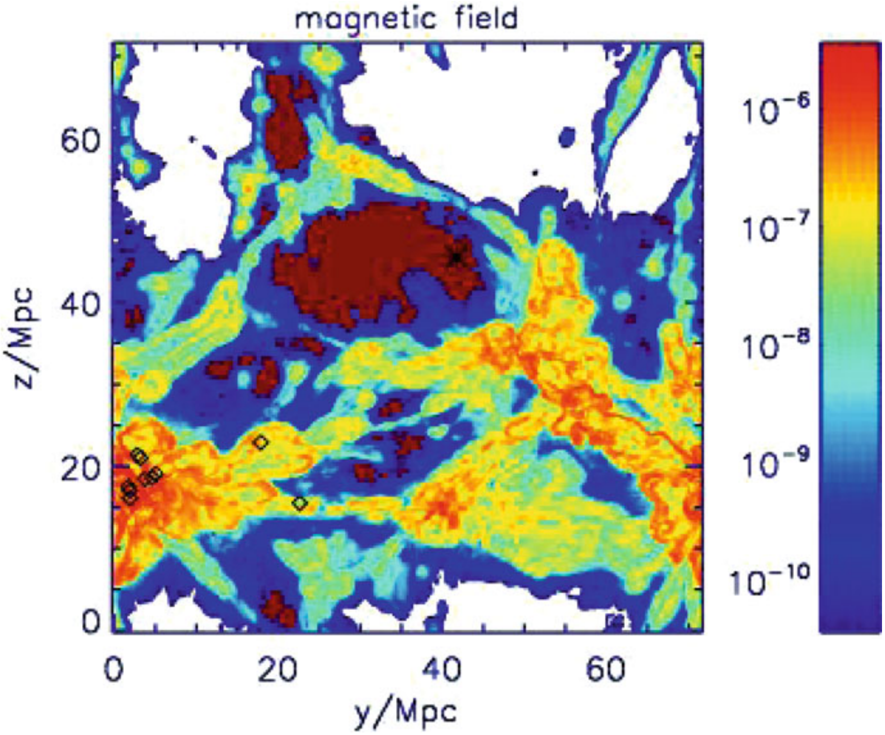


Fig. 4.1 A cross section through a typical large scale structure simulation such as the ones discussed in Ref. [61, 74] on a scale of 70 Mpc in both directions. Ten sources marked with diamonds in the environment of a massive galaxy cluster. The *black cross* indicates the observer. The color contours represent the magnetic field strength in units of Gauss, as indicated

observed in galaxy clusters. Alternatively, it has been assumed that the EGMF follows the local vorticity and turbulent energy density of the matter [27]. These numerical approaches agree on the fact that these fields tend to follow the large scale galaxy structure, i.e. the fields tend to be strongest around the largest matter concentrations. A cross section through one of these simulations [61, 74] is shown in Fig. 4.1. However, they disagree on certain aspects that are relevant for UHECR deflection, most notably the filling factor distributions, i.e. the fraction of space filled with EGMF above a certain strength, as a function of that strength [79]. While this causes considerable differences in the size of the deflection angles predicted between the source and the observed events, the deflections tend to be *along and within* the cosmic large scale structure of the galaxy distribution. This can be seen in Fig. 4.2 where the upper panel shows how the arrival directions relate to the source positions on the sky and the lower panel shows the distribution of the deflection angles between these two directions. In this scenario the deflected UHECR arrival directions tend to follow arc-like structures that result from deflections within the

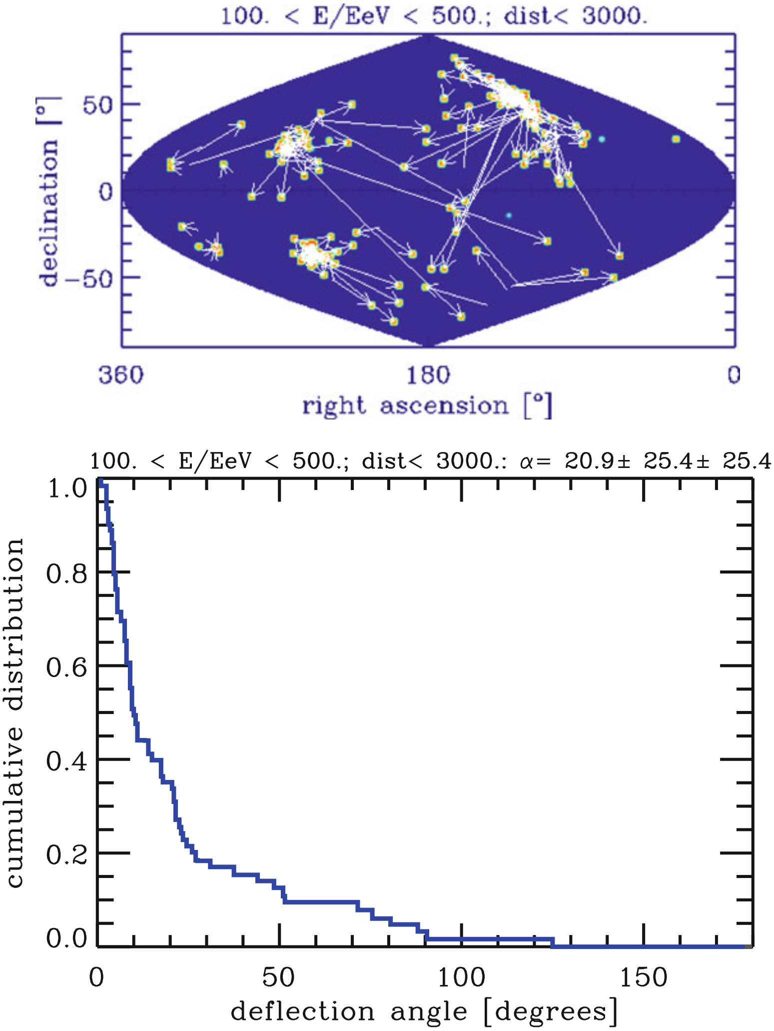


Fig. 4.2 *Upper panel*: Simulated arrival directions of UHECR above 10^{20} eV in a scenario where the sources shown in Fig. 4.1 inject a pure iron composition with an $E^{-2.2}$ spectrum and equal luminosity up to 10^{22} eV. The density of discrete sources in this simulation is $\simeq 2.4 \times 10^{-6}\,\mathrm{Mpc}^{-3}$ and the maximal distance the primary cosmic rays were allowed to propagate is 3,000 Mpc. The arrows point from the source to the detected event. *Lower panel*: Distribution of deflection angles between arrival direction and source position. The average deflection angle is $\simeq 21^\circ$ with a scatter of $\simeq 26^\circ$

large scale cosmic filaments. In other words, as long as the sources are not very nearby, the EGMF is unlikely to deflect UHECRs out of the large scale structure since the fields in the voids are very small. This means that the overall UHECR arrival direction distribution arriving outside the Galaxy is likely to still correlate with the local large scale structure even in the scenarios with large EGMF, heavy nuclei and large deflection angles, although the events do in general not point back to the sources. On the other hand, since deflections in the Galactic field are unlikely to correlate with extragalactic deflections, large deflections of heavy nuclei in the Galactic field are expected to have a much stronger influence on correlations with the local large scale structure.

3 Testing Fundamental Symmetries: Lorentz-Invariance and Cosmic Gamma-Rays

Both loop quantum gravity and string theory often break the Lorentz symmetry or realize it in ways different from special relativity. Typically, such effects manifest themselves through new terms in the dispersion relation, the relation between energy E and momentum p of a particle of mass m, that are suppressed by some power n of the Planck mass $M_{\rm Pl}$,

$$E^2 = m^2 + p^2 \left[1 + \eta \left(\frac{p}{M_{\rm Pl}}\right)^n\right] , \tag{4.3}$$

where η is a dimensionless number (we use natural units in which the vacuum speed of light $c_0 = 1$). Such terms can modify both the free propagation of particles and their interactions.

The propagation velocity now depends on energy in a different way than in case of Lorentz invariance. In fact, in the relativistic limit keeping only terms to first order in m^2 and η, the group velocity for Eq. 4.3 is

$$v = \frac{\partial E}{\partial p} \simeq 1 - \frac{m^2}{2E^2} + \frac{\eta}{2}(n+1)\left(\frac{E}{M_{\rm Pl}}\right)^n \equiv 1 - \frac{m^2}{2E^2} + \delta(E), \tag{4.4}$$

where $\delta(E) \equiv \eta(n+1)(E/M_{\rm Pl})^n/2$ is the deviation from the Lorentz-invariant velocity. For photons, $m = 0$, this can lead to arrival time-delays between photons of different energies emitted by GRBs or by flares of active galactic nuclei. Such time delays have indeed been observed from space by Fermi LAT and Fermi GBM in the 10–100 GeV region [4] and from the ground, for example, by the MAGIC telescope above 150 GeV [13]. They have been used to establish upper limits on the Lorentz invariance violating (LIV) terms. For $n = 1$ these are typically of order one, $|\eta| \lesssim 1$ [4].

Furthermore, the kinematics of interactions can be modified which typically happens when the LIV terms become comparable to the particle rest mass, $E \gtrsim E_{\rm cr} = (m^2 M_{\rm Pl}^{n-2})^{1/n}$. As a result, the larger the particle mass the higher the energy at which LIV effects come into play. Therefore, TeV electrons and positrons, but not protons, can be used to constrain $n = 1$ LIV effects (see, e.g., [59]), and UHE protons are required to obtain constraints on hadronic LIV terms with $n = 2$

scaling. A particularly interesting case is the superluminal motion which occurs for $\delta(E) > m^2/(2E^2)$ or $E > m/(2\delta)^{1/2}$, where for the general case $\delta(E)$ is the difference of the LIV term for the particle and the photon: At such energies a charged particle would emit vacuum Cherenkov radiation, similar to the motion of an ultra-relativistic charge in a medium with index of refraction larger than one. The resulting rapid energy loss would imply that particles cannot reach such energies in astrophysical environments. Their observation in turn allows to rule out the corresponding LIV parameters.

The arguments above make it clear that LIV effects with $n \geq 1$ increase with energy. The highest energies in Nature are observed in high energy astrophysics, in particular TeV gamma-ray astrophysics and UHE cosmic rays and neutrinos. There is thus a new field emerging at the interface of quantum gravity phenomenology, string theory and astrophysics. In fact, many of the LIV terms of the form of Eq. 4.3 have already been strongly constrained (for reviews see, e.g., [17]). We mention in particular constraints based on the flux suppression feature observed in UHECRs that is consistent with the GZK effect: A tiny Lorentz invariance violation with $\delta_\pi(E_\pi) - \delta_p(E_p) \gtrsim 5 \times 10^{-23}$ would lead to a significant shift of the GZK feature and would thus be ruled out (for a review see, e.g., [82]). In terms of η, for $n = 2$, LIV effects should thus be suppressed by a factor $\gtrsim 10^6$. LIV can also lead to spontaneous decay, vacuum Cherenkov-radiation and modified photo-disintegration reactions of very high energy nuclei, thereby influencing UHECR chemical composition. This makes future UHECR composition measurements also relevant for testing Lorentz invariance violation [76].

In the following we will focus on photons for which the most important interaction in an astrophysical and cosmological context is pair production on low energy target photons (for a review see, e.g., [77]). The highest energy photons we know should be produced are the ones resulting from the decay of π^0 mesons produced by the GZK effect. A certain fraction of the UHECR flux should thus be photons. Due to pair production on the CMB and infrared backgrounds and subsequent inverse Compton scattering of the produced electrons and positrons an electromagnetic cascade develops which quickly shifts the electromagnetic flux below the pair production threshold on the CMB, $\simeq 10^{15}$ eV. As a result, the expected photon fraction of the UHECR flux is rather small, less than 10% around 10^{20} eV and less than 1% around 10^{19} eV (see, e.g., [40]). In fact, only experimental upper limits are currently available consistent with the experimental sensitivity [7].

However, a tiny Lorentz symmetry violation can inhibit pair production such that the predicted UHE photon fraction would be much larger, of the order of 20% for $10^{19}\,\text{eV} \lesssim E \lesssim 10^{20}\,\text{eV}$, because any photon produced by pion production, even at cosmological distances, would only be subject to redshift and thus contribute to the local UHE photon flux. This contradicts the observational upper limits and can thus be used to constrain the LIV parameters in the electromagnetic sector. The resulting constraints are very strong, in fact much stronger than the ones obtained from arrival time dispersion of gamma-rays from GRBs [4]: Typically, for LIV terms suppressed to first order in the Planck scale, $n = 1$, values $|\eta| \gtrsim 10^{-14}$ are ruled out, whereas for second order suppression, $n = 2$, values $|\eta| \gtrsim 10^{-6}$ tend to be constrained [38, 39]. Since such dimensionless coefficients would be expected to be of order one if they are not forbidden by some symmetry, this suggests that LIV is most likely absent altogether at first and second order suppression with the Planck scale.

4 Searching for New Light States in Electromagnetic Emission of Astrophysical Sources

Many extensions of the Standard Model of particle physics, in particular scenarios based on supergravity or superstrings, predict a "hidden sector" of new particles interacting only very weakly with Standard Model particles. Such scenarios do not necessarily only contain Weakly Interacting Massive Particles (WIMPs), new heavy states at the TeV scale and above, some of which are candidates for the dark matter, but often also predict Weakly Interacting Sub-eV Particles (WISPs) that can couple to the photon field A_μ (for a recent review see, [51]). The most well-known examples include pseudo-scalar axions and axion-like particles a and hidden photons that mix kinetically with photons.

Axion-Like Particles (ALPs) are described by a Lagrangian of the form

$$\mathscr{L}_{a\gamma} = \frac{1}{8\pi f_a} a F_{\mu\nu}\tilde{F}^{\mu\nu} + \frac{1}{2} m_a^2 a^2 = -\frac{1}{2\pi f_a} a \mathbf{E}\cdot\mathbf{B} + \frac{1}{2} m_a^2 a^2 \,, \tag{4.5}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ the electromagnetic field tensor, $\tilde{F}^{\mu\nu}$ its dual, $\mathbf{E}$ and $\mathbf{B}$ the electric and magnetic field strengths, respectively, f_a a Peccei-Quinn like energy scale and m_a the axion mass. In addition, ALPs in general have similar couplings to gluons giving rise to mixing between axions and neutral pions π^0. The actual axion was proposed to solve the strong CP-problem, a problem of phase cancellation in quantum chromodynamics, and exhibits a specific relation between coupling and mass, $m_a \simeq 0.6\,(10^{10}\,\mathrm{GeV}/f_a)\,\mathrm{meV}$ [67].

A hidden photon field X_μ describes a hidden $U(1)$ symmetry group and mixes with the photon through a Lagrangian of the form

$$\mathscr{L}_{X\gamma} = -\frac{1}{4} F_{\mu\nu}F^{\mu\nu} - \frac{1}{4} X_{\mu\nu}X^{\mu\nu} + \frac{\sin\chi}{2} X_{\mu\nu}F^{\mu\nu} + \frac{\cos^2\chi}{2} m_{\gamma'}^2 X_\mu X^\mu + j^\mu_{\mathrm{em}} A_\mu \,, \tag{4.6}$$

where $X_{\mu\nu}$ is the hidden photon field strength tensor, $m_{\gamma'}$ the hidden photon mass and χ a dimensionless mixing parameter and j^μ_{em} is the electromagnetic current. Typical values for the mixing parameter range from $\sim 10^{-2}$ down to 10^{-16}.

These couplings to photons can induce many interesting effects that are relevant for astronomy and astrophysics: in the presence of electromagnetic fields, in particular of magnetic fields, photons can oscillate into axions and vice-versa, an effect known as Primakoff-effect [70]. In fact, for a while this possibility was even entertained as a possible explanation of the disturbing observation that the explosions of white dwarfs (Supernovae of type Ia – see Chap. 2) which can serve as "standard candles" because of their roughly constant explosion energy are dimmer than expected in a decelerating Universe that would otherwise lead to the conclusion that the expansion of the Universe must accelerate [68, 73]. Although meanwhile this possibility is basically excluded because it predicts other signatures, notably distortions of the CMB, which have not been observed [62], photon-ALPs mixing can still play a role at higher energies.

Photons can also oscillate into hidden photons even in vacuum. These oscillations can be modified in the presence of a plasma which gives the photons an effective mass whereas the WISP mass is essentially unchanged. This can give rise to matter oscillations reminiscent of the MikheyevSmirnovWolfenstein effect for neutrino oscillations [60, 87]. In particular, even if the mixing in vacuum is very small, one can have resonant conversions of photons into WISPs within a plasma. Such photon conversions in vacuum and in matter can have effects both within astrophysical sources and during propagation of photons from the source to the observer.

The coupling of WISPs to photons and (in case of axions) also to fermions can have an influence on the evolution and structure of astrophysical objects. Due to their weak coupling to ordinary matter, once produced, these hidden sector particles can leave most objects without significant reabsorption, providing an efficient cooling mechanism. This has lead, for example, to strong limits on axion masses and couplings from the requirement that core-collapse supernovae should not cool much faster than predicted if their cooling is dominated by neutrino emission, in order to be consistent with the few neutrinos observed from the cooling phase of SN1987A (see, e.g., [53]).

Even if the physics of the astronomical objects is not significantly modified, the photon rates and spectra observable at Earth can be influenced either within the source or during propagation to the observer. A sensitive probe of photon-WISP oscillations requires as detailed an understanding of the emission process as possible. In this context, one of the best understood radiation sources in the Universe is the cosmic microwave background (CMB). Its spectrum deviates from a perfect blackbody by less than $\simeq 10^{-4}$, distortions that have been measured by the COBE-FIRAS experiment [37], and whose deviations from isotropy are of the order of 10^{-5} and have themselves been measured at the percent level by WMAP [54]. This radiation essentially comes from the surface of last scattering, at a distance of a Hubble radius today, and any photon-WISP mixing at a level of $\sim 10^{-4}$ would induce a spectral distortion or an anisotropy in conflict with the observations. This has lead to some of the strongest limits on the parameters of Eqs. 4.5 and 4.6: For $10^{-9}\,\text{eV} \lesssim m_a \lesssim 10^{-4}\,\text{eV}$ one has $f_a \gtrsim 10^{11}(B_{\text{rms}}/\text{nG})\,10^{10}\,\text{GeV}$ which strengthens to $f_a \gtrsim 10^{12}(B_{\text{rms}}/\text{nG})\,10^{11}\,\text{GeV}$ for $10^{-14}\,\text{eV} \lesssim m_a \lesssim 10^{-11}\,\text{eV}$ [63]. Since photon-ALP mixing requires the presence of a magnetic field, the absence of significant effects on the CMB imposes an upper limit on the combination B_{rms}/f_a, with B_{rms} the rms large scale extra-galactic magnetic field. Furthermore, requiring the distortions of the CMB induced by photon-hidden photon mixing to be smaller than the COBE-FIRAS limit leads to a bound on the mixing angle $\chi \lesssim 10^{-7} - 10^{-5}$ for hidden photon masses $10^{-14}\,\text{eV} \lesssim m_{\gamma'} \lesssim 10^{-7}\,\text{eV}$ [64]. In contrast to the case of ALPs, these constraints only depend on the vacuum mixing angle χ since no external magnetic fields are necessary for photon-hidden photon mixing.

Most other relevant astrophysical sources are non-thermal in nature and thus much less well understood. This is the case in particular for X-ray and gamma-ray sources. Still, if the photon spectra from these objects can be well approximated by power laws, photon-ALPs mixing can induce steps in the spectra that may be detectable. Depending on the strength of magnetic field within the sources, for ALP masses $m_a \sim 10^{-6}\,\text{eV}$ significant effects on spectra between keV and TeV energies can occur for $f_a \lesssim 10^{13}\,\text{eV}$ [48,50]. These effects are complementary and potentially more sensitive compared to more direct experimental bounds the best of which come from *helioscopes*: photons from the sun are converted to ALPs in the solar magnetic field which in turn can be reconverted to photons in an artificial magnet in front of a telescope on Earth which then detects these photons. For $m_a \lesssim 0.02\,\text{eV}$ the CERN Axion Solar Telescope (CAST) experiment provided the strongest constraint, $f_a \gtrsim 10^{10}\,\text{GeV}$ [20].

Since photon-ALP mixing is energy dependent, ALP signatures are best revealed when comparing luminosities at different energies. In particular, it has been pointed out that the *scatter* of correlations of luminosities in different energy bands deviates from a Gaussian if photon-ALP mixing occurs. In fact, considerable deviations from Gaussian scatters have recently been found in the correlations between the luminosities of AGNs in the optical/UV and X-rays [25]. If these sources are located in galaxy clusters which are known to contain magnetic fields of micro Gauss strength, photon-ALP mixing could explain this observation if $m_a \ll 10^{-12}\,\text{eV}$ and $f_a \lesssim 10^{10}\,\text{GeV}$. In this case, almost energy independent photon-ALP mixing would occur at energies above $\simeq 2\,\text{keV}$, whereas the mixing would be highly energy dependent at energies $\ll 0.5\,\text{keV}$, thereby inducing non-Gaussian correlations. Similar effects would occur with photon-ALP conversion in magnetic fields within AGNs if $m_a \ll 10^{-7}\,\text{eV}$ and $f_a \simeq 3 \times 10^{8}\,\text{GeV}$. It has been pointed out, however, that the scatter in the correlation between optical and X-ray luminosities observed in AGNs can also be explained by X-ray absorption [69].

Another possible signature for photon mixing with a new light state has been discussed in the context of high energy gamma-ray observations by the ground-based telescopes MAGIC, H.E.S.S., VERITAS and CANGAROO-III. The absorption of such gamma-rays in the infrared background appears weaker than expected based on models for the infrared background [10, 83], although this is currently inconclusive [14, 26]. If gamma-ray absorption is indeed weaker than computed for the real infrared background, this could be explained if part of the gamma-rays are converted into ALPs around the source which in turn are reconverted into gamma-rays in the Galactic magnetic field [80]. This works for ALP parameters $10^{-10}\,\text{eV} \lesssim m_a \lesssim 10^{-8}\,\text{eV}$ and $f_a \sim 10^{9}\,\text{GeV}$ Alternatively, conversion and re-conversion could be induced by the EGMF if $m_a \lesssim 10^{-10}\,\text{eV}$ and $5 \times 10^{10}\,\text{GeV} \lesssim f_a \lesssim 10^{18}\,\text{GeV}$ [28,29]. A recent detailed study on these effects has been performed in Ref. [75]. We note, however, that an apparently reduced absorption of $\gamma-$rays from high redshift sources can also be explained if these $\gamma-$rays are produced near Earth by primary TeV-PeV *cosmic rays* from the same source which interact much less frequently with the low energy target photons than TeV $\gamma-$rays [33]. This is possible provided that cosmic ray deflection is sufficiently small, corresponding to large scale EGMFs of strength $B \lesssim 3 \times 10^{-14}\,$G [34].

5 Conclusions

In this contribution we have discussed three examples in which astronomy plays an interdisciplinary role at the intersection with the neighboring scientific fields of cosmology and particle physics: The nature and origin of the highest energy particles observed in Nature, tests of the Lorentz symmetry which is one of the pillars of modern science tiny breakings of which may yield fundamental insights into Nature and may lead to observable effects at the highest energies, and, at the opposite end of the energy scale, the mixing of photons with new light states such as axion-like particles or hidden photons. While this list is certainly not exhausting and does not include other important topics such as the search for dark matter, it hopefully gives an idea about the role of interdisciplinarity in astronomy. With the first results coming in from the Large Hadron Collider, the most powerful existing particle physics experiment in terms of energy and luminosity, new levels of cross-fertilization between astronomy and particle physics are expected for the near future.

Acknowledgements This work was supported by the Deutsche Forschungsgemeinschaft through the collaborative research centre SFB 676 Particles, Strings and the Early Universe: The Structure of Matter and Space-Time and by the State of Hamburg, through the Collaborative Research program Connecting Particles with the Cosmos within the framework of the Landesexzellenzinitiative (LEXI).

References

1. R.U. Abbasi et al., Astropart. Phys. **30**, 175 (2008) [arXiv:0804.0382 [astro-ph]]
2. R. Abbasi et al. [HiRes Collaboration], Phys. Rev. Lett. **100**, 101101 (2008) [arXiv:astro-ph/0703099]; R.U. Abbasi et al., Astropart. Phys. **32**, 53 (2009) [arXiv:0904.4500 [astro-ph.HE]]
3. R.U. Abbasi et al. [HiRes Collaboration], Phys. Rev. Lett. **104**, 161101 (2010) [arXiv: 0910.4184 [astro-ph.HE]]
4. A.A. Abdo et al. Nature **462**, 331 (2009)
5. J. Abraham et al. [Pierre Auger Collaboration], Science **318**, 938 (2007) [arXiv:0711.2256 [astro-ph]]; J. Abraham et al. [Pierre Auger Collaboration], Astropart. Phys. **29**, 188 (2008) [Erratum-ibid. **30**, 45 (2008)] [arXiv:0712.2843 [astro-ph]]
6. J. Abraham et al. [Pierre Auger Collaboration], Phys. Rev. Lett. **101**, 061101 (2008) [arXiv:0806.4302 [astro-ph]]; J. Abraham et al. [The Pierre Auger Collaboration], Phys. Lett. B **685**, 239 (2010) [arXiv:1002.1975 [astro-ph.HE]]

7. J. Abraham et al. [The Pierre Auger Collaboration], Astropart. Phys. **31**, 399-406 (2009) [arXiv:0903.1127 [astro-ph.HE]]
8. J. Abraham et al. [Pierre Auger Observatory Collaboration], Phys. Rev. Lett. **104**, 091101 (2010) [arXiv:1002.0699 [astro-ph.HE]]
9. P. Abreu et al. [Pierre Auger Observatory Collaboration], arXiv:1009.1855 [Unknown]
10. F. Aharonian et al. [H.E.S.S. Collaboration], Nature **440**, 1018–1021 (2006) [astro-ph/0508073]
11. F. Aharonian et al., Astrophys. J. Lett. **695**, L40 (2009)
12. M. Ahlers, L.A. Anchordoqui, M.C. Gonzalez-Garcia et al., Astropart. Phys. **34**, 106–115 (2010) [arXiv:1005.2620 [astro-ph.HE]]
13. J. Albert et al. [MAGIC Collaboration and Other Contributors Collaboration], Phys. Lett. B **668**, 253 (2008) [arXiv:0708.2889 [astro-ph]]
14. E. Aliu et al. [MAGIC Collaboration], Science **320**, 1752 (2008) [arXiv:0807.2822 [astro-ph]]
15. D. Allard et al., JCAP **0609**, 005 (2006) [arXiv:astro-ph/0605327]
16. D. Allard, N.G. Busca, G. Decerprit, A.V. Olinto, E. Parizot, J. Cosmol. Astropart. Phys. **0810**, 033 (2008) [arXiv:0805.4779 [astro-ph]]
17. G. Amelino-Camelia, [arXiv:0806.0339 [gr-qc]]; D. Mattingly, Living Rev. Rel. **8**, 5 (2005) [arXiv:gr-qc/0502097]; S. Liberati and L. Maccione, Ann. Rev. Nucl. Part. Sci. **59**, 245 (2009) [arXiv:0906.0681 [astro-ph.HE]]
18. L.A. Anchordoqui, H. Goldberg, D. Hooper, S. Sarkar, A.M. Taylor, Phys. Rev. D **76**, 123008 (2007) [arXiv:0709.0734 [astro-ph]]
19. L.A. Anchordoqui, D. Hooper, S. Sarkar, A.M. Taylor, Astropart. Phys. **29**, 1 (2008) [arXiv:astro-ph/0703001]
20. E. Arik et al. [CAST Collaboration], JCAP **0902**, 008 (2009) [arXiv:0810.4482 [hep-ex]]
21. P. Auger, R. Maze, T. Grivet-Meyer, Académie des Sciences **206**, 1721 (1938); P. Auger, R. Maze, Académie des Sciences **207**, 228 (1938)
22. D.J. Bird et al., Astrophys. J. **441**, 144 (1995)
23. R.D. Blandford, Phys. Scripta **T85**, 191 (2000) [arXiv:astro-ph/9906026]
24. P. Blasi, S. Burles, A.V. Olinto, Astrophys. J. **514**, L79–L82 (1999) [astro-ph/9812487]
25. C. Burrage, A.C. Davis, D.J. Shaw, Phys. Rev. Lett. **102**, 201101 (2009) [arXiv:0902.2320 [astro-ph.CO]]
26. L. Costamante, F. Aharonian, R. Buehler et al., [arXiv:0907.3966 [astro-ph.CO]]
27. S. Das, H. Kang, D. Ryu, J. Cho, Astrophys. J. **682**, 29 (2008) arXiv:0801.0371 [astro-ph]; D. Ryu, S. Das, H. Kang, Astrophys. J. **710**, 1422 (2010) [arXiv:0910.3361 [astro-ph.HE]]
28. A. De Angelis, O. Mansutti, M. Roncadelli, Phys. Rev. **D76**, 121301 (2007) [arXiv:0707.4312 [astro-ph]]
29. A. De Angelis, O. Mansutti, M. Persic et al., [arXiv:0807.4246 [astro-ph]]
30. C.D. Dermer, arXiv:1008.0854 [astro-ph.HE]
31. K. Dolag, D. Grasso, V. Springel, I. Tkachev, JETP Lett. **79**, 583 (2004) [Pisma Zh. Eksp. Teor. Fiz. **79**, 719 (2004)] [arXiv:astro-ph/0310902]; JCAP **0501**, 009 (2005) [arXiv:astro-ph/0410419]
32. J.W. Elbert, P. Sommers, Astrophys. J. **441**, 151 (1995) [arXiv:astro-ph/9410069]
33. W. Essey, O. Kalashev, A. Kusenko et al., [arXiv:1011.6340 [astro-ph.HE]]
34. W. Essey, S.'i. Ando, A. Kusenko, [arXiv:1012.5313 [astro-ph.HE]]
35. A. Falcone, H. Hase, C. Pagoni, C. Ploetz [Fermi Collaboration], Astrophys. J. **719**, 1433 (2010) [arXiv:1006.5463 [astro-ph.HE]]
36. Fermi LAT Collaboration, Science **328**, 725 (2010) [arXiv:1006.3986 [astro-ph.HE]]
37. D.J. Fixsen, E.S. Cheng, J.M. Gales, J.C. Mather, R.A. Shafer, E.L. Wright, Astrophys. J. **473**, 576 (1996) [astro-ph/9605054]
38. M. Galaverni, G. Sigl, Phys. Rev. Lett. **100**, 021102 (2008) [arXiv:0708.1737 [astro-ph]]
39. M. Galaverni, G. Sigl, Phys. Rev. D **78**, 063003 (2008) [arXiv:0807.1210 [astro-ph]]
40. G.B. Gelmini, O.E. Kalashev, D.V. Semikoz, J. Cosmol. Astropart. Phys. **0711**, 002 (2007) [arXiv:0706.2181 [astro-ph]]
41. G. Giacinti, M. Kachelriess, D.V. Semikoz, G. Sigl, arXiv:1006.5416 [astro-ph.HE]

42. D. Gorbunov, P. Tinyakov, I. Tkachev et al., JETP Lett. **87**, 461–463 (2008) [arXiv:0711.4060 [astro-ph]]
43. D. Grasso, H.R. Rubinstein, Phys. Rept. **348**, 163–266 (2001) [astro-ph/0009061]
44. K. Greisen, Phys. Rev. Lett. **16**, 748 (1966); G.T. Zatsepin, V.A. Kuzmin, JETP Lett. **4**, 78 (1966) [Pisma Zh. Eksp. Teor. Fiz. **4**, 114 (1966)]
45. M.J. Hardcastle, arXiv:1003.2500 [astro-ph.HE]
46. M. Takeda et al., Phys. Rev. Lett. **81**, 1163 (1998) [arXiv:astro-ph/9807193]
47. A.M. Hillas, Annu. Rev. Astron. Astrophys. **22**, 425 (1984)
48. K.A. Hochmuth, G. Sigl, Phys. Rev. D **76**, 123011 (2007) [arXiv:0708.1144 [astro-ph]]
49. D. Hooper, A.M. Taylor, S. Sarkar, arXiv:1007.1306 [astro-ph.HE]
50. D. Hooper, P.D. Serpico, Phys. Rev. Lett. **99**, 231102 (2007) [arXiv:0706.3203 [hep-ph]]
51. J. Jaeckel, A. Ringwald, arXiv:1002.0329 [hep-ph]
52. M. Kachelriess, S. Ostapchenko, R. Tomas, New J. Phys. **11**, 065017 (2009) [arXiv:0805.2608 [astro-ph]]; M. Kachelriess, S. Ostapchenko, R. Tomas, arXiv:1002.4874 [astro-ph.HE]
53. W. Keil, H.T. Janka, D.N. Schramm, G.Sigl, M.S. Turner, J.R. Ellis, Phys. Rev. D **56**, 2419 (1997) [arXiv:astro-ph/9612222]
54. E. Komatsu et al., arXiv:1001.4538 [astro-ph.CO]
55. K. Kotera, D. Allard, A.V. Olinto, [arXiv:1009.1382 [astro-ph.HE]]
56. P.P. Kronberg, Rept. Prog. Phys. **57**, 325–382 (1994); J.P. Vallee, Fundam. Cosmic Phys. **19**, 1 (1997)
57. J. Linsley, Phys. Rev. Lett. **10**, 146 (1963)
58. R.V.E. Lovelace, Nature **262**, 649 (1976)
59. L. Maccione, S. Liberati, A. Celotti et al., JCAP **0710**, 013 (2007) [arXiv:0707.2673 [astro-ph]]
60. S.P. Mikheev, A.Y. Smirnov, Sov. J. Nucl. Phys. **42**, 913 (1985) [Yad. Fiz. **42**, 1441 (1985)]
61. F. Miniati, Inter-galactic shock acceleration and the cosmic gamma-ray background, Mon. Not. Roy. Astron. Soc. **337**, 199 (2002) [arXiv:astro-ph/0203014]
62. A. Mirizzi, G.G. Raffelt, P.D. Serpico, Phys. Rev. **D72**, 023501 (2005) [astro-ph/0506078]
63. A. Mirizzi, J. Redondo, G. Sigl, JCAP **0908**, 001 (2009) [arXiv:0905.4865 [hep-ph]]
64. A. Mirizzi, J. Redondo, G. Sigl, JCAP **0903**, 026 (2009) [arXiv:0901.0014 [hep-ph]]
65. I.V. Moskalenko, L. Stawarz, T.A. Porter et al., Astrophys. J. **693**, 1261–1274 (2009) [arXiv:0805.1260 [astro-ph]]
66. A. Neronov, I. Vovk, Science **328**, 73 (2010) [arXiv:1006.3504 [astro-ph.HE]]
67. R.D. Peccei, H.R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977)
68. S. Perlmutter et al. [Supernova Cosmology Project Collaboration], Astrophys. J. **517**, 565–586 (1999) [astro-ph/9812133]
69. G.W. Pettinari, R. Crittenden, arXiv:1007.0024 [astro-ph.CO]
70. H. Pirmakoff, Phys. Rev. **81**, 899 (1951)
71. K. Ptitsyna, S. Troitsky, arXiv:0808.0367 [astro-ph]
72. F.M. Rieger, F.A. Aharonian, [arXiv:0910.2327 [astro-ph.HE]]
73. A.G. Riess et al. [Supernova Search Team Collaboration], Astron. J. **116**, 1009–1038 (1998) [astro-ph/9805201]
74. D. Ryu, H. Kang, P.L. Biermann, Astron. Astrophys. **335** 19 (1998)
75. M.A. Sanchez-Conde, D. Paneque, E. Bloom et al., Phys. Rev. **D79**, 123511 (2009) [arXiv:0905.3270 [astro-ph.CO]]
76. A. Saveliev, L. Maccione, G. Sigl, JCAP **1103**, 046 (2011) [arXiv:1101.2903 [astro-ph.HE]]
77. L. Shao, B.Q. Ma, arXiv:1007.2269 [hep-ph]
78. G. Sigl, F. Miniati, T.A. Enßlin, Phys. Rev. D **70**, 043007 (2004) [arXiv:astro-ph/0401084]
79. G. Sigl, F. Miniati, T. Ensslin, Nucl. Phys. Proc. Suppl. **136**, 224 (2004) [arXiv:astro-ph/0409098]
80. M. Simet, D. Hooper, P.D. Serpico, Phys. Rev. **D77**, 063001 (2008) [arXiv:0712.2825 [astro-ph]]
81. F.W. Stecker, Phys. Rev. Lett. **21**, 1016 (1968)
82. F.W. Stecker, S.T. Scully, New J. Phys. **11**, 085003 (2009) [arXiv:0906.1735 [astro-ph.HE]]

83. F.W. Stecker, S.T. Scully, Submitted to: Astron. Astrophys. **478**, L1 (2008) [arXiv:0710.2252 [astro-ph]]
84. R. Ulrich, R. Engel, S. Muller et al., [arXiv:0906.0418 [astro-ph.HE]]
85. R. Ulrich, R. Engel, M. Unger, [arXiv:1010.4310 [hep-ph]]
86. E. Waxman, J. Miralda-Escude, Astrophys. J. **472**, L89–L92 (1996) [astro-ph/9607059]
87. L. Wolfenstein, Phys. Rev. D **17**, 2369 (1978)
88. Y. Xu, P.P. Kronberg, S. Habib et al., Astrophys. J. **637**, 19–26 (2006) [astro-ph/0509826]

Chapter 5
Gravitational Wave Astronomy

Alessandra Buonanno

Abstract Spacetime is a dynamic and elastic entity both influencing and influenced by the distribution of mass and energy that it contains. As a consequence the accelerated motion of mass and energy can generate ripples or gravitational waves in the fabric of spacetime propagating at the speed of light. Those ripples encode unique information about the source, whatever it is a rapidly rotating neutron star, a binary black-hole system, a supernova or a rapidly changing gravitational field. Today, those ripples could be detected for the first time by instruments monitoring displacements on a scale one million times smaller than a single atom. The ongoing research on gravitational waves will improve our ability to detect and extract unique information from the observed waveforms, test fundamental equations of general relativity, and design increasingly sensitive detectors. The direct detection of gravitational waves is now in sight. It will constitute one of the major scientific discoveries of the next decade.

Keywords Gravitational waves • Black hole physics

1 The Search for Gravitational Waves: A Long Journey Started Half a Century Ago

Gravitational waves were predicted by Einstein in 1916 [56, 57], but have never yet been directly observed. However, strong indirect evidence for their existence comes from observations of binary pulsars [65]. In those relativistic two-body systems the variation of the orbital period is measured to exquisite precision and it is found to

---

A. Buonanno (✉)
Maryland Center for Fundamental Physics & Joint Space-Science Institute,
Department of Physics, University of Maryland, College Park, MD 20742, USA
e-mail: buonanno@umd.edu

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_5,

be fully consistent with predictions of general relativity – which are that the orbital period will diminish, and the system will lose energy, as a result of the emission of gravitational waves. Thanks to theoretical and experimental progress made by the gravitational-wave community during the last 40 years, the direct detection of gravitational waves is now in sight. This will constitute one of the major scientific discoveries of the next decade, as it will permit a new kind of observation of the cosmos quite different from today's electromagnetic and particle observations.

The effect of gravitational waves on closely separated free-falling test masses is characterized by a time-dependent tidal force. The gravitational-wave interferometers currently operating are detectors of astonishing sensitivity, designed to detect a passing gravitational wave through almost inconceivably small displacements of the instruments' mirrors. To catch the tiny gravitational-wave signal buried in the instrumental noise of the detector requires both a reliable knowledge of the signal shape and a thorough monitoring of all possible sources of noise mimicking a gravitational-wave signal. This requires a deep knowledge of both how the interferometer functions and of the physical source of the gravitational-wave emission.

The quest for direct observation of gravitational waves began in the 1960s with the pioneering work of Joseph Weber [116] at the University of Maryland, and for almost three decades it has been pursued solely using meter-scale resonant-bar detectors [100, 102]. These detectors are cylindrically shaped bars whose mechanical oscillations can be driven by a passing gravitational wave oscillating at a frequency around 1 kHz. During the last 10 years a network of ground-based laser-interferometer gravitational-wave detectors has been built and has taken data at design sensitivity. It is a worldwide network composed of the Laser Interferometer Gravitational wave Observatory (LIGO) [105], Virgo [111], GEO-600 [103],[1] and TAMA [110], and it has operated in the frequency range of 10–10^3 Hz. Within the next 5 years those detectors will be upgraded to a sensitivity such that event rates for coalescing binary systems will increase by a factor of one thousand, thus making very likely the first detection and establishing the field of gravitational-wave astronomy.

Research and development of more sensitive detectors operating at or below the so-called standard-quantum limit has already started [101], and may well result eventually in gravitational-wave laboratories operating for decades and routinely detecting gravitational waves from several astrophysical and cosmological sources.

Within the next 15 years, the ground-based detectors will likely be complemented by the laser-interferometer space antenna (LISA) [104], a joint venture between NASA and European Space Agency, and the Japanese Deci-hertz Interferometer

---

[1] LIGO is a National Science Foundation (NSF) project operated by Caltech and MIT. In 1999 LIGO and GEO-600 formed the LIGO Scientific Collaboration, bringing together $\sim$400 scientists worldwide. The number of researchers involved is comparable to that participating in high-energy physics collaborations. In 2008, the LSC and Virgo formed the LSC/Virgo Collaboration to share data and software.

Gravitational Wave Observatory (DECIGO). Those detectors will search for gravitational waves in the frequency range 3×10^{-5}–1 Hz, and 10^{-2}–1 Hz, respectively. At much lower frequencies, 10^{-9}–10^{-8} Hz, scientists around the world are currently using millisecond pulsar timing to detect gravitational waves, and the large number of millisecond pulsars that may be discovered in the near future with the square kilometer array [109] would provide us with an ensemble of clocks that can be used as multiple arms of a gravitational-wave detector. Finally, current [108] and future cosmic microwave background probes might detect gravitational waves at frequencies of $\sim 10^{-17}$ Hz by measuring the cosmic microwave background polarization.

Although skewed by the author's own background, expertise and personal experience, this review focuses on areas in which there have been astonishing and exciting developments during the last 10 years.

2 The Production and Typical Strength of Gravitational Waves

Whereas electromagnetic waves are produced by accelerated charges, gravitational waves are produced by accelerated (gravitational) masses. The generated gravitational waves can be expressed in terms of time-changing multipole moments of the source. At linear order in G, the first non-trivial time-changing multipole moment is the second time derivative of the mass quadrupole-moment, instead of the dipole moment as in electromagnetic theory.[2] Thus, if a distribution of mass M has characteristic size r and is subject to periodic motion with period P, the dimensionless gravitational-wave strength or strain is $h \sim \varepsilon\, G/c^4\, (M r^2/P^2)/R$, with ε the deviation from sphericity and R the distance to the source. Introducing the kinetic energy $E_{\rm kin} \sim M r^2/P^2$, we have $h \sim \varepsilon\, G/c^2\, (E_{\rm kin}/c^2)/R$. For $E_{\rm kin} \sim 1 M_{\odot} c^2$, $\varepsilon \sim 0.01$ and $R \sim 33$Mpc which is the distance to the Virgo cluster, we obtain $h \sim 10^{-19}$. Laser-interferometer gravitational-wave experiments monitor the differential displacement ΔL of the mirrors hanging at the extremities of the arm cavities, with $\Delta L \sim hL$, L being the length of the interferometer's arm cavities. For LIGO and Virgo $L = 4, 3$ km, respectively. Thus, to detect strains $h \sim 10^{-19}$ detectors must monitor displacements on a scale one million times smaller than a single atom![3]

If we compute the gravitational-wave luminosity, we obtain $L_{\rm GW} \sim \varepsilon^2\, G/c^5 (M r^2/P^3)^2$ which can be re-written as $L_{\rm GW} \sim \varepsilon^2 c^5/G (GM\omega/c^3)^6 (r c^2/GM)^4$ with

---

[2] At linear order in G, the mass monopole, mass dipole and current dipole multipole moments do not produce gravitational waves because they do not vary in time due to the conservation of energy, linear momentum and angular momentum. This result is also true if one assumes the equivalence principle. In fact, in electromagnetic theory there is no dipole radiation if the charged particles have e/m constant, with e the electric charge and m the mass.

[3] The currently most sensitive LIGO detector has achieved a strain noise of $h = 3 \times 10^{-22}$ in its most sensitive band from 100 Hz to 200 Hz. This result was obtained by employing Fabry-Perot cavities which increase the length of the interferometer cavities by almost two orders of magnitude.

$\omega \sim 1/P$. The factor $c^5/G \sim 3.6 \times 10^{59}$ erg/s is huge (it is 10^{26} times larger than the electromagnetic luminosity of the Sun). However, the dimensionless factor $GM\omega/c^3$ is generally tiny unless the speed of the source is close to the speed of light. For binary systems moving on a circular orbit the characteristic binary velocity is $v/c \sim (GM\omega/c^3)^{1/3}$ and binary separation $GM/rc^2 \sim v^2/c^2$, thus $L_{\mathrm{GW}} \sim c^5/G\,(v/c)^{10}$. For binary systems composed of compact bodies,[4] — including black holes if they carry spin – v/c can reach values of 0.4–0.6 without the bodies being tidally disrupted. As a consequence, the gravitational luminosity in the merger of two black holes can be $\sim 10^{54}$ erg/s, which is only two orders of magnitude smaller than the electromagnetic luminosity of all galaxies in the observable Universe $\sim 10^{56}$ erg/s!

The above description of the generation of gravitational waves applies to binary systems, rapidly rotating neutron stars and collapsing stars, and it is based on classical general relativity. In the early Universe gravitational waves could be produced through classical mechanisms such as the collision of bubbles of true vacuum formed at first order phase transitions or time-changing density inhomogeneities, but also through amplification of quantum vacuum fluctuations, a typical phenomenon of quantum field theory in curved spacetime.

3 Coalescing Compact Binary Systems

Binary systems made of black holes and/or neutron stars, spiraling in toward each other and losing energy through the emission of gravitational waves, are among the most promising detectable sources of gravitational waves. In the case of binary black holes, the most dynamic and non-linear phase of their evolution occurs when the holes end their long inspiral with a plunge, merge with each other, and leave behind a deformed black hole. The latter eventually settles down to a spherical or oblate shape after getting rid of its deformations by emitting gravitational waves in the surrounding spacetime. This final stage is generally referred to as ringdown phase. In the case of binary neutron stars or binary systems composed of a neutron star and a black hole, depending on the mass ratio and the spin, the evolution may end up with the neutron star being tidally disrupted, the subsequent formation of a disk or bar-like structure, the system eventually leaving behind a newborn black hole or neutron star.

The detectors' noise level and the weakness of the waves prevent observing the waveforms directly. For this reason the search for gravitational waves from

---

[4] Compact bodies are bodies for which the compactness parameter $\gamma = GM/Rc^2$ is close to one, with R the characteristic size of the body. For black holes $\gamma = 0.5$, for neutron stars $\gamma \sim 0.2$–0.4. The Sun is certainly not a compact body since its radius is several thousand times larger than $GM_\odot/c^2 \sim 3\,\mathrm{km}$, thus $\gamma \ll 1$.

binary systems and the extraction of parameters, such as the masses and spins, are based on the matched-filtering technique, which requires accurate knowledge of the waveform of the incoming signal.

3.1 Solving the Two-Body Problem and Computing the Waveforms

The post-Newtonian expansion is the most powerful approximation scheme in analytical relativity capable of describing the two-body dynamics and gravitational-wave emission of inspiraling compact binary systems [17]. The post-Newtonian approach expands the Einstein equations in the ratio of the characteristic velocity of the binary v to the speed of light. However, as the black holes approach each other towards merger, we expect the post-Newtonian expansion to lose accuracy because the velocity of the holes approaches the speed of light. Today, the two-body dynamics and gravitational energy flux of non-spinning objects are known up to 3.5PN order (i.e., $(v/c)^7$), allowing the gravitational-wave phasing for circular orbits to be computed at 3.5PN order beyond the leading 2.5PN term [17].[5] Several calculations at high PN order also exist for spinning objects and binary systems moving in eccentric orbits.

The difficulty of solving the Einstein equation analytically lies mainly in its non-linear structure. Solving the Einstein equation numerically can overcome this problem. The last 5 years have seen dazzling breakthroughs in numerical relativity, and today several groups are able to simulate on a computer the evolution of two compact objects during the last stages of inspiral, merger, and ringdown, and to extract the emitted gravitational-wave signal. Furthermore, a new community has formed at the interface between numerical and analytical relativity. The new synergy has led to a deeper understanding of the two-body problem, disclosing unexpected effects such as the 1,000 km/s recoil velocity gained by the newborn hole when two black holes carrying spin merge [37, 61]. The original breakthrough in 2005 was obtained by Pretorius [89] using a particular formulation of the Einstein equation. After 6 months, two groups could independently simulate the merger of two black holes [14,35] – the University of Brownsville group (today at the Rochester Institute of Technology), and the Goddard-NASA group – both using the so-called moving-puncture technique. Thus, after more than 30 years of unsuccessful attempts that nevertheless provided the community with the essential foundations of knowledge, the gold rush in numerical relativity came to an end.

It is interesting to consider that whereas advances in numerical relativity are generally due to the work of several people, the breakthrough was a single-person success [89]. Moreover, three of the people responsible of the breakthrough at the University of Brownsville and Goddard-NASA had participated in the so-called

---

[5] Powers of $(v/c)^n$ correspond to $(n/2)$ PN order with respect to the leading Newtonian term.

Lazarus project in 2001 [12,13]. This project consisted in evolving in full numerical-relativity the binary system for less than an orbit just prior to merger, and then stopping the evolution, extracting from the results of the simulation the spacetime metric of a deformed black hole, and using perturbation theory calculations to complete the evolution during ringdown. It turns out that the waveform of the Lazarus project compared very well with the merger waveform computed in full numerical relativity years later [15].

Furthermore, prior to the numerical-relativity breakthrough and the Lazarus project, a new method was proposed in analytical relativity to describe the dynamics and gravitational-wave emission of binary black holes during inspiral, merger and ringdown: the effective-one-body approach [28, 29, 31, 47, 51]. This approach uses the very accurate results of post-Newtonian theory. However, it does not use those results in their original Taylor-expanded form (i.e., as polynomials in v/c), but instead in some appropriate resummed form. In particular, the effective-one-body approach [16, 29, 47, 51, 52] maps the dynamics of two compact objects of masses m_1 and m_2, and spins $\mathbf{S}_1$ and $\mathbf{S}_2$, into the dynamics of a test-particle of spin $\mathbf{S}_*$ moving in a deformed Kerr[6] metric with spin $\mathbf{S}_{\text{Kerr}}$. The deformation parameter is the symmetric mass ratio $m_1 m_2/(m_1+m_2)^2$ which ranges between 0 (test particle limit) and $1/4$ (equal-mass limit). The effective-one-body approach relies on the assumption that the comparable mass case is a smooth deformation of the test-particle limit case. The resummation of the conservative dynamics is obtained through a canonical transformation and naturally includes in the final stage of the binary evolution a plunge, an innermost-stable-circular orbit, a photon orbit, also called the light ring,[7] and a horizon. The other crucial aspect of the effective-one-body approach is the way it builds the full waveform, including merger and ringdown. Inspired by results in the 1970s on the radial infall of test particles in a Schwarzschild black hole, and by the close-limit approximation,[8] the effective-one-body approach assumes that the merger is very short in time, although broad in frequency, and builds the merger-ringdown signal by attaching to the plunge signal a superposition of quasi-normal modes. This match happens at the light ring where the peak of the potential barrier around the newborn black hole sits.

The analyses of the numerical simulations have revealed that several predictions of the effective-one-body approach are indeed correct. These include the adiabatic transition from inspiral to plunge to merger, the extremely short merger phase, and the absence of high-frequency features in the merger waveforms, the burst of radiation produced at merger being filtered by the potential barrier surrounding the newborn black hole. The flexibility of the model and the fact that the quasi-circular

---

[6] The Kerr metric describes the geometry of spacetime around a rotating massive body.

[7] The light ring is the unstable circular orbit of massless particles such as photons and gravitons in Schwarzschild or Kerr spacetime. The Schwarzschild metric describes the geometry of spacetime around a non-rotating massive body.

[8] The close-limit approximation assumes that the merged object can be approximated as a perturbed black hole or neutron star during the ring-down phase of the coalescence.

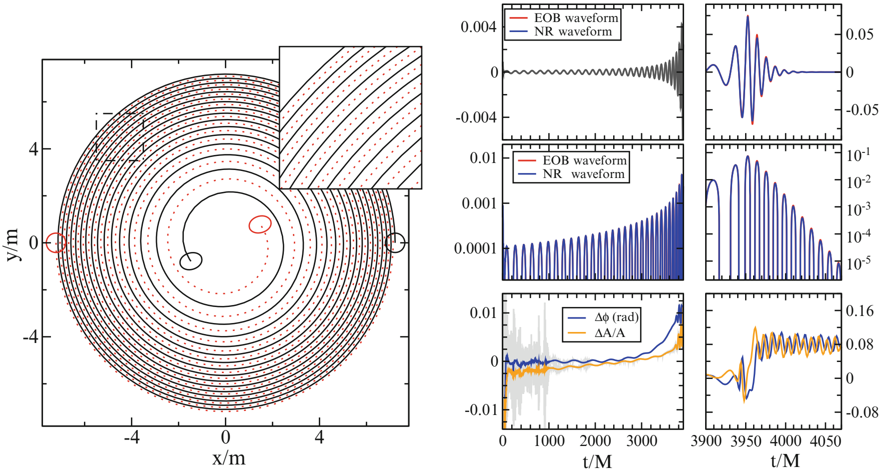


Fig. 5.1 *Left panel*: Trajectories for the evolution of an equal-mass, non-spinning black hole computed by the Caltech/Cornell/CITA group using their spectral code (Reprinted figure with permission from Boyle et al. [20]; http://link.aps.org/abstract/PRD/v76/p124038. Copyright (2007) by the American Physical Society) *Right panel*: Comparison between the numerical-relativity waveform computed by the Caltech/Cornell/CITA group and the calibrated effective-one-body (EOB) waveform. $\Delta\phi$ and $\Delta A/A$ are the phase difference and fractional amplitude difference between the numerical and analytical waveforms (Reprinted figure with permission from Buonanno et al. [34]; http://link.aps.org/abstract/PRD/v79/p124028. Copyright (2009) by the American Physical Society)

motion is described mostly by the radial potential, makes it possible to change the duration of the plunge, the positions of the innermost stable circular orbit, the light ring and the horizon by properly reshaping the potential. This is achieved by including unknown higher-order PN terms in the potential and calibrating them to numerical-relativity simulations. In this way effective-one-body waveforms reproduce numerical-relativity waveforms within the numerical error (see Fig. 5.1).

The analyses and theoretical progress made in Refs. [16, 21, 32–34, 48, 49, 53–55, 86–88, 91], have demonstrated that it is possible to devise and calibrate analytical effective-one-body waveforms for use in detection protocols. This is crucial, since thousands of waveform templates need to be computed to extract the signal from the noise, an impossible demand for numerical relativity alone. The effective-one-body templates developed in Ref. [33] have been used in LIGO to search for the first time for merging black holes. Furthermore, phenomenological templates [9, 94] have also been developed and used in LIGO searches.

When black holes in binary systems carry intrinsic rotation or spin, and the spins are not aligned with the orbital angular momentum, the spins induce precession of the orbital plane. This adds substantial complexity to the gravitational waveforms, extending the parameter space and increasing substantially the number of templates. Given the high computational cost of running numerical simulations, template construction directly based on numerical simulations is currently impractical. The goal is to produce a smaller survey of numerical waveforms and use them to calibrate analytical templates which can then be used to generate efficiently and faithfully tens of thousands of templates. To achieve this goal the numerical and analytical relativity (NRAR) communities have formed the NRAR collaboration [107]. The NRAR has been awarded computational time by the NSF and is currently performing simulations on a large region of the parameter space, focusing on spinning black-hole systems.

The work at the interface between numerical and analytical relativity is also having an impact in astrophysics. In fact, today, thanks to analytical formulae obtained in PN theory and calibrated to numerical results, we can sample a large volume of black-hole parameter space and predict the distribution of recoil velocities attainable in black-hole mergers [36, 37, 61, 62]. This has important implications for the hierarchical growth of supermassive black holes and first galaxies.

3.2 A Simple Picture. Where Are the Non Linearities?

The numerical-relativity results have revealed an intriguing simplicity in the transition from inspiral to merger and from there to ringdown. This simplicity was anticipated by studies in the 1970s, by the close-limit approximation, the effective-one-body approach, and the Lazarus project. However, this expectation of a simple result was not shared by many people. The majority of the gravitational-wave community expected to see complex merger waveforms with high-frequency details.

The real picture turns out to be quite the opposite. After a long adiabatic quasi-circular inspiral, the two objects follow a rather blurred innermost stable circular orbit and make a still adiabatic plunge. As they approach each other, their motion generates rotational gravitational perturbations around the light ring which eventually excite by resonance the quasi-normal modes of the newborn black hole.[9] Once the two black holes are inside the potential barrier which peaks around the light ring, the direct gravitational radiation from the two holes is strongly filtered by the potential barrier. Part of the energy produced in the strong merger-burst remains stored in the resonant cavity of the geometry, i.e., inside the potential barrier, and what is released outside is just the ringdown signal.

Does this simplicity mean that the non-linearities of general relativity are absent? Not at all! Comparisons with analytical PN models and the effective-one-body model during the last 15 orbits of evolution have demonstrated that the best

---

[9] In Schwarzschild, in the eikonal approximation ($\ell \gg 1$, $\ell \sim m$), the frequency of the least damped quasi-normal mode is related to the orbital frequency at the light ring through $\omega_{0\ell m} \sim l\,\omega_{\mathrm{light-ring}}$, and the gravitational modes with $\ell \sim m$ peak around the light ring. However, gravitational modes with $\ell \neq m$ behave differently.

agreement with the numerical-relativity results is obtained when corrections up to the highest PN order available today are included. Thus, as expected, non-linear effects are present and dominant in the strong-field phase. The waveform simplicity is the result of (i) the absence of characteristic scales close to merger when radiation reaction, orbital and precession time scales become of the same order of magnitude; (ii) the formation of a potential barrier filtering the direct radiation from the merger burst, and (iii) the highly dissipative nature of disturbances in black-hole spacetime.

3.3 Evolution of Binary Systems in the Presence of Matter and Electromagnetic Fields

The breakthrough of numerical relativity in 2005 has also opened the possibility of investigating several physical processes produced by binary systems that are coupled to matter and/or electromagnetic fields.

Binary systems containing neutron stars have been simulated for a variety of equations of state, mass ratios and spins [58, 59, 93, 97, 98] (see Fig. 5.2). In the case of a black-hole/neutron-star binary, an accretion disk can form. Depending on the mass of the accretion disk, the disk can engine the production of short gamma-ray bursts [59]. LIGO detectors have searched for short-duration gravitational-wave bursts associated with a large number of gamma-ray bursts, but have found no evidence of gravitational radiation [5]. LIGO detectors have also searched for gravitational radiation from GRB 070201, the electromagnetically determined sky position of which was coincident with the spiral arms of the Andromeda Galaxy [4]. Again, no signal was found.

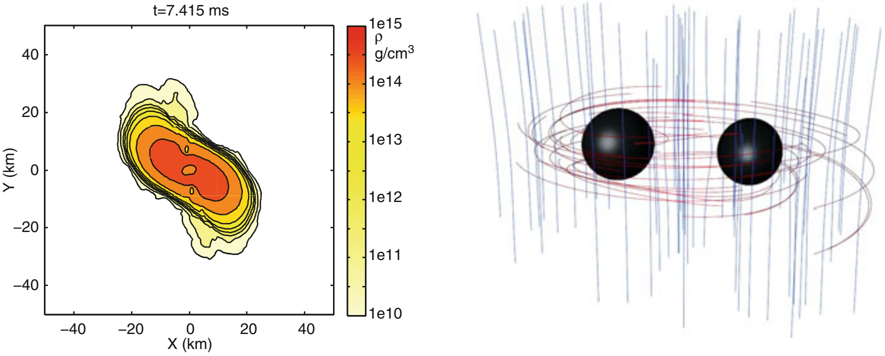


Fig. 5.2 *Left panel*: Isodensity contours during the merger of two neutron stars [93]. *Right panel*: Magnetic and electric field lines around inspiralling equal-mass non-spinning black holes prior to merger (Reprinted figures with permission from C. Palenzuela, L. Lehner and S. Yoshida [85]; http://link.aps.org/abstract/PRD/v81/p084007. Copyright (2010) by the American Physical Society) The electric field lines are twisted around the black holes, while the magnetic fields are mostly aligned with the direction perpendicular to the orbital plane

If a binary black hole is surrounded by an accretion disk, the binary dynamics, notably the energy lost during coalescence and/or the recoil velocity gain by the newborn black hole, can affect the disk – creating shocks and thus electromagnetic signals. The latter will be a counterpart signal to the gravitational-wave signal. These processes are of great interest in astrophysics, but also in cosmology because if observed and if the host galaxy is identified, we could extract the redshift, and measure from the gravitational-wave observation the luminosity distance of the binary.

Quite recently, numerical relativity simulations have contributed to the understanding of mechanisms predicted in the 1970s, such as the Blandford-Znajek effect [18]. In Ref. [84] the authors have simulated a binary black hole coupled to the electromagnetic field of the circumbinary disk and found dual jets. The latter are due to the electromagnetic field extracting energy from the orbiting black holes. Finally, the electromagnetic field can be affected by the orbiting black holes. The binary's dynamics can induce variability in possible electromagnetically induced emissions during the merger epoch [85, 95] (see Fig. 5.2).

4 If Mountains on Neutron Stars Were Centimeters Tall

As reviewed in Sect. 2, at leading order gravitational waves are generated by the variation in time of the quadrupole moment of the source. The rapid rotation of highly magnetized neutron stars (or pulsars) can produce gravitational waves if the pulsar shape deviates from axisymmetry. It is common to quantify the oblateness of the pulsar through the ellipticity parameter $\varepsilon = (I_1 - I_2)/I_3$ where I_i with $i = 1, 2, 3$ are the principal moments of inertia of the rotating body [22].

If the pulsar rotates around one of its principal axes, the gravitational-wave signal is emitted at twice the pulsar rotation frequency. The value of the ellipticity depends on the neutron star properties, in particular the maximum strain that can be supported by its crust. Pulsars are thought to form in supernova explosions. The outer layers of the star crystallize as the newborn pulsar cools by neutrino emission. Anisotropic stresses during this phase could lead to values $\varepsilon \lesssim 10^{-6}$ although with exotic equations of state $\varepsilon \simeq 10^{-5}$–$10^{-4}$. Values of ε on the order of 10^{-6} correspond to perturbations or mountains on the surface of the neutron star a few centimeters tall!

The detection of continuous, monochromatic frequency waves is achieved in gravitational-wave detectors by constructing power spectrum estimators and searching for statistically significant peaks at fixed frequencies for very long time. If T is the observation time, the signal-to-noise ratio grows like $\sqrt{T}$. The detection is complicated by the fact that the signal received at the detector is not perfectly monochromatic due to the Earth's motion. Because of Doppler shifts in frequency, the spectral lines of fixed frequency sources spread power into many Fourier bins about the observed frequency. Given the possibility that the strongest sources of continuous gravitational waves may be electromagnetically invisible or previously undiscovered, an all sky, all frequency search for such unknown sources is very important, though computationally very expensive.

Quite interestingly, LIGO and Virgo have been able to set astrophysically relevant limits on the ellipticity of the Crab pulsar [3], and also to the neutron star in the supernova remnant Cassiopea A [2].

5 Waiting for the Next Supernova in the Milky Way or Nearby Galaxies

Supernovae are triggered by the violent collapse of a stellar core,[10] and eventually form a neutron star, or a black hole. The core collapse proceeds extremely quickly, lasting less than a second and the dense fluid of the core undergoes motions with relativistic speeds. Small deviations from spherical symmetry during this phase can generate gravitational waves (see e.g., Refs. [11, 83] and references therein).

Numerical simulations have predicted strains on the order of $h = 6 \times 10^{-17}\sqrt{\eta_{\rm eff}}\,(M/M_\odot)^{1/2}\,(10\,{\rm kpc}/R)^{1/2}\,(1\,{\rm kHz}/f)\,(10\,{\rm ms}/\tau_{\rm collapse})^{1/2}$ where M is the mass of the collapsed star, R is the distance to the source, and f is the frequency of the burst of radiation. Finally, if a certain amount of energy is released in gravitational waves during the explosion, we have $\Delta E_{\rm GW} = \eta_{\rm eff}\,M\,c^2$, with typically $\eta_{\rm eff} \sim 10^{-9}$–$10^{-5}$. LIGO and Virgo in their advanced configuration could detect signals from core-collapse supernovae in the Milky Way and nearby galaxies. Current results of all-sky searches for gravitational-wave bursts with LIGO-GEO-Virgo are summarized in Ref. [1].

More recent results indicate that gravitational waves could also be produced by neutrino emission during the supernovae explosion. In this case, the gravitational-wave signal would extend toward lower frequencies $\sim 10\,{\rm Hz}$ [79]. Moreover, the superposition of independent gravitational-wave signals from supernovae at cosmological distances may give rise to a stochastic gravitational-wave background. While the estimates remain uncertain within several orders of magnitude, this background may become detectable with future detectors in space [30].

After a supernovae explosion or a collapsar a significant amount of the ejected material can fall back, subsequently heating the neutron star or spinning the black hole. Quasi-normal modes can be excited in this process. There is also the possibility that the collapsed material might fragment into clumps that orbit for some cycles like a binary system or form bar-like structures that also produce gravitational-wave signals.

---

[10] Except for supernovae Ia.

6 Disclosing the Dark Age of the Early Universe

What we know with confidence today about the early Universe goes back as far as the big-bang nucleosynthesis time when light elements first formed, the Universe had a temperature around 1 MeV, it was ~ 1 s old and dominated by radiation. The cosmic microwave background radiation (CMB) that we measure today with amazing accuracy was emitted when the Universe had a temperature around 1 eV, and it was $\sim 10^5$–10^6 years old. We have never detected any relic background produced prior to the time the CMB was generated. Even the background of cosmological neutrinos emitted at the time the Universe had a temperature of 1 MeV has never been observed.

Gravitational waves interact very weakly with matter. If they were produced in the *dark age*[11] – that is, the age prior to big bang nucleosynthesis when we know the Universe was radiation dominated – they would travel unchanged and provide us with a snapshot of the Universe at that time.

A stochastic background of gravitational waves could have been produced by the rapidly changing gravitational field during the inflation stage through the mechanism of amplification of quantum vacuum fluctuations [63, 99]. Today this background would span the frequency range 10^{-16}–10^{10} Hz, which covers the frequency band of current and future detectors on the ground and in space. The cosmological gravitational-wave background could have left signatures in the CMB polarization that could be detectable with future CMB probes.

Gravitational waves could have been generated at the end of inflation during the preheating phase [71]. The latter is a highly non-thermal phase during which transient density inhomogeneities are created whose time-changing mass multipoles emit gravitational waves. Cosmic strings or superstrings could have formed in the early Universe at symmetry-breaking phase transitions or at the end of brane inflation scenarios [43,50]. If so, due to their large tension (e.g., $\sim 10^{22}\,\mathrm{g/cm}$), loops of strings will oscillate relativistically, emit gravitational waves, shrink in size and disappear, but they will be constantly replaced by small loops broken off very long loops (longer than Hubble radius). Moreover, if strong first order phase transition occurred in the early Universe, bubbles of true vacuum will form, travel and collide with each other producing gravitational waves [66].

The signals described above carry information on otherwise unexplored physics between $\sim 10^9$ GeV and $\sim 10^{16}$ GeV. The detection of gravitational waves from the dark age would be a revolutionary discovery.

LIGO and Virgo detectors have achieved sensitivities such that they can start excluding regions of the parameter space of the expected signals and constrain the equation of state of the Universe during the dark age [19] in a physically significant way [7, 8].

---

[11] Note that dark age is also used in astronomy to denote the epoch before stars formed.

7 How to Beat the Standard Quantum Limit in Gravitational-Wave Detectors

Current gravitational-wave detectors are highly accurate position-measurement instruments. They are already so sensitive that limitations from quantum mechanics must be considered when upgrading them. As pointed out by Braginsky in his seminal papers of the 1970s, the Heisenberg uncertainty principle, if applied naively to the test masses in gravitational-wave interferometers, produces a free-mass standard quantum limit on the interferometer's sensitivity: the more accurately one measures the test-mass displacement at a given time, the larger the disturbance that is imposed on the velocity and the less accurately one can measure the test-mass displacement at later times. It is however possible to circumvent the standard-quantum limit by changing the optical design of the instrument and introducing appropriate readout schemes (quantum–non-demolition techniques).

The last 40 years have seen major developments in the area of quantum non demolition. The initial works by Caves and collaborators [38–40] introduced the formalism to describe the quantum optical noise in gravitational-wave detectors, notably shot noise and radiation-pressure noise.[12] In subsequent years scientists have proposed several schemes to beat the standard quantum limit [23, 24, 72], including ways of injecting squeezed light with frequency-dependent squeezed angles in laser interferometer gravitational-wave detectors to reduce the shot noise.

Quite surprisingly for the research community, at the beginning of 2000, Refs. [25, 26] found that the optical configuration of the next generation of LIGO detectors (i.e., advanced LIGO) can already beat the free-mass standard-quantum limit, provided thermal noise can be suppressed sufficiently. This is because of correlations between photon shot noise and radiation pressure noise created by the signal-recycling cavity in the advanced-LIGO optical configuration. The study in Refs. [25, 26] revealed an optomechanical effect subsequently verified experimentally in the 40-m interferometer at Caltech [78] and in table-top optical-cavity experiments at MIT [44]. This effect was termed in Ref. [25] the *optical spring* effect: The dynamics of the system composed of arm-cavity mirrors and optical fields resemble those of a free test mass (mirror motion) connected to a massive spring (optical fields), which can resonate at two pairs of finite frequencies. As the light power is increased the (coupled) mechanical resonant frequency moves away from zero, while the (coupled) optical resonant frequency does not vary much, being present already as pure optical resonance in the limit of low light power (see Fig. 5.3). Near these resonances the noise curve can beat the free

---

[12] Quantum-vacuum *phase* fluctuations in the interferometer result in photon shot noise. Quantum-vacuum *amplitude* fluctuations in the interferometer produce fluctuating radiation pressure noise on the test-mass mirrors. The standard quantum limit characterizes the regime in which the quantum measurement error (shot noise) becomes equal to the back action noise (radiation pressure noise).

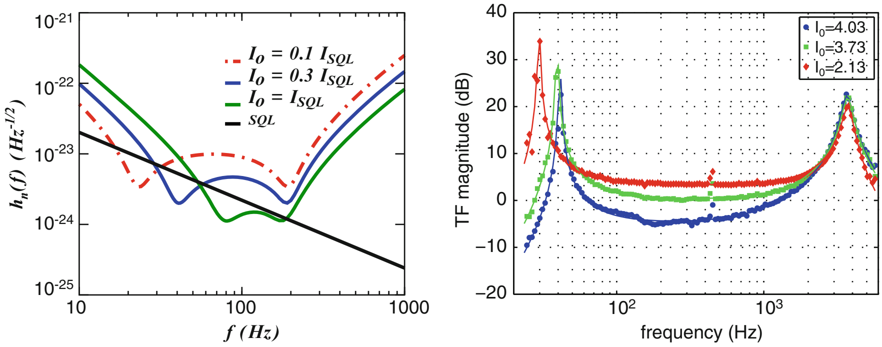


Fig. 5.3 *Left panel*: The square root of the quantum-noise spectral density for advanced-LIGO versus frequency for various choices of the light power. The standard-quantum limit line is also shown for comparison. *Right panel*: The magnitude response of the 40 m interferometer at Caltech in the same optical configuration of advanced-LIGO, for different values of the light power in Watts (Reprinted figures with permission from Miyakawa et al. [78]; http://link.aps.org/abstract/PRD/v74/p022001. Copyright (2006) by the American Physical Society)

mass standard-quantum limit. This phenomenon is not unique to signal-recycling interferometers; it is a generic feature of detuned cavities [67] and was used in proposing the *optical bar* gravitational-wave detectors [24].

During the last several years, a variety of optical configurations have been designed to beat the free-mass standard quantum limit in detectors beyond advanced LIGO [27,41,64,68,69,72,90,92,113]. More recently, theoretical and experimental work has also focused on designing gravitational-wave detectors or table-top experiments that can probe optomechanical effects, work toward the standard quantum limit, and test quantum mechanics with macroscopic objects [6, 45, 46, 70, 77, 80, 81].[13] Finally, several experiments [75, 114, 115] have demonstrated the possibility of creating and controlling squeezed light in the frequency band of gravitational-wave detectors.

Today, research groups working on mechanical systems ranging in size from nanometer-scale oscillators [76, 82], to centimeter-scale optical cavities [10, 45, 60, 73], to kilometer-scale gravitational-wave detectors [6] are all approaching a regime in which either the mechanical system or its interaction with the environment must be described quantum mechanically. Thus, researchers from backgrounds as diverse as astrophysics, mesoscopic condensed matter physics, and quantum optics are converging on common goals related to quantum effects in mechanical systems.

---

[13] A typical LIGO/Virgo mirror weighs ∼10–30 kg, while mirrors employed in table top experiments weigh on the order of a few grams.

8 The Universe Viewed Through Gravitational Waves

What can we learn about astrophysics, cosmology and gravity through the direct detection of gravitational waves? And how will the Universe look like from a gravitational-wave observatory?[14]

Gravitational waves encode detailed information about the sources that have generated them. On one hand, the waveform's sensitivity to the source's parameters increases the waves' complexity and their number, making harder to catch them. On the other hand, the dependence of the signal shape on the source's parameters is a blessing. The detection of gravitational waves from coalescing black holes will tell us how *heavy* each of the black holes was, how *fast* the holes were spinning, the *shape* of their *orbit* (circular? elongated?), *where* the holes were in the sky, and *how far* they were from the Earth [112].

The observation of gravitational signals from binary systems composed of a neutron-star and black hole would allow us to measure the equation of state of the neutron star, because the signal would have specific signatures depending on how the neutron star is tidally disrupted by the companion. By measuring the frequency and decay time of the quasi-normal modes during the ringdown, and the binary parameters during the long inspiral, we could infer whether the object formed through merger was a black hole or not. By localizing the binary in the sky with high precision, and by associating it with an electromagnetic counterpart, we could compute the luminosity distance and use binary systems as standard candles (or sirens) to extract cosmological parameters [42, 74]. In the absence of an electromagnetic counterpart, the detection of a large number of binary systems may still allow us to extract the Hubble parameter and do cosmology [96]. Accurate measurements of gravitational waves would allow to test general relativity in the strong-field regime.

With the formation of large collaborations like the LIGO Scientific Collaboration and Virgo Collaboration, and the efforts of the Numerical INJection Analysis (NINJA) [106] and NRAR, the gravitational-wave community has entered a new era. The construction of gravitational-wave observatories, the solution of the two-body problem in general relativity and its analytic description, and the discovery of new techniques to beat the standard quantum limit in gravitational-wave detectors, have established new synergies and areas of research at the interplay between numerical relativity, data analysis, astrophysics and analytical relativity, and also between high-precision measurements, quantum optics, and experimental gravity.

It is worth considering that the field of gravitational waves is perhaps the only field in theoretical and experimental physics that has achieved such remarkable

---

[14] If gravitational-wave detectors operating in the frequency band $\sim$10–10^2 Hz were converting online gravitational signals from merging binary systems into audial signals, our ears would hear *chirping* signals – that is, a signal with increasing amplitude and frequency – lasting a few seconds or a few minutes. Audial signals from pulsars would be periodic and continuous, lasting the entire life of the observatory. Pursuing these signals would certainly be an entertaining activity.

results, developed such unprecedented and sophisticated technology, and opened new areas of research, before its *holy grail* — the gravitational waves themselves – have even been found. It is the great eagerness of the research community for the knowledge that gravitational waves will provide that keep the community together and drives its research forward.

Whenever research has opened a new window on the Universe, we have found surprises that have revolutionized and enriched our understanding. The vistas we will see through gravitational waves will surely afford similar revelations.

Acknowledgements The author acknowledges support from NSF Grant PHY-0903631 and NASA grant NNX09AI81G. The author thanks Walter Jacob for suggestions which helped to improve the prose of the manuscript.

References

1. J. Abadie et al., All-sky search for gravitational-wave bursts in the first joint LIGO-GEO-Virgo run. Phys. Rev. **D81**, 102001 (2010)
2. J. Abadie et al., First search for gravitational waves from the youngest known neutron star. Astrophys. J. **722**, 1504–1513 (2010)
3. B. Abbott et al., Beating the spin-down limit on gravitational wave emission from the Crab pulsar. Astrophys. J. **683**, L45–L50 (2008)
4. B. Abbott et al., Implications for the origin of GRB 070201 from LIGO observations. Astrophys. J. **681**, 1419–1428 (2008)
5. B. Abbott et al., Search for Gravitational waves associated with 39 gamma-ray bursts using data from the second, third, and fourth LIGO runs. Phys. Rev. **D77**, 062004 (2008)
6. B. Abbott et al., Observation of a kilogram-scale oscillator near its quantum ground state. New J. Phys. **11**, 073032 (2009)
7. B.P. Abbott et al., An upper limit on the stochastic gravitational-wave background of cosmological origin. Nature **460**, 990 (2009)
8. B.P. Abbott et al., First LIGO search for gravitational wave bursts from cosmic (super)strings. Phys. Rev. **D80**, 062002 (2009)
9. P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A.M. Sintes, J.T. Whelan, B. Brügmann, P. Diener, N. Dorband, J. Gonzalez, M. Hannam, S. Husa, D. Pollney, L. Rezzolla, L. Santamaría, U. Sperhake, J. Thornburg, Template bank for gravitational waveforms from coalescing binary black holes: nonspinning binaries. Phys. Rev. D **77**(10), 104017 (2008)
10. O. Arcizet et al., Radiation-pressure cooling and optomechanical instability of a micromirror. Nature **444**, 71 (2006)
11. L. Baiotti, I. Hawke, L. Rezzolla, On the gravitational radiation from the collapse of neutron stars to rotating black holes. Class. Quant. Grav. **24**, S187–S206 (2007)
12. J. Baker, B. Brügmann, M. Campanelli, C.O. Lousto, R. Takahashi, Plunge waveforms from inspiralling binary black holes. Phys. Rev. Lett. **87**(12), 121103 (2001)
13. J. Baker, M. Campanelli, C.O. Lousto, The Lazarus project: a pragmatic approach to binary black hole evolutions. Phys. Rev. D **65**, 044001 (2002)
14. J.G. Baker, J. Centrella, D.I. Choi, M. Koppitz, J. van Meter, Gravitational-wave extraction from an inspiraling configuration of merging black holes. Phys. Rev. Lett. **96**(11), 111102 (2006)
15. J.G. Baker, M. Campanelli, F. Pretorius, Y. Zlochower, Comparisons of binary black hole merger waveforms. Class. Quantum Grav. **24**, S25–S31 (2007)

16. E. Barausse, A. Buonanno, An improved effective-one-body Hamiltonian for spinning black-hole binaries. Phys. Rev. **D81**, 084024 (2010)
17. L. Blanchet, Gravitational radiation from post-Newtonian sources and inspiralling compact binaries. Living Rev. Rel. **9**(4) (2006)
18. R.D. Blandford, R.L. Znajek, Electromagnetic extractions of energy from Kerr black holes. Mon. Not. R. Astron. Soc. **179**, 433–456 (1977)
19. L.A. Boyle, A. Buonanno, Relating gravitational wave constraints from primordial nucleosynthesis, pulsar timing, laser interferometers, and the CMB: implications for the early universe. Phys. Rev. **D78**, 043531 (2008)
20. M. Boyle, D.A. Brown, L.E. Kidder, A.H. Mroué, H.P. Pfeiffer, M.A. Scheel, G.B. Cook, S.A. Teukolsky, High-accuracy comparison of numerical relativity simulations with post-Newtonian expansions. Phys. Rev. D **76**, 124038 (2007)
21. M. Boyle, A. Buonanno, L.E. Kidder, A.H. Mroué, Y. Pan, H.P. Pfeiffer, M.A. Scheel, High-accuracy numerical simulation of black-hole binaries: computation of the gravitational-wave energy flux and comparisons with post-Newtonian approximants. Phys. Rev. D **78**(12), 104020 (2008)
22. P.R. Brady, T. Creighton, C. Cutler, B.F. Schutz, Searching for periodic sources with LIGO. Phys. Rev. **D57**, 2101–2116 (1998)
23. V.B. Braginsky, F.Y. Khalili, Gravitational wave antenna with QND speed meter. Phys. Lett. **A147**, 251–256 (1990)
24. V.B. Braginsky, M.L. Gorodetsky, F.Y. Khalili, Optical bars in gravitational wave antennas. Phys. Lett. **A232**, 340–348 (1997)
25. A. Buonanno, Y. Chen, Signal recycled laser-interferometer gravitational-wave detectors as optical springs. Phys. Rev. **D65**, 042001 (2002)
26. A. Buonanno, Y.B. Chen, Quantum noise in second generation, signal recycled laser interferometric gravitational wave detectors. Phys. Rev. **D64**, 042006 (2001)
27. A. Buonanno, Y.B. Chen, Improving the sensitivity to gravitational-wave sources by modifying the input-output optics of advanced interferometers. Phys. Rev. **D69**, 102004 (2004)
28. A. Buonanno, T. Damour, Effective one-body approach to general relativistic two-body dynamics. Phys. Rev. D **59**(8), 084006 (1999)
29. A. Buonanno, T. Damour, Transition from inspiral to plunge in binary black hole coalescences. Phys. Rev. D **62**(6), 064015 (2000)
30. A. Buonanno, G. Sigl, G.G. Raffelt, H.T. Janka, E. Muller, Stochastic gravitational wave background from cosmological supernovae. Phys. Rev. **D72**, 084001 (2005)
31. A. Buonanno, Y. Chen, T. Damour, Transition from inspiral to plunge in precessing binaries of spinning black holes. Phys. Rev. D **74**(10), 104005 (2006)
32. A. Buonanno, G.B. Cook, F. Pretorius, Inspiral, merger, and ring-down of equal-mass black-hole binaries. Phys. Rev. D **75**(12), 124018 (2007)
33. A. Buonanno, Y. Pan, J.G. Baker, J. Centrella, B.J. Kelly, S.T. McWilliams, J.R. van Meter, Approaching faithful templates for non-spinning binary black holes using the effective-one-body approach. Phys. Rev. D **76**, 104049 (2007)
34. A. Buonanno, Y. Pan, H.P. Pfeiffer, M.A. Scheel, L.T. Buchman, L.E. Kidder, Effective-one-body waveforms calibrated to numerical relativity simulations: coalescence of non-spinning, equal- mass black holes. Phys. Rev. D **79**, 124028 (2009)
35. M. Campanelli, C.O. Lousto, P. Marronetti, Y. Zlochower, Accurate evolutions of orbiting black-hole binaries without excision. Phys. Rev. Lett. **96**(11), 111101 (2006)
36. M. Campanelli, C.O. Lousto, Y. Zlochower, D. Merritt, Large merger recoils and spin flips from generic black hole binaries. Astrophys. J. Lett. **659**(1), L5–L8 (2007)
37. M. Campanelli, C.O. Lousto, Y. Zlochower, D. Merritt, Maximum gravitational recoil. Phys. Rev. Lett. **98**, 231102 (2007)
38. C.M. Caves, Quantum mechanical noise in an interferometer. Phys. Rev. **D23**, 1693–1708 (1981)
39. C.M. Caves, B.L. Schumaker, New formalism for two-photon quantum optics. 1. Quadrature phases and squeezed states. Phys. Rev. **A31**, 3068–3092 (1985)

40. C.M. Caves, K.S. Thorne, R.W.P. Drever, V.D. Sandberg, M. Zimmermann, On the measurement of a weak classical force coupled to a quantum mechanical oscillator. I. Issues of principle. Rev. Mod. Phys. **52**, 341–392 (1980)
41. Y.b. Chen, Sagnac interferometer as a speed-meter-type, quantum- nondemolition gravitational-wave detector. Phys. Rev. **D67**, 122004 (2003)
42. D.F. Chernoff, L.S. Finn, Gravitational radiation, inspiraling binaries, and cosmology. Astrophys. J. **411**, L5–L8 (1993)
43. E.J. Copeland, R.C. Myers, J. Polchinski, Cosmic F- and D-strings. J. High Energy Phys. **06**, 013 (2004)
44. T. Corbitt et al., A squeezed state source using radiation pressure induced rigidity. Phys. Rev. A **73**, 023801 (2005)
45. T. Corbitt et al., An all-optical trap for a gram-scale mirror. Phys. Rev. Lett. **98**, 150802 (2007)
46. T. Corbitt et al., Optical dilution and feedback cooling of a gram-scale oscillator to 6.9 mK. Phys. Rev. Lett. **99**, 160801 (2007)
47. T. Damour, Coalescence of two spinning black holes: an effective one-body approach. Phys. Rev. D **64**(12), 124013 (2001)
48. T. Damour, A. Nagar, Comparing effective-one-body gravitational waveforms to accurate numerical data. Phys. Rev. D **77**(2), 024043 (2008)
49. T. Damour, A. Nagar, Improved analytical description of inspiralling and coalescing black-hole binaries. Phys. Rev. D **79**, 081503 (2009)
50. T. Damour, A. Vilenkin, Gravitational radiation from cosmic (super)strings: bursts, stochastic background, and observational windows. Phys. Rev. **D71**, 063510 (2005)
51. T. Damour, P. Jaranowski, G. Schäfer, Determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation. Phys. Rev. D **62**(8), 084011 (2000)
52. T. Damour, P. Jaranowski, G. Schäfer, Effective one body approach to the dynamics of two spinning black holes with next-to-leading order spin-orbit coupling. Phys. Rev. D **78**, 024009 (2008)
53. T. Damour, A. Nagar, E.N. Dorband, D. Pollney, L. Rezzolla, Faithful effective-one-body waveforms of equal-mass coalescing black-hole binaries. Phys. Rev. D **77**(8), 084017 (2008)
54. T. Damour, A. Nagar, M. Hannam, S. Husa, B. Brügmann, Accurate effective-one-body waveforms of inspiralling and coalescing black-hole binaries. Phys. Rev. D **78**, 044039 (2008)
55. T. Damour, B.R. Iyer, A. Nagar, Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries. Phys. Rev. D **79**, 064004 (2009)
56. A. Einstein, Sitzber. Preuss. Akad. Wiss. **688** (1916)
57. A. Einstein, Sitzber. Preuss. Akad. Wiss. **154** (1918)
58. Z.B. Etienne, Y.T. Liu, S.L. Shapiro, T.W. Baumgarte, Relativistic simulations of black hole-neutron star mergers: effects of black-hole spin. Phys. Rev. **D79**, 044024 (2009)
59. F. Foucart, M.D. Duez, L.E. Kidder, S.A. Teukolsky, Black hole-neutron star mergers: effects of the orientation of the black hole spin. Phys. Rev. D *83*, 024005 (2010)
60. S. Gigan et al., Self-cooling of a micromirror by radiation pressure. Nature **444**, 67 (2006)
61. J.A. Gonzalez, M.D. Hannam, U. Sperhake, B. Brügmann, S. Husa, Supermassive recoil velocities for binary black-hole mergers with antialigned spins. Phys. Rev. Lett. **98**, 231101 (2007)
62. J.A. Gonzalez, U. Sperhake, B. Brügmann, M. Hannam, S. Husa, Maximum kick from nonspinning black-hole binary inspiral. Phys. Rev. Lett. **98**, 091101 (2007)
63. L.P. Grishchuk, Amplification of gravitational waves in an istropic universe. Sov. Phys. JETP **40**, 409–415 (1975)
64. J. Harms et al., Squeezed-input, optical-spring, signal-recycled gravitational wave detectors. Phys. Rev. **D68**, 042001 (2003)
65. R.A. Hulse, J.H. Taylor, Discovery of a pulsar in a binary system. Astrophys. J. **195**, L51–L53 (1975)
66. M. Kamionkowski, A. Kosowsky, M.S. Turner, Gravitational radiation from first order phase transitions. Phys. Rev. **D49**, 2837–2851 (1994)

67. F. Khalili, Frequency dependent rigidity in large scale interferometric gravitational wave detectors. Phys. Lett. **A288**, 251–256 (2001)
68. F.Y. Khalili, V.I. Lazebny, S.P. Vyatchanin, Sub-SQL sensitivity via optical rigidity in advanced LIGO interferometer with optical losses. Phys. Rev. **D73**, 062002 (2006)
69. F.Y. Khalili, H.X. Miao, Y.B. Chen, Increasing the sensitivity of future gravitational-wave detectors with double squeezed-input. Phys. Rev. **D80**, 042006 (2009)
70. F. Khalili et al., Preparing a mechanical oscillator in non-Gaussian quantum states. Phys. Rev. Lett. **105**(7), 070403 (2010)
71. S.Y. Khlebnikov, I.I. Tkachev, Relic gravitational waves produced after preheating. Phys. Rev. **D56**, 653–660 (1997)
72. H.J. Kimble, Y. Levin, A.B. Matsko, K.S. Thorne, S.P. Vyatchanin, Conversion of conventional gravitational wave interferometers into QND interferometers by modifying their input and/or output optics. Phys. Rev. **D65**, 022002 (2002)
73. A. Kleckner, D. Bouwmeester, Sub-kelvin optical cooling of a micromechanical resonator. Nature **444**, 75 (2006)
74. D. Markovic, On the possibility of determining cosmological parameters from measurements of gravitational waves emitted by coalescing, compact binaries. Phys. Rev. **D48**, 4738–4756 (1993)
75. K. Mckenzie, D.A. Shaddock, D.E. McClelland, B.C. Buchler, P.K. Lam, Experimental demonstration of a squeezing enhanced power recycled Michelson interferometer for gravitational wave detection. Phys. Rev. Lett. **88**, 231102 (2002)
76. C. Metzger, K. Karrai, Cavity cooling of a microlever. Nature **432**, 1002 (2004)
77. H. Miao et al., Probing macroscopic quantum states with a sub-Heisenberg accuracy. Phys. Rev. **A81**, 012114 (2010)
78. O. Miyakawa et al., Measurement of optical response of a detuned resonant sideband extraction interferometer. Phys. Rev. **D74**, 022001 (2006)
79. E. Mueller, M. Rampp, R. Buras, H.T. Janka, D.H. Shoemaker, Towards gravitational wave signals from realistic core collapse supernova models. Astrophys. J. **603**, 221–230 (2004)
80. H. Mueller-Ebhardt et al., Entanglement of macroscopic test masses and the standard quantum limit in laser interferometry. Phys. Rev. Lett. **100**, 013601 (2008)
81. H. Mueller-Ebhardt et al., Quantum state preparation and macroscopic entanglement in gravitational-wave detectors. Phys. Rev. **A80**, 043802 (2009)
82. A. Naik et al., Cooling a nanomechanical resonator with quantum back-action. Nature **443**, 193 (2006)
83. C.D. Ott, Probing the core-collapse supernova mechanism with gravitational waves. Class. Quant. Grav. **26**, 204015 (2009)
84. C. Palenzuela, L. Lehner, S.L. Liebling, Dual jets from binary black holes. Science **329**, 927 (2010)
85. C. Palenzuela, L. Lehner, S. Yoshida, Understanding possible electromagnetic counterparts to loud gravitational wave events: binary black hole effects on electromagnetic fields. Phys. Rev. **D81**, 084007 (2010)
86. Y. Pan, A. Buonanno, J.G. Baker, J. Centrella, B.J. Kelly, S.T. McWilliams, F. Pretorius, J.R. van Meter, Data-analysis driven comparison of analytic and numerical coalescing binary waveforms: nonspinning case. Phys. Rev. D **77**(2), 024014 (2008)
87. Y. Pan, A. Buonanno, L. Buchman, T. Chu, L. Kidder, H. Pfeiffer, M. Scheel, Effective-one-body waveforms calibrated to numerical relativity simulations: coalescence of non-precessing, spinning, equal-mass black holes. Phys. Rev. **D81**, 084041 (2010)
88. Y. Pan, A. Buonanno, R. Fujita, E. Racine, H. Tagoshi, Post-Newtonian factorized multipolar waveforms for spinning, non-precessing black-hole binaries. Phys. Rev. D**83**(6), 064003 (2011)
89. F. Pretorius, Evolution of binary black-hole spacetimes. Phys. Rev. Lett. **95**(12), 121101 (2005)
90. P. Purdue, Y.B. Chen, Practical speed meter designs for QND gravitational-wave interferometers. Phys. Rev. **D66**, 122004 (2002)

91. E. Racine, A. Buonanno, L.E. Kidder, Recoil velocity at 2PN order for spinning black hole binaries. Phys. Rev. D **80**, 044010 (2009)
92. H. Rehbein et al., Double optical spring enhancement for gravitational wave detectors. Phys. Rev. **D78**, 062003 (2008)
93. L. Rezzolla, L. Baiotti, B. Giacomazzo, D. Link, J.A. Font, Accurate evolutions of unequal-mass neutron-star binaries: properties of the torus and short GRB engines. Class. Quant. Grav. **27**, 114105 (2010)
94. L. Santamaria et al., Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries. Phys. Rev. **D82**, 064016 (2010)
95. J.D. Schnittman, Electromagnetic counterparts to black hole mergers. Class. Quant. Grav. **28**, (9), 094021 (2011)
96. B.F. Schutz, Determining the hubble constant from gravitational wave observations. Nature **323**, 310–311 (1986)
97. M. Shibata, K. Taniguchi, Merger of black hole and neutron star in general relativity: tidal disruption, torus mass, and gravitational waves. Phys. Rev. **D77**, 084015 (2008)
98. M. Shibata, K. Uryu, Merger of black hole – neutron star binaries: nonspinning black hole case. Phys. Rev. **D74**, 121503 (2006)
99. A.A. Starobinsky, Spectrum of relict gravitational radiation and the early state of the universe. JETP Lett. **30**, 682–685 (1979)
100. The auriga detector. http://www.auriga.lnl.infn.it/
101. The einstein telescope. http://www.et-gw.eu/
102. The explorer and nautilus detectors. http://www.roma1.infn.it/rog/
103. The geo600 collaboration. http://www.geo600.uni-hannover.de
104. The laser interferometer space antenna. http://www.lisa-science.org
105. The ligo scientific collaboration. http://www.ligo.org
106. The ninja collaboration. http://www.gravity.phy.syr.edu/links/ninja.html
107. The nrar collaboration. https://www.ninja-project.org/doku.php?id=nrar:home
108. The planck science team. http://www.rssd.esa.int/index.php?project=Planck
109. The square kilometer array. http://www.skatelescope.org
110. The tama collaboration. http://tamago.mtk.nao.ac.jp
111. The virgo collaboration. http://www.virgo.infn.it
112. K. Thorne, *Black Holes and Time Warps*, 1st edn. (W.W. Norton & Company, New York, 1994)
113. M. Tsang, C.M. Caves, Coherent quantum-noise cancellation for optomechanical Sensors. Phys. Rev. Lett. **105**, 123601 (2010)
114. H. Vahlbruch et al., Demonstration of a squeezed light enhanced power-and signal-recycled Michelson interferometer. Phys. Rev. Lett. **95**, 211102 (2005)
115. H. Vahlbruch et al., Coherent control of vacuum squeezing in the gravitational-wave detection band. Phys. Rev. Lett. **97**, 011101 (2006)
116. J. Weber, Detection and generation of gravitational waves. Phys. Rev. **117**, 306–313 (1960)

Part II
Astronomy in Society

Chapter 6
A History of Observatory Sciences and Techniques

David Aubin

Abstract The aim of this chapter is to present a survey of the issues that have concerned the historians of the observatory sciences over the last few years. As an instituted place of knowledge, the observatory has a longer history than either laboratory or field, but has not been the focus of as many studies. After raising the question "What is an observatory?", some findings about the epistemology of the observatory sciences are discussed. The notion of "observatory techniques" is introduced and discussed in order to account for the diversity of practices to be found in observatories. The history of the observatory in the western world is divided into three periods: (1) the age of the pioneers between the seventeenth and the end of the eighteenth century; (2) a period of triumph and crisis in the nineteenth century during which the observatory was very prominent in the social panorama of science; and (3) the age of specialization that followed. We mostly focus on the first two periods and speculate whether we might recently have entered a new period, especially with the rise of the environmental sciences.

Keywords History and philosophy of astronomy • Sociology of Astronomy

1 Introduction: Visiting the Observatory

King Friedrich Wilhelm III of Prussia was a practical man. While his troops occupied Paris in the spring of 1814, he hastily raced through the Invalides, Notre-Dame, and the Pantheon. His guide was his countryman, the naturalist Alexander von Humboldt who, upon his return from the Americas in 1804, had settled in

---

D. Aubin (✉)
Institut de mathématiques de Jussieu, Université Pierre et Marie Curie, 4,
Place Jussieu, F-75252 Paris Cedex 05, France
e-mail: daubin@math.jussieu.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_6,

Paris to prepare the publication of the wealth of data brought back from his travels. Devoting little more than 15 min to the painting galleries of the Louvre museum, the king was, however, keen to see the scientific institutions that had helped establish French military supremacy in the previous decade. He visited the École polytechnique and the Institut de France. But the doors of the Observatory remained closed to him as, although he was a dear friend of Humboldt's, the astronomer François Arago refused to meet the man whose army had toppled Napoleon I.

One day, Humboldt told Arago he was leaving town. The next morning, he showed up at the Observatory and said to his friend: "I wished to shake your hand one last time; I'm leaving with this man .... My travel companion ... is rather curious to visit the Observatory, can you show him the observation rooms, Lenoir's sextant, etc.?" Wearing a travel cap, the king of Prussia was thus shown around, while Arago, much to Humboldt's alarm, heatedly expressed his disapproval at the occupying forces' policies [23].

The lessons of Jena had been learned by Friedrich Wilhelm: this is why he was willing to go along this humiliating stratagem to be able to visit the observatory. The king was well aware of the importance for the conduct of modern warfare of new geodetic and cartographic techniques developed by French astronomers [21]. In the nineteenth century, however, State leaders rarely needed to put on a disguise to step in the observatory. On the contrary, observatories had become central scientific institutions of the modern states. From Göttingen to Königsberg, from Brussels to St. Petersburg, from Rio de Janeiro to Petchaburi (Thailand), it became increasingly clear that the endowment of expansive observatories was an inescapable requirement for any modern state intent to preserve its political independence and be integrated into the emerging world-system. In the nineteenth century alone, the number of observatories rose exponentially from less than three dozen to more than 200 [24], without taking into account the increasingly large number of observatories devoted, not primarily to astronomy, but to meteorology, geomagnetism, geodesy, navigation, or, toward the end of the century, aeronautics.

Visiting observatories became a favorite occupation for princes, scientists and foreign dignitaries. "The afternoon in Greenwich," the physicist Hermann von Helmholtz wrote to his wife on September 8, 1853, "belonged to the most interesting and agreeable of my trip" to England [29]. Formal inspections of the Royal Observatory at Greenwich were called "visitations." First instituted in 1710, only in the nineteenth century did they become elaborate social events. Before 1830, prominent politicians had frequently figured among the so-called Visitors. Later, formal members of the Board of Visitors were mostly drawn from the Royal Society and the Royal Astronomical Society, whilst up to 200 guests were admitted in the Observatory and were shown the telescopes and other instruments by the staff [33].

Visitors also came from all over the world to see the observatories of Western Europe. The Chinese emissary Li Shuchang who spent 7 years in Europe in the 1880s is one of them. Among other marvels of industrial societies he paid attention to, observatories figured in good position. He visited the observatories in Greenwich and Paris; he was allowed to use the telescope in the observatories of Berlin and Madrid; he observed the moons of Jupiter, Saturn's rings, and the phases of Venus. "I made these observations with my own eyes," he wrote. "Concerning the western theory according to which each planet is an earth, even though one does not want to believe in it, it is a bit difficult to refuse it" ([37], my translation from the French). As the century drew to a close, observatories had become so attractive to the general European public that astronomers feared that this infatuation threatened their work. Ole Molvig has shown how this led the Berlin scientific community to establish new institutions for scientific popularization ([6], pp. 325–343). Very recently, the occurrence count of the word "observatory" in a sizable chunk of the literature in English has been made possible by the use of the Google Books database. The graph produced is striking, showing a marked increase from very low count before the 1790s to a peak before 1820 followed by renewed interest throughout the nineteenth century and a decline thereafter. Clearly, this widespread interest in the observatory is the sign that it played an important part in the science and culture of the nineteenth century.

Over the last decades, historians of science have focused on the material culture in which science was pursued [18]. They have emphasized the importance of instruments, practices and tacit knowledge. In the course of such studies, it was realized that to pay close attention to the places where science was pursued deepened our understanding of the dynamic at play in scientific development [20, 40]. Much work was devoted to the study of the laboratory sciences [27] and of the field sciences [30] demonstrating that very specific epistemologies were rooted in these places of knowledge. The interaction between field and laboratory has also given great insights in the development of biology [28].

In the last few years, historians of science have developed a similar approach to the history of the observatory [6, 7]. As an instituted place of knowledge, the observatory has a longer history than either laboratory or field with which it interacted greatly [2]. The result of such investigation is a reconsideration of the place of the observatory in the history of science. In the following, some findings about the epistemology of the observatory sciences will be discussed. The notion of "observatory techniques" will be introduced and discussed in order to account for the diversity of practices to be found in observatories. The history of the observatory in the western world can be divided into three periods: (1) the age of the pioneers between the seventeenth and the end of the eighteenth century; (2) a period of triumph and crisis in the nineteenth century during which the observatory was very prominent in the social panorama of science; and (3) the age of specialization that followed. We shall mostly focus on the first two periods and speculate whether we might recently have entered a new period, especially with the rise of the environmental sciences. But first, let us ask the question: what is an observatory?

2 What Is an Observatory?

In *The Poet at the Breakfast-Table*, first published in the *Atlantic Monthly* in 1872, the American writer Oliver Wendell Holmes assumed that observatories were well known to his readers, but nevertheless gave an interesting description of the place:

> I suppose everybody who reads this paper has visited one or more observatories, and of course knows all about them. But as it may hereafter be translated into some foreign tongue and circulated among barbarous but rapidly improving people, people who have yet no astronomers among them, it may be well to give a little notion of what kind of place an observatory is.
>
> To begin then: a deep and solid stone foundation is laid in the earth, and a massive pier of masonry is built up on it. A heavy block of granite forms the summit of this pier, and on this block rests the equatorial telescope. Around this structure a circular tower is built, with two or more floors which come close up to the pier, but do not touch it at any point. It is crowned with a hemispherical dome . . . cleft from its base to its summit by a narrow, ribbon-like opening, through which is seen the naked sky . . . . No place, short of the temple of the living God, can be more solemn . . . . [The observatory] is the material image of the Christian; his heart resting on the Rock of Ages, his eye fixed on the brighter world above [26].

Holmes' account is striking because it quickly moves from a material description of the observatory emphasizing the stone pier, the cupola, and the instrument that sat in between to a mystical image of its place in the contemporary imagination. Knowledge of astronomy was a criterion of demarcation between the civilized and the uncivilized as is amply demonstrated by the case of Thailand in the 1860s ([6], pp. 86–117). Observatories had a symbolic function, but this function was rooted in a very specific material culture.

Observatories came in many gazes. There are the monuments established in the seventeenth century near great metropolitan capitals, such as the Paris and Greenwich observatories. The history of such institutions is well known [15, 17]. They depended on the State and generated a vast amount of documents which were generally well preserved. Astronomers working in these observatories were often figures of international stature in their field, members of scientific Academies, and the authors of many textbooks. For all these reasons, we know much about the history of such institutions. Some remarkable monographs have been published recently about the history of other observatories [16, 31].

What has been less emphasized, however, is the fact that national observatories were often concerned with much more than just astronomy. Crucial experiments in physics took place in observatories: Coulomb's work with torsion balances or Gauss' with magnetometers, for example. Mathematical statistics was in part developed by observatory scientists, such as Laplace and Quetelet [5]. In the eighteenth century, the gigantic cartographic project headed by the Cassinis had led to important geodetic operations to determine the meter by Delambre and Méchain, and then by Biot and Arago himself [1]. Geodesy and cartography would continue to play crucial parts in nineteenth-century observatories. Observatory scientists had always paid attention to weather conditions in order to correct the readings of their telescope; they had always been interested in geomagnetism as an offshoot of their official concern for navigation. In the first half of the nineteenth century, they would try to apply observatory routines and techniques to the study of the physics of the earth, launching the modern era of meteorology and geomagnetism.

But national observatories were not alone in the picture. Private observatories started to multiply. The eighteenth century private observatory was usually a room in palace or a rooftop terrace [22]. In the following century, private observatories became much grander in size and ambition. The gentlemanly tradition persisted as can be attested by the famous Leviathan built for Lord Rosse in Parsonstown. But perhaps more interesting from a sociological viewpoint was the proliferation of amateur astronomers able to equip themselves with small working observatories. One of them Hermann Goldschmidt discovered 14 new asteroids from a small observatory in the center of Paris from 1852 to 1861 [25]. Instrument makers also built their own observatories where scientists who were not able to afford them could use a variety of instruments. As the century unfolded, a great deal of observatories were established, temporarily or permanently, in remote places such as in the colonies or on the top of the highest mountains. As the extreme and problematic case of the Mont Blanc Observatory from 1893 to 1909 has allowed to highlight [34], the quality of an observatory depended on a few important characteristics: observation carried out on a routine basis; non-mobile instrumentation around which the observatory was built and about which much technical information was known; and deep insertion in a network through which information transited efficiently.

3 Observatory Techniques: The Tychonic and Hevelian Models

Concerns about the proper location of observatories have always been a pressing matter for astronomers. The first important European observatory was Tycho Brahe’s Uraniborg erected in the late sixteenth century [13]. As has been observed, this place already had some of the crucial characteristics of later observatories, such as the great instruments it held, the interdisciplinary nature of the research pursued there, and the strict division of labor it adopted to carry its routine task of observation. But historians have not emphasized as much the fact that this observatory was extremely isolated. Located on an island owned by Tycho between Sweden and Denmark, it was not built close to the port but stood alone in the middle of the island. It was moreover surrounded by imposing walls.

While the Tychonic observatory cherished isolation, another model emerged in the seventeenth century with the Hevelian observatory. Erected on a platform above three contiguous houses he owned in the center of Dantzig (now Gdańsk), the observatory of Hevelius relied on the set of skills directly available in the city for grinding his glass, sketching his maps or printing his books. This observatory was built in symbiosis with the early modern urban culture. When Imperial powers such as France and England set up their own observatories in Paris and Greenwich later in the seventeenth century, both models were in the minds of the founders.

Both achieved a compromise between isolation and immersion in urban culture by establishing themselves close but outside of the capital city [3]. But what made them "observatories"?

Let us consider what early observatories produced before the beginning of the nineteenth century, which made them so valuable for so many people. The most obvious product of the observatory was the numerical table. Observatories have long before the laboratory for example been specialized in the production of quantitative data. Numbers produced by the observatory could come from instruments specially designed for such purpose, divided circles being the privileged instruments allowing to assign a numerical value to an angle [11]. Various graduated instruments and clocks peopled the observatory and churned out a continuous output of data that were noted, copied, collected, and preserved. But the observatory's numerical tables almost never consisted in raw data arranged in columns, no more than there were simple derivation of predictions from theory. Whether they were from an observational or theoretical origin, the numbers tabulated by observatory scientists almost always were the results of long computational procedures.

Computing therefore was consubstantial with the observatory culture and its emphasis on numerical data. The computing procedures that were developed have several aspects that one may be interested in. First, there are purely mathematical aspects to them, such as the way in which perturbation theory was developed in the eighteenth and nineteenth centuries to turn the intractable analytical three-body problem into a problem that could be approximately solved numerically. Second, there were statistical aspects developed at the very end of the period under consideration: the least-square method introduced by Laplace and Gauss allowed observatory scientists to fix the constants in the theory on the basis of necessarily faulty data. But statistical methods were not always as mathematical as the Laplace-Gauss method and could sometimes involve a practice that is closer to art, such as the spreadsheet method devised and used by Tobias Mayer in the production of his lunar tables [45]. Finally, computing methods were also transformed into social practices whereby rather unskilled computers were asked to carry out specific algorithms developed by observatory scientists [14].

The observatory scientists were also heavily involved in the production of maps. It is interesting to notice that while the Tychonic observatory seemed geared toward the production of numerical tables, the Hevelian observatory was more inclined to produce very elaborate and beautiful maps [47]. Similarly, the Paris observatory, which was closer to the city than Greenwich spent much effort under the Cassinis on the production of maps of France, while Greenwich was more concerned with tables. But the maps produced by observatories were different from all other maps produced before in that they were based on the very same techniques used to produce tables. In fact, one may go as far as saying that maps produced by observatories were just another way of representing numerical tables. The location of every star on Hevelius' maps, the location of every village on the Cassinis' maps were equivalent to their coordinates. They were determined with the same instruments with graduated circles and they were the result of extensive computing.

The products of observatories were therefore highly dependent on this instrumental technology. In fact one may claim that the technology also was a product of the observatory culture. Although instrument makers in Paris and London were nominally independent from the observatories, their work was carried in close interaction with astronomers. As Guy Boistel has shown, the result of this colloboration produced instruments of greater precision as well as computing techniques that soon spread through the world, especially among navigators ([6], pp. 148–173).

Observatories, finally, produced cosmologies. Tycho is famous for having suggested a new geo-heliocentric model to compete with the Ptolemaic and the Copernican models. Of course, the Copernican model was later adopted by astronomers and popularized in a large number of publications. The cosmology of the observatory, however, was not entirely that of the Newtonian *philosophes* of the Enlightenment. A crucial aspect of this cosmology was again the special place it had for numbers. Gravitational theories were not wholly adopted by working astronomers as long as other procedures produced more accurate predictions. In this sense, one may say that the cosmologies of the observatory left room for the pragmatic elements that were missing from more elaborate systems, but better suited to the messiness of dealing with actual data.

4 Triumph and Crisis of the Observatory in the Nineteenth Century

At the beginning of the nineteenth century, as we have seen above, observatories had acquired a renewed importance for the modern State. Some documents attest that astronomers were now on the search for a new model for designing observatories. In 1810, Jean-Dominique Cassini who had been ousted from the Paris observatory by the French Revolution published a memoir titled “Project and description of a new observatory” [9]. Five years earlier, the German architect Georg Henrich Borheck also produced an extensive analysis of *The Principles for the installation of the new observatory … of the university of Göttingen* [8]. What is most noticeable about these two documents is the way in which they insisted upon combining the requirement of architecture with the practice of astronomy (Borheck included in his report a long excerpt from the Gotha astronomer Xaver von Zach). The principles on which the construction of new observatories should be based clearly emphasized the need to put the activities of the professional observer at the center of the builders’ concerns. Everything starting from the size and the location of the building to the materials used were subsumed to those needs. The observatory was being turned into a fully professionalized space.

Of course, this transformation of the observatory hardly occurred in an instant. Often, political demands and matters of convenience forced astronomers to make compromise about the quality of the observation in favor of a location that was closer to political powers, universities, and academies. In the 1840s, the expensive observatory built by the Tsar in Pulkovo became in the eyes of the astronomers the model of what money could buy. As Simon Werrett wrote, even such a model had serious flaws ([6], pp. 33–57). Later in the century the astronomical eldorados moved west, to the United States [41].

Established observatories were hard pressed to adapt to these standards. Some reacted to this pressure by greatly enlarging the scope of their investigations. Observatories became the locus of a great variety of pursuits in positional astronomy, but also in astrophysics which emerged after the experiments of Bunsen and Kirchhoff in 1859. They were the centers of calculation, to use Bruno Latour's phrase [32], around which meteorological as well as oceanographical networks were established. They provided the resources to carry out increasingly extended geodetic surveys. They also were heavily involved in increasing the precision of metrological units (of time, temperature and length). As we mentioned above, observatories were important places where physical experiments were performed and statistical data gathered and analyzed.

The diversity of the pursuits in the nineteenth-century observatory raises the question: what did unite these various activities? Significantly, observatory scientists and others who were close to the observatory culture of the time also raised the question of the unity of science. It was in his course of popular astronomy that Auguste Comte produced his Positivist thought which sought to unify the sciences on the basis of the empirical method. As John Tresch has argued ([6], pp. 253–284), it was in a monumental work titled *Cosmos* that Humboldt suggested that a unified instrumental approach could unlock a new understanding of the whole universe. It was Angelo Secchi, a Jesuit astronomer, director the observatory of the Collegio Romano in the Vatican who suggested that all physical phenomena could be understood as the interaction between force and matter…

To tackle the complex technical space that was the observatory, it is useful to focus on "observatory techniques" [6]. They include the whole set of physical, methodological, and social techniques rooted in the observatory as focus of inquiry. Observatory techniques included the set of practices required to perform successfully at the telescope eyepiece; the calibration, manipulation, and coordination of precision instruments for making observations and taking measurements. They embraced methods of data acquisition, reduction, tabulation, conservation, as well as complex mathematical analyses (error analysis and celestial mechanics). They also included various techniques of representation for the production of maps, drawings, or photographs, but also of material, numerical, and textual – indeed poetic – representations of the heavens and the earth, that ultimately shaped the way in which the world, society, and science itself could be construed. Finally, these techniques incorporated the social management of personnel within the observatory as well as international collaborations.

Observatory techniques were developed inside and outside of observatories – by instrument makers in their workshops, navy officers on ships, civil engineers in the field, or physicists in their cabinets. But in the observatory they were uniquely assembled to form a coherent set of techniques. Thereby these techniques helped define a space of knowledge: the observatory. Observatory techniques reveal the perpetually re-engineered cohesion of the observatory sciences. They formed a consistent foundation to a unified science of the heavens and the earth practiced by observatory scientists in the first part of the nineteenth century and later publicized in widely popular works such as Arago's *Popular Astronomy* and Humboldt's *Cosmos*.

Observatory techniques therefore required that the space be reconfigured to accommodate them. In the middle of the nineteenth century, both the Greenwich and Paris observatories were converted by George Biddell Airy and Urbain Le Verrier, respectively, into fully professional spaces. The model of the factory has often been evoked to describe this new space where division of labor was organized around the main production [42, 43]. At the same time, instrumental technologies were greatly developed [12]. One may here mention the case of the physicist Léon Foucault who was allowed to develop his great 80 cm telescope at the Paris Observatory in the 1860s [44]. All these developments were put at the service of the values of precision characteristic of the age [48].

In a sense, one can say that the observatory was turned into a number factory. As I have argued elsewhere, this led to the development of what can be called "observatory mathematics" [5]. According to Airy, indeed "every part of the operation of an observatory is mathematical." The construction of instruments was reliant on mathematical mechanics, the construction and the proper understanding of the defaults of a telescope required a knowledge of mathematical optics. The discussions and interpretation of observations were done through mathematical astronomy. The higher problems finally, such as the discovery of a comet's orbit from observations, required the high mathematics of gravitational astronomy and mathematical analysis. As a consequence, Airy argued, the hierarchical place one occupied within the observatory was determined by one's mathematical knowledge.

The nineteenth century, however, also saw a great change in the culture of representation in the observatory. In the early part of the century, Humboldt introduced the isothermal lines: a system of lines supposed to show the distribution of mean temperature across the surface of the globe. The image was not the representation of an actual object to be found in the world but completely man made. This system of lines was a simple way to represent pictorially observation points, theoretical speculation and the result of computation. The historians Peter Galison and Lorraine Daston have argued that the introduction of the photography can be taken as exemplary of new period in the history of objectivity [19]. Although this process occurred on a much larger scale than the observatory alone, one cannot but be struck by the important part played by observatory scientists in the establishment of photography as an instrument of scientific investigation [36]. Several approaches coexisted among scientists with respect to the way in which this new means of investigations could be integrated with the quantitative culture that dominated the observatory [39]. In the end, as the use of photography during the transits of Venus in 1874 and 1882 showed, it mostly was a quantitative analysis of the photographic plates that imposed itself among observatory scientists [4].

While the quantitative culture of the observatory was adopted by a great number of practitioners of others fields of science or by civil servants of the technocratic State, observatory scientists pioneered the networking techniques for interrelating the data. Already in the eighteenth century, most notably during the transit of Venus, the letter networks instituted by astronomers were impressive and allowed the rapid exchange and coordination of information [46]. Over the course of the nineteenth century, observatory scientists' networking activities increased significantly. Meaningful only as long as they were inserted in global networks, new observatories always inaugurated their work by measuring their latitude and longitude as precisely as possible. This often involved not only astronomical observations but land surveying as well.

In the 1830s and 1840s, observatory scientists used their networking techniques to establish a research program on geomagnetism. The goal of such program was to produce accurate maps of geomagnetic fields across the globe, using in particular Humboldt's technique of the isolines. Scientists involved included John Herschel, Adolphe Quetelet, Carl Friedrich Gauss and Humboldt. Observation protocols were drafted and simultaneously followed by observers in various observatories. As a result, geomagnetism was established on foundations that were similar to astronomy's: precise quantitative measurement, international coordination and high-level analytical theories [10].

In the 1850s, a new technology became available to them: the telegraph. It was immediately exploited by observatory scientists. Airy sought to determine the difference in longitude between Greenwich and Paris as soon as the submarine telegraphic line was opened, but his attempt was slowed down by the political turmoil in France. The first international telegraphic determination of a longitude difference therefore occurred between Greenwich and Brussels in 1853. The combination of the networking techniques and the telegraphic technology was a powerful force that transformed the study of the weather by the last third of the century [38].

Observatory scientists were led to reflect on the possibilities of international scientific cooperation. In 1853, two congresses were convened in Brussels by Quetelet: one was the first international congress in oceanography, the second the first international congress of statistics. The two congresses established rules for standardizing observations and analysis of data. Observatory scientists' mastery of such techniques for the manipulation of numbers, the production of precise images and the networking of data were precious for nineteenth-century societies. Industrialization and colonialism relied on similar techniques. Simon Schaffer has for example shown how the observatory techniques deployed in the Paramatta Observatory in Australia were very close to those needed for keeping control over an Empire than spanned the planet ([6], pp. 118–147).

Last but not least, observatory scientists pioneered the way in which science was communicated to an increasingly literate public. Not only that, they strongly promoted worldviews in which God only played a small part, or no part at all. As Charlotte Bigg has argued, this effort was not independent from research ([6], pp. 305–324). On the contrary, one can say that the public engagement of observatory scientists was a central concern of theirs which made their scientific investigation possible.

5 The Age of Specialization and Contemporary Neo-Humboldtianism

As the nineteenth century drew to a close, the unified culture of the observatory had broken down. Astrophysical observatories in Meudon, South Kensington and Postdam were established independently of older institutions. Meteorology had given rise to large centralized administrations that kept observatories simply as observing stations, but by and large developed independently from astronomical observatories. Similarly, geomagnetism, vulcanology, mareology, oceanography, geodesy, etc., branches of sciences which had emerged in close connection and often within the observatory culture were endowed with their own institutions, peopled by their own specialists, using their own instruments… The specialization of knowledge at the end of the nineteenth century is a well known story which led to the emergence of the modern scientific disciplines [35].

But would it be too daring to suggest that with the emergence of the environmental sciences one may be witnessing the rise of a "neo-Humboldtianism" of sorts? Three aspects are worth emphasizing here. First contemporary climate science is rooted in the use of representation techniques pioneered by Humboldt, most notably the isoline. Without such a tool, it is difficult to imagine that climate scientist could convey any message at all. The representation techniques they use rely on the combination of techniques for producing and manipulating numbers, representations, and networks that were properly coordinated in the observatory culture of the nineteenth century. The important reliance on various kinds of networking techniques is the second aspect of contemporary neo-Humboldtianism which needs to be emphasized. And the interdisciplinary character of climate science has indeed raised questions about the unity of science that are not unlike those that agitated nineteenth-century observatory scientists. Finally, climate science has also tended to involve the public at large, not in the least because it raises questions about the place of humankind in Nature. In this time when the scientific basis of climate change is so efficiently put in question in the public arena, a second look at the way in which the observatory culture succeeded in leaving such an imprint on western societies might be helpful.

References

1. K. Adler, *The Measure of All Things: The Seven-Year Odyssey that Transformed the World* (The Free Press, New York, 2002)
2. D. Aubin, Orchestrating observatory, laboratory, and field: Jules Janssen, the spectroscope, and travel. Nuncius **17**, 143–162 (2002)
3. D. Aubin, The fading star of the Paris Observatory in the nineteenth century: astronomers' urban culture of circulation and observation. Osiris **18**, 79–100 (2003)
4. D. Aubin, L'événement astronomique du siècle ? Une histoire sociale des passages de Vénus, 1874–1882 [Introduction to the Special Issue]. *Cahiers François Viète* **11–12**, 3–14 (2006)

5. D. Aubin, Observatory mathematics in the nineteenth century, in *Oxford Handbook of the History of Mathematics*, ed. by E. Robson, J. Stedal (Oxford University Press, Oxford, 2009), pp. 273–298
6. D. Aubin, C. Bigg, H.O. Sibum (eds.), *The Heavens on Earth: Observatories and Astronomy in Nineteenth-Century Science and Society* (Duke University Press, Rayleigh, 2010)
7. G. Boistel, *Observatoires et patrimoine astronomique français* (ENS éditions; Cahiers d'histoire et de philosophie des sciences, Paris, 2005)
8. G.H. Borheck, *Grundsätze über die Anlage neuer Sternwarten mit Beziehung auf die Sternwarte der universität Göttingen*, ed. by K. Beuermann (Universitätsverlag, Göttigen, 2005)
9. J.-D. Cassini, *Mémoires pour servie à l'histoire des sciences et à celle de l'Observatoire royal de Paris, suivis de la vie de J.-D. Cassini écrite par lui-même et des éloges de plusiers académiciens morts pendant la Révolution* (Bleuet, Paris, 1810), p. 63ff
10. J. Cawood, Terrestrial magnetism and the development of international cooperation in the early nineteenth century. Ann. Sci. **34**, 551–587 (1977)
11. A. Chapman, *Dividing the Circle: The Development of Critical Angular Measurement in Astronomy 1500–1850* (Ellis Horwood, New York, 1990)
12. A. Chapman, The astronomical revolution. in *Möbius and His Band: Mathematics and Astronomy in Nineteenth-Century Germany*, ed. by J. Fauvel, R. Flood, R. Wilson (Oxford University Press, Oxford, 1993), pp. 32–77
13. J.R. Christianson, *On Tycho's Island: Tycho Brahe, Science, and Culture in the Sixteenth Century* (Cambridge University Press, Cambridge, 2003)
14. M. Croarken, Human computers in eighteenth- and nineteenth-century Britain, in *The Oxford Handbook of the History of Mathematics*, ed. by E. Robson, J. Stedall (Oxford, Oxford University Press, 2009), pp. 375–403
15. S. Debarbat, S. Grillot, J Lévy, *L'Observatoire de Paris: son histoire, 1667–1963* (Observatoire de Paris, Paris, 1984)
16. S.J. Dick, *Sky and Ocean Joined: The U.S. Naval Observatory, 1830–2000* (Cambridge University Press, Cambridge, 2003)
17. E.G. Forbes, A.J. Meadows, D. Howse, *Greenwich Observatory: The Royal Observatory at Greenwich and Herstmonceux, 1675–1975*, 3 vols. (Taylor & Francis, London, 1975)
18. P. Galison, *Image and Logic: A Material Culture of Microphysics* (University of Chicago Press, Chicago, 1997)
19. P. Galison, L. Daston, *Objectivity* (Zone Books, Boston, 2007)
20. T.F. Gieryn, Three truth-spots. J. Hist. Behav. Sci. **38**, 113–132 (2002)
21. A. Godlewska, *Geography Unbound: French Geographical Science from Cassini to Humboldt* (University of Chicago Press, Chicago, 1999)
22. R. Hahn, Les observatoires en France au XVIIIe siècle, in *La Curiosité scientifique au XVIIIe siècle: cabinets et observatoires* (Hermann, Paris, 1986), pp. 653–659
23. E.T. Hamy, *Correspondance d'Alexandre de Humboldt avec François Arago* (E. Guilmotot, Paris, 1908), p. 359ff
24. D.B. Herrmann, An exponential law for the establishment of observatories in the nineteenth century. J. Hist. Astron. **4**, 57–58 (1973)
25. F. Hoefner, Hermann Goldschmidt. Cosmos **4**, 321–324 (1866)
26. O.W. Holmes, The poet at the breakfast-table, V. Atlantic Mon. **29**, 613 (1872)
27. F.A.J.L. James, *The Development of the Laboratory: Essays the Place of Experiment in Industrial Civilization* (American Institute of Physics, New York, 1989)
28. R.E. Kohler, *Landscapes and Labscapes: Exploring the Lab-Field Border in Biology* (Chicago University Press, Chicago, 2002)
29. R.L. Kremer, *Letters of Hermann von Helmholtz to His Wife 1847–1859* (Franz Steiner, Stuttgart, 1990), p. 127
30. H. Kuklick, R.E Kohler, Science in the field. Osiris **11** (1996)
31. J. Lamy, *L'Observatoire de Toulouse aux XVIIIe et XIXe siècles: archéologie d'un espace savant* (Presses universtaires de Rennes, Rennes, 2007)
32. B. Latour, *Science in Action* (MIT Press, Cambridge, 1987)

33. P.S. Laurie, The Board of visitors of the Royal Observatory. Q. J. Roy. Astron. Soc. **7**, 169–185 & **8**, 334–353 (1966) (1966–1967)
34. S. Le Gars, D. Aubin, The elusive placelessness of the Mont-Blanc Observatory (1893–1909): the social underpinnings of high-altitude observation. Sci. Context **22**, 509–531 (2009)
35. T. Lenoir, *Instituting Science: The Cultural Production of Scientific Discipline* (Stanford university Press, Stanford, 1997)
36. T. Levitt, *The Shadow of Enlightenment: Optical and Political Transparency in France, 1789–1848* (Oxford University Press, Oxford, 2009)
37. S. Li, *Carnet de notes sur l'Occident*. Translated by Shi Kangqiang (Ed. de la Maison des sciences de l'homme, Paris, 1988)
38. F. Locher, *Le savant et la tempête: Étudier l'atmosphère et prévoir le temps au XIXe siècle* (Presses univesritaires de Rennes, Rennes, 2008)
39. L. Maison, S. Le Gars, Janssen, Rayet, Cornu: trois parcours exemplaires dans la construction de l'astronomie physique en France (1860–1890). Revue d'Histoire des Sciences **59–1**, 51–81 (2006)
40. A. Ophir, S. Shapin, The place of knowledge: a methodological survey. Sci. Context **4**, 3–21 (1991)
41. A. Saint-Martin, The new astronomical eldorado: the French understanding of American astrophysics, 1900–1920. Nuncius. J Hist. Sci. **23**, 91–113 (2008)
42. S. Schaffer, Astronomers mark time: discipline and the personal equation. Sci. Context **2**, 115–146 (1988)
43. R.W. Smith, A national observatory transformed: Greenwich in the 19th century. J. Hist. Astron. **45**, 5–20 (1991)
44. W. Tobin, *The Life and Science of Léon Foucault. The Man who Proved the Earth Rotates* (Cambridge university Press, Cambridge, 2003)
45. S. Wepster, *Between Theory and Observations: Tobias Mayer's Explorations of Lunar Motion, 1751–1755* (Springer, Berlin, 2010)
46. S. Widmalm, A commerce of letters: astronomical communication in the 18th century. Sci. Stud. **2**, 43–58 (1992)
47. M.G. Winkler, A. Van Helden, Johannes Hevelius and the visual language of astronomy, in *Renaissance and Revolution: Humanists, Scholars, Craftsmen and Natural Philosophers in Early Modern Europe*, ed. by J.V. Field, A.J.L. James Frank (Cambridge University Press, Cambridge, 1993), pp. 97–116
48. M.N. Wise, *The Values of Precision* (Princeton University Press, Princeton, 1995)

Chapter 7
Astronomy and Technical Progress

James Lequeux and Laurent Vigroux

Abstract Astronomy is perhaps the best example of fundamental research aiming at increasing our knowledge well beyond our human neighborhood. But astronomy is also a Big Science, which is partly technology-driven. Progress in observational capabilities is due to progress in detectors, telescopes, satellites, etc. In the first part, we remind of the use of astronomy in the past. Then, we use several examples to describe the complex interactions between astronomy, technology development and industry in the modern world. We conclude by a short description of the global economic impact of astronomy.

Keywords Technology transfer • Adaptive optics • Detectors • Radio astronomy • Astronomy: research and development

1 Introduction

Fundamental research is generally associated with pure science, with the aim of increasing our knowledge of Nature without consideration of possible applications. It is too often considered as an intellectual game played by selfish individuals disconnected from the real life. When considering big science like particle physics or astronomy, one might also conclude that fundamental research uses very expensive

---

J. Lequeux (✉)
LERMA, Observatoire de Paris, 61 avenue de l'Observatoire, F-75014 Paris, France
e-mail: james.lequeux@obspm.fr

L. Vigroux
Institut d'Astrophysique, 98 bis boulevard Arago, F-75014 Paris, France

CEA/DSM.IRFU, Saclay, France
e-mail: vigroux@iap.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_7,

toys and that public money should be spent on more fruitful activities. Conversely, the intellectual and economical impact of fundamental science on society as a whole is recognized by many governments which attempt to use fundamental research as a development tool to improve economy.

Astronomy offers a good case for understanding the complex relationships between science, technique and industry. Its status amongst the different sciences has evolved enormously with time. From Antiquity to the nineteenth century, astronomy was considered as the science *par excellence*. Moreover, its practical usefulness was universally recognized. The situation changed progressively during the nineteenth century and later, due to the emergence of other sciences and to the fact that the applications of astronomy narrowed considerably (see Chap. 6). At present, the situation is completely reversed: astronomy has lost most of its direct applications and has become a pure science. On the other hand, its progresses depend strongly on techniques most often developed for other purposes. However, the return from astronomy to technical developments and to industry is not negligible, although poorly known from the general public. In this paper, we will describe the interactions between astronomy and technique on a few examples, and we will conclude by some considerations on the global impact of this science on the society.

2 The Astronomy of the Past[1]

The astronomical phenomena of alternating day and night and the succession of seasons regulate the life of man since the origins. The development of agriculture, and later of urban civilisation, made necessary the measurement of time throughout the year. The periods of the Sun and Moon were the natural units of time reckoning and were used to define the calendar, depending on the civilization; sometimes, other periods appeared, such as a Venus period with the early Mexicans or a Jupiter period with the Indians. Stars were used by sailors and nomads for orientation: they had to find ways to identify the stars and to determine their positions. The planets were supposed to be of good or ill omens, and it was considered necessary for this purpose to predict their position with respect to the stars. This rather complex task stimulated very early the emergence of specialists, astronomers-astrologers, who knew mathematics and were able to deal with these difficult calculations (see Chap. 17). Indeed astronomy was the first science, and a very useful one. Its developments were triggered by practical considerations, and the results were outstanding: during Greek Antiquity, correct orders of magnitude were obtained for the shape and size of the Earth, the precession of its axis was discovered and the prediction of planetary positions was possible with reasonable accuracy.

This situation remained essentially the same until the seventeenth century. Mathematical and observational techniques and instruments were perfected, more

---

[1] For a more detailed description see [12, 13, 17].

results were obtained, but the nature of astronomy was unchanged. Then its horizon enlarged. Astronomy was now asked to measure longitudes, which requires a comparison of clocks at different places. Good clocks were available thanks to Christiaan Huygens, a physicist who was also an astronomer, but they were not transportable. Fortunately, Galileo had discovered the satellites of Jupiter, whose eclipses in the shadow of the planet are observable from large portions of the Earth surface and provide the needed synchronizing signals for the local clocks, allowing the determination of longitude differences. In view of the importance of this problem for trade and economy in general, considerable efforts were devoted to the observation and prediction of these eclipses, in particular in the newly created observatories in Paris and Greenwich (a by-product of this work in Paris was the discovery of the finite velocity of light). As a result, longitudes could be determined accurately on the continents, allowing the construction of accurate maps. But the motions of the ships prevented observations of Jupiter's satellites at sea. Astronomers were now in charge of the measurement of the Earth – geodesy and to some extent cartography – and they did it well. They also developed the study of the motion of the Moon with respect to the stars, allowing in this way some progress in the determination of longitude at sea because the Moon could then provide universal time signals, but this was difficult to use in practice and not very accurate: the real breakthrough came with the invention by John Harrison of reliable marine chronometers, in the middle of the eighteenth century. The interest of astronomical observations for determining longitudes vanished, except for giving time to the ships at the main harbours: astronomers were only left with the measurement of the stability of marine chronometers. It was the first time when some outside technique superseded the usefulness of astronomy.

During this period, astronomers were the first scientists to organize important expeditions, first to measure longitudes of the main cities of France at the end of the seventeenth century, then for measuring portions of meridians in northern Sweden and in Peru at the middle of the following century. Other countries lead similar expeditions. The preparation and logistics of these expeditions required considerable, coordinated efforts from the governments, the scientific bodies and the scientists themselves. This culminated in the voyages to observe the transits of Venus in front of the Sun in 1761 and 1769. Observations of these transits in different places of the terrestrial globe gave at that time the best possibility to determine the distance of the Sun to the Earth. Not only the phenomena were observed in many places of Europe and North America, but expeditions were sent by nine European countries to China, India, Indonesia, Madagascar, North America including Caribbean islands and Mexico, Siberia and Tahiti. It is remarkable that this was possible in spite of bad relations between some of these countries, requiring the delivery of safe-conducts. Then the results were assembled in order to determine the distance to the Sun, a process which took many years. This is the first example of a developed international cooperation. Later, Astronomy continued to be the leader in such cooperation and can still be considered today as a model international science.

However, astronomy progressively lost ground during the nineteenth century as far as practical applications were concerned. Cartography, and later geodesy, became the prerogatives of specialists, mainly military. Earth magnetism, which was also amongst the attributions of astronomers who had to determine magnetic declination for sailor's compasses, also fell outside their range during the second half of this century. For historical reasons, astronomers were also in charge of meteorology. This culminated with the creation of the International meteorological service by Urbain Le Verrier at the Paris Observatory, but did not survive his death in 1877, after which the service became autonomous. The evolution was similar in the rest of the world. The only remaining practical duty of astronomers was the determination and distribution of time: this was, for example, the main activity of the Paris Observatory until WW2. This is still the case to some extent: astronomers, not alone but now in collaboration with physicists, remain at the forefront of the most accurate determinations of time and also, thanks to very long baseline interferometry, of the measurements of the motions of the Earth and of the continent drifts. The applications of celestial mechanics knew a temporary glory with the launch of artificial satellites and space probes, but the accurate prediction of their motions is now routine and the specialists do not belong anymore to astronomical observatories. What remains to astronomers is the prediction of possible impacts of asteroids on the Earth, something of obvious importance for mankind but a minor activity for them.

With these few exceptions, the direct military-industrial-commercial usefulness of astronomy is over. Astronomy is now essentially a "cultural" science, for which there is, however, still a strong demand. Nevertheless there are still strong interactions between astronomy, technique and industry: we will now discuss a few examples of these interactions.

3 The Development of Radio Astronomy

The first example is the birth and development of a new field: radioastronomy. Searches for a possible radio emission from the Sun were made around 1900 but with no result because the detectors were not sensitive enough and because the ionosphere, which blocks the propagation of long radio wavelengths from outside the Earth, was not yet discovered. The first detection was made fortuitously in 1933 by Karl Jansky. Jansky was an engineer at the Bell Telephone Laboratories which wanted to develop transatlantic wireless communications. He was asked to study parasitic emissions at frequencies around 20 MHz. For this, he built a rotating antenna supported by front-wheels and axles from an old Ford T and started looking for sources of interferences. He actually discovered an emission by thunderstorms, but was surprised to detect a faint periodic signal with a period close to one day. Refining his observations, he found that the period was 23 h 56 min, which was exactly the period of the apparent rotation of stars in the sky: thus the signal was of celestial origin. He also found that this signal came from the Milky Way, in particular its centre in the constellation Sagittarius. Despite his own interest to search for other astronomical sources, he was assigned other tasks by Bell Labs and could not continue. The discovery gained large attention from the public, but astronomers showed no interest with only a few exceptions (see [8]).

It was another radio engineer who took the relay from Jansky: Grote Reber. Reber wanted to work with Jansky at Bell Labs but could not obtain a position there. He obtained instead a full job in a radio company in Chicago and decided to build with his own funds a radiotelescope in his house backyard. This was the first radio parabolic antenna, 9.5 m in diameter. With it he could confirm in 1938 the result of Jansky at a higher frequency and made the first rough map of the radio emission of the Milky Way. The origin of this emission remained a mystery until the early 1950s, when Karl Otto Kiepenheuer in Germany suggested that it was synchrotron radiation, an emission by high-energy electrons moving in the Galactic magnetic field. This emission had been discovered in 1947 in a synchrotron, an accelerator of electrons, hence its name. The theory of synchrotron emission benefited to some extent from the work done to understand the radio emission of celestial sources. Synchrotron beams are now common tools to study the properties of materials, especially in molecular biology.

In these two examples the discovery was made possible by technical progress. That of Jansky was serendipitous, and Reber's map of the Milky Way was obtained just to make better observations, but without supporting theory. Later, the 21 cm line of interstellar atomic hydrogen was discovered in 1951, after Henk Van de Hulst calculated its exact wavelength, a work suggested by the famous Dutch astronomer Jan Oort. However, the line was not found in Netherlands for lack of sensitivity in the radioastronomy receivers, but in the USA where Harold Ewen and Edward Purcell had built a complete receiving system especially adapted to the detection of this line; in particular, they used a new modulation scheme by switching frequencies. The result was confirmed within a short delay by the Dutch group and an Australian one using this modulation technique (see e.g. [19]). In this case, the discovery followed a theoretical prediction and was made by a dedicated instrument.

The subsequent history of radioastronomy abounds in similar examples (see e.g. [12]). We will say a few words on the origin of the French-German-Spanish Institute for Radio Astronomy at Millimeter wavelengths (IRAM) [5]. In the late 1960s, Germany had developed a strong expertise in large parabolic antennas for radioastronomy: German industry had built for Australia in 1963 the best radiotelescope of the time (the 64 m diameter Parkes radiotelescope, still in operation) and was now building a giant antenna for Germany (the 100 m radiotelescope at Effelsberg). The latter antenna included a novel technique invented by the German-American radioastronomer Sebastian von Hoerner: the homologic principle of controlled deformation, which maintained a parabolic shape in spite of gravity when pointing at different inclinations. On their side, French radioastronomers had good expertise in interferometry, a technique that allows to obtain high angular resolving powers by combining the signals from different antennas, and even to make radio images. Not unexpectedly, both sides wanted to push their expertise to millimetre wavelengths, mainly because it was a technical challenge. Their initial scientific motivations were rather weak: it is only in 1970 that Arno Penzias and Bob Wilson (who had obtained a Nobel prize for their discovery of the background radiation of the

Universe) in a few weeks discovered a handful of interstellar molecules through their radio emission lines at millimetre wavelengths because they had the best millimetre receiver of the time. It was now clear that this wavelength range was promised to a bright future. Each country had its own plan: a 30 m millimetre radiotelescope for Germany and a millimetre interferometer with several 10–15 m movable antennas for France. Due to money restrictions and political will, the two projects were forced to merge as IRAM in 1979, while a European laboratory to produce critical detector components (then only available in the USA) was set up in Cork (Ireland). It soon turned out that the two instruments of IRAM, although located in different sites, were perfectly complementary. IRAM was such a success that Spain decided to join: this fostered a strong development of radioastronomy in this country, which is presently at the forefront of research. The scientific results of IRAM are outstanding, but will not be developed here.

IRAM gives a striking example of the development of a new scientific domain. The initial motivation was essentially technical, the only scientific argument for the project being the foreseen potentialities which materialized only progressively. The promoters of the project were convinced of these potentialities from the start, and it is their conviction (not shared by many astronomers of the time) and their stubbornness, which lead the authorities to finance this project in a rather unfavourable time. As a reward, technical progresses in the antenna design and in the electronics acquired during the construction of IRAM gave a strong impetus to the relevant industry and are quite useful in the design of a still more ambitious world-wide instrument presently under construction: the Atacama Large Millimeter Array (ALMA).

4 Adaptive Optics: At the Cross-Roads of Astronomical and Defence Applications

Looking for the first time in the eyepiece of a telescope is disappointing: the images seem blurred and have a fast erratic motion. This is due to atmospheric turbulence, which distorts the images, and does not allow benefiting from the full resolving power of the instrument. A concept to compensate for this phenomenon, called the astronomical seeing, was proposed by Horace Babcock in 1953 [4]. But it was not feasible at the time (see [15]).

Somewhat later, a strong interest developed for spy artificial satellites: the USA launched 146 such satellites from 1960 to 1972, and the Soviets probably a comparable number. It was considered very important to watch carefully the satellites in orbit. Ground-based telescopes for this faced the same image degradation as astronomical telescopes. Military research was also engaged to focus laser beams on distant targets, which is required to compensate for the widening of the beam by atmospheric turbulence, a problem similar to that of image correction. In order to perform real-time compensation of the atmospheric image degradation, American defence engineers developed the active correction of the telescope optics. They developed a novel system based on a sensor to measure the distortions of the incoming wavefront by turbulence, and on real-time correction by a deformable mirror to compensate for these distortions. The first positive results were obtained in 1982, but the system soon was strictly classified. During the same period, astronomers (mainly in France) developed in parallel similar systems, called adaptive optics, but more slowly and with much more limited means. They achieved the crucial step in the early 1990s thanks to the declassification of some adaptive optics components and to the availability of a new generation of infrared detectors (adaptive optics is easier to implement at longer wavelengths). Adaptive optics is presently an essential part of most modern telescopes [11]. The future generation of extremely large telescopes (the European project has a diameter of 42 m) cannot be operated without efficient adaptive optics systems. The needs of astronomy and defence have now diverged: astronomers are interested in correction over a large field, essentially in the infrared, while the military require high adaptive performances in the visible with very fast response time. Defence experts are still working on the development of airborne laser systems to destroy ballistic missiles or ground targets, requiring corrections similar to those given by adaptive optics.

Surprisingly, adaptive optics has now important applications in ophthalmology. Inhomogeneities in the eye crystalline lens and in the cornea cause blurred images on the retina, a phenomenon similar to atmospheric seeing which can be corrected for in the same way. This correction allows a breakthrough in the diagnostic and curing of retina diseases. Several astronomy groups in the world have created spin-off companies to design adaptive optics systems for imaging and laser focusing in collaboration with hospitals.

Adaptive optics gives a good example of a technical development initiated by astronomers who identified the problem and established the theoretical bases for a solution. But they were not able to build an operational system until declassification of crucial components developed, essentially for defence applications.

5 Detectors: A Symbiotic Activity

Most astronomical observations are limited by sensitivity. Progress often comes from an increase of the size of the telescope or of the sensitivity of the detectors. Modern detectors have an efficiency of nearly 100% instead of a few percent for photographic plates: the gain is equivalent to using a 5 m diameter telescope instead of a 1 m one [11]. However, the development of modern mosaic detectors is beyond the capabilities of academic laboratories. Astronomers have to work in collaboration with industry to develop detectors matching their needs. This is not new: in the past, astronomers used to work with the Kodak research laboratories to obtain photographic plates suited to their needs. However, the use of photographic plates stopped in the late 1980s, to be replaced by mosaics of electronic detectors.

The Charge Coupled Devices (CCD) were invented at the Bell laboratories, initially as memory devices, but soon developed for imaging. The optimisation of the performance of imaging CCDs was boosted by the decision of NASA to have a CCD on board the Hubble Space Telescope. The first CCD cameras on telescopes were operational in 1980, and in less than 10 years they replaced almost all previous detectors for ground-based optical astronomy. The first-generation CCDs had many problems: poor transfer efficiency, remanence after saturation, high dark current, etc. All these defects were more apparent in astronomical CCD cameras, which pushed the possibilities to their limits: they became the best testbeds to understand the details of the physics of the CCDs [10]. The work done in astronomy laboratories working closely with industry was essential in order to build new-generation CCDs with improved performances. Thinning the CCDs, a technique invented by astronomers, opened the possibility of using them in the ultraviolet. Similar improvements of X-ray mosaic detectors were obtained thanks to X-ray astronomy: these detectors are presently used for many applications, ranging from control scanners in industry to medical imaging.

In the infrared, the first incentives for detector development were defence applications. The high development cost of these detectors could only be supported by the military, and the detectors used by astronomers were obtained as side-products of the defence projects. As for the CCDs, the progresses in these detectors have benefited from the use of infrared cameras by astronomers, and it is for this reason that some astronomy groups were allowed to use them in spite of military classification. The present generation of infrared detectors is affordable for many civil applications: night vision, security, thermography, medical imaging, etc. One of the most spectacular applications is the possibility to identify in a crowd a person suffering from some disease provoking fever, by looking at the temperature of the body through its infrared emission. It is very likely that in the future, the civil market will become the driver for infrared devices, as it was the case of the development of the CCDs.

A similar development arose for bolometers, which are thermal detectors sensitive to radiations at all wavelengths, especially useful in the far infrared. For ultimate sensitivity, they have to be operated at extremely low temperatures inside cooled enclosures in order to limit the parasitic radiation of the surroundings. Progresses in silicon etching and microelectronics make it possible to design two-dimensional arrays of bolometers. Each bolometer is a very thin membrane of silicon, supported from the main structure by insulating rods. At room temperature, they are now used in thermal infrared cameras, and are available at a price compatible with mass market. They were first developed in the USA but soon after, in France, they have been manufactured by the LETI, an applied research department of the French Atomic Energy Commission (CEA). In 1995, an array of microbolometers was judged necessary for use on the HERSCHEL Space Observatory, a European Space Agency satellite dedicated to far infrared and millimetre astronomical space observations. The Service d'Astrophysique of the CEA, then headed by one of us (LV), was in charge of providing the array [1]. The main technical challenge was to increase the size of each element from 30 to 750 μm, which meant etching very large and thin suspended membranes. This development was made by the LETI, while the testing and characterization of the detectors was entirely done in the CEA at Saclay. HERSCHEL has been now launched and the detectors are a full success (see [6]). Now, the LETI is thinking of building similar microbolometer arrays for operation at millimetre waves at room temperature. They can be used for many applications such as detection of landmines, industrial control or security scanners. Such millimetre wave scanners already exist, but they are very expensive and difficult to operate, while a microbolometer camera would be as easy to operate as a normal TV camera. The main limitation would be privacy control, since a person looks naked when seen by a millimetre scanner: dispatching security scanners everywhere will be limited by the public acceptance.

These developments provide a good illustration of the interaction between applied research, which develops new techniques with immediate goals, and astronomy, which pushes these techniques to their extreme limits. The improvements made by astronomers are used in return by applied research: they help in the development of manufacturing processes or are even at the origin of new applications.

6 Large Projects and Industrial Developments

Large facilities developed for astronomy, as well as for other fundamental science like particle and nuclear physics, are at the edge of new technologies and require specific developments. The design, the construction and the operation of these facilities are under the responsibility of public or international organizations, such as NASA in USA and ESA in Europe for space science missions, or ESO or IRAM in Europe for ground based astronomy. All these agencies have their own technical division to develop specific components and to monitor the technical activities of industrial contractors.

The construction of space observatories in Europe and in USA is under the leadership of a space agency, ESA or NASA, which keeps the responsibility of the observatory in-flight scientific performances. The satellite itself is built by the industry under the management of a Prime Contractor, with specifications provided by the Space Agency. Usually, the whole observatory is split in two main parts: the satellite, which includes the service module providing attitude control, electrical power and Earth communication, and the payload with the scientific instruments. The limit between the two parts is somewhat arbitrary and mission dependent. For example, a telescope can be included in the payload or not. The satellite is always built by industrial contractors while the payload can be built by industry, or by specialized research laboratories. In the USA, these laboratories could be the NASA centres such as the Jet Propulsion Laboratory, in Europe the technology division of the ESA or dedicated research laboratories present in several European countries. To illustrate the sharing of tasks between industry, space agencies and research laboratories, we can again use the example of HERSCHEL launched by ESA in 2009. This space observatory for observations in the far infrared includes the largest telescope ever built for a space mission and three instruments. The 3.5 m diameter telescope in silicon carbide, a new material for telescopes, was a specific development made by a French company under an ESA contract. This company had already built smaller size telescopes in this material, which have been used to validate the different technological steps needed to manufacture the HERSCHEL telescope. The success of the HERSCHEL telescope will be the base for future developments of silicon carbide telescopes. The three instruments were built by consortia of European Laboratories (the main laboratories involved were in Belgium, France, Germany, Italy, Spain, The Netherlands, and the United Kingdom), with participation of teams from the USA and Canada. These three instruments are very innovative, and in many areas unprecedented. The technical skills required for their manufacturing are beyond the capability of industrial companies. Only research laboratories, used to Research and Technology developments, can take the leadership and the associated risk of the design, the manufacturing and the assembly of such instruments. The HERSCHEL instruments are cooled by superfluid helium to temperature below 4 K. The superfluid helium tank and the heat exchanger with the instruments were produced by a industrial company in Germany A company in Italy was in charge of the service module. The industrial prime contractor was a French company that was also in charge of the assembly and the verification of the whole satellite before launch, as well as the in flight commissioning of the spacecraft. The ESA project team had the key role of ensuring the final performance of the observatory. This has required taking the responsibility of the industrial procurements and the validation of the deliveries at each step of the integration. ESA closely monitored the progress of the instruments in the participating laboratories, and was responsible for the coupling of the instruments to the systems delivered by industry. In parallel, ESA has developed, with the help of the instrument consortia the ground segment needed for the flight operation of the observatory and the technical and scientific data processing on ground. ESA is also the point of contact with the astronomical community using the HERSCHEL observatory: selection of observing proposals, execution of observations, delivery and archive of scientific data.

This very complex organization requires a strong management and good understanding of the different practices in research laboratories and industry. This can be achieved only by a long tradition of work in common between the industry, the laboratories and ESA. This example shows that the technology transfer between the research world, including the leading agency, and industry could be sorted in three categories:

- Novel technologies originated in research laboratories, or novel association of existing technologies that are used later by industry to push beyond customary limits.

- Technologies developed in collaboration between industry and research laboratories through development within a contract or collaboration agreement.
- Technologies developed or extended by industry through the execution of a procurement contract

We have used HERSCHEL as an example, but similar conclusions could be drawn for all large facilities in space or on ground. All these projects are the results of common effort between research laboratories, industry and scientific agencies. Each of the different parties has in charge a part of the project, corresponding to their skill and capabilities.

7 The Impact of Astronomy on Economy

In the preceding sections, we have discussed the complexity of the relationships between technical developments and astronomy. Astronomy has contributed to technical advances in many areas: antennas, telescopes and optics in the whole electromagnetic spectrum, cryogenic and vacuum techniques, detectors, signal and image processing, communication techniques, etc. Astronomy is technically driven, but it also drives the technique. We read in the recent Millennium Report from the US National Academy of Science [2]: “In some areas, astronomers have pioneered the technology, while in the others we have worked symbiotically with industry and the defence sector in developing and perfecting the appropriate technologies.”

Another aspect of the impact of astronomy on economy simply comes from the fact that it is big science where investments are accompanied by large industrial returns. In Table 7.1, the cost of some recent astronomical projects is compared to that of some other equipments.

A large fraction of the construction costs of these facilities goes to industry as contracts for goods and services. This is a direct transfer, which is easy to quantify. A good estimate of the overall budget of European astronomy has been obtained in

Table 7.1 Costs of large astronomical projects

Approximate costs (M€) of large astronomical facilities		
ALMA	750	Ground based mm interferometer
Herschel + Planck (ESA)	1,100	Cost for ESA only
European extremely Large Telescope (ESO)	900	Cost for ESO only
Costs for comparison		
Aircraft carrier	300	
Fighter aircraft	300	
Airbus A380	250	
Airbus A320	60	
100 km of highway	600	

Table 7.2 Annual budget for European astronomy

Annual budget for European astronomy (M€/year)	
European Southern Observatory (ESO) (member state contributions)	150
Cost of National ground based facilities	250
European Space Agency (ESA) scientific program (member state contributions)	400
Cost of national contributions to ESA payload and national space projects	200
Cost of scientific staff and laboratories	1,000
Total	**2,000**
Expenses in industrial contracts	500

2008 during the elaboration of the ASTRONET Infrastructure Road Map (see [7]; see also Chap. 9). It is summarized in Table 7.2.

Dealing with the most costly infrastructure, ESO and ESA are the main providers of large industrial contracts in astronomy. For both organizations, about half of the contributions of the member states are returned to their industry as high-value activities. one-thirds of the budget corresponds to running costs to operate existing facilities. The situation differs for national astronomy institutes where the main expenses are associated to manpower and running costs. But even in these cases the return to industry is far from negligible.

Ground-based and space astronomy have a strong impact on the economical activity in general. Detailed studies have been performed in Hawaii [9] and in Arizona [18], where astronomy is very important; they show that aside the observatories and public research organizations, many small companies dedicated to optics and instrumentation developed to respond to the needs of the research institutes. In Arizona, for ten direct jobs created in public astronomical organizations, another six jobs are generated in the economy, and for every dollar of direct wages and salaries in these organizations, 1.3 $ is generated in the state economy. Similar studies done on the CERN [16] and ESA [14] have shown that the overall gain for the economy is 2–3 times what is directly generated by the contracts, or 1.2–1.6 times the total budget of the organizations. Using data from Table 7.2, we see that some 1,000 M€/year are generated in Europe in this way aside 500 M€/year through contracts.

On the other hand, the importance of the stimulus created by astronomy for obtaining products at the ultimate limits of the possibilities cannot be overstated. More unexpected are the improvements of quality insurance or project control, marketing benefits in terms of the image of industrial companies, and the increase of the staff motivation through new challenges and non-conventional work. It might well be that the main advantages for industrial companies of cooperation with astronomers are not the financial gain (which in any case would be difficult to quantify), but the benefits to their image and the overall improvement in skill and motivation of the workers [3].

8 Conclusion

Although astronomy was a directly useful science in the past, it is at present essentially pure science. But astronomy is also big science, with a strong dependence on technique. Progresses in astronomical observation capabilities are driven by technique. However, the interactions between astronomy, technique and industry are more complex, as demonstrated by the examples developed in this paper. Beyond its main fundamental goals of astronomy, most of its advocates insist on its importance in raising the interest of public and students for science. We hope to have shown that astronomy is also important in stimulating advanced developments in industry and has a large impact on economy in general (see also [20]). Whether of not this means a progress for mankind is another point, which is beyond the scope of the present article.

References

1. P. Agnese, C. Cigna, J.L. Pornin, R.L. Accomo, C. Bonin, N. Colombel, M. delcourt, E. Doumayou, J. Lepennec, J. Martignac, V. Reveret, L. Rodriguez, L. Vigroux, Proc. SPIE **4855**, 108 (2003)
2. *Astronomy and Astrophysics in the New Millenium: Panel Reports. The Practical Contributions of Astronomy to Society* (The National Academies Press, Washington, DC, 2001)
3. E. Autio, A.P. Hameri, M. Nordberg, J Eng. Technol. Manage. **13**, 301 (1996)
4. H. Babcock, Publ. Astron. Soc. Pac. **65**, 229 (1953)
5. P. Encrenaz, J. Lequeux, J. Gómez-González, W. Orchiston, The Genesis of the institute of radioastronomy at millimetre wavelengths. J. Astron. Hist. Heritage (2011) in press
6. http://herschel.esac.esa.int
7. http://www.astronet-eu.org
8. http://www.nrao.edu/index.php/learn/radioastronomy/radioastronomyhistory
9. *Innovation and Technology in Hawaii: An Economic and Workforce Profile* (The Hawaii Science and Technology Institute, October 2008), see http://www.hiscitech.org
10. J.R. Janesick, T. Elliot, S. Collins, M.M.M. Blouke, J. Freeman, Opt. Eng. **26**, 692 (1987)
11. P. Léna, F. Lebrun, F Mignard, *Observational Astrophysics* (Springer, Berlin, 1998); P. Léna, D. Rouan, F. Lebrun, F. Mignard, D. Pelat, *L'observation en astrophysique* (EDP Sciences & CNRS Éditions, Paris, 2008)
12. J. Lequeux, *L'Univers dévoilé* (EDP Sciences, Les Ulis, 2005)
13. J. Lequeux, *Le Verrier* (EDP Sciences, Les Ulis, 2009)
14. C. Llewellyn Smith, The use of basic science: basic versus applied science (1997), http://public.web.cern.ch/public/en/About/BasicScience1-en.html and references herein
15. C. Max (2001), http://cfao.ucolick.org/EO/resources/History_AO_Max.pdf
16. M. Nordberg, *Contract Benefits and Competence-Based Supplier Strategies: CERN as a Case Study* (Institute of Technical management, and Espoo, University of Technology, Helsinki, 1994)
17. A. Pannekoek, *A History of Astronomy* (Dover, New York, 1989)
18. V. Pavlakovich-Kochi, A.H. Charney, L. Mwaniki-Lyman, *Astronomy, Planetary and Space Science Research in Arizona* (Economic and Business Research Center, Eller College of Management, The University of Arizona, Tucson, 2007)

19. H. van Woerden, R.G. Strom, J. Astron. Hist. Herit. **9**, 3 (2006)
20. L. Vigroux, Astronomy, technology development and industry, in *The Role of Astronomy in Society and Culture*, ed. by D. Valls-Gabaud, A. Boksenberg, *Proceedings IAU Symposium 260* (UNESCO, Paris), pp. 120–128, (2009) in press

Chapter 8
Up the Decade! Predictions, Prescriptions, and International Collaborations by the American Astronomical (and Other) Communities

Virginia Trimble

Abstract Astronomers have a 200+ year history of international collaborative projects, some more successful than others. Section 1 looks briefly at earlier ones of these, and we return to the topic at the end. The International Astronomical Union remains unique among more than a dozen such organizations in that its primary members are individual scientists (currently about 9,000) rather than nations, academies of science, and other institutions. In addition, American astronomers appear to have been the first community to engage in deliberate surveys of the health of their subject, predictions for its future, and prioritization of widgets required to make their predictions come true. Sections 2.1–5 look in detail at those decadal reviews, their content, procedures, successes, and failures. Section 2.6 addresses very briefly the sixth survey and some of its implications for international astronomy.

Keywords Astronomy: decadal surveys • Astronomy: publications • Bibliography • Sociology of Astronomy • Scientometrics

1 International Introduction

The astronomers brought together at Seeger Observatory (outside Gotha, now a small dot on the map of Thuringia, about three-quarters of an inch west of Weimar) by Franz Xaver von Zach in 1798 may well have been the first modern international scientific meeting [6]. Among its goals was the establishment of a “celestial police”, selected at Lilienthal in 1799 [5], to hunt for the planet between Mars and Jupiter predicted by Bodes Law. Giuseppi Piazzi, observing from Palermo, found Ceres before he had been told he was one of the “policemen”.

V. Trimble (✉)
Department of Physics and Astronomy, University of California, Irvine, CA 92697-4575, USA
e-mail: vtrimble@astro.umd.edu

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_8,

Jump ahead to 1887, when Admiral E.A.B. Mouchez (director at Paris) convened the first meeting of representatives of, eventually, 22 observatories, none in the US. Their goal was to take advantage of the new technology of dry-plate photography to compile a *Carte du Ciel* and *Astrographic Catalogue* of the entire sky. The photographs were completed over several decades, and can now serve as first epoch plates for proper motion studies with a long base line, and the Permanent International Committee of the Carte du Ciel was an important part of the world community until 1914. The measurement of the plates and publication of the *Catalogue* was finally completed in 1964, by which time the publication of the Palomar Observatory Sky Survey (as glass and paper copies of the 48" Schmidt plates) had largely rendered them unnecessary.

At the beginning of the twentieth century came George Ellery Hale's International Union for Cooperation in Solar Research. It met in Oxford in 1905, Meudon in 1907, and Pasadena in 1910, where Schwarzschild's motion to expand its territory to "Astrophysik in Allgemeinen" passed with no opposition [6].

Also before 1914, Jacobus Kapteyn proposed that the best way to determine the overall structure of the "stellar system" (then generally thought to be the entire universe) would be to focus measurements of positions, proper motions, radial velocities, and brightnesses and colors of stars in certain key parts of the sky. His 1906 *Plan of Selected Areas* was a sound one. Although his goals were eventually overtaken by Harlow Shapley's use of globular clusters to map out the Milky Way and the discovery of galactic rotation by Bertil Lindblad and Jan Oort, one still sees studies of various stars and other sources focused in particular Selected Areas.

> When the War was over,
> With the Kaiser out of print,
> So many bought some tortoises
> To watch the beggars sprint.

But not George Ellery Hale who returned at once to the international fray. The treaty of Versailles had officially abolished all previously-existing international scientific organizations (including the Solar Union and the Carte du Ciel Committee), but urged the prompt establishment of new ones – exclusively by and for the victors, with neutrals to be considered later, and the losers much later. Hale was there in Brussels in 1919 for the establishment of what became ICSU and participated in the organization of the first Triennial General Assembly of the International Astronomical Union in Rome in 1922. Commission 32 of the new union provided a home for the Selected Areas program, though the *Carte du Ciel* had included too many German observatories to be fully incorporated into the IAU. Germany, Austria, and Hungary were finally admitted after World War II, though The Netherlands, the first neutral member, adhered from 1922 onward. Commission 32 eventually self-destructed, but Commission 33 is Structure and Dynamics of the Galactic System.

Another relative failure belongs to the post-war period. While the plates for the POSS were being exposed, Fritz Zwicky examined pairs systematically and thereby discovered about 100 supernovae. As the survey wound down, he attempted to persuade colleagues at a number of observatories to carry on supernova surveys in a uniform fashion, and he established a Working Group (under the Commission on Galaxies) of the IAU to coordinate the efforts. But, very shortly before his death [19], he opined that knowledge of rates vs. host galaxy type etc. had not improved in 20 years, because the searches were not sufficiently similar or systematic. The IAU working group died with him; was re-established in 1982 in hopes of coordinating follow-up studies so that the supernovae found would be maximally informative; voted itself out of existence in 1991 because SN 1987A had made the world adequately supernova-conscious; and was re-established at the Prague IAU in 2006. I chaired the group in 1982–1991 and am (probably) still a member but am not quite sure what the primary goals now are.

Several current supernova surveys worry mostly about following up on Type Ia events for use as cosmological probes and, secondarily, learning about nucleosynthesis by all types, so that the statistical issues are no longer regarded as terribly important. The operation of LSST should, in due course, produce numbers vs. anything you might care about for which statistical errors will be vanishingly small (though, of course, we can still make systematic errors!).

It is difficult, I think, to declare any of these intended international projects from before 1920 an unqualified success. In contrast, the IAU itself and the two treaty-based European collaborations, ESO (European Southern Observatory) and ESA (European Space Agency, originally ESRO, the European Space Research Organization) have a great many accomplishments to their credit.

2 Predictions, Explanations, and Advice

Bagehot [3], has written that political commentary can serve three purposes, (1) prediction of what is likely to happen, (2) explanation of the meaning of what has happened, and (3) advice on what ought to be done in the future. These are not independent. If your advice is taken (and you have understood the situation) then your predictions have a fair chance of coming true; while the clarity with which you explain the meaning of recent past events will affect the willingness of taxpayers and their representatives to take your advice. The situation is quite similar for the six decadal surveys of American astronomy and astrophysics, the last of which was released even as I was writing. They have attempted to predict future discoveries, to explain the significance of recent ones, and to urge funding agencies to go ahead with the facilities the advisors regard as most important, with the third function generally emphasized.

The following sections discuss how the process started with the 1964 Whitford report [7] and how it has developed since; relationships between what they asked for and what we got; commentary and predictions about the astronomical community contained in the reports; and similar attempts at reporting and prioritizing for astronomy in other countries and other scientific communities in the USA. In addition to prioritizing facilities (which necessarily also to a certain extent prioritizes the science that can be done) the decadal reports have attempted, to varying degrees (a) to estimate the costs of what they want to do (usually underestimates, of course), (b) to assess the size of the community available to do the work and its likely near-term development, based on graduate enrollments, (c) to explain why the science of the recent past and near future is exciting, even when the impact on GNP is likely to be negligible, and (d) to find examples of contributions from astronomy and astrophysics to science education and technology transfer (formerly "spin-offs") that may actually have enhanced the GNP or may do so in the future.

There seems to be a widespread opinion in the community that the earlier reviews were more successful than the later ones, in the sense of their high-priority items having gone forward. The data do not entirely support this impression (Tables 8.1–8.6). The implication that the current review may be pointless could, therefore, be (a) false, (b) true, or (c) true for different reasons.

A conflict of interest statement: I chaired the panel addressing topic (d) for the fourth review and served on that same panel for the fifth, but have never been part of any of the main committees or of the panels and working groups prioritizing facilities or science. There have been some far more significant members of the community never involved in the decadal process. S. Chandrasekhar (Nobel Prize, Physics 1983) was active in teaching and research at the University of Chicago for roughly half a century, and edited our most prestigious publication, the Astrophysical Journal, for about 20 years. Near the end of his life, I asked him (casually, at a conference reception) why he had never been a member of any of the panels or committees. "No one ever asked me," he responded, going on to quote a verse of an English folk poem, "The Fair Young Maid". His ApJ successor, Helmut Abt, who carried the journal forward for another 20 years, was a member of the main committee for the second review and chaired its panel on optical astronomy. He was also a member of the topic (b) panel (Status of the Profession) for the fourth review.

There has been a certain amount of continuity in the process, with at least a few people having participated in three reviews.

2.1 An Outline of History

In late 1962, the US National Academy of Sciences, through its Committee on Science and Public Policy (COSPUP), established a Panel on Astronomical Facilities whose 1964 report, *Ground-Based Astronomy: A Ten Year-Program*, [7] then fed back in the reverse direction, from the Panel (chaired by Albert Whitford of Lick Observatory) to COSPUP (chaired by George Kistiakowsky, the Harvard physicist) to NAS President Frederick Seitz (also a physicist). The product was almost always called the Whitford Report, though his name appears nowhere on the cover or title page.

This same pattern, a commissioning from NAS to a subsidiary body to a decadal committee and a report making its way back up the line for publication was followed for subsequent reviews, given various titles as follows, but all generally called by their chairs' names:

1972: *Astronomy and Astrophysics for the 1970s*: Jesse L. Greenstein (Caltech) [8]
1982: *Astronomy and Astrophysics for the 1980s*: George B. Field (Harvard) [9]
1991: *The Decade of Discovery in Astronomy and Astrophysics:* John N. Bahall (Inst. for Advanced Studies, Princeton) [10]
2001: *Astronomy and Astrophysics in the New Millenium:* Christopher F. McKee (UC Berkeley) and Joseph H. Taylor, Jr. (Princeton University) [11]
2010: *New Horizons in Astronomy and Astrophysics:* Roger Blandford (Stanford University) [12]

All are Academy members, and their institutions, you will notice, are not a random sample of universities, colleges, observatories, and research laboratories where astronomers are employed. This is also true, though to a lesser extent, of the full memberships of the main panels, their subcommittees, and working groups. All reports after the first were reviewed by some other body of the NAS before publication, though McKee-Taylor was the first to list them by name. Not all reviewers are Academicians, though they are widely known in the community.

The review processes have become gradually broader and more consultative over the decades. From Greenstein onward there were subsidiary specialized panels and/or working groups. From Field onwards there have been open meetings and/or open letters to the community, inviting input. The number of people directly involved peaked at 300+ for the Bahcall report, but the smaller numbers for McKee-Taylor and Blandford have been balanced by more interaction with the community at AAS meetings, etc., and, for the 2010 project, an opportunity for any astronomer or group to submit white papers making the case for specific instruments and other programs. For instance, folks hoping for the International X-Ray Observatory for somewhere around 2021 submitted a couple of white papers to every disciplinary panel from solar system to cosmology.

I will return to demographic issues in Sect. 2.5.

2.2 *What They Asked for and What We Got*

In order to determine these items, I dug out and re-read my own copies of Greenstein [8], Bahcall [10], and McKee-Taylor [11]; borrowed Field [9] from our library, and asked a handful of senior colleagues if by any chance they had a copy of Whitford [7] I could borrow. Most said no, mentioning possible library copies, but an utterly charming answer came from Helmut Abt, who said he had had a copy until about half an hour before, at which time he had put it in the mail to me. And indeed he had.

What I learned is contained in Tables 8.1–8.5, with a summary in 6. The "what they asked for" columns come as directly as possible from the prioritized lists

Table 8.1 The Whitford report (1964)

What they asked for	What we got
3 × 150–200″ telescopes	KPNO 150″, 1973
one in south	CTIO 150″, 1976
48″ Schmidt with it	(UKST, 1973)
4-year study for "largest possible" optical telescope	On-going – now a 50-year study
4 × 60–86″ telescopes at Universities or research institutions on fairly good sites	U Tx 107″ 1969; Las Campanas 100″ 1975; Steward 90″ 1969; UHi 88″ 1970; KPNO 84″ 1984 Catalina/UAz 5 × 60″ 1965–1970, various owners; 4 × 60″ 1967–1971, various places and owners
8–12 38–48″ telescopes at universities	At least 18 built in US 1966–1971; ten owned by universities
Large array of pencil beams to 3 cm eg 100 85′ dishes	VLA 1978, 27 × 25-m dishes
Owens Valley expanded from 2 dishes to 6 or 8, 130′ each	OVRO eventually; BIMA; fused to CARMA 2007 more dishes; various sizes
2 × 300′ fully steerable paraboloids usable down to 3 cm	Greenbank 300′ 1962; 140′ 1965; 3 × 85′ 1965; Goldstone 210′ 1967, but mostly satellite tracking
15 smaller, special purpose radio facilities to "redress balance" with NRAO	Clark Lake 1968; other U Md 1969–1970; Haystack 1964; Hat Creek 1966–1968; Stanford 1970; U Ill 1970; U Iowa 19687–1970; FCRAO 1970; Cornell 1970; Penn State 1968–1972; Harvard 1971
Design study for largest steerable radio paraboloid	No (Arecibo 1963 + upgrades; nothing larger than rebuilt 300′ GBT)
Improved detectors and other peripherals	
Infrared	Leighton-Neugebauer survey 1969, Caltech 60″
Radio receivers for shorter wavelengths	NRAO 12-m mm dish, 1967
Better plates	Most of these things eventually happened, though not all in a decade
Photoelectric cathodes	
Image tubes	
Bigger gratings	
Fabry-Perot's	
Fast cameras (to f/1)	
Improved seeing (AO not mentioned)	
Automation and data reduction	

Table 8.2 The Greenstein report (1972)

What they asked for	What we got
VLA	Yes, 1978
Electronic detectors etc. for 15 "large" optical telescopes	Eventually (not all federal funding)
Test of multielement optical array	Six 1.2 spy spares became MMT, 1979
3 × 100″ optical telescopes	Las Campanas 100″ 1976
Large optical array or 200″	Keck I, 10 m, 1999
Double IR expenditure: ground, air, lab, rocket, balloon	KAO 1975–1995, WIRO 1975; IRTF 1979; IRAS 1983 Learjet ongoing
Design for very large stratospheric tel	SOFIA 2010
Four HEAO's	Three, 1977, 78, 79 (Gamma, x-ray = Einstein, cosmic rays)
Large mm array or 10 m dish	Arizona Radio Observatory; ALMA 2010
More aircraft, balloons, rockets	Not really
Improve existing ground solar facilities	No
Continue OSO series, L, M, N	No, but Skylab 1973; Solar Max 1980–1989
Theory and computation facilities	Rather little; ranked too low, said Field report
Optical astronomy in space, leading to LST for next decade	IUE 1978; HST 1991
Large cm array	No, though VLA pushed to 0.7 cm, and other facilities improved
New astrometric facilities	No (61″ @ Flagstaff was 1963; FGS on HST 1991; HIPPARCOS = European 1989; Gaia – European 2012)
Space launches in 1970s Uhuru 1970 OSO-7 1971 Copernicus 1972 (with UK; UV) SAS-3 1975 OSO-8 1975 ISEE3/ICE 1977 HEAO-1,2,3 1977, 1978, 1979 IUE 1978 SMM 1979	

emphasized in the reports, though in some case the groups also very much wanted something on a longer time scale, for instance a large, space-based optical telescope in Greenstein, though what they asked for was "studies leading toward."

The "what we got" has larger ambiguities. In some cases, the relationship is rather distant in both capabilities and time frame (JWST and Herschel, for instance); for others nothing of the sort ever happened (more large steerable radio paraboloids, and repeated requests for more "medium to large", for the time frame, publicly available optical telescopes). Some happened with little or no federal funding and little public access (the Kecks). In a few cases what we got was very much better

Table 8.3 The Field report (1982)

What they asked for	What we got and when
Major	
Advanced X-ray Facility in Space (originally envisaged as upgradable)	Chandra 1999
VLB Array	1992
15 m new technology telescope	Kecks 1999, 2001
Large Deployable reflector (10 m) in space for FIR, mm	Herschel 3.5 m 2009, ESA leading partner (WFIRST in 2010 report, 1.5 m, NIR)
Moderate	
More explorers, 1–2/year	IRAS 1983, LDEF 1984, COBE 1989
FUSE	1999
Space VLBI	Early trials with TDS; Japanese Halca 1997 task to Japan in 2010 report
Multiple 2–5 m for OIR	No
Advanced Solar Obs. In Space	SDO 2010
Cosmic rays from space	No (Sampex 1992)
SETI	Private/university only
Small (not exhaustive)	
10 m for sub-mm	CSO 1988; H Hertz SMT 1994 (now part of ARO)
MIR interferometer	Townes at Mt. Wilson
High precision optical astrometry	No
10–20 five-year grants for young astronomers per year	Arguably subsumed in NSF CAREER grants, though they argued for separate astro program

than the prioritization implied. This was probably true for the requested near-infrared survey, which became 2MASS, and the lower-priority optical survey with a one-meter dedicated telescope, was utterly outshone by the actual (not federally-funded) Sloan Digital Sky Survey.

The “trees” of Tables 8.1–8.5 make it a little bit hard to see any general picture of the success rate of prioritized requests, which is the reason for the “forest” summary on Table 8.6. Overall, about one third of the requested items came about under (mostly) federal funding within about 15 years of the request, one third happened later and/or with other funding, and one third never happened at all. Indeed eventually the community stopped asking for some kinds of things, like large steerable radio paraboloids and more intermediate-sized optical telescopes. There does not seem to be any strong temporal trend (though the arbitrary 15 year limit has not yet quite expired for the items put forward in the McKee-Taylor report). With the crudest possible metric of Yes/No ratio, Bahcall would seem to have been the most successful and Field the least.

Certainly there does not seem to be a sufficient temporal trend to encourage us to forecast how much of the Blandford request list will eventually come into existence. The reader is invited to formulate assorted “post hoc” explanations, based on speed of change in scientific issues, economic conditions, or whatever else you feel might be important.

Table 8.4 The Bahcall report (1991)

What they asked for	What we got and when
Major	
S(Shuttle)IRTS	S(Space) IRTF = Spitzer Space Tel. 2003
No. Hemisphere IR 8 m	Gemini North, 1999
Millimeter array	ALMA in progress 2010; APEX & ASTE operating
So. Hemisphere 8 m	Gemini South 2001
Moderate	
Adaptive optics	Gradual, many sponsors
FUSE	1999
SOFIA	May 2010
More explorers (8–12)	11 launches (not all explorers) below; UIT, WUPPE, HUT on Shuttle 1995
Optical, IR interferometry	PTOI 1998–2008; NPIO 2001-present; Chara 2004-Present
Several shared 4 m	ARC 3.5 (1984); MDM 2.3 (1986); WIYN 3.5 (1994); ARC 2.5 (1994)
Astrometric Interferometer Mission	To SIM, to SIM-lite; not ranked in 2010
Fly's Eye	1994 (and other UHE gamma and particle arrays)
LEST (large earth-based solar tel.)	AST in McK-T; some AoOs in July 2010, ARRA money
Extended VLA	Endorsed in McK-T; on time on budget for 2012
Collaborations and special instruments	A few
Small (subset)	
2-micron survey	2MASS 1997–2001
IR Interferometry	PTOI 1998–2008, NP10 2001-present, Chara 2004-present
Cosmic background imager	2002 (also DASI, Maxima, Boomerang, etc.)
Lab support	
Astrometric factory	No
300 m radio telescope in Brazil	No
Space radio interferometry	HALCA, Japan 1997; assigned to Japan in 2010
Solar oscillation interferometer	No
Optical survey	SDSS 2000-present
Supernovae neutrino watch	Facilities in Europe, Canada, Japan
Space launches in 1990s	
CGRO 1991 (Great Observatory)	
AXAF/Chandra 1999 (Great Observatory)	
EUVE 1992	
SAMPEX 1992	
RXTE 1995	
ACE 1997	
SWAS 1998	
TRACE 1998	
WIRE 1999 (optical monitor only)	
FUSE 1999	
HETE-1 1996 (failed)	

Table 8.5 The McKee-Taylor report (2001)

What they asked for	What we are getting
Large	
NGST (UVOIR)	JWST (IR only) 2017??
GSMT	GSMT = third large ground priority in 2010 HET (1998), SALT (2003), E-ELT
Con-X	IXO (with ESO & JAXA) = 3/4 large space priority in 2010 report, for 2025
Expanded VLA	On time, on budget for 2012
LSST	Top large, ground priority in 2010 report
TPF	Debudgeted; no target launch date
Single Aperture Far IR (80 m for 30–300 µm)	CALLISTO groups at JPL & GSFC; not in 2010 report
Medium	
TSIP instrument initiatives	In operation
GLAST	Fermi, launched 2008
LISA	3/4 large space priority in 2010 (for 2025) requires collaboration with ESA
Advanced Solar Telescope	Money from ARRA, some instrumentation AoO's out; site on Haleakala
SKA (Square Kilometer Array)	MeerKAT (So. Africa) and ASKAP (Australia) prototypes; US involvement only if reduce other radio projects
Solar Dynamic Observatory	Launched May 2010
CARMA (Fusion of BIMA & OVRO)	2007
Energetic X-ray Imaging Survey Tel.	Explicitly dropped in 2010
VERITAS	2007
Advanced Radio Interferometry between Space and Earth	Explicitly dropped in 2010 (VSOP program in Japan)
Frequency Agile Solar Radio Telescope	Consortium exists; site – VLA or Owens Valley? One of 8 "compelling" mid-scale programs, not prioritized in 2010 report
South Pole Submm Telescope	Not prioritized in 2010 report
Small	
National Virtual Observatory	Has website and some data
LOFAR	European; NL site dedicated May 2010, construction continues
Advanced CR Comp. Experiment for Space Sta.	Not in 2010 report; last Shuttle flight to Space Station likely to carry Alpha Magnetic Spectrometer
Ultra-long-duration balloon flights	Getting longer
Lab and theory	Hard to track

(continued)

Table 8.5 (continued)

What they asked for	What we are getting
Launches in 2000's	Plausible Launches in 2011–2015
HETE-2, 2000	IRIS (Sun)
WMAP, 2001	Nu-STAR (X-rays)
RHESSI, 2002	GEMS (Gravity and Extreme Magnetism)
GALEX, 2003	
SST, 2003	
IBEX, 2005	
WISE, 2009	
SWIFT, 2003	

Table 8.6 Scorecard

Report	Number of identifiable items requested	Number in operation $\leq$15 year after report with mostly federal funding	Number in operation $\leq$15 year after report with mostly other funding (state, private, foreign)	Number eventually built with mostly federal funding	Number eventually built with mostly other funding	Never/very unlikely
Whitford	13	6	0	0	1	6
Greenstein	21	5 (+1 similar)	2	4	1	8
Field	21	3 (+2 similar)	2	3	2	9
Bahcall	29	11	6	5	0	7
McKee-Taylor	23	8	1	5	3	6
Total	**106**	**33** (**+3 similar**)	**11**	**17**	**7**	**36**

2.3 Things We Never Asked For

Uhuru (launched in December 1970) was already flying when the Greenstein report (which mentions some of the early results and asks for four HEAO's as follow-ups) was published in 1972, and Whitford had not addressed space missions. But something with Riccardo Giacconi behind it was going to fly one way or another.

The absence of COBE is more curious. The actual background to the satellite was a NASA announcement of opportunity in 1974, for which three of the responses requested satellite-based measurements of the microwave background to clarify its spectrum and, with luck, find fluctuations other than the $\Delta T/T = 10^{-3}$ dipole due to our motion. These were merged into a single 1976 proposal, with launch in 1989 and exciting results almost immediately (the perfect black body spectrum) and additional ones as time went on (fluctuations on the angular scale of about 10° to which the data were sensitive).

But where is COBE or its ancestor in the Greenstein report? The discussion of radio astronomy and cosmology addresses only radio and other active galaxies. Radio from space mentions only the possibility of extending the baseline for interferometry by launching largish dishes. The main report said that studying the spectrum of the 3 K radiation is very important but advised only the use of balloons and rockets. It also gave significant attention to the possibility of non-cosmological redshifts for quasars and non-cosmological origins for the 3 K radiation.

What went wrong? Well, nobody on the "space" panel was particularly interested in cosmology, while the main survey committee and the "radio" panel each had one cosmologist, and it was the same one, Geoffrey R. Burbidge, whose life-long opposition to a conventional, hot, big bang universe with cosmologically large redshifts was well established by that time.

By the time of the Field report, the need for COBE was generally acknowledged, though the committee expressed fear that, if there was not increased funding for small missions in the 1980s, only three launches would occur, IRAS, COBE, and XTE. In the event, only IRAS (1983) and COBE (1989) went (along with LDEF), with RXTE postponed until 1984.

2.4 More Recent Very Productive Facilities

Well, let us take 3 years of publications (2001–2003) and citations in the 3 years after publication, because the data happen to be available [15], and an arbitrary cutoff of at least 2,000 citations. It is, then, I think fair to say that nearly all the US facilities had been asked for in at least one decadal review.

The items in space were Chandra, and RXTE (plus the non-US XMM and ROSAT) looking at X-rays and indeed both requested; the ultraviolet FUSE, check; the optical HST (multiply requested, only bigger) and the non-US HIPPARCOS (but several committees wanted improved astrometry); in the infrared IRAS (yes) and the non-US ISO (and a later sample clearly brings the Spitzer Space Telescope into the fold [16]. No recent gamma-ray missions reach the 2,000 citation threshold, but FERMI (requested as GLAST) surely will in the next few years. WMAP, the cosmology satellite, was firmly requested as soon as preliminary COBE results appeared.

Of ground-based facilities, the VLA, which was the first priority in the 1972 Greenstein report, has been the most productive and influential radio facility in each of three decades. Next is the JCMT (a UK owned and operated millimeter and submillimeter dish located in Hawaii), which filled a niche addressed by Greenstein and Field, but left to European and university facilities.

The ground-based optical case is perhaps the diceyist. The current superstars (in some combination of papers and citations) are the Kecks, 2MASS, SDSS; plus the European Very Large Telescopes and the 48$''$ Schmidts, the northern one of which predates even the Whitford report and the southern one of which (requested by Whitford) was eventually provided as an Anglo-Australian joint effort. The ranking of facilities by production of papers and citations per paper will, of course, change with time, but even for the most recent sample we have (papers from 2008 cited in 2009, [16]) the two Gemini's together (priorities one and three of the Bahcall report) are outscored not only by the Japanese 8 m Subaru but also by the 3.5 m CFHT. A decade earlier, the CFHT and comparable-sized Anglo-Australian Telescope both outpaced the 4 m at Kitt Peak [14].

2.5 Demographics of the Review Process

The founders of most American scientific societies and other structures in the nineteenth century were white, protestant males. It is instructive to sit down with the lists of members and officers of, for instance, the American Physical Society, the American Astronomical Society, the American Association for the Advancement of Science, and Phi Beta Kappa and notice that the order of incorporation of other groups (first among the members and later, sometimes much later, among the officers) is typically Catholics, women and Jews, Asian and Hispanic Americans, and blacks come last. The decadal panels and committees are, I suppose, more prestigious than the American Astronomical Society but less prestigious than the National Academy of Science, and I thought it might be interesting to look for similar effects among their memberships.

Tables 8.7 and 8.8 display the results and also take a stab at classifying the membership by subfields (optical, radio, theory, and all) and types of institutions (private observatories and high-profile universities; institutes supported by the federal government; 4 year colleges and outreach organizations; industry; other). I am reasonably sure of the numbers, except the Blandford population (the first of the decadal groups where I cannot claim to know fairly clearly who they all are). The Blandford subgroups included several additional hispanic and African Americans, because NAS policy requires at least one per panel. Numbers for subfields, institutions, minorities, and women for the American Astronomical community as a whole in recent years are given in the Blandford report, and earlier reports have somewhat similar information for the 1960s, 1970s, and 1980s.

Table 8.7 Demographics of the survey committees and their panels and working groups, numbers of people

REPORT/Number of Subgroups	Total	Women	Blacks	Hispanics	Asian-Americans	Jews
Whitford – 0	8/0	0	0	0	0	0
Greenstein – 11	23/89	0/1	0/0	0/1	0/0	4/7
Field – 13	21/110	2/8	1/1	0/0	0/1	1/12
Bahcall – 16	15/300	1/31	0/3	0/2	0/4	2/41
McKee-Taylor – 13	15/112	3/13	0/1	0/1	0/4	2/7
Blandford - 9	23/129[a]	6/32	1/0	0/2	0/8	2/15

[a] Plus 71 not named in main report on six Infrastructure Study Groups

Table 8.8 Affiliations and subdisciplines of committee members

Report	Research Univ. Private Obs.	Govt. Obs. Labs	Industry	EPO	Non-US	Opt	Rad	Sp	Th	Inst. Other
Whitford	8	0	0	0	0	4	3	0	0	1
Greenstein	17	6	0	0	0	9	3	2	5	4
Field	16	5	0	0	0	8	1	4	5	3
Bahcall	12	2	1	0	0	3	2	3	4	3
McKee-Taylor	12	2	1	0	0	5	3	1	3	3
Blandford	15	2	2	2	2	7	2	2	6	6

Statements well supported by the data include (a) the panels etc. have never been entirely representative as to institutional affiliations (over-representing the high-profile universities and private observatories), (b) the incorporation of women was slow but (with 20% in the Blandford population) has probably reached equilibrium with the senior membership of the community, (c) even by the dismal standards of science in general and astronomers in particular, black, Hispanic, and even Asian-American colleagues are not getting their fair say, and (d) Jews, who make up something like 3% of the American citizenry are considerably over-represented, as will be seen in just about any academic, intellectual sample you care to collect.

2.6 The Decadal Dinner Cub and the Sixth Report

At the January, 2010 meeting of the American Astronomical Society, the editors of *Nature* convened a group of seven astronomers (whose names, affiliations, and so forth appear in the 18 February issue of *Nature* (vol. 463, p. 868). Given the small number, I think their institutions and subdisciplines are about as representative as the 2010 Blandford group. Table 8.9 shows their priority list. Curiously (as can be seen from the details of the voting procedure described in the *Nature* article) not even LSST was ranked as high as "everybody's second choice." Of their top seven, LSST, GSMT, IXO, and LISA are priorities in the 2010 report. The others are not (but the diners could probably not have been expected to guess that the number two priorities for both space and ground would be several instruments chosen by competitions within the agencies).

It will be 2025 before we can add the Blandford report to Table 8.6, and the only prediction that I am prepared to make is that, because they have asked for very few specific facilities, we will get very few corresponding to their requests!

Table 8.9 The decadal dinner club

Prioritized list	Status in 2001	Status in 2010
1. LSST	No. 5 in "large"	No. 1 "large" ground
2. GSMT	No 2 in "large"	No. 3 in "large" ground
3. TPF	No. 6 in "large"	Not ranked or discussed
4. Con-X/IXO	No. 3 in "large"	Equal 3/4 in "large" space
5. LISA	No. 3 in "moderate"	Equal 3/4 in "large" space
6. JDEM	Not mentioned	Possible NASA/DOE/ESA collaboration; not ranked
7. SKA	No. 5 in "moderate"	Probably not

3 Predictions About the Astronomical Community and Funding

The Whitford report was prepared and published during the post-Sputnik surge of enthusiasm for "space," which very much spilled over into increased funding for astronomy, widely defined, and students wanting to become astronomers. In the decade before the Greenstein (1972) report, the number of AAS members, IAU members from the US, and numbers of astronomy students in graduate programs all roughly doubled. But signs of leveling off were already apparent in first-year graduate enrollments, and over the longer baseline, 1960–1964 to 2009, numbers of AAS and American IAU members have grown at a bit less than 4% per year, somewhat below the "minimum" 4.5%/year estimate in the Whitford report.

Current numbers of graduate students and new PhD's are less well determined, because a good many live in physics or joint physics and astronomy departments. For what it is worth, 47 US institutions currently offer astronomy Ph.D.'s (according to a compilation by the American Institute of Physics) compared to 40 around 1970 and 26 in 1960. There was a good deal of optimism at that level in the Whitford report, which expected graduate enrollments to grow by 19% per year, perhaps tapering to 7%/year by 1973. Particularly optimistic institutions included Georgetown, which expected 45 students by 1966 (but closed its astronomy degree program not long after that) and Maryland, expecting 55, the same number as Harvard. Harvard actually just about reached its goal, awarding 35 astronomy Ph.D.'s between 1968 and 1972 (and finally climbed back up to that level with 35 receiving degrees dated 2005–2010). I think that no other institution can quite make this claim. Caltech, with a more modest goal of 35 students for 1966 awarded 24 astronomy PhDs between 1968 and 1972 (including mine). These numbers come from a combination of alumni directories and web sites and are not entirely consistent in their inclusion/exclusion of folks studying astronomy/astrophysics within physics departments. Leveling and tapering are probably not quite the right description; while fluctuations in numbers of students, numbers of PhDs, numbers taking first jobs in astronomy, and numbers eventually landing in tenured positions are subject to Poisson statistics, real variations, correlated with funding (and perhaps just as important, perceived funding) of pre- and post-doctoral fellowships and numbers of "real" jobs available, seem to be larger.

Because it takes 5–6 years to produce a PhD and another 6–12 to get her into a permanent job, availability of person power has nearly always been out of phase with opportunities. Harvard, Illinois, and Caltech produced their peak numbers of PhDs in astronomy in 1968, 1969 and 1970 respectively, just after some rapidly expanding departments had filled their ranks with whoever was available from a smaller pool (and no, I will not provide examples!)

In the funding department, of course we have always asked for more than was going to materialize (hence the one third success rate shown in Table 8.6). More striking, every report from Greenstein onward has called for more "balanced" programs, meaning more ground-based astronomy, and, within the ground-based priorities, more federal funding for optical astronomy. To achieve this sort of balance within a fixed budget means, of course, less for space, and, on the ground, less for radio, though the language hasn't always been that blunt.

This is perhaps also the right place to provide a warning that we should be careful what we ask for, especially if there is no mechanism for re-prioritizing. Blandford suggests some sort of mid-term review. Once upon a time, a Next Generation Space Telescope, to operate across the full range of ultraviolet, optical, and infrared wavelengths, with an 8 m mirror and total cost significantly less than that of HST up to launch time, seemed like a very good investment. Over the years, the mirror shrank; the UV and most of the optical (and longest IR) wavelengths were removed from the requirements; and the pre-launch costs rose by factors of two or three, or four, depending on where you look. Because the initial number was 10^9 dollars or thereabouts, the balloon has soaked up a very large fraction of the total science support available from NASA. In this regard, the recommendations of the most recent UK planning exercise (Sect. 4) are perhaps of interest.

4 Other Countries[1] and Other Branches of Science

These remarks are based simply on which reports happened to be readily available and are not complete. American astronomy appears to have been the first community to engage in exercises of this sort. Other nations began their prioritizing in the 1980s (and I have to hand Australia [3], Germany [4] in English translation, and the United Kingdom [13]). Among American sciences, the physicists and geophysicists have produced reports, but were slow to adopt prioritized lists, simply declaring an assortment of projects as worth having. This should perhaps become a more general strategy, now that what we are likely to get is more dependent on agreements with other nations and multi-national organizations.

The current Australian Decadal Plan (Australian Academy of Science, 2005) covers 2006–2015 under the title *New Horizons* [2]. It is striking for its emphasis on people, noting early on that the number of employed Australian astronomers

---

[1] See Chap. 9 for the EU.

has been roughly constant 1995–2005, but with permanent positions fading and being replaced by temporary ones (true in a number of other countries as well). And the strategic plan puts students and professional astronomers and engineers first and new facilities last. Among those facilities SKA ranks first on the ground (with the recognition that gradual reduction in funding for the very productive Australian Telescope National Facility will be required), and partial share of some extremely large optical telescope second (they currently support Gemini at the level of 6.2%). The major scientific questions to be addressed include the customary "origins" items (galaxies, stars, supermassive black holes, habitable planets) but also include origin of cosmic magnetic fields and complex interstellar chemistry.

The German report (Deutsche Forschungsgemeinschaft 2003; English translation 2008 [4]) credits by name only a six-member editorial committee, covers 2003–2016, and begins with a set of 23 scientific issues, including cosmic rays, the sun, and final stages of stellar evolution, which have pretty much been squeezed out of American lists in the last report or two. The scientific priorities emphasize the need for on-going involvement in major international projects. Their small/medium/large divisions are in terms of German input to funding with dividing lines at 10 and 25 million Euros. The "demographics" appendix shows the distribution of permanent positions among the states making up the Republic. The top scorers are Bavaria (mostly Garching), Baden-Wurttemberg (mostly Heidelberg), North Rhine-Westphalia (mostly Bonn), and Brandenburg (mostly Potsdam).

The highest-ranked large projects were SOFIA, Herschel, XEUS (now IXO) Darwin (spectra of potentially habitable planets), and ALMA. The "mediums" were LISA, instrumentation for a solar orbiter, access to space telescopes, the VLT, and HESS/MAGIC. SKA appears, but below these others. High priority for small investments starts with shares of Planck, GAIA, NGST, a space-UV facility, and GLAST. Among the existing facilities with major German support, the report recommends holding steady on IRAM[2] (as the major millimeter telescope in the Northern hemisphere), re-evaluation in a few years of Effelsberg (the largest single, steerable dish in the world at 100 m), and planned withdrawal from Calar Alto (in favor of greater support from Spain and by German universities acting individually).

The report issued from the United Kingdom in 2009–2010 (STFC, 2010 [13]) was prepared in the wake of a major reduction in funding available for astronomical (and some other kinds of) research, driven partly by world economic events, partly by mergers and rearrangements of funding agencies, and partly by a drop in the value of the British pound relative to other currencies in which obligations are owed. The combination of pounds down and expenses of new facilities up resulted in a report that, more clearly than any other, recommends withdrawal from both existing and planned projects. The intention is to be part of the European Extremely Large Telescope collaboration (and ESO in general) and SKA, to hang on for GAIA, Planck, JWST, and Herschel (but not IXO), and to move out of Gemini, XMM, LOFAR, and a number of other programs. Funding for the JCMT through

---

[2] See Chap. 7.

2012 was agreed (with the future TBD) but with pullouts from Gemini as soon as possible and the United Kingdom Infrared Telescope (still quite a productive Hawaii-based mirror) in 2014 or 2015. The planned withdrawal from the Anglo-Australian Observatory was completed in 2010. The Blandford report dropped a few projects that had received high priorities in McKee-Taylor, but none that were actually under construction or in use. The UK report is, therefore, precedent-setting. The panel had access to the results of Trimble and Ceja [16]. These data are included as an Appendix. The UK report also has a list of successful technology transfer from astronomy to industry, medicine, defense, and so forth of the type that became customary in the US starting with the Bahcall report.

The interested reader is invited to go web-crawling for recent exercises in prioritization undertaken (I think somewhat unwillingly) by the physics community. A geophysics effort dates from 1991 and devotes a proportionately much larger fraction of text to scientific priorities (at least somewhat ranked and with global paleo-environments and biological evolution at the top). But there is also a section on the facilities, equipment, and data bases that will be needed (they came to the idea of an NVO somewhat ahead of astronomy, I think) and another with suggestions for future funding (more, of course).

5 Interdisciplinarity and International Collaborations in the Future

Disciplines come and go, and it was not silly for one pundit, making forecasts for the year 2100 to ask, “What are the major branches of physics, if physics still exists.” After all, a century ago, the important branches of theoretical physics were mechanics, thermodynamics, acoustics, electricity, and electromagnetism [18]. So while there have been recognizable astronomers for a few hundred years, some of them were trained in medicine (Copernicus, Rhetticus, and Henry Draper), some in mathematics (Eddington, Milne, and McCrea), and some in other things (Galileo, Newcomb, Curtis, and Shapley). And astronomy has tended to swallow people whose skills were needed, spectroscopists and what we would now call atomic physicists from about 1880 on, nuclear physicists from 1950 on, and most recently particle physicists.

Mergers in progress probably include astrochemistry (with ever-increasing complexity of interstellar molecules and pathways for forming them) and astrobiology[3] (said by some to be a subject with no subject matter to study). But the “interdiscipline” that is changing most rapidly and most rapidly changing our landscape is undoubtedly computational astrophysics, essential for handling ODARs (overwhelming data arrival rates) as well as for turning equations of cosmology, star formation, radiative transfer, and all the rest into “simulations” to

---

[3] See Chap. 20.

be compared with those data (which can typically require another giant computer program to pretend to be a telescope looking at the data). Any number I might give you for lines of code, flops, bytes, cores or any other measure of computing power and complexity will be out of date before this volume appears, but for a snapshot of one topic in July, 2010, see [1]. Incidentally, every decadal report has asked for additional support of computational facilities, right back to Whitford, where it was called automation of acquisition and reduction of data and information storage, but never, as “Field” said about “Greenstein”, with very high priority (Section 5 of 7 in the Blandford report).

We began by looking at attempts at international collaborations, back when nations were both much smaller and many fewer, none of which was overwhelmingly successful, so perhaps it makes sense to end with a look toward international collaborations of the future. The early US decadal reports carry a strong flavor of “If the US doesn’t do something, it won’t get done.” This was already not entirely true in the Whitford era. The Southern 48” Schmidt was built as the UKST (United Kingdom Schmidt Telescope); repeated requests for better astrometry were largely met by the European HIPPARCOS satellite (whose public data base was never much used by the American astronomical community, present company excepted, of course; [17]), and the again European Gaia will presumably be the next step toward better positions, parallaxes, and proper motions. Much of the millimeter and submillimeter radio astronomy from the 1970s onward was done with European facilities, including the British JCMT.

Having survived WWII and the Cold War, the IAU has been a considerable success, now involving more than 60 countries, including two Chinas, and more than 9,000 individual members, and sponsoring nine symposia per year as well as a triennial General Assembly. These symposia are chosen by agreement among the Presidents of its 12 (mostly subdisciplinary) Divisions and the 12-member executive committee of the Union itself. Attendance at the GAs has run around 2,000 people in recent years. In contrast, the General Assemblies of the International Union of Pure and Applied Physics involve only about 150 people, representing member countries and the subdisciplinary Commissions. But no Commission has more than 15 members, and only the chairs attend the GAs. The number of sponsored conferences is larger, but they are chosen by individual Commissions, with no general discussion of what might be good for physics as a whole.

What has become increasingly clear in the last 15 years is that no one nation or continent can afford the full suite of telescopes, satellites, and all the rest needed to push astronomy forward at all wavelengths and all resolutions on all kinds of sources. Already noted above are the Anglo-Australian and Canada-France-Hawaii optical telescopes, which yielded the largest numbers of papers and citations for about a decade (when the Palomar 200” and the KPNO and CTIO 4 m mirrors were also collecting photons). Perhaps the most spectacular international satellite has been the International Ultraviolet Explorer, launched in January 1978 and operated until September 1996. It was a NASA, ESA, SERC (UK) joint mission, operated in real time 16 h a day from Goddard Space Flight Center and 8 h from the

ESA tracking station near Madrid. Early on, it was very much like a ground-based telescope, in that the observer went to the tracking station and watched the spectrum gradually build up on a video screen.

The IAU has had, from time to time, a Working Group on Future Large Scale Facilities, but in practice this has served only to make sure that everybody coming to its sessions at the GA found out what everybody else was planning. It could conceivably develop into a consultative body of the sort recommended in the Blandford report to think, at least every 5 years, about international collaborations. US membership of the IAU amounted to about one third for many years. It is now closer to one-quarter.

At the moment, I am not quite sure that the US yet has the right attitude toward these matters. The Blandford report clearly acknowledges that the choice of LISA vs. IXO as a major space mission for 2020 or beyond must be coordinated with the European choices within its Cosmic Visions program (Anderson), but digging into small print one finds phrases like "nearly all of this report's ranked recommendations have opportunities for contributions – often substantial – by foreign partners" (p. 304) and, concerning some merger of the American JEDEM mission and the European Euclid, "if… the arrangement is consistent with the US playing a clear leadership role."

The situation on the ground is not much better. If there are to be two Terribly Large Telescopes, it seems reasonable to put one north of the equator and one south. Now the Europeans have decided to put their E-ELT in Chile (the second choice site was La Palma). The US has two competing projects – with the decadal report [12] strongly urging NSF to get behind one or the other as soon as possible, to give other international partners confidence. Of the two, the GMT is definitely aimed toward Chile and the TMT toward Hawaii, but the Blandford report doesn't seem to feel that this should be a consideration in choosing between them. If I had my druthers, I would put the TMT in Chile and the E-ELT in the Canary Islands, which happens to be a pleasant place to visit, but, like Chandra, I wasn't asked, which is perhaps a good place to stop.

Acknowledgments I am most grateful to Michael Rowan-Robinson for requesting from me data on citation rates to papers reporting observations taken with various telescopes for consideration by the most recent UK decadal review, to Ron Ekers for a copy of the Australian New Horizons: A Decadal Plan for Australian Astronomy 2006–2015, and especially to Helmut Abt for the loan of his copy of the Whitford report and his response to an email query about whether he might have one, saying that he had had one until about half an hour before, but he had just put it in the mail to me! The keyboarding of text and tables was expertly done by Alison Lara.

References

1. J. Alves, B.E. Elmegreen, J.M. Girart, V. Trimble (eds.), *Computational Star Formation*, in Proceedings of IAU Symposium 270 (Cambridge University Press, Cambridge, 2011)
2. Australian Academy of Sciences, National Committee for Astronomy, *New Horizons: A Decadal Plan for Australian Astronomy 2006–2015*, (2005)

3. Bagehot, *The Economist*, 3 July 2010, p. 56
4. Deutsche Forschungsgemeinschaft (*English translation 2008) Status and Prospects of Astronomy in Germany 2003–2016* (2003)
5. M. Hoskins (ed.), *The Cambridge Illustrated History of Astronomy* (CUP, Cambridge, 1997), p. 188
6. R.D. Jarrell, J. Astron, Hist. Herit. **12**, 127 (2010)
7. National Academy of Sciences, *Ground-Based Astronomy: A Ten-Year Program* (Whitford Report, 1964)
8. National Academy of Sciences, *Astronomy and Astrophysics for the 1970's*, vol. 2 (Greenstein Report, 1972)
9. National Academy of Sciences, *Astronomy and Astrophysics for the 1980's*, vol. 3 (Field Report, 1982)
10. National Research Council, *The Decade of Discovery in Astronomy and Astrophysics*, vol. 2 (Bahcall Report, 1991)
11. National Research Council, *Astronomy and Astrophysics in the New Millennium*, vol. 2 (McKee-Taylor Report, 2001)
12. National Research Council, *New worlds, New Horizons in Astronomy and Astrophysics Prepublication Copy* (Blandford Report, 2010)
13. STFC (2010), http:///www.stfc.ac.uk/Resources/PDF/GBFRFinal.pdf (the Robinson Report)
14. V. Trimble, Scientometrics **84**, 21 (2010)
15. V. Trimble, J.A. Ceja, Astron. Nach. **329**, 632 (2008)
16. V. Trimble, J.A. Ceja, Astron. Nach. **331**, 338 (2010)
17. V. Trimble, A. Kundu, Astron J. 115, 358 (1998)
18. F. von Hippel, Phys. Today, 41 (June 2010)
19. F. Zwicky, in *Supernovae and Their Remnants*, ed. by C.B. Cosmovici (Riedel, Dordrecht, 1974), p. 1

Chapter 9
Building a Strong, Unified European Astronomy

Johannes Andersen

Abstract European astronomy owes its present positions of leadership to the development of pan-European cooperation. For many years, this happened mainly through a few international organisations, chiefly the *European Southern Observatory* (ESO) and the *European Space Agency* (ESA). Their success highlights the potential of the much greater resources invested in European astronomy through national programmes, especially when including university institutes and staff. From 2005, the ASTRONET consortium of funding agencies for astronomy plus ESO and ESA has worked to develop a comprehensive strategic plan for coordinating these investments. The aim is to globally optimise their scientific returns and include all of Europe. Based on a long-term *Science Vision* (2007), the ASTRONET *Infrastructure Roadmap* (2008) describes a coherent investment plan for new infrastructures at all wavelengths, on the ground and in space; the necessary underpinning of theory, computing and human resources; and initiatives to maximise their benefit for society in general. This article describes the ASTRONET process, its status as of 2010, and the strategy for implementing its plans over the next several years.

Keywords Sociology of Astronomy

1 Background

Over the past half century, European astronomy has progressed from backwater to front-row player. This has primarily been achieved by pooling national human and financial resources to create organizations and facilities beyond the capability

---

J. Andersen (✉)
Nordic Optical Telescope Scientific Association, Apartado 474, ES-38 700 Santa Cruz de La Palma, Canarias, Spain

The Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark
e-mail: ja@not.iac.es; ja@astro.ku.dk

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_9,

of any single European country. The outstanding examples are, on the ground, the *European Southern Observatory (ESO)* which operates the world's leading optical observatory, the Very Large Telescope (VLT) at Cerro Paranal in Chile, and is the European partner in the global radio telescope project ALMA; in space, the astrophysics missions in the science programme of the *European Space Agency (ESA)*.

But European astronomy is much more than these international organisations: several national facilities, e.g. in radio astronomy, are also of world-class standard and complement the capabilities offered by ESO and ESA. Moreover, the first-class engineers and scientists who design and build all these facilities and – crucially! – conduct the actual research with them are distributed in universities and research institutes all over Europe – and not only in the Western half. In the future, front-line research infrastructures will be even larger and more expensive than today, and demands on cost-effectiveness will no doubt increase. In order to remain globally competitive, we must learn to coordinate the use of *all* the technical, financial, and human resources of European astronomy more comprehensively and effectively than ever before, based on a science-driven long-term plan.

A model exists in the Decadal Surveys, which have been conducted in the USA for half a century (see Trimble) under the aegis of the National Academy of Sciences (NAS). After years of fruitless searches for an equivalent European host organisation, the European funding agencies for astronomy created the ASTRONET consortium in 2005 to address this task. The European Commission (EC) supported this initiative with a 4-year ERA-NET grant of 2.5 MEuro under Framework Programme 6 (FP6) for 2005–2009, later extended until the end of 2010. It was the consensus opinion of the partners, however, that an initiative such as ASTRONET was a necessity whose time had come, with or without the EC grant.

The following is an account of the activities and results of the first 5 years of ASTRONET. More comprehensive and up-to-date information is maintained at the ASTRONET web site at http://www.astronet-eu.org.

2 Organisation of ASTRONET

ASTRONET began on September 1, 2005, as a fairly small consortium of funding agencies signing the EC contract: The Ministries of Education and Science of Germany (BMBF) and Spain (MEC, now MICINN), ESO, the Centre National de la Recherche Scientifique (CNRS, France), the Particle Physics and Astronomy Research Council (PPARC – now STFC, UK), the Netherlands Organisation for Scientific Research (NWO), the Instituto Nazionale di Astrofisica (INAF, Italy), and the Nordic Optical Telescope Scientific Association (NOTSA, representing the five Nordic countries). For formal administrative reasons, Projektträger DESY (PT-DESY) – the executive arm of the BMBF – also joined as a contractor; ESA and the Max Planck Gesellschaft (MPG, Germany) as Associate Members.

The Institut National des Sciences de l'Univers (INSU) of CNRS serves as Coordinator of ASTRONET. Drs. Anne-Marie Lagrange and Fabienne Casoli were the Scientific Coordinator and Project Scientist who established the whole project.

2.1 Structuring the Task

The tasks to be performed under the EC contract were divided into a number of Work Packages. Each contains several tasks, and one contractor is responsible for each Work Package and task. A Consortium Agreement was concluded between all contractors to regulate the governance structure of ASTRONET, which is headed by a Board consisting of all partners. An Executive Committee consisting of the workpackage and task leaders and the Chairperson of the Board, chaired by the Scientific Coordinator, monitors the project regularly and prepares the decisions of the Board. Finally, a Project Office at INSU manages day-to-day business.

This organisation has worked very well, even as ASTRONET has grown, partner organisations have changed names, and individuals have been replaced for various reasons. Dr. Jean-Marie Hameury is Scientific Coordinator since 2006; the present author has had the privilege of chairing the Board since the beginning.

2.2 Membership

A specific objective of ASTRONET was to engage *all* European astronomical communities in its endeavour in such ways as may be possible here and now. To this end, three levels of membership in ASTRONET were devised:

Full Contractors are national or regional organisations that fulfil the formal EC criteria for participants in an ERA-NET and are responsible for meeting the goals and providing the deliverables specified in the contract. Contractors (except BMBF) manage a share of the EC funding, with associated obligations as regards financial reporting – not a trivial matter. Thus, only one contractor has joined ASTRONET since the start, the National Research Agency (NCBiR) of Poland.

Associate Members are national or regional organisations that are *de facto* responsible for the development of astronomy within their geographic or scientific domains. Associates commit to supporting the goals and recommendations of ASTRONET and contribute to some of its administrative tasks, but have no formal responsibility towards the EC and do not themselves manage funds from the contract.

Associate Members participate fully in the Board meetings (with travel support from ASTRONET), except for formally voting on matters directly related to the EC contract, and they participate in as many ASTRONET activities as their status and human resources allow. Hence, their communities remain as fully integrated in the ASTRONET programme as those of the Contractors.

Forum Members are national research, funding or educational institutions that want to stay fully informed of ASTRONET's activities, but make no commitments to the work programme and have no obligations to the EC. Forum members are invited to Board meetings with speaking, but no voting rights.

This membership structure has proved very effective. The formal rules for acceding to an ERA-NET contract are fairly strict and, in practice, exclude many of the types of organisation that *de facto* represent astronomy in many European countries. Keeping the number of signatories of the EC contract down has the considerable advantage of minimising the bureaucratic overhead of the project, and the way ASTRONET operates has been designed to maximise the involvement of all interested partners through a pragmatic interpretation of formal rules.

3 Developing the Strategy

The top-level aspirations of ASTRONET were sixfold: *(i)* to pioneer science-based long-term strategic planning for the development of *all* of European astronomy, including new research infrastructures on the ground and in space, at all electromagnetic wavelengths and including particles; *(ii)* to include both new and existing infrastructures in the scientific and financial planning; *(iii)* to base the financial planning on project lifetime costs; *(iv)* to include theory, computing and archiving, training and recruitment of the all-important human resources, and relations to industry and society; *(v)* to include *all* of the new Europe in this endeavour; and *(vi)* to establish joint science-based long-term planning, followed by corresponding common actions, as a permanent feature in European astronomy.

3.1 Previous Experience

The Decadal Surveys of US astronomy have long been the standard example of a science successfully developing its own comprehensive plans and priorities. They were the obvious inspiration for ASTRONET, but importing the concept to Europe is not simple: First, the political, administrative, financial and cultural structures underpinning European astronomy are diverse and vastly different from those of the USA. Second, the whole concept of comprehensive strategic planning is new in many European countries, already at the national level – let alone for Europe as a whole.

Successful European cooperation was, of course, demonstrated already by ESA, ESO and CERN. Since 2000, the discipline-oriented *Integrated Infrastructure Initiatives* (aka I3s) OPTICON for optical-infrared astronomy (http://www.opticon-eu.org), RadioNet for radio astronomy (http://www.radionet-eu.org), and ILIAS for astroparticles (http://www-ilias.cea.fr) have also worked to promote planning and coordination within their respective fields, supported by substantial grants from the EC (~ 40 MEuro total in FP6 alone).

The I3s have greatly stimulated networking, coordinated common technology development and provided trans-national access to all modern European mid-size facilities in their fields. Thus, the I3s have made even more European astronomers discover the advantages and joys of working together more closely. This experience undoubtedly prepared the ground for the far more ambitious ASTRONET initiative, even though ASTRONET cannot fund development of anything other than plans. OPTICON and RadioNet have provided invaluable input to the ASTRONET planning process, and their Coordinators are invited to all Board meetings.

3.2 Planning the Planning

The first step in preparing the ASTRONET work programme was to “Plan the Planning” – i.e. devise a process that would lead to high-quality planning documents that would be credible to the funding agencies as well as the scientific community. Fusing lofty scientific aspirations with sound technical assessments and realistic financial constraints into a single plan that is feasible in a continent of great political and cultural diversity is a tall order. Yet, the ASTRONET partners – the agencies that pay for it all – were convinced that a good common plan would lead to better science and better use of their money than the previous fragmented approach.

A two-stage process was devised:

A Science Vision for European Astronomy would be developed first to define top-level and secondary priorities for the main scientific questions that European astronomy should address over the next 10–20 years. It would also define the tools needed to answer those questions. Such new facilities would be described in generic terms, but specific projects would not be discussed: To facilitate the start of an unfamiliar process, thorny issues of competing projects, priorities and funding were deferred to the next stage, the *Infrastructure Roadmap*.

The Infrastructure Roadmap, in turn, would bite all those bullets and assemble a comprehensive plan for coordinated investments in infrastructures at all wavelengths, in space and on the ground. Its remit comprised theory and computing, networks and archives, laboratory astrophysics, and human resources, including training and recruitment of scientists and engineers, education and outreach, relations to industry and benefits for society as a whole.

In order to gather general support for these documents by users and agencies alike, both were developed in similar two-stage processes. A Working Group (WG) was appointed for each report by the ASTRONET Board, with full independence regarding its recommendations. Each WG was supported by a number of topical panels with remits decided by the WG itself. Potential lobbyists for specific projects were not appointed to the WGs, but substantive discussions of each project were needed for the panels to develop informed recommendations. These were subsequently integrated into the full report under the responsibility of the WGs.

The WGs and Panels were instructed to make maximum use of recent long-term plans and roadmaps by national and international agencies: Nothing is gained by reinventing well-designed wheels, especially not wheels that are perpendicular to existing ones designed by similar groups of people. Developments and plans by potential global partners in large projects would be taken into account as well. However, each WG had sole responsibility for the content of its report.

In total, the two WGs and their panels included well over 100 of the best scientists in Europe. It was gratifying to find that so many were willing to devote their precious time to this task, which sceptics had declared impossible from the start.

After a sanity check by the ASTRONET Board, each draft report was posted on the WWW for comment by the community during 1–2 months. The draft was next discussed at an open Symposium attended by 250–350 persons, substantial travel support being provided by ASTRONET. In allocating travel support, priority was given to securing the attendance of representative individuals from all Member and Associated States, regardless of their level of membership (see Sect. 2.2). At each Symposium, reports were also presented on recent experience with the implementation of the previous US Decadal Survey. After the symposium, further input from the community was also collected via the WWW. Based on this comprehensive advice, the WGs then finalised and published their reports; the Roadmap WG even posted detailed replies on the WWW to the many comments received.

This somewhat elaborate procedure was designed to maximise community support for the compromise recommendations that are a fact of life in such efforts. Experience shows that if ample opportunity has been provided to present all good ideas and all valid criticism, scientists accept more readily that not all their wishes can be fulfilled in the real world.

3.3 The Science Vision

The *Science Vision* WG was led by Prof. Tim de Zeeuw, then of Leiden University, Netherlands, subsequently Director General of ESO. It published its report *A Science Vision for European Astronomy* (see the ASTRONET web site) in September 2007. Its four panels addressed the key unanswered questions within the vast field of astronomy under the following headlines:

- *Do we understand the extremes of the Universe?*
  This panel covered the Big Bang, dark matter and dark energy, black holes and neutron stars, γ-ray bursts and supernovae, and cosmic rays.
- *How do galaxies form and evolve?*
  Subtopics under this heading included the emergence of the first stars and galaxies, the formation of large scale structure, the origin of the heavy chemical elements, and the assembly of galaxies over time, including or own Milky Way.

- *What is the formation and evolution of stars and planets?*
  This panel addressed the formation of stars from gas and dust, stellar structure and evolution, the life-cycle of interstellar matter and stars, the formation and evolution of planetary systems of great diversity, and the evidence for life elsewhere in the Universe.
- *How do we fit in?*
  The final group of topics concerned the Sun as an (astro)physical laboratory; the Solar-terrestrial relations; the complementary information available from comparison of extrasolar planets and our own Solar System, also on the history of the latter; and the search for life elsewhere in the Solar System.

Within each topic, the questions and subtopics were developed in sufficient detail to define realistic research projects likely to lead to an answer within the period considered, and the types of tools that would be needed. It should be emphasized that, although the Science Vision report did not discuss specific projects, it defined the comprehensive and balanced scientific basis on which the development of all other ASTRONET activities rests: The overarching goal is *Science*, not management.

The Science Vision represents a first interesting departure from the US model: It was developed as a self-contained process and report, separate from the technical and financial issues covered by the Roadmap. In addition to providing a cleaner structure and start of the planning process, this opens the possibility of revising the two documents separately, on the timescales of scientific vs. technical progress.

3.4 The Infrastructure Roadmap

The *Infrastructure Roadmap* WG was chaired by Prof. Michael F. Bode of Liverpool John Moores University, UK, and accomplished a Herculean task to publish its report (http://www.astronet-eu.org/IMG/pdf/Astronet-Book.pdf) in November 2008. The ambition was not only to compile a list of potential projects, such as the Infrastructure Roadmap of the government-level *European Strategy Forum for Research Infrastructures* (ESFRI). The ASTRONET Roadmap would also address the thorny issues of scientific priorities and independently and realistically assess the schedule, technological readiness and budget of individual projects, including space and astroparticle projects. It would further consider the basis in theory, computing and human resources that is required to build, operate and underpin the new large facilities, and do the science that is the goal of it all. The relations to industry and the societal benefits of astronomy are also covered. None of these links in the complete "food chain" of science is addressed in the ESFRI Roadmap.

All this was to be assembled into a coordinated investment plan that included realistic timescales and costs of constructing and operating new facilities alongside existing ones, as well as the other aspects listed above, within a plausible overall budget envelope. And there could be no cheating on the numbers: ASTRONET itself consists of the agencies that pay for it all!

Given its comprehensive scope, the *ASTRONET Infrastructure Roadmap* is in fact a blueprint for a *European Research Area* in astronomy, rather than merely an *infrastructure* Roadmap. It is an impressive achievement, which has earned much deserved praise and has given astronomy a gratifying reputation in Europe as a truly well-organised science.

The Roadmap WG formed panels to review and develop recommendations on the following five areas within its remit. The panels were asked to not only review and rank existing proposals for new projects, but also to identify any significant overlaps or gaps in the overall complement of facilities. To this end, it collected information in a uniform format through questionnaires sent to well over 100 projects, over 90% of which responded. The final list of projects contains those requiring new European funds of ten million Euro or more, and for which spending decisions were needed after 2008. They were divided into small, medium and large (10–50, 50–400, and >400 million Euros, respectively), and also into short-, medium- and long-term projects (time to operation of ~2015, 2016–2020, and >2020).

- *High Energy Astrophysics, Astroparticle Physics and Gravitational Waves*
  Panel A reviewed requirements and developed priorities for ground- and space-based facilities within this wide area, in close cooperation with our "sister" organisation in astroparticle physics, ASPERA, and with the I3 ILIAS. It was reassuring to find that, within the areas of overlap, there was general agreement on the top-priority projects, even if relative priorities could differ due to the different perspectives of pure physics vs. ASTRONET's focus on astrophysical sources.
- *Ultraviolet, Optical, Infrared and Radio/mm Astronomy*
  Panel B had the task of reviewing and prioritising the facilities of the future across all electromagnetic wavelengths above X-rays, on the ground and in space – a scope that includes essentially all the largest new ground-based facilities. Input from the ESO long-term plan and assistance from the discipline-specific infrastructure networks OPTICON and RadioNet were essential in this task. Again, as with Panels A and C, it was satisfactory to find that, although defined through an independent procedure, the resulting recommendations agreed very well with those developed at the same time for the ESA *Cosmic Vision*.
- *Solar Telescopes, Solar System Missions, and Laboratory Astrophysics*
  Panel C had a relatively easy task as regards ground-based Solar telescopes, as the European Solar physics community had already converged on a single new major facility, the 4m European Solar Telescope (EST). The EST is intended to replace the existing facilities within a decade or so. Solar System missions are quite another matter, due to the diverse scientific aims and generally very high cost of such projects, which essentially makes them feasible through global cooperation only. Here, input from the EuroPlaNet network was invaluable. Finally, the Panel considered laboratory astrophysics, which provides vital physical data underpinning most of the facilities under Panels A and B, but also includes curation of samples returned by interplanetary space missions.

- *Theory, Computing Facilities and Networks, Virtual Observatory*
  The topics of Panel D included theory development as well as what is now called astronomy-related e-Science. Developments in this area happen at breathtaking speed and typically involve close networking comprising at least all of Europe. Coordination is needed to steer them in the most constructive directions.
- *Education, Recruitment and Training, Public Outreach*
  Science is done by humans, with theory developed by humans and facilities built and operated by humans. The review of human resource issues by Panel E is what turns the ASTRONET Roadmap into a comprehensive strategy for the healthy, organic development of European astronomy as a whole. Thus, the recommendations of Panel E address the whole "food chain" of astronomy, from public interest in science, attracting school children to the sciences, training and recruiting the highly skilled staff needed to build and operate the next generation of large research infrastructures, and the returns to the society that supports it all. Relations to industry were covered under the last item. Drawing on the intellectual resources of the New Member States is a vital element in the strategy.

The detailed individual recommendations are described in the 175-page Roadmap report and will not be summarised here: Astronomers will want to consult the report itself, while the scope and structure of the process are perhaps more interesting to the lay reader than the specific details. Moreover, regardless of ASTRONET plans and priorities, major projects live, change or die on their own as a function of scientific, technical, political and financial circumstances at critical times. Indeed, the global financial landscape has already changed markedly since 2008.

What appears important in the long-term perspective is, first, that a document now exists to describe the comprehensive background on which decisions on specific projects are taken – often by completely unrelated organisations or individuals. Second, the proof has been made that at least a peaceful science like astronomy is able to sort out its differences and present an agreed, coherent plan for the future to the funding agencies. To be sure, ASTRONET has no authority to force national agencies to take any specific decisions, but the common parenthood to its recommendations is a strong signal that "common sense" should prevail.

4 Engaging all of Europe

Strategic planning tends to be associated with the usual few large, centrally organised, wealthy Western European countries that tend so set the course on the international scene. Measured in financial capacity in the short term, this may be true. Measured in intellectual resources it is not, even in the short term. For the long term the ambition for astronomy must be, as for the European Union itself, to liberate the financial and human resources of the newer Member States, for the benefit of all. Astronomy cannot accomplish this by itself, but it can perhaps lead the way.

The flexible, pragmatic membership structure of ASTRONET was designed to facilitate this task. At the end of 2010, largely due to the efforts of Dr. Birgitta Nordström on behalf of NOTSA, ASTRONET comprises 10 Contractors, 20 Associates and 6 Forum Members representing a total of 29 European countries with a combined population of just over 550 million inhabitants. Among astronomically developed countries, only Ireland is still missing. ASTRONET has indeed become fully European.

5 Coordination of Resources and Procedures

The financial resources for European astronomy are predominantly national when all is included. Two necessary conditions for pooling and coordinating the use of national resources are *(i)* an approximate, but consistent inventory of those resources, and *(ii)* a minimum degree of similarity in the scope, allocation procedures and schedules of national research grant programmes. Tasks to improve the situation on both fronts are included in the ASTRONET work programme.

5.1 Inventory of Resources

The task to establish a comprehensive and complete inventory of the financial and human resources for astronomy, and their organisation, started out with great ambitions. It soon turned out that only for a few, large countries with a strongly centralised structure (e.g. France, Italy) was this possible at all, and their funding and staffing structures were generally found to be incompatible. In less centralised countries (e.g. Germany, Switzerland), no single government agency possesses all the relevant facts. In both cases, official organisations such as those adhering to ASTRONET are only allowed to release exact, official numbers, while for planning purposes an accuracy of, say, 10% is often quite satisfactory. This exacerbates the general obstacle that national agencies in small countries are often unable to shoulder the effort needed to provide such statistical information.

Based on the experience from this exercise – ironically thought to be among the easier of ASTRONET's tasks! – two pilot projects were launched:

One, led by NOTSA, was designed to answer the question, "Can a set of questions be designed to provide the minimum information needed for long-term planning, while keeping the effort involved to a level where the questions are actually answered by all?". The scope of this pilot project comprises the five Nordic countries. Documenting the corresponding effort by all involved is part of the project, which should be complete by end of 2010.

The other, led by NCBiR of Poland, aims to provide a similar body of information about astronomy in the New Member and Associated States, with a content and procedures adapted to the circumstances of this part of Europe. This task will also be largely complete by the end of 2010.

5.2 The Common Call

The ultimate goal of the EC for the ERA-NET scheme is to encourage national funding agencies to pool their resources and develop mechanisms for funding joint research projects that draw on the experience and facilities of scientists in several European countries. A so-called "Common Call" is therefore a mandatory part of any ERA-NET contract.

In a Common Call, a consortium of funding agencies agrees on a common research theme, and the participants pledge funds to support research projects within that theme. Funding proposals from project teams comprising researchers from at least three Member or Associated States are submitted to a single address by a single deadline. They are then peer reviewed by a single evaluation committee appointed by the consortium.

Funding for the successful projects may be provided in two different ways, through a "Real Common Pot" or a "Virtual Common Pot". In the former model, all funding is pooled and allocated to the teams without regard to the nationality of the individual scientists; in the latter, each national agency funds only its own participants in each project. In an ideal world, the "Real Common Pot" is clearly preferable – supporting the best science and the best scientists in Europe without regard to nationality. In the real world, the "Virtual Common Pot" requires minimal surrender of sovereignty and departure from established national principles and regulations, and is therefore the most common funding model so far.

ASTRONET had only committed to prepare a Common Call during its first 4 years, but already in early 2008 announced an actual call with the theme "Common tools for future large submm facilities". The proposal and evaluation procedure went smoothly, and the successful projects received their funding effective January 1, 2009. Preparations for a second Common Call started in the last half of 2010.

6 Current Status and Initiatives

At the end of 2010, just over 5 years after the start, ASTRONET must qualify as a resounding success. To mention a few highlights:

- Despite much initial scepticism as regards the feasibility of such common planning in Europe, the *Science Vision* and *Infrastructure Roadmap* have been completed and published, essentially on schedule.
- As detailed in Sect. 4, ASTRONET today includes representation of virtually all European countries with significant activity in astronomy, with a combined population of 550 million people.
- The strong and comprehensive Roadmap has given astronomy a reputation in European research infrastructure management circles as a science that has "really got its act together" in a way that serves as a model for others.

- Given that some of the proposed large projects will require global cooperation, it is noteworthy that the most recent US Decadal Survey report (2010) systematically refers to what "Europe" plans and intends. Presenting a common strategy and priorities has clearly made Europe a more credible and influential partner in global projects. The Decadal Survey report also recommends that close cross-ocean contacts be maintained over the decade to coordinate global planning even better – a proposal that ASTRONET warmly welcomes.
- As noted above (Sect. 5.2), the first Common Call was launched ahead of schedule. The funded projects are ongoing, and a second call is in preparation.
- Finally – but most importantly – ASTRONET has been awarded a new ERA-NET contract for 2011–2014 under FP7 to follow up the implementation of the Roadmap and further strengthen inter-agency coordination in Europe. The final goal for the second contract period is to establish such coordination as a self-sustaining activity and make EC funding superfluous by 2015.

Overall, the ASTRONET initiative has been far more successful that anyone dared dream when the proposal was prepared in early 2005, given its ambitious programme and unfamiliar nature at the European – and often even national – level. To be sure, a couple of tasks remain unfinished, notably the systematic inventory of resources for astronomy in Europe and the embryonic coordination of the aims, procedures and deadlines of national funding agencies. However, a sound foundation has been laid for resolving those issues as well.

7 Maintaining Momentum

The teams behind the ASTRONET Science Vision and Infrastructure Roadmap can take justified pride in the substance and attractive appearance of their reports and the reaction from all sides. However, *action* on the recommendations is what counts in the end. To quote the succinct, proud – possibly arrogant – motto of Danish astronomer Tycho Brahe (1546–1601), "Non haberi sed esse".

Keenly aware of this, the ASTRONET Board is not leaning back and waiting for action to happen: The Board is initiating a continuing programme of Review Committees and Working Groups to develop recommendations and implementation plans for specific areas highlighted in the Roadmap. It will then initiate action on the recommendations. Several such initiatives have already been launched, and others are being planned, as outlined below.

Within a European scope, the following initiatives are under way or included in the ASTRONET work plan for 2011–2014, in cooperation with OPTICON and/or RadioNet as appropriate:

- A prominent recommendation of the Roadmap was to optimise the scientific impact and cost-effectiveness of the European 2–4 m (optical) telescopes through improved coordination. A *European Telescope Strategy Review Committee*

(ETSRC) was appointed in September 2008 to review the options and propose an implementation plan. Its report, submitted in May 2010, recommends to equip and operate all the European 2–4 m telescopes as a single system with globally optimised instrumentation and a single time allocation procedure.

In response, the owners of these telescope decided in September 2010 to approve the principle of allocating all (=trans-national *and* currently national) observing time on all the telescopes through a single proposal mechanism by $\sim$2015. Coordinating their instrumentation will be done within a similar time frame. It was recognised that many formal and practical obstacles must be surmounted to implement this historic decision, and a specific plan for this process was commissioned from OPTICON before the end of 2010.

- The need for a variety of *wide-field spectroscopic surveys* was highlighted in both the Science Vision and the Roadmap; key science cases are cosmological surveys and ground-based support for Gaia science. A Working Group to review the possible options and recommend priorities was appointed in October 2009 and is expected to submit its report in the first half of 2011. Given that Europe possesses no 8m telescopes with fields of the order of 1°, this WG needs to liaise closely with the ETSRC (see above).
- A *European Radio Telescope Strategy Review Committee (ERTSRC)* was appointed in September 2010, with a remit similar to that of the ETSRC for the optical telescopes. It is expected to deliver its report at the end of 2011.
- *Laboratory Astrophysics* is an important cross-cutting topic. Applications range from basic nuclear, atomic and molecular physics data for astrophysical spectroscopy at all wavelengths to curation of samples of matter returned by interplanetary space missions. *A European Task Force for Laboratory Astrophysics (ETFLA)* was established in September 2010 to make strategic recommendations for the development of this field in Europe. It will take note of the international situation, notably that in the USA where similar recommendations were made by the recent Decadal Survey. Its report is expected during 2012.
- *Astrophysical Software* and associated issues of computing paradigms, networking and archiving are assuming ever greater importance in astronomy, as emphasised in the Roadmap. Accordingly, in July 2010 the ASTRONET Board appointed an *Astrophysical Software Laboratory Committee (ASLC)* to draft a development plan for the Astrophysical Software Laboratory that was recommended in the Roadmap.
- Following up the implementation of the Roadmap will be an ongoing activity over the next 4 years (and no doubt beyond). Some large projects have an established host organisation, such as ESO for the European Extremely Large Telescope (E-ELT). ASTRONET then has no role beyond placing the project in the overall context as summarised in the Roadmap. Other, essentially global projects, such as the Square Kilometre Array (SKA) radio telescope or the Cerenkov Telescope Array (CTA) for cosmic rays, have no single European host organisation, and ASTRONET may be helpful here.

Other actions will be needed to turn the recommendations of the reviews listed above into reality. Close contacts will also be maintained to ESFRI within the general context of European investment in research infrastructures.

- A mid-term review and update of the implementation plan for the Roadmap is foreseen for ∼2015. It will be prepared during the FP7 contract, based on a review of progress made by then. Because technological and financial developments are a function of other forces than scientific progress and occur on different timescales, a review or update of the Science Vision may or may not be needed at the same time. This freedom of choice is an added benefit of the two-stage approach taken in the first ASTRONET planning cycle.

Within a global scope, ASTRONET will maintain appropriate contacts to potential partners in future global projects. This is particularly relevant for the USA, as the 2010 Decadal Survey report made a number of very similar recommendations to those above, including a standing implementation monitoring and advisory committee and a mid-term review. Internal procedures will be quite different on the two sides, given the great differences in their structure. It is recognised that ESO and ESA have their own established contacts to their US counterparts, but the participation of both in ASTRONET ensures that no confusion will arise. The OECD Global Science Forum will be another important partner in future discussions with a larger global forum. With luck and determination, all parties may perhaps meet for a round of global strategic planning around 2020.

8 Epilogue

ASTRONET has come a long way in its first 5 years. The partners realise that the road ahead is even longer. But it already seems safe to conclude that ASTRONET, in its first period, has firmly established the feasibility as well as the desirability of joint end-to-end strategic planning and coordination for a global science in Europe. This will benefit not only the scientific community and the funding agencies *per se*, but also make Europe a stronger and more credible partner in the global astronomy projects of tomorrow. Given the sound underlying logic of this approach, continued progress in this direction should be an irreversible process.

Acknowledgements The pioneering efforts of the first ASTRONET Coordinator and Project Manager, Drs. Anne-Marie Lagrange and Fabienne Casoli, defined and largely established the scope and structure of the project that have proved so successful. It is appropriate to also salute the dedication and hard work of the task leaders for the two cornerstone actions of ASTRONET, Tim de Zeeuw for the Science Vision and Mike Bode for the Infrastructure Roadmap, and their more than 100 Panel members who gave freely of their time and wisdom for the common cause. The effective and always pleasant cooperation of the present coordinator, Dr. Jean-Marie Hameury, the Project Office staff, and all the Board members is gratefully acknowledged, both by the project itself and by this author personally. I also thank Jean-Marie Hameury and Mike Bode for valuable comments on a draft of this article. ASTRONET is supported through ERA-NET contracts from the European Commission, no. 026075 under FP6 and no. 262162 under FP7.

Chapter 10
The Future of Space Astronomy

Fabienne Casoli

Abstract Access to space in the past 50 years has revolutionized astronomy. The wavelength range accessible to astronomers has expanded into the ultraviolet, X-ray and gamma-ray regions, as well as the infrared, millimeter and submillimeter domains. Robot probes have started the exploration of our Solar System and wandered around the inner and outer planets, asteroids and comets. A man-made probe, Voyager-1, has even reached the frontiers of the solar system. The way our Sun works, its gigantic eruptions, its wind that pervades interplanetary space, and its relationships with the solar system planets and with planet Earth have been revealed by the remote sensing and in-situ measurements. This harvest of outstanding results is not finished since more than 40 space missions are currently in operation, and a dozen launches are foreseen before 2015. Although space astronomy has to face several challenges: increasing complexity and cost, technical developments, global co-operations to put in place, the landscape of the 2015–2025 decade is almost defined, while astronomers are already sketching the missions of the mid 2020s.

Keywords Space vehicles • Space vehicles: instruments • Sociology of Astronomy

1 Why Should Astronomers Go to Space?

The first reason for which astronomy needs space is the Earth atmosphere. Observing from the bottom of a warm and turbulent atmosphere is not the best way to access the Cosmos. The Earth atmosphere is opaque to most of the electromagnetic spectrum; only few "windows" are accessible from ground-based telescopes, mainly

---

F. Casoli (✉)
Centre National d'Etudes Spatiales (CNES), Paris, France
e-mail: fabienne.casoli@cnes.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_10,

the visible and radio parts of the spectrum, up to the millimetre wavelength range, with some windows in the near infrared. Even at wavelengths where the atmosphere is transparent, its turbulence affects astronomical images and limits the resolution of images to a fraction of an arcsecond, ten times more than the theoretical diffraction limit of modern telescopes of 10 m class. This is one of the reasons for the impressive achievements of the Hubble Space Telescope (HST), which is a small telescope (2.4 m) by modern standards, but has the great advantage of being at 600 km above the Earth surface.[1] The sky background is also much darker when seen from outside the atmosphere, which makes the detection of faint objects easier.

But being in space offers many other advantages for astronomy. For example, space provides infrared astronomy with the cold environment it needs: the thermal emission of ground-based telescopes at room temperature makes them very bright in the infrared, while in space, telescopes can be passively cooled to temperatures of some tens of Kelvin.

Cloud coverage and the alternation of day and night prevents from achieving the long series of observations from the ground which are essential for solar and stellar physics, as well as for the search for exoplanets by transit photometry. Spacecrafts in Earth orbit can observe the same sky patch for continuous series: more than 150 days in the case of the CNES minisatellite CoRoT devoted to stellar physics and exoplanet hunting. In the field of solar physics, the ESA/NASA SOHO observatory which is positioned on a halo orbit around the first Lagrange point of the Earth-Sun system observes the Sun 24 hours a day.

Last but not least, the knowledge of the solar system has made giant leaps with robotic exploration of the planets, satellites and small bodies that began with the Moon in the 1960s. Even with the constraints inherent to space missions, limited power and weight in particular, in-situ studies are for planetary science an essential tool. The same statement can be made for plasmas of the solar system: Earth magnetosphere, the magnetospheres of the giant planets, and the heliosphere. Indeed, these two fields, plasma sciences and Solar system studies, have played a great role in the history of the development of space astronomy, so that for space sciences, astrophysics is generally considered as a separate field from plasma and planetary sciences. For example, the scientific structure of COSPAR (Committee on Space Research, created in 1958 by the International Council of Scientific Unions) comprises several Committees, among which one for Earth-Moon, Planets and small bodies, one for Space Plasmas, and another one for Research in Astrophysics from Space.

Finally, let us mention that the free fall environment of satellites in Earth orbit offers a fantastic laboratory to conduct fundamental physics experiments: very sensitive tests of general relativity such as tests of the equivalence principle or measurements of the gravitational redshift have to be conducted in space.

---

[1] Modern ground-based telescopes are equipped with adaptive optics systems, which helps to overcome atmospheric turbulence, but cannot fully compensate for it.

While space science was clearly not the main driver of the huge investments made to develop space activities, the role of scientific research (including Earth observation) in the convention of the European Space Agency (entered in force in 1980) clearly states ESA's role for scientific research: "The purpose of the Agency shall be to provide for and to promote, for exclusively peaceful purposes, cooperation among European States in space research and technology and their space applications, with a view to their being used for scientific purposes and for operational space applications systems". Indeed the science program is one the few mandatory programs of ESA, to which all member states must participate.

2 The Successes of Space Astronomy

2.1 Exploring the Solar System

A large fraction of our knowledge of the solar system bodies: planets, satellites, comets and asteroids, comes from space probes. Let us give a few examples.

Exploration has proceeded in successive steps: flying by, orbiting, landing, roving, and finally returning samples to Earth. Flybys give the first glimpse on solar system bodies. For example, in the 1980s, the "Grand Tour" of the giant planets and their satellites by the NASA probes Voyager 1 and 2 have revealed the surprising diversity of the giant planets and their satellites: icy moons that may hide oceans below an ice-shelf, the volcanic moon of Jupiter Io, tenuous rings, a diversity that is out of the reach of terrestrial telescopes. On the "small bodies" side, the observation of Halley's comet by the European probe Giotto in 1986 showed for the first time the shape of a comet nucleus and found the first evidence of organic material in a comet.

The next step was to orbit the planet or the satellite. This has been achieved for five solar system planets: Venus, Earth, Mars, Jupiter, Saturn, and for the Moon as early as in 1966. To this list Mercury will be added in 2011, when the NASA probe Messenger will be inserted in a Mercury orbit, and the asteroid Vesta that will be visited by the DAWN probe. Orbiters with their payload of instruments such as cameras, spectrophotometers, plasma sensors, are essential to build a global knowledge of the object; for example, a series of martian orbiters, among which the highly successful European orbiter Mars Express, have allowed to reconstruct the history of Mars, to map its mineralogy, and to establish that the Red Planet has been rather wet in its first billion years, with shallow salty seas and lakes. But it is also from the observations from the orbit that scientists have been able to witness the runaway green house effect that has made Venus inhabitable, or to study in depth the thick and turbulent atmospheres of the giant planets Jupiter and Saturn.

Landing on a planetary body was the next step. This has been attained for a handful of solar system bodies: Moon, Mars, Venus, Saturn's satellite Titan, and the Eros asteroid. The European mission Rosetta should achieve the first landing on a comet nucleus in 2014. In addition to the classical tools of astronomers: imagery and spectroscopy, this allows in-situ analysis of the soil, rocks, and of the atmosphere, in the immediate vicinity of the lander. Images of the soil of Titan, its rocks of water ice and frozen methane, and its hydrocarbon lakes are a great achievement of the European probe Huygens. As for Mars, one of the very exciting discoveries of NASA landers is the presence of water ice. Moving on the surface has been achieved for even less bodies: only Moon and Mars have been explored by rovers. However, explore is not exactly the right word since the longest travel from the landing site is 23 km in the case of Mars (NASA's rover Opportunity) and 37 km for the Moon (Lunokhod 2). Despite these limitations, in-situ analysis of the soil of Mars is a mandatory step to establish whether life has appeared and developed at some moment of the history of the Red Planet.

Given the limited resources that can be allocated on instruments onboard planetary landers and rovers, and the fact that they cannot be adjusted to the actual samples, in-depth analysis of the extraterrestrial samples can only be done by returning these samples to the Earth in order to study them with the sophisticated tools of geosciences and geobiology. This step of returning samples has been achieved for the Moon, with about 0,3 kg returned by the soviet robotic missions and of the order 382 kg returned by the astronauts of the Apollo program. The analysis of Moon samples has been essential in establishing the history of our satellite and dating its terrains, thus giving access to the whole solar system history. The analysis of samples of the coma of the comet Wild2, returned by the Stardust mission, have shown an unexpected mixing of material in the presolar nebula. Particles originating from the Sun have been returned by the NASA probe Genesis; some dust particles from the Itokawa asteroid have been returned by the Japanese mission Hayabusa. Except for the Moon, the mass of these samples is of the order of micrograms, but they are unique in being much less altered than meteorites, among which some are known to come from the Moon or Mars.

2.2 Sun and Heliospheric Physics

The knowledge of our star the Sun and its connection to planets, especially the Earth, has been gradually built up by a long series of space missions. Space is indeed a privileged vantage point to observe the Sun 24 h a day to monitor its energy output and its variations. As for planetary physics, the tools of the heliospheric physics include not only remote sensing but also in-situ measurements.

At the dawn of the space age, the earliest experiments discovered the strong links between the Sun and the Earth: Explorer 1 discovered radiation belts (charged particles from the Sun trapped by the geomagnetic field around Earth) in 1958, Mariner 2 (1962) showed that the Sun is at the origin of a flow of particles: the solar wind. From observations on Skylab (1973), scientists have discovered that the Sun is very bright in extreme ultraviolet and X-ray wavelengths, which means that the upper solar atmosphere of the Sun is quite hot, more than one million degrees.

It is also from space that coronal mass ejections, these huge ejections of matter originating form the Sun, have been discovered. This led to the understanding that stars interact with the universe not just through gravity and light but also through electromagnetic fields and particles. The study of how solar wind and solar transient events impact Earth is now known as Space Weather and has led to an entirely new science discipline called “Heliophysics”.

Some of the solar physics missions have been very long-lived and have observed the Sun for more than a solar cycle (about 11 years). This is the case of the joint ESA/NASA Solar and Heliospheric Observatory (SOHO) mission (1995), still in operation in 2010. SOHO is at the Lagrangian L1 point of the Sun-Earth system, 1.5 million kilometres from the Earth in the Sun direction. With SOHO’s instruments, scientists demonstrated that they could detect coronal mass ejections at the sun two to 3 days before they reach the Earth potentially causing damage on e.g. electric power distribution. Instruments on board SOHO were also able to detect acoustic pressure waves on the Sun opening the way to understanding how the interior of our star is organized, and thus how it works. This technique known as helioseismology has provided important clues on the origin of the solar magnetic cycle.

Our planet is immersed in a seemingly invisible yet exotic and inherently hostile environment. Above the protective cocoon of Earth’s atmosphere is a plasma soup composed of electrified and magnetized matter entwined with penetrating radiation and energetic particles. Inflated by the solar wind, a colossal magnetic bubble, the heliosphere, stretches far beyond the orbit of Pluto. This extended atmosphere of the Sun drives some of the greatest changes in our local space environment and affects the magnetosphere, ionosphere, and atmosphere of the Earth.

2.3 Astrophysics

Our understanding of the Universe would be very poor without space observatories. We would be unaware of the torrents of X-rays that escape some stars. Black holes would still be a theoretical concept and we would not know that there is probably a giant black hole at the heart of every large galaxy. We would be completely ignorant of the formation of stars inside their opaque clouds of gas and dust. Most of these phenomena need multi-wavelength observations to be understood properly, and astrophysicists crucially need access not only to visible light and radio waves, which often are only a tiny fraction of the luminous output of most stars and galaxies, but also to infrared and submillimeter waves, as well as X-ray and gamma rays. It is from the combination of all these wavelengths that scientists can get a complete picture of astronomical objects in order to understand their formation, their evolution, and their death.

The discovery of the violent Universe testified by X-ray observations goes back to the early ages of space science since it is in 1962 that a small rocket carrying an X-ray detector was sent above the atmosphere and detected intense emission from the constellation Scorpius. Many celestial bodies emit X-rays, even comets do. X-ray emission is mainly a tracer of very hot plasmas such as the hot gas found in the central regions of galaxy clusters, or in the accretion disks around compact objects such as neutron stars or black holes.

Gamma ray observations also began early in space astronomy, first with balloon-borne experiments and the Explorer 11 satellite in 1961. Celestial gamma-ray sources comprise supernova explosions, black holes, and even the decay of radioactive material such as cobalt, aluminium and iron isotopes. It is in gamma-rays that the most energetic events in the whole Universe, gamma-ray bursts, are detected. These events which have a very short duration by astronomical standards, less than a few seconds, are believed to be linked to highly focused electromagnetic emission during some supernova explosions. Gamma-ray bursts can be observed to very large distances and are thus used as probes of cosmological star formation history.

In the optical range, the atmosphere is transparent to visible light but its turbulence deteriorates the image quality and stability. Space observatories benefit from the full resolving power of the telescope and of the sky darkness. It is this combination of high spatial resolution and high sensitivity that made possible the detection of thousands of very distant galaxies in the deep fields observed by the Hubble Space Telescope, as well as the detailed observation of background galaxies lensed by huge lumps of dark matter in galaxy clusters.

In the infrared, astronomers witness the birth of stars and their planets. It is also the realm of cosmology, since the ultraviolet and visible light emitted by starbursting galaxies in the early epochs of the Universe is redshifted to infrared wavelengths.

Precision measurements of the cosmic microwave background in the millimetre and submillimeter domains have enabled astronomers to determine the age, size, and shape of the Universe. The submillimeter range is also a predilection domain of spectroscopists and astrochemists, who study the complex and surprising chemistry that takes place in interstellar clouds and may lead to the building bricks of prebiotic molecules.

Together with ground-based telescopes, space observatories have thus been fundamental to build the modern understanding of our Universe. It has to be noted that a very large fraction of the huge amount of data gathered by space missions, as well as ground-based observatories, is accessible to the whole scientific community after a short proprietary time during which only the astronomers who have proposed the observation or built the instrumentation can use them. The legacy of the great space observatories is thus shared by astronomers worldwide.

3 Space Astronomy 2011–2025

With more than 40 operating astronomy space missions at the end of 2010 (see Table 10.1), the first decade of the twenty-first century is a golden age for space astronomy. Since NASA foresees the launch of eight missions before 2015, ESA three, and including launches by China, France, India, Japan and Russia, the portfolio of operating astronomy missions could reach 50, taking into account that some of the current missions will be terminated.

Table 10.1 Operating astronomy space missions

	Astrophysics	Sun and heliophysics	Solar system, including Moon
NASA	Fermi (2008); Gamma-Ray Observatory GALEX (2003): GALaxy Evolution EXplorer Hubble Space Telescope (1990) Kepler (2009): exoplanets RXTE (1995): Rossi X-Ray Timing Explorer Spitzer (2003): Infrared observatory Swift (2004): Gamma-ray Burst Explorer WISE (2009): Widefield Infrared Survey Explorer	ACE, Advanced Composition Explorer (1997) Geotail (1992) with JAXA: magnetotail of the Earth IBEX (2008), Interstellar Boundary Explorer RHESSI (2002), Reuven Ramaty High Energy Solar Spectroscope Imager: solar flares Solar Dynamics Observatory (2010) Stereo (2006) : Solar Terrestrial Relations THEMIS (2007): magnetic storms Timed (2001): energy transfer in the upper Earth atmosphere Wind (1994): solar wind	Cassini (1997): Saturn system Dawn (2007): asteroïds Ceres and Vesta Deep Impact/EPOXI, Lunar Reconnaissance Orbiter (2009) Mars Exploration Rovers (Opportunity/Spirit) (2003) Mars Odyssey (2001): Mars orbiter Mars Reconnaissance Orbiter (2005) Messenger (2004): Mercury New Horizons (2006): Pluto and Charon Stardust Next (1999): comets
ESA	Herschel (2009): far-infrared and submillimeter space observatory INTEGRAL, gamma-ray observatory Planck (2009), cosmic microwave background XMM (1999): X-ray observatory	Cluster (2000): solar wind and the Earth SOHO, SOlar and Heliospheric Observatory (1995, with NASA)	Mars Express (2003): Mars orbiter Rosetta (2004): Comet 67P/Churyumov-Gerasimenko Venus Express (2005): Venus orbiter
Japan (JAXA)	Akari (2006): Infrared Imaging satellite Suzaku (2005): X-ray observatory	Hinode (2006): Sun magnetic field and outer atmosphere	Akatsuki (2010): Venus Climate Orbiter
France (CNES)	CoRoT (2006): stellar physics and exoplanets	Picard (2010): solar physics	
Italy (ASI)	AGILE (2007): Gamma-ray astrophysics		
China			Chang'e 2 (2010): Moon orbiter
Canada (CSA)	MOST (2003): stellar physics and exoplanets		

The landscape of missions that will be launched up to 2015 is rather well defined, and space agencies are in the final steps of selecting missions for the 2015–2020. This selection process is highly competitive and begins with a call for proposals to the scientific community, with often several tens of answers and ends with the implementation of one mission only (or sometimes with no mission at all). Let us describe this selection process with the example of the Cosmic Vision program of the European Space Agency. The planning of Cosmic Vision missions for the 2015–2025 decade started with a call for proposing science themes, issued in April 2004. 151 ideas were received and led to a "Science Vision document" organised along four major questions:

What are the conditions for planet formation and the emergence of life?
How does the Solar System work?
What are the fundamental physical laws of the Universe?
How did the Universe originate and what is it made of?

Next a call for proposals for Cosmic Vision missions aimed at answering some of these questions was issued in March 2007. The call's intention was to find candidates for one medium-sized mission (M class) for a launch in 2017 and one large mission (L class) for a launch in 2018. Fifty mission concept proposals were received, and from these, five M-class and three L-class missions were selected by the Science Program Committee (SPC) of ESA in October 2007 for assessment or feasibility studies. For the candidate M-class missions, these studies were completed by the end of 2009 and in February 2010 the SPC decided to advance three missions to the definition phase, with the aim of selecting two of these three in 2011 for launches in 2017 and 2018.

The reason for this change was that in the course of the assessment phases it quickly appeared that L-mission candidates were not mature enough, both from the technical and programmatic points of view (these three missions involve cooperation with NASA and/or JAXA), to be ready for a launch in 2018. The current plan is then to select in mid-2011 two L-mission candidates for a definition phase, aiming at the final selection of one mission in 2012 for a target launch date not earlier than 2020.

The three Medium size mission candidates are two astrophysics missions: EUCLID and Plato, and a solar physics mission, Solar Orbiter. EUCLID is a cosmology mission which aims at understanding the origin of the acceleration of the expansion of the Universe identified as "dark energy". Plato will search for telluric exoplanets in the habitable zone of their stars. Solar Orbiter will observe the Sun from an inclined and elliptical orbit that will allow to observe regions far from the equator, and will come closer to the Sun than any other probe before (0.28 astronomical units).

The three Large size mission candidates are IXO, Laplace/EJSM and LISA.

IXO (International X-ray Observatory), a joint ESA-NASA-JAXA project has three major science drivers: the study of black holes and matter under extreme conditions; the formation and evolution of galaxies, clusters and large scale structures in the Universe and the "life cycles" of matter and energy. The IXO optics will have 20 times more collecting area than any previous X-ray telescope, unprecedented polarimetric sensitivity, and microsecond spectroscopic capability.

Laplace/EJSM (Europa Jupiter System Mission) will be launched towards the Jupiter system and in particular its moons Europa and Ganymede. It will composed of two probes launched separately, JGO: Jupiter Ganymede Orbiter (ESA-led) and JEO: Jupiter Europa Orbiter (NASA-led). Key questions are: What have been the conditions for the formation of the Jupiter system? How does Jupiter work? Is Europa habitable?

LISA (Laser Interferometer Space Antenna), a joint ESA-NASA mission, will detect and observe gravitational waves that are emitted during the most powerful events in the universe. LISA will observe galaxies far back in time and test the fundamental theories of gravitation. LISA consists of three spacecrafts that act as an interferometer with an arm length of 5 million kilometres. There are numerous technology challenges associated with LISA, and ESA will launch in 2014 a dedicated mission, LISA Pathfinder, to test some of these.

This long and complex planning process comes from space agencies' vital needs for long-term plans: given that development times from the first proposal to a mission's results often span 15–20 years, it is necessary to start as soon as 2005 the planning of the 2015–2025 decade. For what concerns the missions of the 2020–2025 period, that will follow the two Medium missions and the Large mission, their selection process has also began: a call for proposals for a Medium-size mission (M3) has been issued in July 2010 with a target launch date in 2022. For this mission the competitive pressure will also be high since 47 proposals have been received from the European scientific community.

On the North-American side, the planning process follows a rather well-defined path with the Decadal Surveys under the auspices of the National Academy of Sciences, that aim at planning both ground-based and space activities, starting with a determination of the current state of knowledge and the identification of the most important scientific questions (see Chap. 8). For the 2012–2021 decade, three of these surveys are either completed or being completed, in Astrophysics (draft report issued in 2010), Planetary Sciences (due in 2011), and Solar and Space Physics (due 2012). A similar exercise has been finalized in Europe in 2007, encompassing the whole field of astronomy, on the ground and in space, with the Astronet Science Vision report followed in 2008 by the Infrastructure Roadmap (see Chap. 9).

3.1 Mars Exploration

Mars exploration will be a prominent feature of the space program in the 2010–2025 period. The key question is that of the past and present habitability of Mars: can we find signs of extinct or present life? Here the activities up to 2018 are rather well defined. In the fourth quarter of 2011,[2] NASA will launch the Mars Science

---

[2] Minimum energy launch windows to Mars occur approximately every 26 months.

Laboratory, a 900 kg rover aimed at investigating the past or present potential of Mars to support microbial life. MSL is a key feature of future Mars exploration since it will use a powerful new device, the "sky crane", to land the huge mass of the rover as close as possible of the area of interest. This period will also see the return of Russia to the scene of interplanetary missions with Phobos Grunt, which aims at returning to Earth samples of Phobos, one of the two satellites of Mars, after landing on this asteroid. Next is MAVEN (Mars Atmosphere and Volatile Evolution Mission), a NASA mission which will study the atmosphere of the Red Planet.

Next steps should be a cooperative endeavour between Europe and the United States. Indeed, the 2016 and 2018 missions are foreseen as a two-mission cooperation between NASA and ESA. This joint program is called ExoMars, and should provide Europe with the flight and in-situ enabling technologies that are necessary for future exploration missions, such as an international Mars Sample Return mission. The science aims are to search for signs of past and present life on Mars, to investigate how the water and geochemical environment varies, and to investigate Martian atmospheric trace gases and their sources.

The ExoMars program will start in 2016 with the launch by a NASA rocket of an orbiter provided by ESA, carrying a payload of five instruments (4 with US leadership and one with European leadership) to study in the Mars atmosphere trace gases of possible biological importance, such as methane and its degradation products. This orbiter will provide next missions with a telecommunications asset. The 2016 mission will also carry an Entry, Descent and Landing Demonstrator Module, which will provide Europe with the technology for landing on the surface of Mars with a controlled landing orientation and touchdown velocity.

The 2018 mission will use the MSL skycrane to land two rovers on the Mars surface, one European, one US. The ExoMars Rover will provide key mission capabilities: surface mobility, subsurface drilling and automatic sample collection, processing, and distribution to instruments. It will host the Pasteur payload, a suite of analytical instruments dedicated to exobiology and geochemistry research.

Missions after the ExoMars programme are not yet selected. The final goal is clearly to return samples of Mars rocks to Earth by the mid 2020s. This project known as Mars Sample Return (MSR), but it is a rather complex endeavour that probably requires three separate launches, and still poses some technological challenges. Given that it is also very expensive, budget profiles makes it unlikely that MSR can begin in 2020. Therefore it is likely that an additional Mars mission, not yet chosen, will take place in between the ExoMars program and Mars Sample Return. Among the projects studied for this 2020 mission is a network mission that would land several probes on the Mars surface, allowing exciting science such as Mars seismology or meteorological studies.

The Mars program is thus in a situation quite similar to that of the science program at ESA: rather well defined until 2020, even if some choices are still to be made; and with several perspectives opening in 2020–2025 period, prepared by active technology development programs.

4 The Future of Space Astronomy: Challenges

The wealth of present astronomy missions and of projects is such that it seems that the future of space astronomy can only be limited by the budgets that space agency can devote to it. However, space astronomy is facing a number of challenges specific to its nature. Some are technical and give rise to extensive technology developments. Some are programmatic and related to the complex and ambitious nature of future projects. Let us describe some of these challenges.

4.1 Planetary and Solar Physics Missions: Some Technical Challenges

For space exploration keywords are autonomy, miniaturisation, radiation hardened devices. For Solar physics and heliophysics missions additional difficulties are related to the harsh environment and the large temperature variations encountered by solar probes.

An important question is that of energy sources. The probes that explore the outer planets of the solar system or the rovers that try to survive the cold nights of planet Mars need compact energy sources. Solar powered batteries are often not sufficient and one has to implement nuclear power sources such as Radio isotope Heater Units (RHUs) or Radio isotope Thermoelectic generators (RTGs). RHUs produce several Watts of heat from the radioactive decay of Plutonium. RTGs produce a few hundreds of electric Watts, generally also from the 238 isotope of Plutonium. RHUs and RTGs are relatively simple devices and do not really represent technological challenges. However, access to the nuclear material with the necessary purity is not easy and for the time being only USA and Russia can provide deep space missions with such radioisotope powered energy sources. Implementing these devices on a spacecraft and launching them is also difficult since one has to follow complex and restrictive regulations.

As well as energy, mass is scarce on spacecrafts exploring the Solar System. The mass allowed to each science instrument is measured in hundreds of grams, kilograms in the best cases. On Mars Science Laboratory, the largest rover ever sent to Mars, the mass allocated to the science payload is less than 50 kg, while the rover mass is around 900 kg. This combination of low mass and low power consumption makes planetary science instruments very specific. In addition, as most space science missions, they must be able to operate in harsh environments: very cold (outer planets), or very hot (Mercury, Venus). In this respect, the exploration of Jupiter's moon Europa, which is supposed to host a water ocean below its ice shelf appears very difficult because of the high radiation environment, comparable to that inside Tchernobyl's plant after the catastrophe: electronic devices do not survive long in this environment and the nominal lifetime of the Europa orbiter which is part of the EJSM project, currently in assessment phase both at ESA and NASA, is only a few weeks. This probe is supposed to observe Europa from orbit; we see that we are far from being able to send a rover on Europa's surface, and even more to drill the surface of Europa in order to study the ocean that is likely below.

Missions to explore the inner regions of the solar system, and the Sun itself, face a constant problem of high temperatures and/or large temperature variations. Projects like Solar Probe plus, which will approach the Sun at a distance of 8.5 solar radii, needs a heat shield able to survive a temperature higher than 2,000°C. The instruments of Solar Orbiter, an ESA/NASA project currently proposed to ESA science program, will observe the Sun from a highly elliptical orbit that will bring it at distance about 62 solar radii from its surface; the instruments will experience more than a 100°C temperature change along the orbit.

Mars exploration gathers a large fraction of the difficulties of solar system missions. The emblematic mission of the mid-2020s, Mars Sample Return, aims at returning half a kilogram of martian rocks and soil to the Earth. This may seem a modest achievement compared to the hundred of kilograms returned by the Apollo missions, but the MSR program will need at least three different launches to be accomplished. Among the difficulties one can quote the high precision landing needed to land in the region of interest, the small rocket that will carry the samples from the Mars surface to the capsule that will return them to Earth, and finally the ground-based facility for the analysis of samples that may be hazardous.

Mars Sample Return is a program that will be composed of at least three missions. Some of these could be launched inside the same launch window; however, one probably needs to have the orbiter with the rendez-vous and capture system and the return capsule successfully put in orbit around Mars before launching the spacecraft carrying the lander, rover and rocket that will retrieve the samples and bring them to the return capsule.

4.2 Astrophysics Missions: Some Technical Challenges

For astrophysics missions, in all wavelength ranges, there has been at least one generation of space missions, and several in most cases. A good example is the infrared range: after the first pioneering observations by sounding rockets, the IRAS mission, a joint US-Netherlands-UK endeavour has been followed by the ESA project ISO, then by the NASA Explorer Spitzer, and now by the Herschel ESA mission. Each of these missions has yielded order of magnitude improvements over the previous one, either in telescope size (from 0.85 m for Spitzer to 3.5 m for Herschel) or in detector performance, or both. But making now a significant step forward, while being in line with ground based large infrastructures like the ALMA interferometer, implies order of magnitude improvements in angular resolution and sensitivity. This means either a very cold telescope, which is the purpose of the Japanese project SPICA (a 3.5 m telescope cooled to 5 K), or a much larger telescope, 6–10 m in diameter, or even a space interferometer. This was the purpose of the project named FIRI (Far InfraRed Interferometer) that ESA has studied in 2006. It has been concluded that FIRI posed several difficult technology challenges, specifically in the field of cooling, optics and detector technology; moreover, the mass of the spacecraft was probably larger than what could be accommodated by present-day launchers.

More generally, technical challenges specific to astrophysics missions are related to telescope sizes, cooling of the detectors or the telescope, and finally data analysis.

Increasing diameter is a general trend for both ground based and space telescopes, with the goal to increase the sensitivity to faint sources, as well as the angular resolution (for ground-based telescopes, this implies the use of adaptive optics). Specific problems encountered by space missions is that it is not possible to launch monolithic telescopes with diameters much larger than 3.5 m (the diameter of the Herschel telescope, the largest ever flown). One then has to fold the mirror before launch, as will be done for the 6.5 m dish of the James Webb Space Telescope, and then to unfold and align its 18 segments with exquisite accuracy. Building interferometers is one solution for ground based telescopes to improve angular resolution, however space interferometry has never been attempted (except in the radioastronomy range with a Japanese space VLBI experiment): it will probably need to master techniques such as formation flying, where the positions, distance and relative orientation of two (or more) spacecrafts are controlled to a very high degree of accuracy. First steps in this direction are being made by the Swedish project PRISMA, in cooperation with France, Germany and Denmark, which is currently in operation.

In the high energy range, increasing the collecting surface of the telescope and the angular resolution imply a large focal length of the telescopes, which is not easy to accommodate on a launcher. A good example is the ESA/NASA/JAXA project IXO (International X-ray Observatory), that will have a launch mass of around 6,600 kg and will be about 10 m long and 4 m in diameter in its launch configuration. A focal length of 20 m has been selected for IXO as a balance between science requirements and engineering constraints. As no current launch vehicle is capable of accommodating a payload that is 24 m long, IXO will have a deployable structure to position the instrument module at the mirror focus after launch. IXO mirrors are also a challenge: the mirrors for XMM-Newton, the ESA X-ray observatory that is currently in operations, were manufactured from gold-plated nickel. Extending this technique to a diameter of 3.8 m would result in a mirror assembly too heavy to be launched, and two new technologies are being investigated: mirror shell segments manufactured from slumped glass (NASA) and silicon pore optics (ESA).

All these examples point out that future astrophysics observatories need extensive technology developments. In addition, these missions are generally intended to be positioned not in Earth orbit, but in the second Lagrange point of the Sun-Earth system, where WMAP, Herschel and Planck are already positioned. This is the case for the JWST and it is clear that at 1.5 million of kilometres from the Earth, repairing faulty instruments is out of question. These requirements do not mean that these missions are not feasible, but imply that they will be expensive, will need long developments times and are doable only in international collaboration. This leads us to the second part of challenges that future astronomy missions will encounter, maybe the most difficult ones: programmatic challenges.

5 Future Astronomy Missions: Programmatic Challenges

Many of the future astronomy space missions will be very ambitious missions involving a large international cooperation with very long timescales. For example, the Mars Sample Return, which has been the goal of the solar system scientist for many years, will be a global endeavour involving at least Europe and the United States. All of the Large Missions candidates of ESA's Cosmic Vision are foreseen to be built in cooperation with NASA and in one case JAXA. Given that on both sides of the Atlantic, space science programs proceed through separate prospective surveys and successive down-selections, it is not obvious that a given project will be the first priority of both communities and proceed at the same pace while surviving all financial ups and downs that are common in space missions. This does not mean that these global cooperations cannot succeed: many successful examples can be given such as the Cassini/Huygens mission to Saturn and Titan, but it is clear that they need long times to be realized.

Maintaining momentum over the long timescales of the largest space missions is in itself a challenge. For example, the European mission to a comet, Rosetta, was decided in 1993, launched in 2004, and will arrive on its target in 2014: this is typical for missions to the outer solar system. If a mission to the Jupiter system such as EJSM/Laplace is selected in the 2010s, it may be launched in 2020–2022 and will arrive 5 years later in the Jupiter system. For solar physics missions, the development time to the cruise duration of a project like Solar Orbiter is more than 10 years, while this mission was first proposed in the 1990s. Astrophysics missions are not much easier: the medium-size cosmology mission Planck was first proposed in 1992 as a CNES "small mission", launched by ESA in 2009; first cosmology data will be released in 2012. The duty cycle of a space mission, from the proposal to data release, is thus of the order of 15–20 years. Space astronomy must therefore maintain the teams and technical skills over these durations. This is also a problem for scientists who are involved in mission development since this activity are not very effective in terms of publications.

A more fundamental question is that space science programs must be able to react to new discoveries, such as exoplanets or dark energy in the 1990s. Space agencies must then keep opportunities in their programs and try to achieve a balance between small and medium-size missions, which should be easier and quicker to implement, and large missions.

However, the possibilities offered by space observatories make them unique even when unexpected discoveries occur. A good example is the use of the Spitzer space observatory to detect the light of transiting exoplanets. Spitzer, launched in 2003, has been the first telescope to detect light from a planet outside our solar system. Before that, astronomers have always used indirect methods to detect exoplanets by the effects they induce on their parent star's light, such as Doppler velocimetry or transit photometry. With Spitzer, the difference between the star spectrum when the planet is behind the star with its spectrum when the planet is in front the star gives direct access to the planet spectrum in the infrared. This allows very interesting investigations of the planet's atmosphere such as the detection of water, methane, CO or CO2. What is even more remarkable is that these observations have been proved possible even with a "warm" Spitzer, i.e. after the mission has exhausted all its liquid helium and only the short wavelength instrument is functioning.

The idea of a space-based infrared observatory had been discussed in the US since the 1980s under the name of SIRTF; the overall characteristics of Spitzer, its mirror size in particular, were frozen in the years 1993–1994, before the (ground-based) discovery of the first exoplanet around a main-sequence star, 51 Peg. It is thus clear that observing exoplanets was not part of the initial science case of Spitzer…

Another balance that has to be achieved is that between the budget devoted to missions in operation, which are more and more numerous given the high reliability of space missions and the development of new missions. For example, it has been decided that the budget of mission operations should not exceed 20% of the total budget of the science program of ESA in order to leave room for the development of the new Cosmic Vision missions.

Finally, the huge costs of some of the most ambitious astronomy missions are indeed a difficulty. Total lifecycle costs of several G$ or G€ are difficult to accommodate in the flat budget profiles space agencies devote to astronomy. The Hubble Space Telescope has probably been the most expensive astronomy mission up to now (with a cost to launch of about 5 G$), but projects like the James Webb Space Telescope or Mars Sample Return are likely to exceed this. In these conditions, cost overruns, although not more frequent in space astronomy than in other complex projects, can put a whole program in danger. Together with technology and programmatic, and closely linked to them, this management issue is then the third challenge that space astronomy must face.

6 Conclusion: A Bright Future for Space Astronomy?

Astronomy is a science based on observations, and space observations have proven to be vital for its advancement. Future astronomy space missions are challenging but also exciting; preparing them implies careful planning of technical developments that take the best out of the inventiveness of the scientific laboratories and boost the capacities of the industry. The global cooperation endeavours that will be needed to develop them are indeed complex but they also make these enterprises attractive. One must be confident that space astronomy will continue developing and producing fascinating results about the Universe.

Part III

The Tools of Observation and the Profession of Astronomer

Chapter 11
Small Telescopes and Planets

Andrzej Udalski

Abstract In spite of their small size, 1 m class telescopes still play an important role at the frontier of discoveries in many astrophysical fields. The best examples of their achievements are microlensing surveys – large scale, long term sky surveys that, since the beginning of the 1990s, have provided unique photometric data that has been widely used by the astronomical community for hundreds of scientific projects. In particular, small telescopes and ultra small cameras have, in the last two decades, provided crucial data for development of the two key fields of modern astrophysics – gamma ray burst (GRB) astrophysics and searches for extrasolar planets. Here, we describe the most important applications and discoveries made by small telescopes during the last two decades, focusing mostly on the searches for exoplanets conducted with this class of instruments.

Keywords Astronomical instrumentation • Methods and techniques • Telescopes • Planetary systems

1 Introduction

In the last decade of the twentieth century the fate of small astronomical telescopes seemed inevitable. One after another, the operation of telescopes with primary mirror diameter smaller than 2 m was discontinued at all large observatories worldwide. The European Southern Observatory, La Silla, CTIO at Cerro Tololo or Kitt Peak Observatory in Arizona, USA are the best examples of these actions. The reason was simple – the era of a new generation of large telescopes had just begun and the huge costs of operation of these giants required considerable savings from

---

A. Udalski (✉)
Warsaw University Observatory, Al. Ujazdowskie, 400-478 Warszawa, Poland
e-mail: udalski@astrouw.edu.pl

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_11,

existing practices. Small telescopes became natural victims of the new strategy in observational ground-based astrophysics. Their old-style operations were no longer the most effective and, though previously considered irreplaceable as scientific instruments, they could no longer compete with the new telescopes in making top rank discoveries. Only a revolution could change their status. And the revolution came.

In 1986 Bohdan Paczyński published his seminal paper [15] on the basics of gravitational microlensing and proposed to use these phenomena to search for dark matter. Later, in the beginning of the1990s, the same phenomena were proposed for studies of the Galactic structure [16] and the search for extrasolar planets [7, 13].

Microlensing phenomena are very rare. Even in the most microlensing efficient regions of the sky, only one star per million can be significantly magnified at the time. Thus only a huge, long-term sky survey, regularly monitoring millions of stars, could ensure the detection and characterization of the larger sample of microlensing events that will be necessary for studies of the astrophysical problems mentioned above. At the time when Paczyński's paper was published, his idea of microlensing surveys sounded like a sort of science fiction.

However, in the early 1990s a few groups of astronomers started thinking about how to bring this science fiction to reality. Larger format CCD detectors, more powerful computers/workstations, and modern image processing techniques made the idea of regular microlensing monitoring of millions of stars more feasible. Almost at the same time in 1991/2, three groups – French EROS, Australian-American MACHO and Polish-American OGLE – began the first generation microlensing surveys. Only 1 year after – in 1993 – the first microlensing phenomena in the line-of-sight toward the Large Magellanic Cloud [2] and the Galactic center [24] were discovered, thereby starting a new field of modern astrophysics. Paczyński's dreams had come true.

Microlensing surveys re-vitalized small telescopes. Sky targets appropriate for microlensing surveys require a large number of stars so that light can be magnified by lensing objects passing in front of them. The most promising places in the sky are the dense stellar regions of the Galactic center and two nearby galaxies: the Large and Small Magellanic Clouds.

All these targets are within the reach of small telescopes. With 1–1.5 m class telescopes one can monitor photometrically, with good photometric accuracy, millions of stars. However, conducting a microlensing survey required also a non-typical observing strategy contrary to the one used with telescopes of this class at that time. Instead of granting the telescope time for many different projects in usually short – a few days long – time slots, a telescope for a microlensing survey had to be dedicated to the project. Microlensing searches require long term (months/years) regular monitoring of the densest stellar regions of the sky.

Thus, the idea of microlensing surveys forced a new approach in observation strategy. This new strategy not only revolutionized the ways of conducting observations but also opened new completely unexplored niches of astrophysics that could be studied with relatively small telescopes. From the microlensing point of view, the most interesting regions of the sky – the Galactic center and disk and the Magellanic Clouds – are the main "laboratories" of modern astrophysics. They

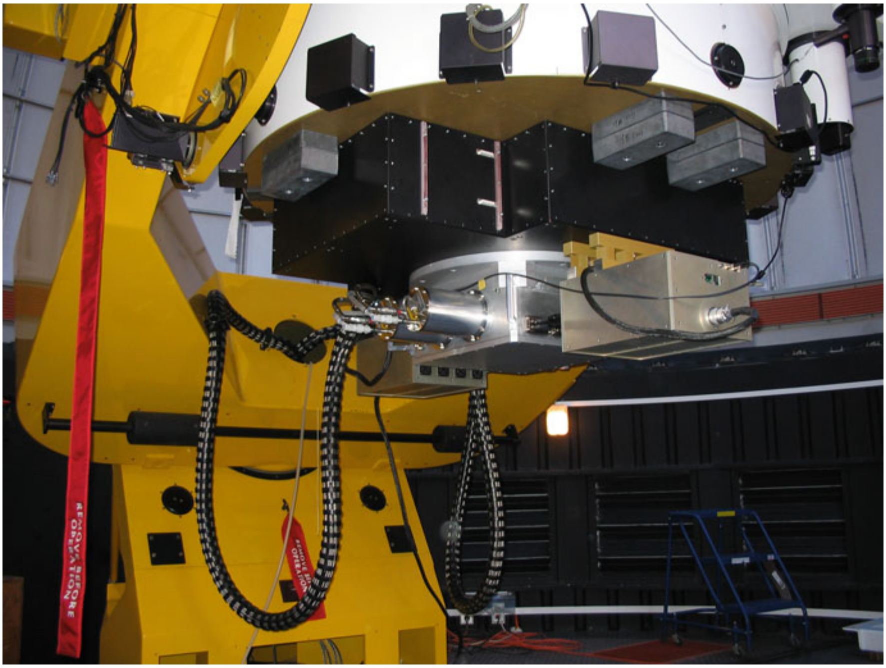


Fig. 11.1 1.3 m Warsaw telescope at Las Campanas Observatory, Chile with OGLE-IV 256 Megapixel mosaic CCD camera

were observationally neglected before 1990s, in particular with regard to modern observing techniques. Thus, the data collected in the course of microlensing searches provided a unique opportunity for studying not only microlensing events but also large samples of many classes of objects of main astrophysical interest. Hundreds of new discoveries – often changing the existing knowledge in the field – have been made during the last couple of years thanks to microlensing survey databases.

The evolution of microlensing surveys during the last 20 years can best be seen by taking the OGLE project as an example. The OGLE project became a world leader among microlensing surveys. With time it changed its profile from a focused microlensing program to a multi-field sky survey, providing significant contributions to many astrophysical fields. The project evolved with time, gradually increasing its observing capabilities. Each of the instrumental upgrades started a new phase of the project. The OGLE observing facilities are located at Las Campanas Observatory, Chile (owned by the Carnegie Institution of Washington) which is one of the best astronomical observing sites worldwide. During the phase called OGLE-I (1992–1995), a 1-m Swope telescope was used [23]. Since 1996 observations have been carried out with the 1.3 m Warsaw telescope (Fig. 11.1), dedicated exclusively for the OGLE project. The second OGLE-II phase [25] lasted up to the end of 2000 when its first generation CCD camera was replaced with an eight chip, 8192×8192 pixel wide field mosaic camera starting the third phase OGLE-III (2001–2009; [22]).

Although the OGLE-III phase ($>$200 million stars regularly observed) would have continued to be competitive even now, the new scientific challenges resulting from, among others, the OGLE-III discoveries, made it necessary to upgrade the observing capabilities by another order of magnitude and to progress to the OGLE-IV phase.

To achieve this goal a new 32-chip mosaic 256 Mega-pixel camera was designed for the OGLE project and installed on the 1.3 m Warsaw telescope. It is worth noting that the OGLE-IV new mosaic belongs to the largest instruments of this class in the world. The new mosaic camera fills the entire wide field of view of the 1.3 m Warsaw telescope (1.4 square degrees in the sky) covering an area four times larger than the OGLE-III instrument. With much shorter reading time this leads to almost an order of magnitude larger data flow compared to OGLE-III.

The OGLE-IV instrumental setup enabled observations of much larger areas of the sky and observations of selected fields with much better time resolution. The estimated number of regularly observed objects in OGLE-IV is of the order of one billion. The total size of collected raw images is larger than 30 Terabytes per year. With these capabilities the OGLE-IV survey will be among the largest optical surveys worldwide for the next couple of years. OGLE-IV began regular monitoring of the sky on the night of March 4/5, 2010.

2 Science from the Microlensing Surveys

Gigantic databases of photometric measurements of millions of stars collected during microlensing surveys with small telescopes became "gold mines" for modern astrophysics. Usually made public to the wide astronomical community, they allowed studies in many astrophysical fields. The term "data mining" became commonly used by astronomers worldwide.

Examples of the variety of top rank astrophysical science resulting from microlensing surveys can be seen again in the OGLE project example. During all of its phases the OGLE project contributed significantly in many fields of observational astrophysics:

Gravitational Microlensing. The empirical side of this field largely depends on the OGLE project. From several thousands of microlensing events discovered so far, the vast majority was detected by OGLE. The OGLE project discovered the first microlens toward the Galactic bulge, the first binary microlens, derived the first estimate of the optical depth to microlensing toward the Galactic bulge and participated in virtually all important discoveries in this field, including widely publicized planetary microlensing. Also the original puzzle about the form of dark matter that was the driver of the first generation surveys seems to have been solved now: a very little fraction of the mysterious dark matter exists in a form able to produce microlensing.

Extrasolar Planets. The OGLE project pioneered two new photometric methods of detection of extrasolar planets: transits and microlensing. The first four transiting planets were selected by OGLE from its dedicated photometric campaigns and then confirmed spectroscopically by other teams. For a few years the OGLE planets were the only transiting planets known, except for HD 209458. The OGLE team discovered the vast majority of planetary microlenses and actively participated in their precise characterization.

Variable Stars. Microlensing surveys revolutionized this field by the detection of large samples of new variable stars. OGLE survey is a world leader in this field providing large and homogeneous samples of variable stars of all types. In particular, it should be stressed that the OGLE samples of Cepheids from the Magellanic Clouds and OGLE Period-Luminosity relations are the base of the modern extragalactic distance scale. They were used, for example, by the Hubble Key Project aiming at the determination of the Hubble constant. Hundreds of other analyses and discoveries are based on OGLE variable stars.

Photometry, Galaxy, Magellanic Clouds structure. OGLE Photometric Maps contain the calibrated mean photometry and astrometry of all stellar objects detected in the fields observed by OGLE. They are widely used by the astronomical community not only for scientific projects but also for calibrating photometry, as this is a huge set of secondary BVI standards.

Astrometry. The OGLE images can be also used for astrometric purposes. Proper motion catalogs of the stars in the OGLE fields were built and used for analyses of the dynamics of the Galaxy.

Miscellaneous. Many other projects based on OGLE data were completed during the last couple of years. Among others they include studies of stellar clusters (Omega Cen, 47 Tuc, many open clusters), dwarf galaxies (e.g., Sculptor, Sagittarius, IC1613), interstellar extinction towards the Galactic bulge, characterization of many optical counterparts of X-ray sources from the Magellanic Clouds or search for binarity of the planetary nebulae nuclei. OGLE long term monitoring of QSO2237 + 0305 (Einstein Cross) provides a unique, ∼12 years long light curve of all four components of this gravitational lens [28]. The OGLE project has even successfully started supernova monitoring.

3 Even Smaller Telescopes

Microlensing surveys put new life into the small telescopes of the 1 m class mirror diameter. The results of these surveys have become important sources of targets for a variety of follow-up programs, very often feeding the largest class telescopes. However, the last 15 years were even more profitable for small telescopes.

Fig. 11.2 ASAS project cameras at Las Campanas Observatory, Chile

Paczyński [17] advocated that the variability of the brightest objects in the sky, in spite of the fact that they had been observed for a thousand years, is actually poorly known. Encouraged by the success of microlensing surveys, he proposed to implement shallow wide-angle surveys of the whole "bright sky". With the advent of CCD techniques and cameras, shallow surveys of the whole sky could be carried out with small instruments, namely photographic objectives of 10–20 cm diameter and commercially available CCD cameras working in automatic mode.

Paczyński's idea was implemented for the first time in 1997 when the prototype instrument of the All Sky Automatic Survey (ASAS) started regular monitoring of large parts of the southern and equatorial sky at Las Campanas Observatory, Chile [18]. It soon turned out that indeed the number of known variables in the General Catalog of Variable Stars is only a small fraction of all varying bright objects. In the following years the ASAS project increased significantly its observing capabilities by installation of additional units (Fig. 11.2) and upgrading hardware and software, as well as installing two new instruments at the Mount Haleakala Observatory on Maui Island, Hawaii, USA, both of which regularly monitor the northern sky. The ASAS catalog of variable bright objects containing about 50 thousands stars has become an important source of data for many astrophysical projects and statistical studies of variable stars of all kinds.

Two unexpected findings in 1999 converted the ultra-small telescopes into mature members of the observational astrophysics instruments capable of making frontier discoveries. At the end of the 1990s, several groups of astronomers were involved in the hunt for optical afterglows of gigantic cosmic explosions called gamma ray bursts (GRBs). These rare phenomena were recorded for the first time in the 1960s from space. In the 1990s, gamma ray satellite missions became capable of catching such outbursts in almost real time. Typically, the gamma ray astronomers, upon observing an ongoing outburst, alerted their collaborators in ground-based observatories who pointed the telescopes into the gamma burst direction looking for an optical counterpart. This strategy finally triumphed when the first optical afterglow of GRB 970228 was detected with the 4.2 m WHT telescope at Canary Islands [29]. Within months, the next optical counterparts were found with even smaller 1 m class telescopes.

The main trouble in the identification of optical counterparts at that time was the poor accuracy of positions of the outburst in the sky. This limited severely the searches for optical afterglows with large, usually narrow-field telescopes. Also the coordinates were usually obtained after a shorter or longer while after a short-lived gamma ray burst, delaying the optical follow-up. Therefore another strategy for the hunt was developed – the search for GRB afterglows with very small telescopes, actually photo-lenses with CCD cameras – working in an automatic way, fast and automatically responding to satellite information on the on-going burst. Such instruments could have wide field of view, although of course at the cost of the magnitude range. To facilitate such strategy a network "Gamma-Ray Burst Coordinates Network (GCN)" was created that distributed appropriate information on the detection of GRBs over the Internet.

The ROTSE project was the first to implement this mode of operation. However, its success was not certain at all as the astronomers expected that optical afterglows should be faint, rather below the 16th magnitude limit of the first generation ROTSE instruments.

On January 23, 1999, a long gamma ray burst, lasting about 1.5 min, was registered by the Compton satellite. ROTSE instruments responded automatically on the alert and collected a series of images of the sky in the direction of GRB 990123. It turned out later that it not only detected the optical counterpart of GRB 990123, but also imaged in optical range the entire burst with 25 s cadence until it faded below the ROTSE instrument range. This was the first observation of the optical flash of a GRB [1]. To everybody's great surprise it turned out that the maximum brightness of the optical afterglow reached ninth magnitude, indicating that the optical outbursts can be much brighter than originally thought.

Encouraged by this great success, similar projects with small automatic instruments started regular monitoring of the sky during the last decade looking for bright afterglows of GRBs. A few of such flushes up to 13th magnitude were registered. However, so bright afterglows as that of GRB 990123 are extremely rare. The next spectacular case was caught on March 19, 2008. The optical afterglow of GRB 080319B reached less than 5.5 magnitude becoming for a short while a "naked eye" object. Due to very fortunate circumstances (almost simultaneous eruption of another GRB in the same direction of the sky) several small optical telescopes covered the event with high cadence simultaneously with the gamma ray SWIFT satellite instruments providing the first extremely precise optical light curve of the GRB outburst [20].

The GRB field, one of the most active fields of modern astrophysics, profits a lot from the science done with small and ultra-small telescopes. The progress in this field during the past several years was enormous thanks to these hard-working instruments.

The second very important and breathtaking discovery with ultra-small telescopes was also made in 1999, again in the one of the most important fields of modern astrophysics – the extrasolar planet field. After decades of fruitless searches for extrasolar planets, the second half of the 1990s finally witnessed the first discoveries of exoplanets orbiting solar type stars [14]. The spectroscopic Doppler shift technique reached at that time the accuracy of radial velocity measurements of several meters per second. This was good enough to detect variations of radial velocity of host stars with an amplitude of a few hundreds m/s caused by orbiting planets of Jupiter size mass. Therefore, several new detections soon followed the first discovery.

The spectroscopic method allows determination only of the lower limit of the planetary mass (m sin i), because inclination, i, of the orbit is unknown. On the other hand if the inclination is high enough, a planet can transit regularly in front of the host star causing the drop of its brightness at the 1–3% level in the case of a Jupiter size planet. Such a micro-eclipse could be in principle detected from the ground with precise photometric measurements. Not surprisingly then – each freshly detected planet-hosting star was almost always monitored photometrically after the discovery for small depth transits. The trouble was that the spectroscopic discoveries were originally made for the brightest stars of 6–8 magnitude. Such objects are usually too bright for typical small telescopes of 1 m class equipped with modern CCD detectors, opening an observing niche for ultra-small telescopes.

The first positive detection of an exoplanet transit was made in November 1999 by two instruments: a small instrument equipped with just a 10 cm diameter photolens and a CCD camera [5] and a 0.8 m telescope with a photomultiplier [9]. Star HD 209458 was known to host an exoplanet discovered spectroscopically. Photometric monitoring discovered shallow 2% deep transits in the light curve of this star at the correct moment of the spectroscopic orbit. This discovery provided additional strong confirming evidence that indeed an exoplanet orbits HD 209458. As the inclination of this orbit could now be derived from the photometric light curve ($\sin i \sim 1$), the mass of the planet could be therefore unambiguously calculated. Moreover, the size of the planet could also be obtained from the photometric transit shape. Thanks to observations with ultra small and small telescopes, HD 209458 became the first planet beyond the Solar System with all most important parameters, such as radius, mass, density etc., known. Again the smallest instruments provided a huge contribution to the progress of modern astrophysics.

4 Small Telescopes and Extrasolar Planets

4.1 Planetary Transits

Encouraged by the successful detection of the first exoplanet transit in HD 209458, astronomers began intensive preparations for the reverse approach – the search for extrasolar planets with the photometric transit method. The main idea of such a search was simple – one should monitor a large number of stars with high photometric precision looking for shallow, a few percent deep periodic dimmings of brightness of expected shape – so-called transits. If several such transits were photometrically covered, one could derive the photometric orbit (time of transit, period) and estimate the size of the transiting body. If small enough – one could suspect that the effect is caused by an exoplanet orbiting the observed star. Thus, it seemed that the method should provide fast and numerous detections of the new worlds.

The advantages of the discovery of a transiting exoplanet are obvious. In these cases the orbit must always be highly inclined – close to $i = 90^\circ$ – so the mass of the planet from additional spectroscopic observations can be unambiguously determined. This makes transiting planets the only exoplanets where all the most important parameters: radius, mass, density, can be directly derived from observations, leading to full characterization of the planet. These objects are the most valuable exoplanets for studying their structure. Not surprising then that scientists eagerly waited for a larger sample of transiting planets.

Unfortunately, the practice turned out to be more complicated. The size of a transiting body is not an unambiguous clue that allows one to distinguish exoplanets from other possible bodies orbiting the host star. Two other objects – brown dwarfs and very small M-type stars of less than 10% solar mass can have dimensions comparable with giant, Jupiter-sized planets. Therefore they both can produce transits practically indistinguishable from transits of exoplanets. So photometry alone cannot firmly claim the discovery of an exoplanet when transits in the light curve are detected. Additional follow-up observations are needed for confirmation of the planetary origin of the candidate.

These follow-up observations are spectroscopic radial velocity measurements, identical to those in the case of the classical search for exoplanets with the spectroscopic method. They allow determination of the candidate mass which finally makes it possible to classify the transiting object as exoplanet, brown dwarf or small-mass star. Spectroscopy also provides an additional cross check as the radial velocity variations must occur in appropriate phase with a photometric orbit. In principle the spectroscopic follow-up requires much less observations than regular spectroscopic discovery of an exoplanet, actually just two measurements at quadratures (photometric phase 0.25 and 0.75) should be sufficient to estimate the mass of the companion. However, in practice this is never the case – ruling out non-planetary scenarios for the observed system as well as the determination of additional information about the system usually requires much denser spectroscopic orbit coverage.

Another obstacle that must be overcome when a transiting object is detected is a possibility of the so-called blending with neighboring star(s) unresolvable in the sky. Such an object can be either a by-chance star lying in the same direction in the sky or a physically bound object in the observed system. If this extra star is bright enough it can suppress the depth of the regular deep eclipse of, for example, a typical eclipsing binary system, so it may look like a planetary transit. Also the radial velocity variations can be affected and suggest at first glance the companion of planetary mass. Thus, a careful analysis of the light curve, profiles of lines in the spectra etc. is always necessary in each case to rule out the possible blending scenarios.

A photometric search for an extrasolar planet with the transit method requires a large number of monitored stars for transit detection. This can be achieved by two strategies. First, following the success of Charbonneau et al. [5] in the case of HD 209458, one can monitor large areas in the sky with ultra small, wide angle instruments, concentrating on bright (typically brighter than 12 mag.) stars. On the other hand, one can use larger 1 m class telescopes and monitor much smaller areas in the sky but where the stellar density is much larger, providing a huge number of potential targets, for example the fields in the Galactic disk. Good quality precise photometry for stars of even 16–17 mag. can be obtained with 1 m class telescopes and modern CCD cameras. The only limitation in this strategy is presently the range and precision of spectroscopic follow-up observations necessary for confirmation of the planetary status. The largest optical telescopes with the most efficient spectrographs may reach presently 16th magnitude objects with required accuracy of radial velocity measurements, making deeper photometric search scientifically unjustified.

In spite of great expectations from wide field surveys which were formed almost immediately after the HD 209458 transit detection, the first successes in the search for transiting exoplanets came from the second approach. It was the OGLE project which pioneered this field. The first transiting planets discovered with a transit technique approach came from the OGLE-III survey for low-luminosity and planetary objects [26, 27]. Dense stellar fields in the Galactic center and disk were observed during several-weeks long campaigns and the lists of potential transiting exoplanet candidates were made public in the Internet for spectroscopic confirmation. The OGLE project also pioneered the techniques widely used now in transit photometric data processing, like application of de-trending algorithms to the data to achieve the best possible accuracy or commonly used now the BLS transit finding algorithm of [11]. Soon after the photometric discoveries, the first OGLE candidates were confirmed spectroscopically [4, 10, 19], forming for a long time the largest group of transiting exoplanets (Fig. 11.3).

The wide-field surveys conducted with ultra small instruments and concentrated on bright stars for a long time were doing much worse. The main reason was certainly the much worse accuracy of photometry obtained with such small objectives and wide field, only in theory reaching one percent. Thus only single successful detections of transiting exoplanets were announced up to the end of 2006 by the wide field projects like TRES, HAT or XO. However, when the

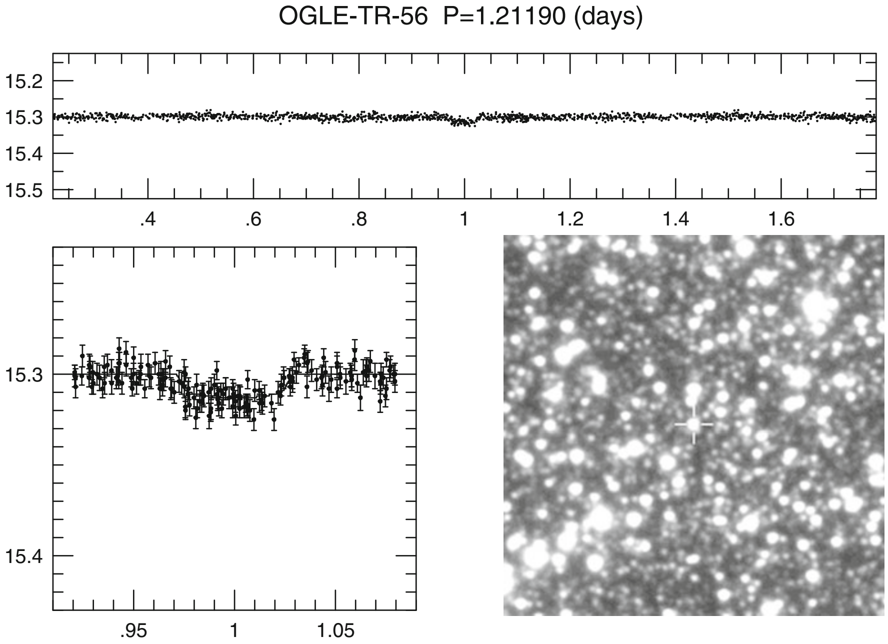


Fig. 11.3 OGLE light curve of the transiting planet OGLE-TR-56B

photometric techniques were improved and new generations of wide field surveys like Super-WASP, HATNet became operational, the number of positive ground-based detections started increasing rapidly, reaching presently about 70 objects. (see http://exoplanet.eu/catalog-transit.php for up-to-date statistics).

An additional sample of transiting exoplanets consists of systems detected spectroscopically and then monitored photometrically, usually with 1 m class telescopes, for transits, as was done in 1999 for HD 209458. This group counts now several systems including one of the smallest transiting planets – hot Neptune GJ 436b [6].

While the first OGLE transiting planets certainly made a breakthrough for the field, its rapid development in the last few years has been mostly due to the detections of bright transiting exoplanets from small instrument, wide field surveys (HATNet, Super-WASP). Bright objects can be observed more precisely both photometrically and spectroscopically. Very often they do not need the precious observing time of the largest optical telescopes for spectroscopic confirmations. Also they can be much better targets for additional follow up observations both from the ground and space. For example, such systems were very often observed by the Spitzer infrared satellite for secondary eclipses to allow for the measurements of the planetary emission in the infrared. Other follow up observations include studies of planetary atmospheres from spectra taken during transits or determination of the spin-orbit inclination of the planet from spectroscopic radial velocity measurements. The transiting planets can also be used for the detection of additional planets in the system. This can be achieved by measuring the precise timing of transits – the deviations can indicate additional exoplanets in the system.

The possibilities of detection of smaller size planets from the ground are limited by the accuracy of photometry that can be achieved from the ground. It is not possible to detect transits shallower than a few millimagnitudes. Therefore high expectations in the planetary transit field are now connected with the space missions – COROT and KEPLER. They should detect much smaller transiting planets, as compared to ground-based detections, due to much more precise photometry from outside the atmosphere. Moreover, KEPLER, for example, will observe selected fields in the Milky Way for the whole mission life of about 4 years. Thus, it will be able to detect transiting planets with longer orbital periods. The KEPLER mission's main goal is to discover the first Earth size planets in the habitable zone and to provide the census of the exoplanets down to Earth size and the distance smaller than about 1 AU from the host stars. The main problem here may be, however, related with reliable spectroscopic confirmation of such small planets which is at the limit of the present spectroscopy. Several positive detections of transiting exoplanets have already been announced from both space missions but so far, except for single cases, only of objects that could have been detected from the ground as well. This is, however, just the beginning of the KEPLER mission.

It should be noticed that the telescopes of both space missions COROT and KEPLER can be included in the small telescopes category as well – the diameter of the COROT telescope is just 30 cm while that of KEPLER is 95 cm. Thus, the entire field of transiting exoplanets and all the breathtaking science relies here indeed on small size instruments.

4.2 Planetary Microlensing

The original idea of using microlensing events as a tool for the search for extrasolar planets comes from Mao and Paczyński (1991) who analyzed microlensing of a distant star caused by a binary lens. They concluded that even in the case of extreme mass ratio (such as a binary system consisting of a star and planet) the effects distinguishing such a microlensing from a standard one caused by a single star should be, in a favorable configuration of a planetary system in the sky, easily detectable.

Microlensing occurs when two stars, a distant "source" star and closer "lens" star come into close alignment. The physical basis of this phenomenon is the deflection of light by a massive body which results from the general theory of relativity. The lens bends the light from the distant source star and magnifies it when the impact parameter is comparable to or smaller than the characteristic value called Einstein ring radius. The magnification grows with decreasing projected separation. The size of the Einstein ring is dependent on the mass of the lens and the geometry of the observer-lens-source system.

A planetary companion to the lens star located in the sky reasonably near one of the microlensing images of the source distorts the image and so changes the magnification. This perturbation of the regular microlensing light curve, called a "planetary anomaly", is short-lived and lasts from a few hours for an Earth-mass planet to a few days in the case of Jupiter-mass planets. Thus, to detect such a planetary anomaly, high cadence photometric observations of the ongoing microlensing events are necessary [7]. Planets can also be detected at very high magnification where the gravitational field of the planet destroys the symmetry of the Einstein ring [8].

The microlensing events caused by a single lens are extremely rare phenomena. A "planetary anomaly" can occur only in a favorable configuration of the planet-lens system in the sky, so that planetary microlensings are orders of magnitude less likely. Therefore, attempts to detect exoplanets via microlensing had to wait until the third phase of the OGLE project when typically up to 650 microlensing events per season had been detected in real time.

The first microlensing planet OGLE-2003-BLG-253/MOA-2003-BLG-53 was discovered in 2003 (Fig. 11.4) [3]. Since then the microlensing planet searches are maturing rapidly. So far about 20 exoplanets were discovered with this method. This number is relatively small compared to the number of discoveries from, for example, spectroscopic searches. Nevertheless the microlensing planet detections already contain a few major discoveries, for example that the "cold Neptunes" or "coldSuperEarths" (exoplanets of a mass of 5–15 Earth masses) at or beyond the so-called "snow line" are common. The detections of the first Jupiter/Saturn "solar system analog" and that of one of the lowest mass planet orbiting a substellar host are other examples of important discoveries. This high ratio of major discoveries to the number of detected planetary microlensing cases results from the favorable situation that microlensing probes the parameter space that is mostly not accessible to other techniques – the cold outer regions of extrasolar planetary systems.

The analysis of the microlensing light curve caused by a binary/planetary lens routinely provides the planet/star mass ratio and star-planet projected separation. Possible ambiguities in the interpretation can be resolved with good quality, continuous light curves. To constrain the mass of the lens star (and, thus, to assess the planet's absolute mass) and its distance, it is in general necessary to derive the probability distribution based on a model of the Galaxy if the lens light cannot be separated. However, in some cases subtle effects in the microlensing curve, such as "finite-source effects" or "parallax effects", can constrain or even allow direct determination of the mass of the host. The former effect is almost always detectable in the case of planetary microlensing. Moreover, complementary high angular resolution observations, either with HST or with adaptive optics, allow one to obtain additional constraints on the parameters of the system and determine masses to within better than 20% by directly measuring the lens and source relative proper motion.

The planetary hosts discovered via microlensing are located several kiloparsecs from the Sun. These are usually faint stars so they are too distant for direct imaging or radial velocity follow-up. Microlensing is sensitive to a wide range of host type stars. They can include G, K, and M-dwarfs and also white and brown dwarfs. Other methods are usually most sensitive to solar type hosts. Thus, microlensing is an independent and complementary detection method that has large potential in

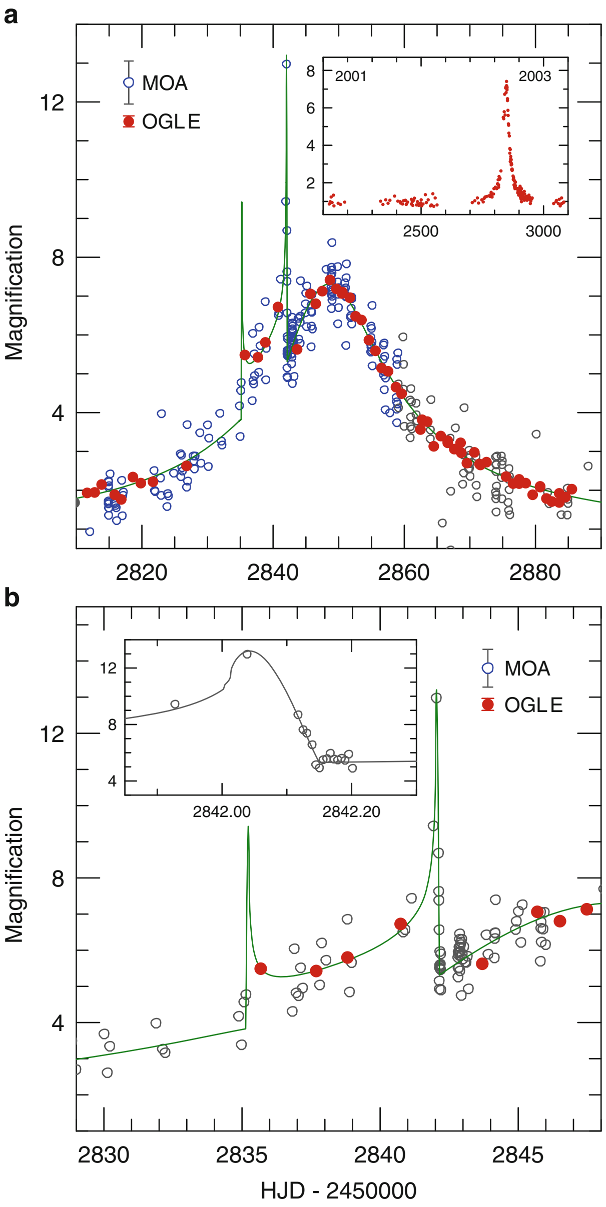


Fig. 11.4 Light curve of the first planetary microlensing event OGLE-2003-BLG-253/ MOA-2003-BLG-53

providing homogeneous data for comprehensive understanding of planet formation processes. For example, because microlensing probes mostly exoplanets outside the snow line, where the favored core accretion theory of planet formation predicts a larger number of low-mass exoplanets, the statistics provided by microlensing should make it possible to test the core accretion model. Microlensing surveys offer the best prospects of completing a full census of planets down to an Earth mass over separation scales of 1–5 AU, i.e., in the part of the parameter space not currently accessible to other methods. This technique will be complementary to the COROT and KEPLER space missions which should provide a comparable census for separations below 1 AU using the transit method.

Microlensing planet searches are currently still in their "discovery phase". The observing strategy of this phase follows an early proposition by Gould and Loeb [7] and is a two-step process. In the first step the microlensing events are found by two survey groups (OGLE and MOA) that have capabilities of continuous monitoring of large areas of the sky and implemented real time detections systems. So far almost five thousand microlensing events have been detected since 1992.

The survey groups alert the microlensing community on the on-going microlensing events. The so-called "follow-up" groups, such as PLANET/Robonet or MicroFun, are performing high time resolution photometry on selected promising events. Follow-up groups include observing sites on all continents, usually equipped with small 1 m class telescopes with CCD cameras. Sometimes even well equipped amateur astronomers participate in these follow-up networks (and discoveries – MicroFun). Also the survey groups are presently capable of observing part of their surveying area with sufficiently high cadence to discover, from their sites, planetary deviations coming from even the smallest mass exoplanets. Such a strategy allows continuous 24 h coverage of these selected events.

Typically a planetary event is recognized as such within a few days after the planetary anomaly, although the full detailed analysis is much more time consuming. As can be seen from this description the discovery and good observing coverage of a planetary microlensing event requires good coordination, international collaboration and fast information and data exchange between different microlensing groups. Microlensing is a great example of a success of such a model of open international collaboration.

4.3 Second Generation Planetary Microlensing Survey

Unfortunately, the current mode of detection of planetary microlensing events cannot guarantee the main goal of planetary microlensing surveys – the full characterization of the frequency and properties of exoplanets in the range where the microlensing method is sensitive. This is because even a fully optimized follow-up network cannot expect to detect all ongoing planetary anomalies. First of all, the capabilities of follow-up projects are limited and allow effective monitoring of only a fraction of the list of ongoing events. Secondly, there is a quite large number of missed microlensing events or events alerted too late – near or after maximum. Also the current "survey/follow-up" model is focused mostly on the "easiest channel" of microlensing planet detection: the rare very high magnification events where the detection of planets has the highest probability or caustic crossing planetary (binary) events which give very characteristic photometric signatures of a planet's presence. In many cases the presence of a planet can manifest itself by only a small short-lived disturbance on the light curve which is easy to overlook with the current model of observations.

The next generation planetary microlensing survey must correct these drawbacks. It is proposed that such a second generation survey will involve a network of 1–2 m wide field telescopes located on at least three continents (South America, Australia/New Zealand, South Africa). Each telescope should be capable of surveying several square degrees containing a few hundreds of millions of the Galactic bulge stars with cadences of 15–20 min. Such a network will then enable uninterrupted, round-the-clock monitoring of all observed objects.

The advantages of this approach are obvious. The efficiency of exoplanet detection will not be dependent anymore on the ability of the detection and alerting of ongoing events. All events located within the surveying area will be well sampled with a frequency sufficient to detect any Earth-mass planetary anomaly.

Extensive simulations of such next generation planetary microlensing surveys indicate that the network consisting of three nodes of 1.3–2 m class telescopes in three continents, covering about ten square degrees in the sky should detect thousands of microlensing events of the sources down to $I = 21$ mag and about 1–4 Earth-mass planets, 10–15 super-Earths and about 100 Jupiter-mass planets per year. Thus, the 5–7 years long survey should allow a complete census to be made for planets down to Earth masses orbiting the host's stars in the range of distances where the planetary microlensing is the most sensitive: 1–5 AU.

It is worth adding here that the second generation microlensing survey should also provide the first and the only possible information on the so-called free-floating planets. Theories of planetary system formation often predict that a fraction of planets can escape their systems during planet formation and evolution. Microlensing is the only technique that can detect free-floating planets and can probe this channel of planet detection. The detection and characterization of the free-floating planet population would then provide important constraints on planet formation theories.

The typical time scales of microlensing events caused by such planets are very short: shorter than a day or so. Therefore they cannot be effectively discovered, alerted and probed by the current survey projects. With the cadence of observations of 20 min, the second generation survey should provide rich observing material for the first census of free floating planets.

The first steps toward establishing a second-generation microlensing network have already been undertaken. The Japanese/New Zealand MOA-II survey operating at Mt. John Observatory in New Zealand carries out observations using a 1.8 m telescope with a CCD mosaic camera covering 2 square degrees in the sky.

In March 2010 the fourth phase of the OGLE project began regular observations. With its 256 Mpixel camera and location at one of the best astronomical sites – Las Campanas, Chile – OGLE has become one of the main nodes of the forming network.

Favorable location of these two observatories (separated by 8 h) already allows 16 h of continuous monitoring of the Galactic center in the Galactic bulge observing season. A third node that can partially fill the gap also already exists – the 8192 pixel mosaic camera and 1 m telescope at the Wise observatory (Israel). Although this site is located not so optimally as the remaining nodes (Northern Hemisphere), the period of the Galactic center visibility can still be about 5 h there. There are plans to establish a permanent node in South Africa (KMTNet – Korean Microlensing Telescope Network) making the OGLE-MOA network closed as well as including a few other existing or planned to be built 1 m class telescopes with large field cameras in the next few years. All these additional nodes can significantly extend the existing skeleton of the network and make it less prone to unexpected events like instrument failures, bad weather etc. It seems then almost certain that in the next couple of years the second generation microlensing survey, operating with small class telescopes, will collect unique observational material for the extrasolar planet field.

4.4 Exoplanets by the "Light Effect"

There is one more potential channel of exoplanet detection with small telescopes. This is the so-called timing method in a variety of its applications. Generally speaking, if the source of a precise signal is located in the binary (planetary) system, the time of the signal arrival to the observer on the Earth will periodically vary as the distance to the source changes due to orbital motion of the source. The delay or advance of the signal could be potentially measured providing information on the components of the system.

To measure the light effect one needs a precise astrophysical clock in the observed system. This can be, for example pulsations. Indeed, the so-called pulsar planets [30] were discovered using this method and very precise pulsar signal as a clock. However, stellar pulsations can be used as well. Other examples of such a clock can be eclipses in stellar binary systems. In these cases a circumbinary planet, brown dwarf or stellar companion could be found.

In the case of a planetary system the light effect is very small as the mass ratio between the planet and star is extreme. Therefore the photometric measurements must be extremely precise to detect this effect. Practically, the method is sensitive only on massive planets of Jupiter mass or larger.

On the other hand the light effect is larger for planets located at larger distances. This, however, means that the orbital periods must be long. Thus, for sound detection covering several orbital cycles, observations must span several years.

Small telescopes of 1–2 m mirror diameter class are perfectly suited for the timing method as the targets are usually quite bright and photometry can be done with millimagnitude accuracy for such stars. So far a few tentative detections of massive exoplanets with this method have been reported. For example a planetary companion to V391 Peg pulsating star was proposed by [21]. Also the discovery of planetary companions in eclipsing system HW Vir was claimed by Lee et al. [12].

All these discoveries require, however, further observational confirmation. There are high expectations that the Kepler mission will provide a bunch of discoveries with this method. Much more precise photometry should allow detection of smaller amplitude "light effects"s than is possible from the ground.

References

1. C. Akerlof, R. Balsano, S. Barthelmy et al., Nature **398**, 400 (1999)
2. C. Alcock, C.W. Akerlof, R.A. Allsman et al., Nature **365**, 621 (1993)
3. I.A. Bond, A. Udalski, M. Jaroszyński et al., Astrophys. J. **606**, L155 (2004)
4. F. Bouchy, F. Pont, N.C. Santos, Astron Astrophys. **421**, L13 (2004)
5. D. Charbonneau, T.M. Brown, D. Latham, M. Mayor, Astrophys. J. **529**, L45 (2000)
6. M. Gillon, F. Pont, B.-O. Demory et al., A&A. **472**, L13 (2007)
7. A. Gould, A. Loeb, Astrophys. J. **396**, 104 (1992)
8. K. Griest, N. Safizadeh, Astrophys. J. **500**, 37 (1998)
9. G.W. Henry, G.W. Marcy, R.P. Butler, S.S. Vogt, Astrophys. J. **529**, L4 (2000)
10. M. Konacki, G. Torres, S. Jha, D.D. Sasselov, Nature **421**, 507 (2003)
11. G. Kovács, S. Zucker, T. Mazeh, Astron Astrophys. **391**, 369 (2002)
12. J.W. Lee, S.-L. Kim, C.-H. Kim et al., Astron. J. **137**, 3181 (2009)
13. S. Mao, B. Paczyński, Astrophys. J. **374**, L37 (1991)
14. M. Mayor, D. Queloz, Nature **378**, 355 (1995)
15. B. Paczyński, Astrophys. J. **304**, 1 (1986)
16. B. Paczyński, Astrophys. J. **371**, L63 (1991)
17. B. Paczyński, in *Variables Stars and the Astrophysical Returns of the Microlensing Surveys*, ed. by R. Ferlet, J.P. Maillard, B. Raban (Frontiers, Gif-sur-Yvette Cedex, 1997), p. 357
18. G. Pojmanski, Acta Astron. **47**, 467 (1997)
19. F. Pont, F. Bouchy, D. Queloz, N.C. Santos, C. Melo, M. Mayor S. Udry, Astron Astrophys. **426**, L15 (2000)
20. J.L. Racusin, S.V. Karpov, M. Sokolowski et al., Nature **455**, 183 (2008)
21. R. Silvotti, S. Schuh, R. Janulis et al., Nature **449**, 189 (2007)
22. A. Udalski, Acta Astron. **53**, 291 (2003)
23. A. Udalski, M. Szymański, J. Kalużny, M. Kubiak ,M. Mateo, Acta Astron. **42**, 253 (1992)
24. A. Udalski, M. Szymanski, J. Kalużny, M. Kubiak, W. Krzemiński, M. Mateo, G.W. Preston, B. Paczyński, Acta Astron. **43**, 289 (1993)
25. A. Udalski, M. Kubiak, M. Szymański, Acta Astron. **47**, 319 (1997)
26. A. Udalski, B. Paczyński, K. Zebrun, M. Szymański, M. Kubiak, I. Soszyński, O. Szewczyk, L. Wyrzykowski, G. Pietrzyński, Acta Astron. **52**, 1 (2002a)
27. A. Udalski, O. Szewczyk, K. Zebrun, G. Pietrzynski, M. Szymański, M. Kubiak, I. Soszyński, Wyrzykowski, L., Acta Astron. **52**, 317 (2002b)
28. A. Udalski, M.K.Szymanski, M. Kubiak, G. Pietrzyński, I. Soszyński, K. Zebrun, O. Szewczyk, L. Wyrzykowski, K. Ulaczyk, T. Wieckowski, Acta Astron. **56**, 293 (2006)
29. J. van Paradijs, P.J. Groot, T. Galama et al., Nature **386**, 686 (1997)
30. A. Wolszczan, D.A. Frail, Nature **355**, 145 (1992)

Chapter 12
Large and Very Large Telescopes

Phil Charles

Abstract The last three decades have seen almost exponential growth in the numbers of large and very large telescopes, with the unanticipated current situation that there are almost as many VLTs as there are 4 m-class telescopes. This growth in numbers has been curiosity-driven, but obviously technology-led. After an initial plateau of 4 m telescopes in the 1980s, the cost-scaling with aperture (typically $\sim D^{2.7}$) has required dramatic changes in crucial fundamentals of large telescope design in order to be both technically feasible and politically affordable. However, to meet astronomers' demands for larger amounts of extremely large telescope time will require exploring new paradigms in large telescope construction. HET/SALT present such an opportunity. Their current status and potential will be described, and used as lessons to be learnt when high technology projects attempt to "break the cost-curve".

Keywords Telescopes • Astronomy: surveys • Astronomical observatories

1 Introduction: Growth in Numbers of Large Telescopes

While telescopic viewing of the sky began almost exactly 400 years ago, the modern era of "large" telescope useage is usually considered to have been ushered in a century ago with the completion of the Hooker 100-inch (2.54 m) glass reflector at Mt Wilson Observatory in California. The dramatic gain in size and sensitivity that this telescope afforded was to revolutionise our understanding of the cosmos, and

---

P. Charles (✉)
South African Astronomical Observatory, P.O. Box 9, Observatory 7935, South Africa

School of Physics & Astronomy, University of Southampton, SO17 1BJ, UK

Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa
e-mail: pac@saao.ac.za

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_12,

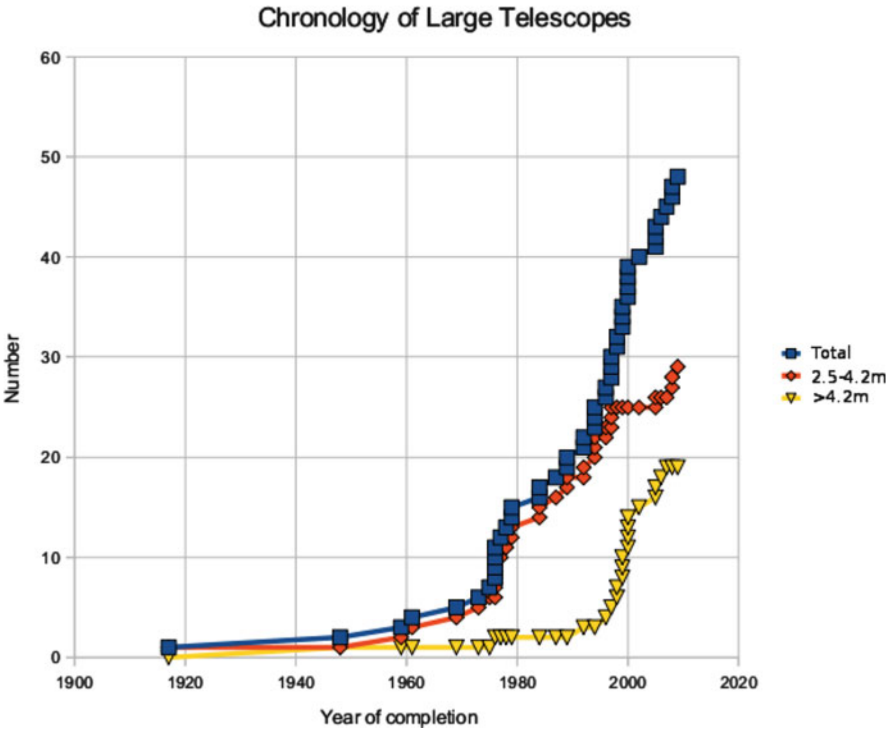


Fig. 12.1 Growth in number of large (>2.5 m diameter) telescopes over the last century. The lower curves have these divided into "large" (2.5–4.2 m) and "very large" (>4.2 m). Data points are mostly extracted from [12]

of our place within it, with Hubble's 1929 discovery of the expanding Universe and the realisation that the "nebulae" were other galaxies like our own, each containing billions of stars. It clearly demonstrated the benefits of curiosity-driven research, when the *status quo* for physics and astronomy at the turn of 1900 had classical physics "explaining everything" and the Sun and our Galaxy apparently dominating the Universe!

Hubble's discoveries fired the imagination of scientists, and more importantly those of the funding bodies. The latter included (particularly in the US, as continues to this day) wealthy philanthropists, such as Hooker, Lick and more recently Keck and Allen. Subsequent large telescopes would be driven by the huge technological gains that were made in the mid and late twentieth century. It is very instructive to see (Fig. 12.1) how the numbers of large telescopes then increased from the Hooker to the present day.[1]

In fact, the Hooker reigned as the world's largest telescope for the next quarter century, until the Hale 200-inch (5.08 m) was completed in 1948. But what is

---

[1] For a comprehensive historical review of the growth in telescopes see [24].

remarkable about Fig. 12.1 is the almost explosive growth in telescope numbers since the mid-70s which is showing no signs of slowing. Indeed, when that growth is examined in terms of telescope apertures in the 2.5–4.2 m and >4.2 m ranges (effectively "large" and "very large" telescopes respectively), the slackening in the construction of large telescopes has been compensated by the recent growth in the very large category, bringing us to the totally unexpected consequence that there will soon be as many VLTs as there are 4 m telescopes! [2]

There is a noticeable break in Fig. 12.1 around 1980, where [38] point out that it represents a fundamental change in the basic parameters and approach to large telescope design. Prior to 1980, virtually all telescopes were equatorially mounted, had slow (≥ 3) f-ratios and were single, thick mirrors. Post-1980 mountings became almost universally alt-az, the optics were much faster (f-ratios ≤ 2) and thin meniscus mirrors were being produced, along with segmented mirrors for the very largest apertures (>8 m). How had this come about?

2 Technology Drivers

In fact, the twentieth century's acceleration in telescope construction was very much technology driven, albeit justified with the latest "big questions" in astronomy. While these are always important, and are crucial in defining the range of instruments and their capabilities on large telescopes, it must be recognised that here there is a fundamental difference in philosophy of operation between astronomy, and the other "big science", particle physics. The unravelling of the nature of nuclear matter over the last century was achieved via particle accelerators, whose growth (in size, complexity and cost) has very much paralleled that of astronomical telescopes. However, accelerators were (and are) always constructed with the aim of reaching the "next" energy level to test and measure the predictions that resulted from lower energy experiments. Once reached, that accelerator's job is essentially complete, and it is usually retired, converted to other activities or upgraded for the next experiment. i.e. they are not general-purpose research tools.

---

[2] However, what is not represented in this figure is the enormous growth since 1990 in the number of small (<2 m) telescopes. While many major observatories have closed or retired their original suite of small telescopes, in order to release resources for supporting their larger facilities, the *new* small telescopes being constructed are very different in nature and operation. Rather than being general-purpose telescopes to undertake a wide range of scientific programs and allocated via classical scheduling to visiting observers, the latest small telescopes are almost entirely dedicated to a single observing mode and are operated robotically. The availability of low-cost computing power and astronomical-grade CCD detectors has made this possible (see Chap. 11). Hence total costs of these projects (e.g. those hosted by SAAO at our Sutherland observing station, such as SuperWASP and KELT) are measured in tens of thousands of dollars, not millions). It also makes scientific programs viable that really could not possibly be executed via classical means.

Operation and use of large telescopes has been almost completely different. By the time telescopes are built (and this is particularly true for space astronomy), most of the “big questions” used to justify their construction are becoming “history” and the field has moved on. (HST was initially proposed in 1962, but not launched until 1990!) But unlike accelerators, telescopes can, and usually do, have capacities that go way beyond the aspirations of their original designers. They are general-purpose research tools that can (and do) make observations and discoveries totally unanticipated when first proposed. Of course, this is helped by the continuing technological advances, particularly in advanced electronic devices (like CCDs) and the high power computing to process what is becoming unimaginably vast quantities of data. An excellent example of the longevity in scientific productivity of astronomical facilities is telescope number 2 in Fig. 12.1. This is the Hale 200-inch which was completed in 1948 yet is still a powerful and important research tool today.

What is clear from Fig. 12.1 is the craving for larger telescope apertures in order to collect more light. This enables observation of fainter (and hence more distant objects), giving “look-back” times that are currently taking us closer to the epoch of re-ionisation [33]. However, larger apertures invoke significant construction challenges for telescope designers and builders, such as

- with image recording limited to photography (whose sensitivity was low, and approximately constant during the twentieth century), the only significant gain possible was through increasing the telescope aperture;
- unfortunately, a larger mirror had to be *thicker* in order to be stiff enough to retain its shape against gravity as the telescope moved around the sky. Hence the weight gain increased as $\sim D^3$, requiring dramatic increases in engineering costs in order to support the massive mirror and move it accurately anywhere in the sky;
- this could be (at least partially) compensated with faster optics which allowed smaller (and hence cheaper) domes to house the telescope, but these provided challenges to the optical designer and constructor;
- additionally, larger mirrors had larger thermal capacity and hence took longer to respond to external temperature changes. This led to temperature gradients with respect to the ambient sky, which frequently led to poor image quality, thereby (at least partly) negating the gain in aperture.

The Hale 200-inch was effectively the last of the “old-school” approach to telescope design (although some attach this epithet to the Russian 6 m mirror with its poor thermal properties as a result of its huge mass). The Hale used the low thermal expansion coefficient (CTE) material pyrex (then just appearing) to try and minimise these effects, but advances in glass technology brought even better, ultra-low CTE materials such as Zerodur into the telescopes of the 1960s, such as the KPNO and CTIO 4 m designs [16]. Nevertheless, they were still “thick” in order to maintain their accurate shape against gravity, and this meant that they were 300–400 tonne monsters, with enormous heavy engineering required to move them quickly and accurately around the sky. The last of these “heavies” was the Anglo-Australian Telescope (AAT) which began operations in 1974, but it also ushered in the first modern-era, totally computer-controlled telescope. Indeed, the AAT exploited the computer control by developing a sophisticated numerical model of the telescope's pointing properties that would allow it to be moved around with hitherto unprecedented precision [40].[3]

However, it was the increase in weight resulting from larger diameter mirrors and associated engineering that led to the cost of telescopes up to $\sim$1980 scaling with diameter as $\sim D^{2.7}$. The relationship is not in fact straightforward, and has been discussed much in the literature (see [29] and references therein). Nevertheless, simple engineering extrapolations of earlier telescope designs were not feasible when a doubling of telescope aperture carried a price tag of a factor 6 or 7 increase in cost. New approaches were needed, or what [39] has referred to as the *innovation factor*, I_F, in order to make very large telescopes affordable (or at least politically acceptable).

The 4 m-class telescopes since 1980 (e.g. NTT, TNG, SOAR) are all thin, "meniscus" mirrors with drastically higher aspect ratios (diameter/thickness) of $\sim$40 compared to earlier generations (the KPNO and CTIO 4 m have a ratio of 6). This gives far superior thermal performance, but the mirrors cannot support themselves, and require sophisticated active mechanical support in order to maintain their shape as a function of elevation angle. The honing of this technology was a key component in making 8 m single mirrors technically feasible. Compared to pre-1980 telescopes, the following areas were essential in bringing the 8m era into being [38]:

- thinner, lighter mirrors encompassing active mirror support (including novel mirror-casting technologies, such as Roger Angel's spinning furnaces [30];
- alt-az mounts with computer-controlled pointing (significantly lower mass than equivalent equatorial mounts);
- faster f-ratios (allowing smaller buildings; the Gemini 8 m is in a building of the same size as the KPNO 4 m);
- diffraction-limited mirror surfaces to exploit adaptive optics atmospheric correction;
- thermal control of building enclosures and telescope equipment;
- segmented mirror designs that allow construction of almost any desired telescope aperture, independent of glass furnace limitations.

A key enabling technology of virtually all the above developments was the advent of powerful, but cheap, computing capacity. Initially through mini-computers (used at the large telescopes of the 1970s and 1980s), it was the evolution of the

---

[3] I recall observing at the AAT in the early 1980s and, after moving between targets, I asked the Telescope Operator whether the telescope was there yet and if we should move to acquisition, rather than slit view, in order to find our target. Without saying a word, the TO simply moved the telescope a couple of arcseconds, and lo and behold my target popped out of the spectrograph slit and into view. The telescope had moved to the requested position so accurately, that the target had landed right in the middle of the spectrograph's 1-arcsec slit (thereby making it invisible in the slit viewer)! Such pointing accuracy would have been unthinkable ten years earlier, but the AAT's performance set new pointing standards that are now expected for all large telescope projects.

humble PC into what is true "super-computer" number-crunching capacity that has revolutionised telescope control, operation and design. My first encounter with the simple application of a collection of PCs to operate the telescope and its instruments was at the Nordic Optical Telescope (NOT) in the late 1980s, where their focus on the building's thermal design and ventilation properties was able to bring its natural imaging performance (without AO) to outstanding levels [23].

So, how have all these developments impacted on the costs of very large telescopes?

2.1 The Cost-Curve

A concept first introduced by Meinel [14], it is much easier to see how telescope costs have evolved by using a plot of their cost *per unit area* as a function of their aperture (Fig. 12.2). This removes most of the steep aperture dependency of the simple cost-aperture scaling.

Clearly the effect of the technology gains on the cost-effectiveness of the current generation of very large telescopes (Kecks, Gemini, VLTs) was not huge, but they are still ∼1/2–2/3 of the costs that might have been expected based on the standard scaling relation, a result predominantly due to their much lighter weight mirrors and alt-az mountings. That the gain was not larger is due to demands on their performance that are greatly reduced in smaller telescopes. As pointed out by Mountain and Gillett [16] the signal-to-noise ratio (SNR) of optical imaging data can be expressed as being $\sim(D/\theta)(\eta/F_{BG})^{0.5}$ where D is the telescope diameter, η the overall throughput of the optical system, θ is the angular size of a point source and F_{BG} is the background flux. With careful design, η and F_{BG} are essentially the same across all telescopes. Which means that the way to maximise the SNR is by maximising D/θ, and that means for any D you always want to minimise θ, which in turn means minimising the telescope contribution to the atmospheric "seeing". Better still, reduce it below the seeing level and aim for the mirror's diffraction limit by using natural and laser guide stars as part of adaptive optics (AO) systems [2]. These are technically challenging (they have been in development for three decades) and expensive.

Since ESO's VLTs cost ∼\$100M (2005) each, then application of the scaling relation would imply that a next generation "extremely large telescope" (ELT) of $D = 42$ m (the European ELT goal) would cost ∼\$9B, an amount comparable to 5 HSTs! The target is then to reduce an ELT cost to a more acceptable ≤\$2B. As emphasised by [39], this means an innovation factor, I_F, of in the region of 5 to 10 in terms of making the technology affordable. Both the TMT and E-ELT design teams are aiming for at least these levels of improvement, which is indicated by their positions in Fig. 12.2.

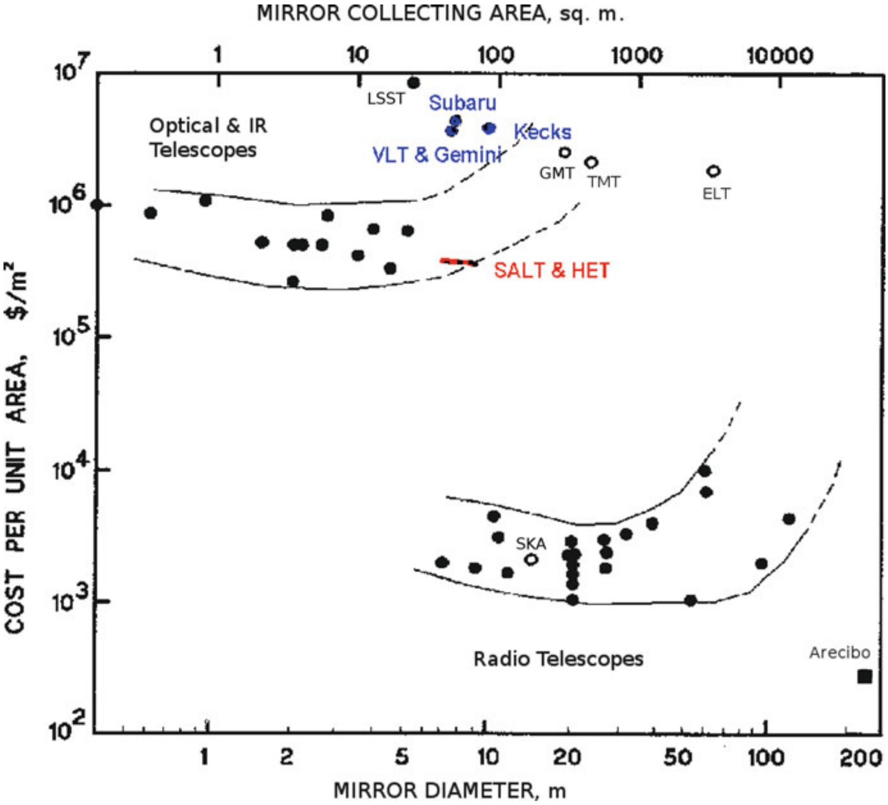


Fig. 12.2 Cost per unit area for optical and radio telescopes as a function of telescope aperture. Solid symbols denote facilities that are completed or under construction. Arecibo is a fixed dish. Open circles represent telescopes still in initial design phase, and so costs are approximate, and in $(2005) (Based on an original concept in [14])

3 Breaking the Cost-Curve

All these developments of the last 30 years have concentrated on achieving the highest technical and hence scientific performance from this new breed of very large telescopes. They have definitely not been “cost no concern” projects, but have still cost substantially more than their earlier brethren. However, in the early 1990s, this approach was challenged. Is there a way of achieving very large collecting area for optical telescopes, but at greatly reduced cost compared to the VLTs? The clue to how this challenge was taken up is actually contained in Fig. 12.2, but it is in the lower part of the plot which addresses *radio* telescopes. That they are cheaper per unit area than optical telescopes is not in the least surprising given the technology of their reflecting surfaces. However, the largest single radio dish is *Arecibo*, and it is a factor of several times cheaper per unit area than any other. Why? It is fixed (a natural crater produced by sinkholes) and so cannot be moved (it accesses only the band of the sky that passes overhead, see [1]). That means there is no elaborate (and expensive) pointing and guiding system needed for the primary mirror. The high technology component for *Arecibo* is in the prime focus platform, which is held almost 500 ft above the dish by cables connected to high concrete towers (a major engineering accomplishment in the early 1960s). The dish is actually spherical and so a secondary/tertiary system is used to bring the light to a focus.

The *Arecibo* concept inspired Dan Weedman and Larry Ramsey at Penn State University in the mid-80s to wonder if this approach could be adapted for an optical telescope [25]. Focussing on the need for a "Spectroscopic Survey Telescope", this concept was taken up by Ramsey et al. [26] and funded as the Hobby-Eberly Telescope (HET), intended as a prototype for constructing 10 m-class optical/NIR telescopes for a fraction of the cost of Keck, VLT, Gemini, etc, and with a spectroscopic survey capability as its main science instrument. Smaller aperture projects, but with similar science aims of undertaking large field spectroscopic surveys are SDSS [28] and LAMOST [13].

3.1 The HET/SALT Paradigm

The Hobby-Eberly Telescope (HET) was completed in the late 1990s at McDonald Observatory in Texas, but before it entered "normal" operations, a southern copy (SALT) had already been adopted by a large, multi-national consortium. The HET/SALT paradigm for very large telescope construction is still evolving, but its key elements are shown schematically in Fig. 12.3. The most radical deviation from traditional optical telescope design (and the single largest factor in its low cost) is the selection of a 91-element segmented *spherical* primary mirror array operating at a *fixed* elevation angle and with the mirror stationary during observations (as at *Arecibo*).

The trade-offs in the HET approach are:

- additional optical components are necessary, i.e. a SAC (spherical aberration corrector) which is an integral component in the HET/SALT design as a result of its spherical primary mirror array (1 m spherical segments are relatively simple and cheap to cast and polish to high accuracy);
- operating at a fixed elevation angle (albeit with selectable azimuth angle) restricts access to $\sim$70% of the sky, and only $\sim$15% at any given moment (thereby greatly simplifying the mirror support function, as the primary mirror array does not change its position relative to gravity);
- the primary mirror is stationary during observation of the target, which requires (as for *Arecibo*) a Tracker to carry the instrument package across the sky. The Tracker is the most complex piece of technology in the HET/SALT design, as it must move in six axes: x, y, z(focus), tip-tilt (θ-ϕ) and rotation (ρ). The Tracker is the *only* moving part of the telescope during an actual observation;

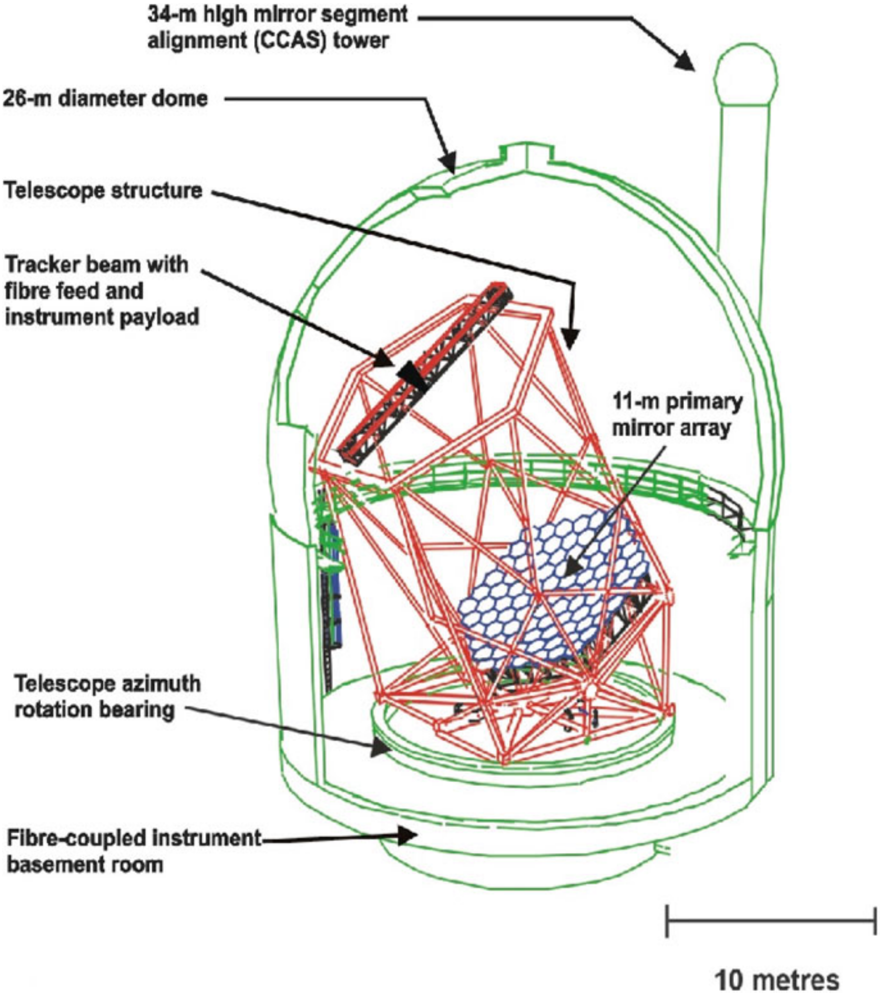


Fig. 12.3 Cutaway schematic showing the key components of the HET/SALT paradigm for cost-effective very large telescope construction

- observing windows on targets are (mostly) restricted to two approximately 1-hour intervals per night, although this can be up to 3 hours at the declination extremes;
- instrumentation is constrained to the volume and mass permitted by the Tracker design at prime focus (or must be fibre-fed below the telescope)

However, against this must be noted the *benefits* that accrue from these physical limitations:

- totally Q-scheduled mode of operation (its constrained sky-access makes HET/SALT more akin to space-based observatories), which allows the observing program to be optimised at all times as a function of the observing conditions [6];

- only staff astronomers undertake the observing programs, leading to more efficient and effective use of telescope time, plus greater control of instrumental setups and configurations, which makes pipeline processing of data more straightforward;
- enables synoptic monitoring programs to be scheduled naturally;
- target-of-opportunity (ToO) programs can be executed without impact on "normal" programs;
- proposers receive their allocated time independent of weather conditions, only the overall efficiency is affected.

With a cost target of just \$14M, HET was actually built for \$16M, and was operating with its first generation instruments by 1999 [9]. However, HET did not meet its original design specifications and suffered from a number of unanticipated problems, which is perhaps not surprising for such a radical, pioneering design (at least in this wavelength range), and one where minimising the cost envelope was such a driving force. Chief amongst these was its poor image quality of 2.5–3 arcsecs, and a major completion project was required to bring HET's performance within specification [3,4]. This was achieved by tackling the four main contributors to HET's image quality (IQ):

- develop a new (Shack-Hartmann) system for alignment of the mirror array (which is located in the CCAS tower, see Fig. 12.3);
- develop a segment alignment maintenance system (SAMS). This uses inductive edge-sensors on the mirror segments to keep them in alignment, once the mirrors have been "stacked" at the commencement of operations. The degradation of the "stack" with time is a result of ambient temperature changes during the night. SAMS had to be developed from scratch, as it was not a part of the original design;
- modify the dome environment with louvres so as to minimise temperature effects within the building;
- improve the model which controls the movement of the Tracker (with the SAC) across the sky. Only x and y were actively guided, yet the SAC needs to be maintained perpendicular to the mirror array to <25 arcsecs and in focus (z) to $<10\,\mu$. The remaining axes had been left open-loop, and this was the major problem. A distance-measuring interferometer and an autocollimator were developed and installed.

The combination of these factors, executed in the early years of the last decade, brought HET's median seeing to $\sim$1.5–1.7 arcsecs, and greatly improved its efficiency of operations.

3.2 SALT: A 2nd Generation HET

Originally the SALT consortium was formed with the intention of making a southern hemisphere copy of HET. But soon after completion and early operation of HET

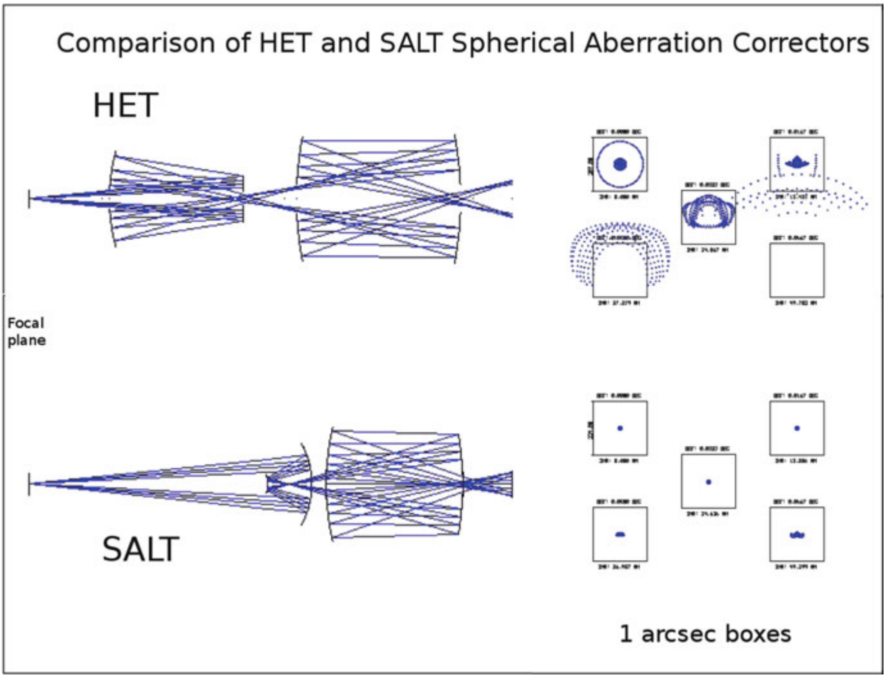


Fig. 12.4 Optical ray-tracing for the HET and SALT SAC designs (from [19]). Spherically aberrated light enters from the right, encounters the four mirrors of the SAC, and emerges corrected at the left. Note the much larger (and more practically useful) back-focal distance of the SALT SAC. To the right of each optical layout are calculations of their image quality, on-axis and at various positions in the field of view. SALT has a field of view of 8 arcmins, approx four times the area of the HET field. The boxes are 1 arcsec on a side

in the late 1990s it became clear that there were a number of areas where further development was required, and alternative solutions needed as just described. Key amongst these was the spherical aberration corrector (SAC), which goes to the very heart of the large spherical mirror paradigm. At SAAO, Darragh O'Donoghue [19] discovered an alternate optical solution for the SAC which had substantially improved IQ performance, over a much larger field-of-view, and provided a much larger back-focal distance (see Fig. 12.4).

In parallel, the SALT Project Team undertook a full review of the original HET system, aiming to benefit as much as possible from the lessons learnt. Consequently, they made a number of changes to the HET design, including all those listed above, apart from that of opting for the inductive sensor version of SAMS. SALT initially adopted a capacitive version, but has had to abandon this due to its humidity sensitivity, and is now investigating an inductive solution [15]. So the key changes incorporated into SALT were:

- greatly improved SAC optical design with IQ specification of <0.2 arcsec (EE50);
- improved mirror support and precision actuation system;

- extended instrumentation suite to exploit the increased space available and larger field-of-view;
- improved building design incorporating an active (pumped glycol system, with heat exhaustion 50 m away from the SALT building) thermal control of all internal heat sources. It is possible to set the telescope enclosure temperature during the afternoon to be close to that predicted for the early evening when operations would begin (and for which a site weather station takes data for input into a specially developed meteorological model).

With an aggressive and well-focussed project management approach, SALT was completed in a remarkably short period of time, going from its Ground-Breaking Ceremony in August 2000 to its Inauguration (by State President Thabo Mbeki) in November 2005. Figure 12.5 shows what had been accomplished in such a short period of time (then a record for 10 m-class telescopes).

However, this schedule was a punishing one, and while it had been achieved within the original construction budget of $20M, it had not yet undergone the testing and commissioning that would demonstrate whether it was meeting its specifications. Unfortunately, the commissioning year was to reveal two major problems [5]: the telescope's crucial image quality (IQ) performance and its main spectrograph's throughput in the blue spectral region. Solving both of these problems demonstrates how SALT holds valuable lessons for future extremely large telescope projects:

- **Fixing SALT's IQ**: full details of the IQ problem and its solution can be found in [20]. Manifested as a serious focus gradient across the 8 arcmin field (thereby preventing the taking of in-focus images across the full 8 arcmin field), this could have been due to optical flaws in any of three areas: the primary mirror array (e.g. in stacking), the SAC or the final imaging optics. It took an extended and detailed investigation led by O'Donoghue to (a) identify that the problem was in the SAC, (b) understand the complex behaviour as being due to a badly designed mechanical interface with the Tracker/payload which transmitted mechanical and thermal forces that could move and distort the SAC mirrors, (c) design a kinematic-mount interface to remove this problem, and (d) prepare the optical test procedure and handling equipment prior to dismantling the SAC, testing all its components, reassemble and realign them, and finally test the SAC back on the telescope. All of this work was undertaken by SAAO staff at Sutherland;
- **Improving RSS' Blue Throughput**: in parallel, the main spectrograph (RSS) was removed from the telescope in late 2006, and its optical lens components were disassembled and returned to the US manufacturer for testing and analysis. The blue throughput problem was identified as originating in the lens-coupling fluid used in the collimator and camera optical systems. While originally of essentially 100% transmission throughout the entire atmospheric optical range, the fluid had reacted chemically with the surrounding polyurethane bladder, thereby changing its optical properties to make it opaque in the UV/blue [5]. What should be of interest to future designers is that the fluid/bladder

Fig. 12.5 SALT, as completed in 2005, showing the primary mirror array, its mirror support truss, and the dome louvres open. At top is the Tracker, which carries the SAC (Spherical Aberration Corrector) and instrument payload

combination had been recommended by the manufacturer, and was (still is) in use in other astronomical spectrographs. However, what was novel about RSS/SALT is its use in the UV/blue, and the combination had not been tested in that wavelength region, nor do existing instruments operate there. Everything had been chosen because it was believed that it had been used before and was therefore "known to work". This turned out not to be the case.

Solving these problems has extended the "effective" construction period by ~4 years and the budget by ~\$5M, but even at those increased levels they represent an outstanding accomplishment (in terms of the optical performance from a 10 m spherical primary for the funds expended). More importantly, this is now demonstrating that the HET/SALT paradigm is a viable route for attaining low-cost, very large apertures, but within the constraints noted earlier, and as proposed in the late 1990s by Sebring et al. [27]. Indeed, the ELT community has expended considerable energy in investigating which technology to scale up to ≥ 30 m telescope diameters, for which there are two major projects currently underway, TMT [17] and E-ELT [34].

It is worth highlighting now some important lessons learnt from SALT, many of which are applicable to a wide range of circumstances:

- the construction schedule was set when SALT was intended to be a *copy* of HET. Major changes were subsequently made to the design (described above), but the schedule (and budget) were not adjusted accordingly. The problem of course is that the original schedule and budget were at a level which made the project politically supportable for a number of the international partners. Yet the changes that occurred were unarguable given what was known by then about the problems with HET;
- no allowance was made for the substantial tightening of technical specifications (IQ requirement of $<0.65''$ across the much wider SALT field-of-view) that resulted from the improved SAC optical design of [19]. This had opto-mechanical implications for the SAC-payload interface that were ignored by the SALT Project Team, and the SAC was mechanically designed in isolation from the requirements of the Tracker/payload interface;
- the operations budget for SALT was estimated according to what had been planned for HET operations, but this was done *before* HET was completed. Once HET operations did begin, their operating budget increased rapidly as the underestimates were recognised, but SALT never changed its estimates during the construction phase. With such a large consortium of partners (13) it was felt better to wait until the two major problems had been solved and science-quality data were flowing before approaching their funding agencies with these revised figures.

In spite of these problems, solving them is already generating benefits for SALT's prototype, HET. Recognising its potential for dedicated survey work, the University of Texas have successfully secured the necessary funding for a major 2nd-generation instrument, the *HET Dark Energy Experiment*, or HETDEX [10]. To accomplish this required a sophisticated, wide field, multi-object, multi-spectrograph instrument. However, the original HET SAC was *not* wide field, and so HETDEX also incorporates in its optical design a replacement for their SAC, and that is based on SALT [19].[4] This therefore highlights the benefits of a multi-site

---

[4] HETDEX also demonstrates an aspect of VLT design and construction that is not susceptible to the cost-savings inherent in the HET/SALT paradigm. The principal scientific instrument on SALT is the multi-mode Robert Stobie Spectrograph [5] which, with a price-tag ~$6M, is comparable to

design where SALT was very much a 2nd generation HET, and HET is now reaping the benefits of SALT by retro-fitting SALT upgrades. It is a supreme example of the value of a truly global collaboration that brings together scientists with a wide variety of backgrounds and experience.

To have done all this with a largely South African team of engineers in the first decade of the "new" South Africa is, in spite of the extended timescale, a considerable accomplishment, and one that has attained a high profile within the country. But SALT is more than just a "cost-effective 10 m telescope". It is an icon for driving science and technology education in a developing nation.

3.3 SALT: Collateral Benefits

SALT was driven in South Africa in the 1990s by my predecessor as SAAO Director, Bob Stobie, who was seeking to enhance the optical suite of telescopes at Sutherland. The largest telescope then operating there was the Radcliffe 1.9 m reflector, which when completed in the 1940s had held the title of "largest telescope in the southern hemisphere", a title that SALT has allowed South Africa to regain. The earliest concept developed was based on a copy of ESO's NTT, but taking the HET route was much more appropriate. It was a pioneering approach, and represented an (almost) affordable entry into the 10 m club. However, its approval by the South African government was conditional upon securing 50% funding from international partners. Helped by the inspirational support of the then President of South Africa's National Research Foundation, Dr Khotso Mokhele, this was quickly achieved in the late 1990s, and the SALT Project began.

However, Khotso insisted that, unlike in other large telescope and research projects, SALT must have an extra dimension. He devised a SALT "Collateral Benefits Plan" whose aim was to maximise benefits from South Africa's investment in SALT for the country as a whole. This was to include both industrial and educational benefits, so as to help overcome and redress the consequences of the policies of past South African governments that excluded the majority from science, engineering and technology education, training and careers.

The resulting SALT collateral benefits program has focussed on three areas: education in mathematics, science, engineering and technology to supply the country and the wider African continent with well-trained and motivated professionals in substantially increased numbers; science communication and awareness to effectively engage with the public in order to disseminate relevant information in the fields of astronomy and space science; and socio-economic development in order to contribute to a better quality of life for all people, especially the disadvantaged.

---

that of general purpose instruments on Keck and ESO's VLT. As a multi-spectrograph instrument, HETDEX is even more expensive, ~\$25M, which is close to double the original price-tag of the telescope itself!

Integral within this plan was NASSP, the National Astrophysics and Space Science Programme [18], which recognised the dearth of astronomy PhDs being produced in South Africa, and especially the lack of involvement of black South Africans. Since its inception in 2003, NASSP has brought astronomy lecturers from around the country to Cape Town (where it is hosted by UCT) to train graduates to Honours and Masters level, after which they are eligible to undertake PhDs at any of the participating institutions [42]. Almost a hundred students have now gone through this programme, which is attracting applicants at double the initial level.

The SALT collateral benefits activities have grown substantially over the last 7 years, reaching an ever-growing fraction of the public and school populations in the Western and Northern Cape provinces of South Africa [8]. But it was the International Year of Astronomy in 2009 that provided the springboard to take these activities to a much wider audience. Their success in doing so was recognised by the IAU in 2010 with the award to SAAO of the contract to host their *Office for Astronomy for Development* (OAD). This office is to coordinate such activities and grow them globally, maintaining the momentum that was begun during the IYA. There will be similar expectations and programs associated with future large science projects, such as South Africa's bid to host the SKA in Africa [31].

4 Large Survey Telescopes

All of the telescopes discussed so far are instrumented and operated for multiple scientific programs, which are selected based on competitive time-assignment procedures. However, target selection for most astronomical programs requires access to sky surveys and catalogues that are generated by very different types of telescopes. The best examples of those are the Palomar and UK Schmidt sky surveys which were completed in the 1950s to 1980s, and are now available online in digital form [7]. They were critical to huge numbers of programs, and the modern successors to these 48-inch wide-field (6.5 degrees square) Schmidt telescopes promise a new revolution in research opportunities.

While the last 20 years have seen a number of highly successful survey-style instruments developed for existing telescopes (such as 2dF for the AAT, [37], and WFCAM for UKIRT, [41]), these have been used for very specific research projects, rather than as general resources in their own right. However, the technological developments described earlier have led to the dedicated telescope survey projects SDSS [28] and 2MASS [35] becoming the modern successors to the 50-yr old Schmidt surveys. More importantly, the project costs of such facilities have evolved so as to make the data-processing and distribution as large, if not larger, than the hardware and basic operational costs. That is not surprising when it is recognised that SDSS used a 120-megapixel camera (covering 1.5 degrees square at a time) and took spectra of 600 objects at a time. These databases are as essential now as the Schmidt surveys were. And unlike the targeted surveys, these databases have become public almost as soon as they were completed. Combined with online access to space-based X-ray, γ-ray, UV, optical and IR surveys, these huge data archives are becoming very important research tools in their own right, and have led to the development of the Virtual Observatory (VO) concept ([11]; and Allen).

The next generation survey telescopes are even more challenging in their grasp. Just completed at ESO are the VST and VISTA, 2.5 m optical and 4 m IR respectively, wide-field survey telescopes, designed to complement the VLT [32]. In the IR, VISTA has the largest (67-megapixel and 3 tonnes in weight) IR camera yet constructed, which will produce a survey $\sim$40 times more sensitive than 2MASS. But even these pale next to the LSST, Large Synoptic Survey Telescope [36]. An 8.4 m telescope destined for completion in the coming decade, its 3,200-megapixel camera (consisting of 189 4 k^2 CCDs) will weigh 3 tonnes, and generate a 165 Pb data archive in its 10 year survey. For the first time in such survey telescopes it will provide both depth and temporal information on its entire accessible sky. For a glimpse of the range of science that this will make possible, see [21], where the Palomar 48-inch Schmidt telescope has been put to use as a *transient factory*, in order to survey the variable sky (see also Pan-STARRS, [22]). The temporal domain is one area of astrophysics where systematic surveys of this form are expected to revolutionise the field.

Acknowledgements I am grateful to David Buckley and Shireen Davis for assisting in the preparation of some of the figures used here and in providing some of the references.

References

1. Arecibo Observatory – Puerto Rico, http://www.naic.edu/index.php
2. J.M. Beckers, Adaptive optics for astronomy: principles, performance, and applications. Annu. Rev. Astron. Ap. **31**, 13–62 (1993)
3. J.A. Booth et al., The Hobby-Eberly telescope completion project. Proc. SPIE **4837**, 919–933 (2003)
4. J.A. Booth et al., The Hobby-Eberly telescope: performance upgrades, status, and plans. Proc. SPIE **5489**, 288–299 (2004)
5. D.A.H. Buckley et al., Commissioning of the Southern African large telescope (SALT) first-generation instruments. Proc. SPIE **7014**, 07–15 (2008)
6. P.A. Charles, Science with SALT in the ELT era: a "low-cost" option. in *Proceedings "Science with 8–10 m Telescopes in the Era of ELTs and the JWST"*, ed. by C. Warden (Fundación Ramón Areces, Madrid, 2010), pp. 124–141
7. Digital Sky Survey, http://archive.stsci.edu/dss/
8. K. Govender, Socio-economic impact of astronomy in South Africa. in *"Communicating Astronomy with the Public" Proceedings from the IAU/National Observatory of Athens/ESA/ESO Conference*. ed. by L.L. Christensen, M. Zoulias, I. Robson (Eugenides Foundation, Athens, 2008), pp. 160–164
9. G.J. Hill, P.J. MacQueen, L.W. Ramsey, M.D. Shetrone, Performance of the Hobby-Eberly telescope and facility instruments. Proc. SPIE **5492**, 94–107 (2004)
10. G.J. Hill et al., The Hobby-Eberly telescope dark energy experiment (HETDEX): description and early pilot survey results. in *Panoramic Views of Galaxy Formation and Evolution, ASP Conference Series*, vol. 399, ed. by T. Kodama, T. Yamada, K. Aoki (AstrThe Palomar Transient Factoryonomical Society of the Pacific, San Francisco, 2008), p. 115

11. International Virtual Observatory Alliance, http://www.ivoa.net/
12. Large Reflecting Telescopes, http://abell.as.arizona.edu/~hill/list/bigtel99.htm
13. Large Sky Area Multi-Object Fiber Spectroscopic Telescope, http://www.lamost.org/website/en/
14. A.B. Meinel, Cost scaling laws applicable to very large optical telescopes. Proc. SPIE **172**(2), 2–7 (1979)
15. J. Menzies, H. Gajjar, S. Buous, D. Buckley, P. Gillingham, SALT segmented primary mirror: laboratory test results for FOGALE inductive edge sensors. Proc. SPIE **7739**, 27–31 (2010)
16. C.M. Mountain, F.C. Gillett, The revolution in telescope aperture, Nature **385**, A23 (1998)
17. J. Nelson, TMT: the next generation of segmented mirror telescopes. in *Proceedings "Science with 8–10 m Telescopes in the Era of ELTs and the JWST"*. ed. by C. Warden (Fundación Ramón Areces, Madrid, 2010), pp. 52–63
18. National Astrophysics and Space Science Programme, http://www.star.ac.za/
19. D. O'Donoghue, Correction of spherical aberration in the Southern African large telescope (SALT). Proc. SPIE **4003**, 363–372 (2000)
20. D. O'Donoghue et al., Saving SALT: repairs to the spherical aberration corrector of the Southern African large telescope (SALT). Proc. SPIE **7739**, 21 (2010)
21. Palomar Transient Factory, http://www.astro.caltech.edu/ptf/
22. Panoramic Survey Telescope & Rapid Response System, http://pan-starrs.ifa.hawaii.edu/public/
23. Properties of the Nordic Optical Telescope, http://www.not.iac.es/telescope/tti/proptxt.html
24. R. Racine, The historical growth of telescope aperture. Publ. Astron. Soc. Pac. **116**, 77–83 (2004)
25. L.W. Ramsey, D.W. Weedman, The penn state spectroscopic survey telescope. in *Very Large Telescopes, Their Instrumentation and Programs; (A85-36926 17-89). ESO Proc.*, Garching, 1984, pp. 851–860
26. L.W. Ramsey, T.A. Sebring, C.A. Sneden, Spectroscopic survey telescope project. Proc. SPIE **2199**, 31–40 (1994)
27. T.A. Sebring, F.N. Bash, F.B. Ray, L.W. Ramsey, The extremely large telescope: further adventures in feasibility. Proc. SPIE **3352**, 792–800 (1998)
28. Sloan Digital Sky Survey, http://www.sdss.org/
29. L. Stepp, L. Daggert, P. Gillet, Estimating the costs of extremely large telescopes. Proc. SPIE **4840**, 309–321 (2003)
30. Steward Observatory Mirror Lab, http://mirrorlab.as.arizona.edu/MISC.php?navi=histo
31. Square Kilometre Array, Africa, http://www.ska.ac.za/
32. The ESO Survey Telescopes, http://www.eso.org/public/teles-instr/surveytelescopes.html
33. The Epoch of Reionization, http://www.astro.rug.nl/~LofarEoR/scibgd/scibgd.html
34. The European Extremely Large Telescope, http://www.eso.org/public/teles-instr/e-elt.html
35. The Two Micron All Sky Survey at IPAC, http://www.ipac.caltech.edu/2mass/
36. J.A. Tyson, New frontiers with LSST: leveraging world facilities. in *Proceedings "Science with 8–10 m Telescopes in the Era of ELTs and the JWST"*. ed. by C. Warden (Fundación Ramón Areces, Madrid, 2010), pp. 160–177
37. Two-Degree Field, http://www.aao.gov.au/AAO/astro/2df.html
38. G.T. van Belle, A.B. Meinel, M.P. Meinel, The scaling relationship between telescope cost and aperture size for very large telescopes. Proc. SPIE **5489**, 563–570 (2004)
39. M.S. Longair, Very large optical-infrared telescopes: today and tomorrow. in *Proceedings IAU Symposium 232*, ed. by P.A. Whitelock, M. Dennefeld, B. Leibundgut (Cambridge University Press, Cambridge, 2006), pp.13–24

40. P.T. Wallace, The pointing and tracking of the Anglo-Australian 3.9 metre telescope. in *Conference on Optical Telescopes of the Future, Geneva, Switzerland, Proceedings. (A79-14001 03-89)* (European Southern Observatory, Geneva, 1978), pp. 123–131
41. WFCAM - The UKIRT Wide Field Camera, http://www.jach.hawaii.edu/UKIRT/instruments/wfcam/
42. P.A. Whitelock, Astronomy and Development in Southern Africa. in *Proceedings of "Accelerating the Rate of Astronomical Discovery" at the 27th IAU General Assembly*, Rio de Janeiro, Brazil. Published online at http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=99, id.19 (2009)

Chapter 13
The Challenge of Optics in Future Extremely Large Telescopes

Eric Ruch

Abstract The development of astronomical telescopes has largely been driven by the capability to produce accurate and complex surfaces that are necessary to correct the geometrical aberrations existing in any optical instrument. Although the optical configurations used in the most recent telescopes were discovered almost 400 years ago, the technical difficulties and challenges to manufacture these optics did not allow one to contemplate the successful building of such mirrors until modern technologies such as computer controlled polishing and the laser interferometer were available. In this chapter we will show that extremely large telescopes of the future will require much more accurate optics than any other telescope in the past. The need for these highly accurate mirrors is driven by the most ambitious scientific goals, such as Earth-like planet detection and spectroscopy. To meet these challenges, several techniques for polishing and figuring mirrors have been developed in Europe and in the United States and some of the most promising will be reported in this chapter. The first results in development of prototypes demonstrate that necessary polishing technology is now available to meet the most stringent requirements set by the astronomers.

Keywords Astronomical instrumentation • Methods and techniques • Instrumentation: adaptive optics • Telescopes

1 Introduction

At the end of the next decade, a number of extremely large telescopes (ELT) will be put into operation and will open a new window in the sky. Those telescopes will not only be significantly larger than the current generation of telescopes but

---

E. Ruch (✉)
Sagem Défense Sécurité, Le Ponant de Paris, 27, Rue Leblanc, 75513 Paris Cedex 15, France
e-mail: eric.ruch@sagem.com

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_13,

they will require significantly better optics. In the European ELT or in the 30 m Telescope (TMT) project in the United States, the primary mirror will consist of several hundreds of hexagonal segments of 1.4 m in size. Here the challenge of the optical industry will be to produce more mirrors for these telescopes in a period of 5 years than in the whole history of astronomy, and much more accurately polished than they were in the past. But this is not the only challenge: large convex secondary mirrors – almost as large as the primary mirrors of the current generation of 8 m telescopes, a large adaptive mirror that is needed to correct the atmospheric turbulence, and large tip – tilt mirrors, are all intrinsic parts of the new generation of telescopes. Most of those mirrors will be three or four times larger than the existing ones.

In a brief survey of the history of astronomical telescopes, we will see that astronomy not only has driven the development of large optics but also of complex shapes of surface profiles, such as aspherical surfaces and was also a major driver of significant development in substrate technology, including of course glass and glass ceramics technology, deformable mirrors, active and adaptive mirrors, etc.

2 The Early Development of Astronomical Mirrors

The history of mirrors used for observing the sky dates back to four centuries ago during the years 1608–1610, when Galileo suggested replacing the lens of his refracting telescope by a concave mirror made of bronze. He observed the direct image produced by his new instrument using an eyepiece, but in order not to obstruct the direct light coming from the stars – which is still a recurring problem in modern telescopes – he had to tilt the mirror. Unfortunately, the skills of the craftsmen were not adequate to polish a mirror good enough to compete with the images produced by refractive telescopes, lenses being less sensitive than mirrors to manufacturing errors. Moreover, the necessity to tilt the mirror to avoid obstruction had also a very negative influence on the image quality due to the aberrations generated in such a telescope due to that tilted mirror.

Significant progress, too often neglected, was achieved by Marin Mersenne, a French monk who published *L'Harmonie Universelle* in 1637. In this book, Mersenne, a disciple of René Descartes, introduced several innovating concepts that can be considered as the basis of modern reflecting telescopes:

- Instead of using an eyepiece, as did Galileo who had to tilt the mirror to have easy access to the image, Mersenne introduced the revolutionary idea of a second mirror that would reflect the light coming from the first mirror. This allows one to focus the image behind the primary mirror in which a hole is drilled at the center to unblock the rays.
- Mersenne invented the afocal telescope and the beam compressor that is so useful in many multiple-mirrors telescope designs.
- Mersenne recognized also that he could correct the spherical aberration of the telescope by using nonspherical mirrors and that in the particular case of the afocal arrangement he could do this correction by using two parabolic mirrors.

– Much earlier than Laurent Cassegrain, he found the fundamental arrangement of the two-mirrors telescope combination, a concave primary mirror associated with a convex secondary mirror and discovered the telephoto effect that is so critical in reflecting telescopes, although it is obvious that he was far from having understood all the implications of that discovery.

Unfortunately, because of the harsh criticism that he encountered, especially that of René Descartes, he made no attempt to build a telescope of his own invention.

Probably being unaware of the work of Mersenne, James Gregory proposed in 1663 a different arrangement using a concave primary and a concave secondary mirror, both having elliptical shapes to correct for the spherical aberration. But once again the challenge of mirror manufacturing was beyond the available skills of the seventeenth century. Newton was the first to propose a solution to overcome both the problem of the tilted mirror in the Galileo telescope and the manufacturing challenges of the small aspherical secondary mirror. He used a flat folding mirror to bend the rays 90° apart from the incoming beam. Then in 1672, a letter of the prior of Bercé described a novel solution of a two-mirror reflecting telescope made of a concave primary mirror and a convex secondary mirror, discovered by Laurent Cassegrain, under whose name this solution is still known today, unfortunately for Marin Mersenne.

So in a few decades of the seventeenth century all forms of modern telescopes were discovered, but it was only at the beginning of the twentieth century that the technological challenges of their manufacturing could be overcome, allowing further development in telescope construction. The main reasons for that were linked to the aspherical secondary mirror and also to the poor reflecting properties of metal substrates that were the only choice available to manufacturers of mirrors.

In 1721, Hadley presented to the Royal Society of England a Newton type telescope of 150 mm aperture (F/10.3) and the images produced by the reflecting telescope were compared to the images of a refractor: although the optical quality was similar, the reflector could not deliver the same brightness of image. James Short, in the middle of the eighteenth century, built quite a large number of Gregory type telescopes and mastered the manufacturing issues of aspherical mirrors for telescopes at F/8 to F/4.

A significant milestone was achieved at the turn of the century by William Herschel who built the largest telescopes of that time with apertures of 500 and 1,220 mm. Built in 1784, the 500 mm telescope would become the instrument Herschel used for his most prestigious discoveries. But he was unable to solve new problems that appeared for the larger mirror he attempted to built, especially the compensation of the gravity deflection of the mirror.

Two major discoveries in the first half of the nineteenth century would be necessary to make further developments and credit should be given to William Parsons (Lord Rosse) and William Lassell. They both improved the composition of the alloy used to build the mirrors and the polishing techniques of the mirrors. However, to increase the diameter of the mirrors they invented two different support concepts that are still used in many modern telescopes: the whiffletree support designed by Thomas Grubb for the 1820 mm primary mirror of the Lord Rosse telescope and the astatic lever support developed by William Lassell for the 1220 mm Maltese telescope.[1] All these telescopes were based on the Newton type configuration.

The first Cassegrain telescopes were built by Thomas Grubb in 1835 and by James Nasmyth in 1845.

The switch from metallic to glass mirrors was only possible after the discovery of the chemical silvering process by Leon Foucault and his brilliant application on the 800 mm telescope that he built in 1862. Unfortunately, further developments of reflecting telescopes were delayed for several decades due to a major problem that arose during the manufacturing of the metallic primary mirror of the Melbourne telescope. This swan song of large metallic mirrors, instead of enabling the development of glass mirrors, resulted in abandoning the whole family of reflecting telescopes and for the last time in telescope history the refractors had the advantage.

At the beginning of the twentieth century, in 1901, George Ritchey built the first reflecting telescope of the modern era: using a glass substrate for the mirrors, fast primary mirror (F/3.9) (which allowed the use of photographic plates at the primary focus), a Cassegrain optical configuration, an open tube and a German equatorial mount. Credit must be given to Ritchey, as an optical and mechanical manufacturer, so well illustrated by the 1200 mm and the 2500 mm telescopes of Mount Wilson. Of course he opened the path to aplanatic configurations and for the first time in history, for his 1 m aplanatic telescope, he used a mirror that was not parabolic: it was the first Ritchey – Chretien configuration now so commonly used in modern telescopes.

3 The Current Generation of Telescopes

The current generation of telescopes in operation since the 1980s can be characterized by the use of glassy material with low or ultra low (near zero) coefficient of thermal expansion, by the use of actively controlled primary mirrors, and for some of the most recent ones by the use of a segmented pupil.

Another trend is the use of faster mirrors, especially for the primary mirror, leading to a shorter telescope tube: in 1789 the Herschel telescope had a primary mirror with an f-number of 10, the primary mirror of the Ritchey telescope at Mount Wilson had an f-number of 5, the primary mirror of the ESO 3.6 m La Silla had an

---

[1] This telescope is a major landmark in telescope history: for the first time an equatorial mount was used which was a major improvement compared to the mounts used by Herschel and Rosse, and large holes were drilled in the telescope tube to insure better ventilation and avoid air stratification so detrimental to the image quality.

f-number of 3, the primary mirrors of the ESO Very Large Telescope (VLT) had an f-number of 1.8 and the primary mirrors of the Large Binocular Telescope (LBT) had an f-number of 1.14.

Most of those telescopes use a Ritchey-Chretien configuration and the primary mirror has a hyperbolic profile close to a parabolic shape. The "asphericity" measured as the maximum departure from the actual profile with respect to the best fit sphere varies as the inverse of the cube of the F-number and can be as large as several hundreds of microns for the most recent mirrors. Polishing such mirrors with an accuracy never reached before could only be contemplated with the simultaneous development of polishing and testing techniques such as computer controlled polishing, active and deformable polishing tools, use of laser and interferometric measurements etc. The correction loops of polishing are driven by accurate measurements of the mirror surface errors: since the invention of the Foucault test, which was one of the major improvements in nineteenth century metrology techniques to the most recent progress in computer generated holographic null tests, everything tends to prove that the ultimate performance of any optic, not only in the field of astronomy, are limited by the metrology and that progress in metrology leads to progress of the optical quality of the mirrors that are produced.

The idea of using an active mirror is not new; it dates back to André Couder who was first to have the idea of correcting the astigmatism of a mirror by introducing a set of forces in 1931. The New Technology Telescope (NTT) of ESO was the first telescope designed with an active mirror and all telescopes built since then have been based on the same principle. The introduction of an active support enables the reduction of the thickness of the mirror for a given diameter and thus allows an increase of the mirror aspect ratio. For a passive mirror the aspect ratio is about 6 and was increased to about 15 for the NTT and to about 50 for the VLT. Reducing the thickness of the mirror was a key parameter to allow the casting of large monolithic blanks (up to 8 m) and at the same time it allows one to reduce the thermal mass of the mirror in order to improve the local seeing of the telescope. But reducing the thickness of the mirror had a detrimental effect on the polishing process: the problem of the high flexibility of the mirror had to be solved and the polishing errors could only be partially compensated by an active correction capability. To ease the polishing of those mirrors, some active correction capability was devoted to the correction of polishing error, but this was mainly limited to the first elastic modes of the mirror corresponding to the low-order aberration correction, such as astigmatism and triangular astigmatism. Other modes, especially axial symmetry aberrations such as spherical aberration, require a significant amount of force to be corrected and therefore the allowed corrections are quite limited in amplitude. The way a thin mirror is supported, both during polishing and testing, is much more critical than for a thick mirror: issues such as gravity effect of the mirror (especially the local effect of the mirror sagging between support points) and local print-through of the support points could only be solved by dramatic improvements in polishing and testing techniques.

At the same time the way of defining polishing specifications had also to take into consideration active corrections capabilities and the limitations imposed by atmospheric turbulence on optical image quality. For example in the case of the VLT primary mirrors, the aberrations generated by the polishing errors should not impact the image quality in the best possible seeing conditions (about 0.2 arc second). A new type of specification was introduced, the Central Intensity Ratio (CIR) which is defined as the ratio between the Strehl coefficient of the image provided by the telescope, including the loss of image quality due to any polishing errors in the primary mirror, and the Strehl coefficient of the perfect telescope (with no polishing errors but with perturbations induced by the atmosphere). In computation of the image quality, the first sixteen modes of the mirror deformation could be used to compensate for mirror polishing errors (of course in a limited range) and all errors with a spatial period higher than 25 mm had to be included in the error map. The specification was a maximum allowed loss of intensity of 18%.

In the case of the Gemini telescopes, the requirement included a traditional specification of encircled energy (more than 80% in 0.1 arc second at a wavelength of 500 nm and for errors having a spatial period larger than 8 mm) and a specification on the intensity of the satellite images that would be formed outside of the central core of the image.

It is therefore clear that the new mirror-supporting technology resulted in a new method of specifying polishing errors that were not only different from the previous ones but also more stringent for the polishers. They had to improve their polishing techniques and measurement accuracy.

In parallel to the large monolithic mirror telescope, a few precursors explored the possibility of using segmented mirrors, some using spherical segments to ease the polishing process (the Hobby Eberly Telescope is the first example of this kind of telescope) and some others using aspherical primary mirrors in the classical Ritchey – Chrétien configuration (Keck I and Keck II telescopes). They are the pathfinders for the next generation of extremely large telescopes.

4 Future Giant Telescopes

4.1 New Needs

In order to better understand the needs and requirements of the future generation of extremely large telescopes that will be in operation at the end of this decade, we have first to understand better what are the open scientific questions such telescopes are supposed to find answers to. In the domain of fundamental physics, are the laws of nature universal or not? How can we better observe the earliest epoch of the universe? How do black holes shape and influence the universe? How do galaxies form and evolve? And closer to us are the exoplanets and the possibility to observe the first sign of life on other planets than Earth.

One of the major differences between the current generation of telescopes and the future generation of giant telescopes is the need for high contrast images, which will be required particularly in the field of exoplanet detection. In order to obtain such high contrast images, the suppression of stray light of any kind must be achieved at a level never obtained so far in optical instruments of such a nature as a giant telescope. In order to detect a planet like a warm Jupiter, the image contrast must be about 10^{-6}, but this level should be about 10^{-10} in order to detect an Earth-like planet at an angle of 1 arc second from its sun.

Such an extreme level contrast can only be obtained if the telescope has the capability of adaptive correction (Adaptive Optics – AO) but extreme adaptive optics (XAO) has to be also implemented in the telescope instruments. As will be shown, this imposes new and more stringent requirements on the mirror specifications. But despite all these correction capabilities, not all aberrations can be corrected or compensated for: in general the aberrations that produce a differential variation of the amplitude of the wavefront cannot be compensated (in opposition to the aberrations that induce phase variation). There are quite a number of aberrations of this nature, but in a segmented telescope only a few have to be taken into account: the type of segmentation of the mirror, the gap between segments, the light obstruction by the secondary mirror and the variation of reflectivity from segment to segment are the most critical. On the other hand the corrections are very efficient for phase errors produced by segment misalignment (piston and tip-tilt errors), mismatch in the segments radius of curvature and the manufacturing errors of the segments themselves. In order to achieve the highest possible level of contrast, several techniques will have to be combined including extreme adaptive optic and image analysis. But first of all the mirror, and especially the primary mirror segments, will need to be produced with requirements never achieved so far for such large optics, especially in a spatial frequency domain where the correction is either not applicable or not very efficient.

4.2 Much More Severe Requirements

The demand of high contrast imaging is the key element that explains the need for tuning the optical requirement as a function of the spatial frequency of the aberration. Whatever method is used to specify these requirements, the high spatial frequency errors have to be as small as possible; more specifically those that are beyond the correction limit of all techniques described in the previous section. Most of the developments made recently in the context of the demonstration phase of extremely large telescope projects, and related to mirror manufacturing, have been devoted to demonstrating that *very smooth surfaces free of high frequency errors can be achieved not only at a prototype level but also in a serial production process at an affordable cost.*

What makes this goal quite difficult to achieve is that the segments of the primary mirror of a large telescope are particularly sensitive to high frequency errors as shown hereunder:

- In order to reduce the overall size of the telescope structure, dome and building, the f-number of the primary mirror has to be small; this is why all future giant telescopes will have a primary mirror with an f-number around f/1 or even slightly less. All current designs are based on a pure Ritchey – Chrétien configuration (either Cassegrain type with a convex secondary mirror or Gregory type with a concave secondary) or a modified Ritchey – Chrétien type (such as the 5-mirror configuration adopted for the European Extremely Large Telescope – E-ELT project). But unlike the 100 m OWL – Overwhelmingly Large Telescope – ESO project and adopted by some of the current 10 m class telescopes such as HET and SALT (see Chap. 12) which are based on spherical primary mirror segments, all future generation giant telescopes will have aspherical primary mirror segments. The main reason for that choice is linked to the difficulty of correcting the huge spherical aberration introduced by a spherical primary mirror, although the OWL design has shown some nice possible optical configurations. Therefore the segments will have to be aspheres and the shorter the F-number the more difficult they will be to manufacture.
- The aspherical departure of a segment is generally the limiting factor of the level of high frequency contents that one can ultimately achieve, supposing that one does not have any limitation in the metrology of the mirror and no limitation in the measurement inaccuracy and noise in the frequency domain that one needs to correct – this is of course generally not the case! The most significant criteria for assessing the polishing complexity are the local slope differences between the mirror's aspherical profile and the closest possible spherical surface. The rationale behind this criterion is the assumption that a spherical surface would be the smoothest surface that can be polished and that the slope difference would give the maximum area on which the surface can be considered to be a sphere and therefore give the size of the polishing tool that can be use in a "small tool polishing process". Based on this criterion it is easy to understand why some advocate the Stress Mirror Polishing (SMP) technique where the aspherical segment is polished as if it were spherical (see hereunder) while others use deformable polishing tools to be compliant with the aspherical surface on an area larger than with a classical rigid tool. Although this criterion is still useful to give a rough estimate of the surface complexity, it does not take into account the recent progress made by "small tool" polishing techniques such as MRF (Magnetorheological Finishing) or other methods. Nevertheless, whichever polishing technique is to be used, the aspherical departure has to be maintained in a range that would allow achievement of very smooth surface while having an f/1 primary mirror; this constrains the segment size (there are of course other criteria in such a choice, such as the cost of the substrate, the handling and transportation issues, the coating facility, etc.. . .). This partly explains the choice of the segments of the E-ELT and the TMT projects that have been limited to 1.4 m leading to a relatively modest aspherical departure of 200 µm for the outermost segments.
- Segmentation creates new problems for polishing. The first one is that the segments no longer have axial symmetry and that the aspherical profile is

dominated by an astigmatism shape. Those types of surfaces are obviously more difficult to produce than the more classical axisymmetric optics. The other issue is that all segments are no longer identical and have slightly different radius of curvature and aspherical profile. In the currently selected paving geometry using hexagonal segments, there is only a six-fold symmetry and the same segment appears only six times in the primary mirror. Other configurations could have more symmetry axes and therefore a larger number of identical segments, but other considerations linked to segmentation, such as for example its influence on the diffraction pattern, are more important and the hexagonal paving has finally been selected by both the E-ELT and the TMT. From a manufacturing point of view this implies that 164 different types of segments have to be produced for the E-ELT, which is of course from a serial production point of view, far from being optimal.

- Segmentation induces another issue that is the mirror edge effect. Most polishing techniques induce a surface error at the edge of the mirror which is generated by the non-continuity of the surface and the boundary conditions of the pressure field applied to the surface under polishing. This edge effect is localized in the last millimeters of the mirror surface and this is why most optical components have a useful clear aperture slightly smaller than the mechanical surface of the optics. Unfortunately this is not possible in a segmented configuration where the whole surface is used to collect light and the optical prescription has to be met over the full surface up to the very edge. This is needed in order to minimize the scattered light produced by the segment's edge. One could oversize the segment during the polishing process and cut it to the hexagonal shape once the optical prescription is met. But this introduces a high risk process step at the end of the polishing when the segment has its most added value, risk not only of breakage but also of chipping and distorting the surface during cutting and thus losing the surface accuracy. It will be necessary to be able to correct the optical surface once the segment is cut to its final hexagonal shape and this implies using a polishing or a figuring technique that does not generate any edge effect.

For example the high frequency errors of the E-ELT segments have to be lower than 7 nm RMS in average over all the 1148 segments including the edge effect. This gives an excellent idea of the challenge that will face the optical industry in producing these mirrors.

4.3 Dedicated Polishing Techniques

Different polishing techniques are contemplated for the production of the primary mirror segments and one of them – the Stressed Mirror Polishing or SMP – has been advocated by Jerry Nelson from the Keck Observatory and implemented when figuring and polishing the segments of the Keck I and Keck II telescopes. This technique is nowadays in competition with small tool computer controlled polishing

(CCP) or deformable tool polishing for the early phase of segment production and the open question is whether SMP is more accurate and more cost effective than the other techniques. As far as the process used for the final figuring or polishing is concerned, there seems to be a general agreement that Ion Beam Figuring (IBF) would be the best candidate.

The SMP combines two advantages that make it a serious candidate for segment polishing. It is based on a spherical mirror polishing technique using a full size polishing tool. It is well known that polishing with a large tool is the way to produce a surface free of high frequency defects or at least with defects of very low amplitude. In the case of a spherical mirror, the tool can be as large as the surface to be polished itself: unlike CCP the tool always remains in contact with the surface to be polished and this explains the smoothness of the surface. Using a full size tool is also more cost effective than using a small tool, since the removal rate is proportional to the surface of the tool. This is compensated by the higher removal rates that can be obtained by some CCP techniques and by their better convergence rate, thus using less iteration to obtain the final mirror figure.

Unfortunately the full size polishing tool cannot be used in the case of an off-axis aspherical segment and therefore one has to use a "trick" to polish the segment as if it were a sphere. The idea is to bend the segment so that it is deformed to the opposite aspherical shape that the segment shall have, polish it spherical so that, after releasing the deformation the mirror will have the exact required aspherical profile. For using this technique several constraints have to be taken into account. First the segment must not be too stiff so that the required deformation can be produced without overstressing the glass and exceeding the allowable load limits. The drawback is that a thinner blank requires a more complex support system to compensate for gravity deflection during observations. The same argument is valid for the amount of aspherical departure of the segments: the selected value (about 200 µm maximum for the outer segment) resulting from the mirror f-number and the segment size is about the maximum a segment can be bent without again exceeding the load limits. But even with these values, a complex behavior of the glass – the delayed elastic effects – have to be taken into account in the polishing strategy.

The deformation has to be applied by a complex bending fixture that is self-contained, portable and stable to better than 1 µm over the working temperature range. The actuation is generally achieved by hydraulic support coupled with a lead-screw and spring loads. A feedback is provided from real time load and temperature sensors and an on-board real time computer closes the control loop. The complexity of the bending fixture is increased by the fact that it must be able to produce more than a hundred different types of aspherical shape.

The applied moment and forces to distort the optical surface are predicted by a Finite Element Model (FEM). The model shows that this deformation can only be obtained on a circular segment that will need to be hexed after polishing. This geometry is also required to avoid the edge effect mentioned earlier. Another limitation in the segment geometry is given by the blind central hole used to attach the lateral support interface to the mirror. This hole needs to be machined after the SMP has been completed. Finally the model shows better accuracy if the segment has a meniscus shape (curved back surface) whereas the preferred geometry for mirror support and integration would recommend a flat back.

There are two limitations on the accuracy of the surface that can be achieved. The first one is that, despite the complexity of the bending fixture, only low order deformation can be applied (the higher orders are either two small to be implemented within the range of accuracy that is possible to achieve and/or require too much force to remain in the load limits of the glass). Practically only astigmatism, coma and triangular astigmatism are used to bend the mirror. Moreover, the prediction of the FEM is not accurate enough and the deformation needs to be refined by actual measurement using either a large coordinate measurement machine or an in-situ measurement device directly implemented on the polishing machine. The latter is highly recommended in order to allow closed-loop control of the mirror bending.

The results so far obtained on demonstration prototype segments are a surface error around 800 nm RMS before hexing the segment. If IBF is to be used for achieving the final mirror requirement (this will be discussed in the next section), this value will probably need to be improved by a factor of about 2, otherwise the IBF runs will probably be too long to be cost effective. In order to make a comparison, the segments after CCP have a surface error between 100 nm and 200 nm, surface errors which are similar to the amount of surface deformation that has been measured by the segment hexing. But the most important question is the comparison between the frequency content of a segment processed by SMP and the high frequency content of a segment processed by CCP after the final IBF corrections have been applied.

4.4 Ion Beam Figuring Final Corrections

Once the segments have been polished by SMP or CCP and cut to the final hexagonal shape, corrections will need to be applied to the optical surface in order to achieve the final requirements. At that stage, any technique that will be used has to control the edge effect and the only one that has proven to be able to do so is the IBF. This process has been successfully used on previous projects using segmented mirrors (Keck and GTC – Gran Telescopio Canarias telescopes) and it has been demonstrated that the IBF has the capabilities to figure segments without generating edge effects. Unfortunately IBF is not able to produce a segment from scratch: it is not a polishing technique (this means that the segment has first to be polished to the correct texture, micro roughness and cosmetic defect level before going to IBF) and the removal rate is low compared to other methods. This is a key element explaining why the surface figure has to be close to the final requirement before starting IBF if the correction time has to be maintained below an acceptable limit.

Unlike the polishing method presented in the previous sections which are based on a chemical - mechanical abrasion process, the IBF can be considered as a non-contact figuring process: the ion gun projects a beam of particles on the optical surface and glass material is removed by the collision of these particles.

This technique was introduced in the optical industry about 20 years ago and has been used for producing the most stringent optical components. The major difference between IBF and polishing techniques like CCP is a better controlled "tool profile" resulting in a stable and well-controlled removal rate, and the end result is that the convergence rate, defined by the ratio of the actual measured surface map after correction and the hit map, is typically around 80% whereas the same ratio would be around 50% for conventional polishing techniques. In other words the IBF needs less time to correct a surface figure error and does it more accurately.

The other main advantage of IBF versus CCP is the absence of edge effects. This can be easily understood by the fact that unlike all other techniques, the IBF does not apply a mechanical pressure on the surface and therefore there is no singular effect at the edge. This is of course the most important reason for selecting IBF for such a finishing stage.

4.5 Results on Prototype Segments

In the frame of a development program of the E-ELT sponsored by ESO, Sagem – Reosc has produced several prototype segments using CCP polishing techniques and IBF for final correction. The segments were produced and first tested on a so-called "metrology support" having the same interface as the mirror in its final unit support.

The overall surface error without any active support compensation on the mirror was 26 nm RMS, to be compared to a maximum allowable error of 50 nm RMS and an average error of 25 nm RMS (the average value will be applicable for the serial production of the 1148 segments and corresponds to the mean result of all segments produced so far). Simulations indicate that after applying active correction (this cannot be done on the metrology support), the high frequency contents will be 10 nm RMS to be compared to a maximum allowable error of 15 nm RMS and an average error of 7.5 nm RMS. The result obtained on the first prototype segment already meets the maximum allowable error and the mean overall error, but improvements have been implemented on the other prototype segment to meet also the mean high frequency error requirement. It should also be noted that these surface errors include the overall surface up to the edge of the segment and therefore any edge error is included.

For the remaining prototype segments, there will be an additional step, the integration of the mirror support, a 27 points whiffletree support system. This integration will be performed after the segments have been polished by CCP and hexed and prior to IBF. In that case the integration errors resulting in distortion of the optical surface (it should be kept in mind that the segments are relatively thin and sensitive to integration errors) can be corrected and the integrated segment will be delivered in the same conditions as it will be used on site. This will also allow physical correction of low order aberrations by an active support and not only a correction by software.

5 Conclusions

The future extremely large telescopes projects will be driven by very ambitious scientific goals and some of them will set the requirement of the primary mirror segments and the other mirrors of the telescope to a level never achieved so far. Development works on prototype segments have been conducted in the optical industry in Europe and in the United States, using different approaches, and the most promising ones have been described in this chapter.

The results obtained so far on the first prototype meet most of the ESO's requirements for the E-ELT M1 segments and ongoing work also includes a demonstration that the mirror requirements can be met at the mirror assembly level.

Achieving this goal is the one of the keys to the success of future generation telescopes.

Chapter 14
Virtual Observations

Mark G. Allen

Abstract The Virtual Observatory is a framework for interoperable and efficient access to world wide astronomical data and services. This review describes the global nature of the VObs initiative, and then briefly outlines the architecture of the core interoperability standards. Data Centres provide the content of the VObs via publishing their data and services using the standards, and a set of tools allow these resources to be searched, combined and analysed. Common themes for using the VObs are searching by object name or coordinates, combining multi-wavelength data, and cross-matching. Scientific use of the system is ramping up, and some early results are described.

Keywords Astronomical data bases • Virtual observatory tools

1 Introduction

The images, spectra, light curves, maps, and a multiplicity of other kinds of data collected by telescopes, and stored in databases and archives across the world, together, form a kind of 'digital sky' that can be 'observed' with software 'instruments' and combined to enable multi-wavelength and multi-epoch analyses of astronomical sources. This concept of a Virtual Observatory (VObs) for Astronomy emerged some 10 years ago to address the burgeoning data collection rates of current and future telescopes, and as a way of opening up new scientific possibilities with interoperable and high performance access to astronomical data and services.

Since that time the VObs projects have made much progress on the definition of the underlying framework of a core set of standards for making astronomy data

---

M.G. Allen (✉)
Observatoire de Strasbourg UMR 7550, Strasbourg, France
e-mail: mark.allen@astro.unistra.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_14,

and services available in a uniform way. Many Data Centres have taken up VObs standards, new data discovery and visualisation software provide a set of tools for astronomers, and the list of VObs enabled papers is growing.

The development of the VObs is part of a sea change in the way the internet is used for science, computing and communication. The VObs is in-step with, and in some areas is leading the field of e-science, 'the use of distributed networks for computationally intensive science or use of large data volumes'. The emerging fields of 'X-Informatics', following Bio-Informatics, promises to build further on the computational resources, data access and interoperability provided by these e-infrastructures bringing 'data mining' techniques, and statistical methodologies for extraction of knowledge from large data sets. In this context the VObs is the e-science initiative of Astronomy, and Astro-Informatics is the 'data-science' for Astronomy.

In a wider view, the VObs also resides within the fast moving environment of the World Wide Web. In a very general sense we see the web as connecting computers together, with many contemporary developments such as cloud computing and the semantic web having synergies with, and benefits for e-science. The web also connects people and the wave of social networking and sharing of data (photos, maps, blogs etc.) on the web has altered the way scientists work, with interesting new possibilities for collaboration, and has allowed the development of VObs portals, not just for the scientists, but to bring the possibility of viewing the digital sky to education and public outreach audiences.

In this short review I provide an updated view of the VObs in terms of the scientific and technical issues driving its development. I outline the role of the Data Centres which provide the data 'content' of the system, and how improved data access is leading to new innovations in scientific tools for finding and using the data. I describe some of the early scientific results enabled by the VObs, covering a wide range of astrophysics.

2 VObs: A Global Initative

Data, everyone has got it, and there's a lot more coming! The archives of the European Southern Observatories (ESO) for example, currently hold some 100s of TB of images and spectra, with wide field imaging about to greatly increase this. The archives of space missions held by the European Space Agency comprise some 70+ TB of data. Future telescopes will take us into a completely new level of data, as we move toward an era dominated by wide and deep surveys. The LSST plans to start imaging the entire southern sky every three nights leading to a data rate of some 6.5 PB per year. The SKA pathfinder experiments expect to keep 70 PB per year, with some 100–1,000 times this for an operational SKA. The SKA data rates present serious operational challenges, but otherwise technical solutions for managing such data within archives already exist, employing lots of disks and air conditioners. The more serious challenge is delivering, or providing access to the

Fig. 14.1 The members of the International Virtual Observatory Alliance – IVOA

data to astronomers. To address the need to perform analysis of PB data sets, the conventional wisdom is to geographically co-locate the computing resources with the data.

In addition to data access, it is also vitally important that data from different telescopes be compatible. Today's astronomy research requires a multi-wavelength approach, using information from across the electromagnetic spectrum to build up a physical picture of astrophysical objects and processes. More and more scientific papers use data from multiple telescopes, so that interoperability between the services providing the data, the tools for using the data, and the formats of the data, are all necessary in order to combine data in scientifically meaningful ways. Interoperability standards provide the key for astronomy data and services to work together, and it is clear that this is only possible with globally agreed standards.

Upward spiralling data rates, and the interoperability of astronomy data across the spectrum were the initial motivators for the VObs, and they remain so today. These issues are faced by astronomy data centres across the world, and VObs initiatives have come together in an alliance to co-operate on the development of common standards, and to share best practices. The International Virtual Observatory Alliance (IVOA) was formed in June 2002 with a mission to "facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems, and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory". The ring of logos in Fig. 14.1 shows the international members of the IVOA.

The concept of a Virtual Observatory has been endorsed by the astronomy community via the ASTRONET 'Strategic Plan for European Astronomy' [2]. The ASTRONET Infrastructure Roadmap makes a number of recommendations for the VO compliance of archives and tools, and to prepare for large surveys and multiwavelength astronomy. Virtual Observatory was also a top recommendation of the 2000 US Decadal Review, with the US Virtual Astronomical Observatory now funded, and the 2010 Decadal Review recognising the potential to substantially enhance the collective value of archival data sets.

2.1 Components of the VObs

The standards that are developed by IVOA form the framework of the VObs. Like the standards that make the internet work, VObs standards are basically a set of protocols for how to expose different kinds of data and services, so that tools that understand those rules can find that data, understand the formats, and use that data for science. As such the VObs is a distributed and open system, whereby data centres are fully responsible for their holdings and can control the way their data is accessed from the VObs. The IVOA standards are domain specific to Astronomy, and all the participants in IVOA are also astronomy data and service providers, and scientists, ensuring that the standardisation effort matches the real needs of the astronomy community.

The VObs standards developed by the IVOA are organised within an overall conceptual architecture. The architecture diagrams in Fig. 14.2 show two views of this. The *Level 0* view highlights the core functions of the VObs. At the base of the diagram is the RESOURCE LAYER which comprises the images, catalogues, spectra, simulations and other data that are stored and made accessible by data centres. The USER LAYER at the top represents the astronomers who access the data for doing their science. The VO CORE concerns all the necessary standards required for the Astronomers to use the resources in a transparent and interoperable manner.

The *Level 1* view of the architecture indicates how the VObs components fit within this structure. REGISTRY provides the function of finding resources, and DATA ACCESS PROTOCOLS allow data to be obtained from data providers. The VO CORE components that make the system interoperable are the query language, the standardised semantic descriptions, the detailed data models describing the data, and the format standards for the data. The details of the RESOURCE LAYER concern the data and its descriptive metadata, this is the basic content of the VObs, and its standardisation is key to making the system work. The USER LAYER involves the different ways the users will interact with the VObs through browser based applications, standalone tools and also programatic access to the VObs.

Some of these components are described in more detail below.

- Registry
  A registry is a kind of yellow pages of the VObs, providing the ability to search and find resources. Registries provide standardised descriptions of the data and

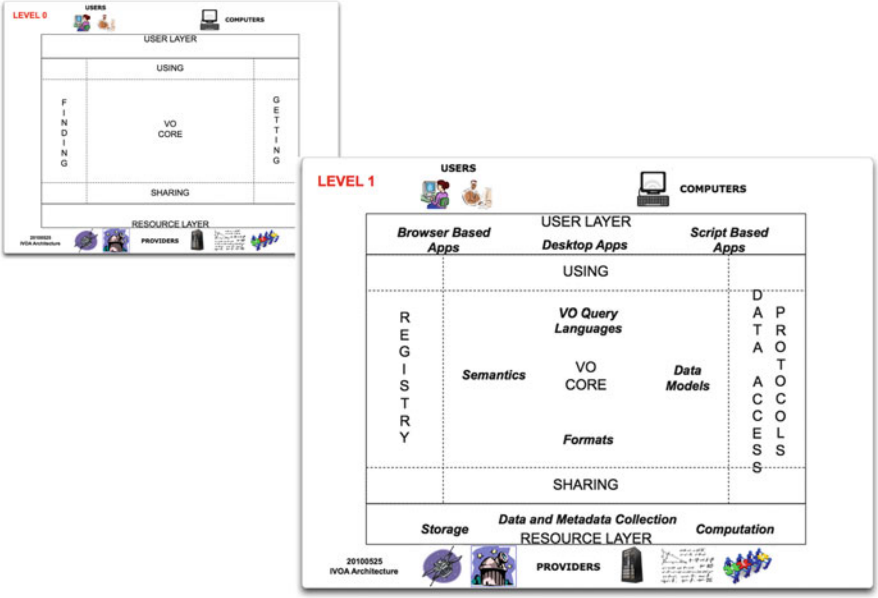


Fig. 14.2 IVOA architecture diagrams

services available in the VObs. Data providers register their services by providing the description in a standard way, and then tools that use registries can then easily find that resource and know how to use it. It is not a centralised system of a single registry, but rather there are many registries that can intercommunicate.

- Data Access Protocols
  The ‘getting’ of data in the VObs is facilitated by standard data access protocols. Data Centres ‘publish’ their data to the VO via standard access layers that are built on-top of their archives and databases. These expose data such as images, or spectra to the VO in a way that can be understood by applications and other services.
- VO Query Language
  A standard way of expressing queries within the VO. Such as a query to an archive for available data, or for selection of targets from catalogue services.
- Data Models
  Describes the data in terms of general abstract concepts. For example images are characterised by the essential information about their axes, the pixel size and sampling, exposure time etc. Data models allows complex data to be mapped to standard concepts in order to enable basic understanding of what the data is and how it may be used.
- Semantics
  A hierarchical set of descriptive words that can be combined in order to express a general, ‘fuzzy’ meaning (Uniform Content Descriptors). This allows columns of

tables, or axes of data sets to be identified as similar things. This aids searching for information in the VO, such as a 'B magnitude' whose column name would otherwise have hundreds of variations.

- Interoperable Applications
  The VO standards for sending 'messages' between software tools allows the various tools to 'talk' to each other, for example 'load this image', or 'highlight these rows'. This allows applications to work together in an interactive way that is much more efficient that saving and re-loading files into multiple applications.

The IVOA Working Groups also use a *Level 3* view of the architecture, where the structure is populated with the various IVOA standards showing how they fit together. This detailed view is not necessary for use of the system, rather the *Level 3* is used as a way of managing the various standards.

3 The Data Centres – Content for the VObs

Astronomy Data Centres are an essential component of the VO as they provide the scientific content, and a wide range of services. In addition to observationally based services, there are also Data Centres that provide theoretical models and results of simulations, reference and bibliographic services, and services centred on a given scientific theme.

A census of Astronomy Data Centres in Europe [1] collected information about Data Centres in a uniform way, characterizing them, and providing a snapshot of the Data Centre community. The census shows a diverse community, covering all scientific areas of astronomy, and with a large variety of different approaches to delivery of data and services. A number of large Data Centres host multiple archives and services, but a significant number of small and less well resourced Data Centres provide important often specialised archives, services and tools. A wide range of data types are provided by the Data Centres, the majority of these can however be described as images, catalogues and spectra. Current interfaces to archives, services and tools are predominantly Web-interfaces.

The census shows a high level of interest in Virtual Observatory methods, with many Data Centres already making some use of VO access protocols, and indicating intent to implement IVOA standards. In the following section we describe how one data centre, the CDS interfaces its services with the VObs, and the importance of curation of astronomy resources.

3.1 The Centre de Données de Strasbourg

The Centre de Données de Strasbourg (CDS) is a data centre that has been making information on astronomical objects available to the astronomy community since

1972 well before the internet era. Starting with collections of stellar data for the study of galactic structure, the CDS now provides information covering all areas of astronomy collated from the literature, surveys and telescope observing logs [11]. The data are curated, critically evaluated and are provided to the community via a range value-added services. In a sense the CDS itself is a kind of early, self contained Virtual Observatory. Today many of these services are now provided through VObs interfaces alongside the dedicated CDS services, and the CDS is deeply involved with development of the VObs. The three services that form the CDS hub are Simbad, Vizier and Aladin.

Simbad is a reference database for astronomical objects [20,21]. It contains identifications, basic data, and bibliographic information for some five million astronomical objects. The database contents are built by scanning and extracting information on astronomical object identifications in the literature by a team of professional 'documentalists' and scientists. The cross-identifications of objects provides the basis for 'name resolving' services provided by SIMBAD, so that any object can be 'looked up' by name or coordinate, or object type. Interfaced to the VObs as a web service with VOTable output, SIMBAD name resolving is employed by many VObs tools as the first step of a query for data on a given astronomical object. Links to the SAO/NASA Astrophysics Data System (ADS) provide access to the original papers via the familiar ADS interface, and conversely the 'SIMBAD Objects' in ADS provides SIMBAD data on the astronomical objects cited in a given paper.

Vizier is a database of thousands of astronomical catalogues [15]. The collection is built upon the major astronomy catalogues built from large surveys, and also those systematically obtained from journal publications. The catalogues are integrated into the Vizier system in more or less their original form, but 'marked up' by use of a standard description. The standard descriptions of the catalogues, the metadata, allows this large set of heterogeneous information to be used together, and provides the flexibility for the catalogue data to be accessible by a wide range of user and machine interfaces. Vizier catalogues are registered as catalogue services in the VObs, allowing them to be queried by cone search queries, with links to their provenance and full metadata, and output in VOTable among many other possibilities. The contents of Vizier are enhanced by VObs standard UCDs (Unified Content Descriptors) which are semantic descriptions of what the columns of the catalogues actually represent, for example that the data represent a sky position, or a flux measurement in particular photometric band.

Aladin is an interactive sky atlas for visualization of astronomical images superimposed by layers of catalogues and other information that can be projected onto sky coordinates. The images come from the dedicated Aladin image server, and a large number of other external image servers, as well as image servers available via the VObs. As such Aladin handles a multiplicity of data sources, and also supplies images via VObs protocols that can be queried via the image access protocol. Queries to 'All VO' from Aladin will return lists of images, but also of catalogues that can be overlaid as image planes, and also spectra.

The data the CDS publishes to the VObs is carefully curated with an emphasis on the quality of the metadata, and preserving the information about the provenance of the data. The original journal articles describing the data are directly linked to the catalogues, and conversely one may search on individual astronomical objects to find all their bibliographic references. This is done via collaborations with the journal publishers, and also though interaction with the authors where necessary. Recently the CDS added the ability for users to add extra relevant information to the system via ‘annotations’ on on Simbad objects.

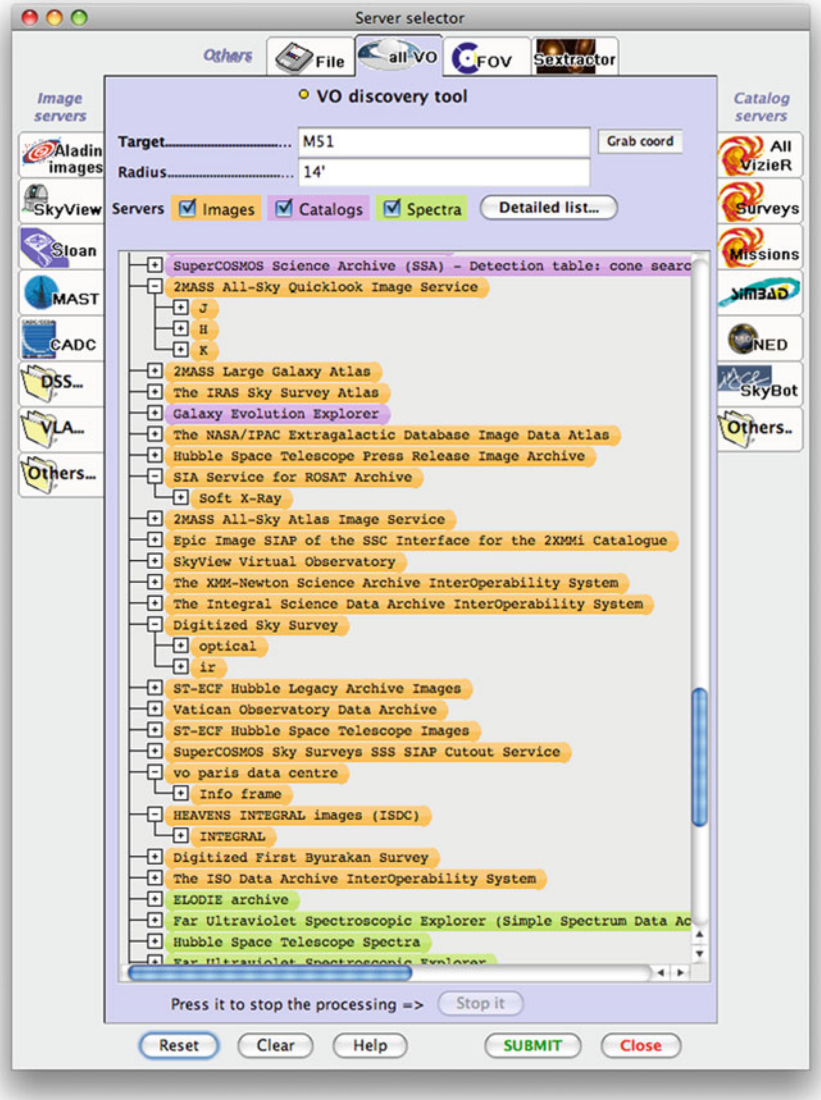
4 Science Tools and Beyond

There are many software tools that can be used to search and use data and information available in the Virtual Observatory. Some of these are featured in the IVOA newsletters [12], and national VO projects maintain lists of the various tools and their functions (e.g. Euro-VO Science Software pages [8]). The developers of these tools naturally have different approaches, some of the tools are ‘VO enabled’ versions of exiting tools, whereas some others were built with VO in mind from the outset. Also, as these new technological capabilities are explored, many of the available tools are prototypes.

To provide an overview of some of the software tools, it is useful to consider some common themes for using the VObs.

4.1 Search by Object Name or Coordinate

Searching by sky position is the most common type of query one can make to the VObs. There are a number of tools that provide this capability, for example, Aladin, US VAO Data Scope, AstroGrid Astroscope, and Topcat. These generally require the input of a sky coordinate ‘Right Ascension and Declination’, or an object name (which is then resolved to a coordinate), plus a radius to define a circular region on the sky about the position of interest, and then the request can be made to VObs resources found in the registry. If you already know which archive you want to search in, most of the tools allow you to go directly there, but if your search is more general then it is possible to query *all* available resources. For example, making a query of 14 arcmin radius around the bright galaxy M51 using Aladin queries some 70 image servers, 130 catalogue servers and 50 spectral data servers (Fig. 14.3). This initial ‘all services’ query indicates whether there is data available from each of the services, but no data has been downloaded yet! The initial query gives you the metadata such as the field of view outlines. The second part of the query actually gets you the data, or in some tools, allows you to send it somewhere. In the case of Aladin, data can be downloaded into the application and saved as files, with VOExplorer it is possible to avoid saving the file locally by using a virtual storage location (VOSpace).



○ VO discovery tool

Target........................ M51

Radius........................ 14'

Servers ☑ Images ☑ Catalogs ☑ Spectra

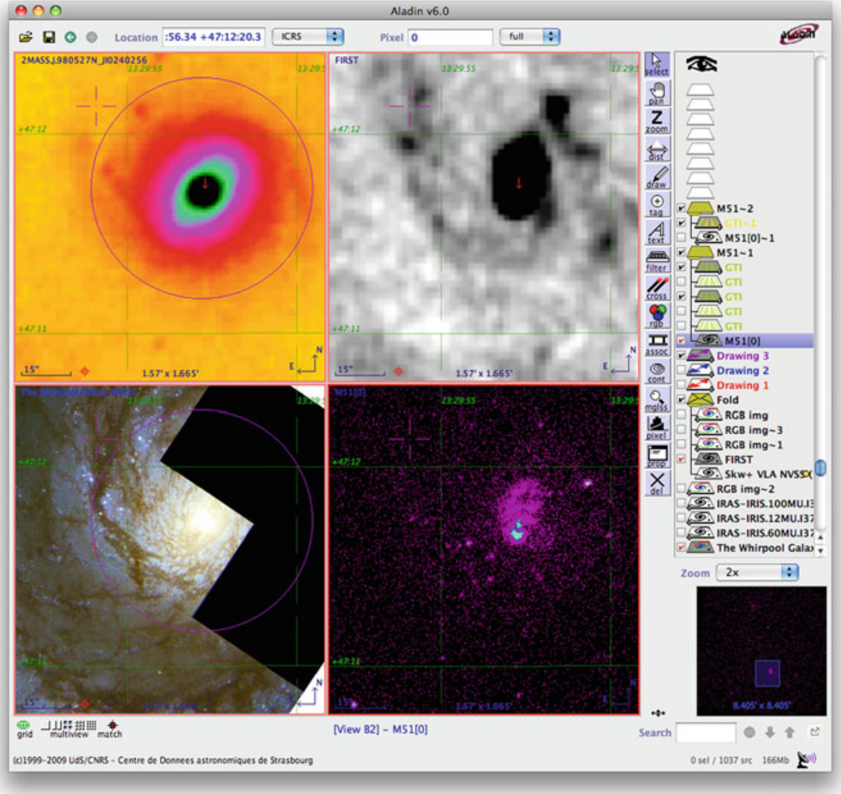
```
[+] SuperCOSMOS Science Archive (SSA) - Detection table: cone searc
[-] 2MASS All-Sky Quicklook Image Service
    [+] J
    [+] H
    [+] K
[+] 2MASS Large Galaxy Atlas
[+] The IRAS Sky Survey Atlas
[+] Galaxy Evolution Explorer
[+] The NASA/IPAC Extragalactic Database Image Data Atlas
[+] Hubble Space Telescope Press Release Image Archive
[-] SIA Service for ROSAT Archive
    [+] Soft X-Ray
[+] 2MASS All-Sky Atlas Image Service
[+] Epic Image SIAP of the SSC Interface for the 2XMMi Catalogue
[+] SkyView Virtual Observatory
[+] The XMM-Newton Science Archive InterOperability System
[+] The Integral Science Data Archive InterOperability System
[-] Digitized Sky Survey
    [+] optical
    [+] ir
[+] ST-ECF Hubble Legacy Archive Images
[+] Vatican Observatory Data Archive
[+] ST-ECF Hubble Space Telescope Images
[+] SuperCOSMOS Sky Surveys SSS SIAP Cutout Service
[-] vo paris data centre
    [+] Info frame
[-] HEAVENS INTEGRAL images (ISDC)
    [+] INTEGRAL
[+] Digitized First Byurakan Survey
[+] The ISO Data Archive InterOperability System
[+] ELODIE archive
[+] Far Ultraviolet Spectroscopic Explorer (Simple Spectrum Data Ac
[+] Hubble Space Telescope Spectra
```

Press it to stop the processing =>

Fig. 14.3 Searching the VObs for data available for the galaxy M51, using Aladin

4.2 *Searching for Multi-wavelength Data*

Having located data via a VObs search, various tools are available to visualise, combine and analyse the data. Aladin allows viewing of images, each of which may have been observed with different instruments sensitive to different wavelengths across



Location :56.34 +47:12:20.3 ICRS

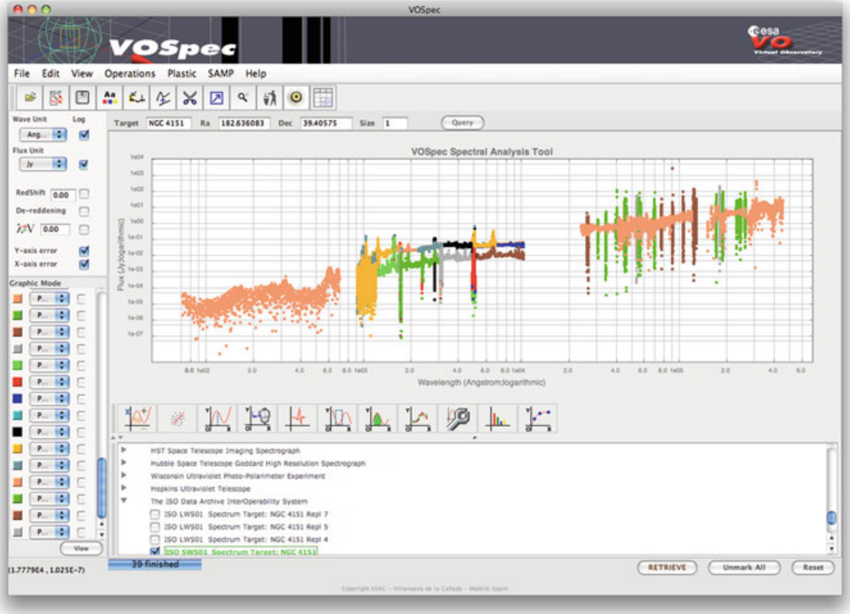
Pixel 0 full

Zoom 2x

Fig. 14.4 Multi-wavelength data showing M51 in Aladin

the electromagnetic spectrum, and with different resolutions and sky projections. The multi-panel view in Fig. 14.4 shows some of the results of a search for data of M51. The infrared, radio, optical and X-ray images shown are registered to the same projection and scale, and can be individually manipulated and combined to bring out the features of interest. Combinations, either in colour space, blinking frames, or as transparency overlays allow the set of multi wavelength images to be treated as a integrated data set. There are a number of such tools in use in astronomy with sophisticated capabilities, here the emphasis is on the ease of finding and quickly visualising the data and meta-data via the VObs. Interoperability via standard formats also allows the obtained images to be loaded into most astronomy visualisation tools.

In spectroscopic observations, the emission is dispersed as a function of wavelength, and different wavelength regimes are sensitive to different physical processes in astrophysical objects. Combining spectroscopic observations from different instruments is often difficult as many observational parameters need to be taken



Wave Unit: Ang... Log ☑

Flux Unit: Jy ☑

RedShift 0.00 ☐

De-reddening ☐

A/V 0.00 ☐

Y-axis error ☑

X-axis error ☑

Target NGC 4151 Ra 182.636083 Dec 39.40575 Size 1

- HST Space Telescope Imaging Spectrograph
- Hubble Space Telescope Goddard High Resolution Spectrograph
- Wisconsin Ultraviolet Photo-Polarimeter Experiment
- Hopkins Ultraviolet Telescope
- The ISO Data Archive InterOperability System
  - ☐ ISO LWS01 Spectrum Target: NGC 4151 Repl 7
  - ☐ ISO LWS01 Spectrum Target: NGC 4151 Repl 5
  - ☐ ISO LWS01 Spectrum Target: NGC 4151 Repl 4
  - ☑ ISO SWS01 Spectrum Target: NGC 4151

(1.7779E4 , 1.025E-7)

39 finished

Fig. 14.5 Spectra of NGC 4151 from multiple archives displayed in VOSpec

into account, and spectroscopic observations can be encoded with a wide range of units and different wavelength/frequency samplings. VObs data models aim to characterise the complexity of spectroscopic observations so that they can be combined in a scientifically meaningful manner. As a first step in this direction, it is necessary that spectra from different origins be convertible into common physical units. The VOSpec tool is able to combine such data allowing for spectra from different archives to be found, and then plotted on a common scale, as shown in Fig. 14.5 for the Seyfert Galaxy NGC 4151.

4.3 *Cross Matching*

Combination of data from different archives is a central theme of VObs science, and there are many levels of cross matching depending on the required accuracy and the details of the analysis being performed. At the most basic level we have positional cross-matching, where astronomical sources detected in an observation or survey, and listed in a catalogue with their sky position, can be cross-matched with other catalogues to find their potential counterparts that lie at the same sky position with a given accuracy. VObs tools such as Topcat and Aladin can perform positional cross matches between catalogues available in the VObs, with good performance for catalogues containing up to millions of sources. For larger catalogues there are a number of services (e.g. Vizier) that allow various types of cross-matching of lists of sources with the largest catalogues of hundreds of millions of sources. Matching of these very large catalogues against each other does however present difficulties, and is currently done via dedicated projects, for example [17] where the 2XMMi catalogue is matched with the SDSS DR6.

Positional cross matching provides a very powerful tool, but sky position alone does not guarantee that the sources detected in one catalogue physically correspond to those detected in another. Astronomers must take care of detection limits for faint sources, the positional accuracies and resolutions of the instruments used for the measurements, and the possibilities of chance alignments and the effects of source confusion. VObs cross match tools allow other parameters, such as colour, to be taken into account via filtering catalogues based on combinations of catalogue columns. Going beyond this there are detailed probabilistic methods that have been developed. Many ideas and requirements for cross matching and combination of multiwavelength data are described in the proceedings of the Euro-VO Workshop on 'Multi-wavelength Astronomy and the Virtual Observatory' [3]. Furthermore, [5] have presented a general probabilistic formalism for cross-identifying astronomical point sources in multiple observations using a Bayesian approach. In [4] the applications of Data Mining and machine learning in astronomy are considered in the VObs context.

4.4 Browsing All 4π of the Sky

From one image in a display tool, to multiple images in a stack of panels, to the display of multi-resoloution maps of the whole sky – the visualisation of astronomy images has greatly improved since the days of the first digital images. Wide field instruments, large area surveys, and professional astronomy image mosaic software (e.g. Montage [14]) have driven aspects of astronomy visualisation. Moreover full sky visualisation has undergone a revolution with Sky in Google Earth [18] and the World Wide Telescope [13]. These visualisers, designed largely for public outreach and education provide an intuitive new way of browsing all 4π of the sky, with the ability to change resolution when zooming in. These ideas and the technologies for hierarchical segmentation of the sky have brought these capabilities into VObs tools like Aladin. Figure 14.6 shows interactive full sky maps of catalogue densities from the Vizier service, the footprint of the SDSS, and a projection of the USNOB1 catalogue.

The VObs compatibility of the World Wide Telescope allows it to interact with other VObs tools. Figure 14.7 shows an overlay of a VOTable containing Ultra Luminous X-ray sources selected via a VObs workflow detailed in one of the Euro-VO Science tutorials (http://www.euro-vo.org/pub/fc/workflows.html). Interaction with

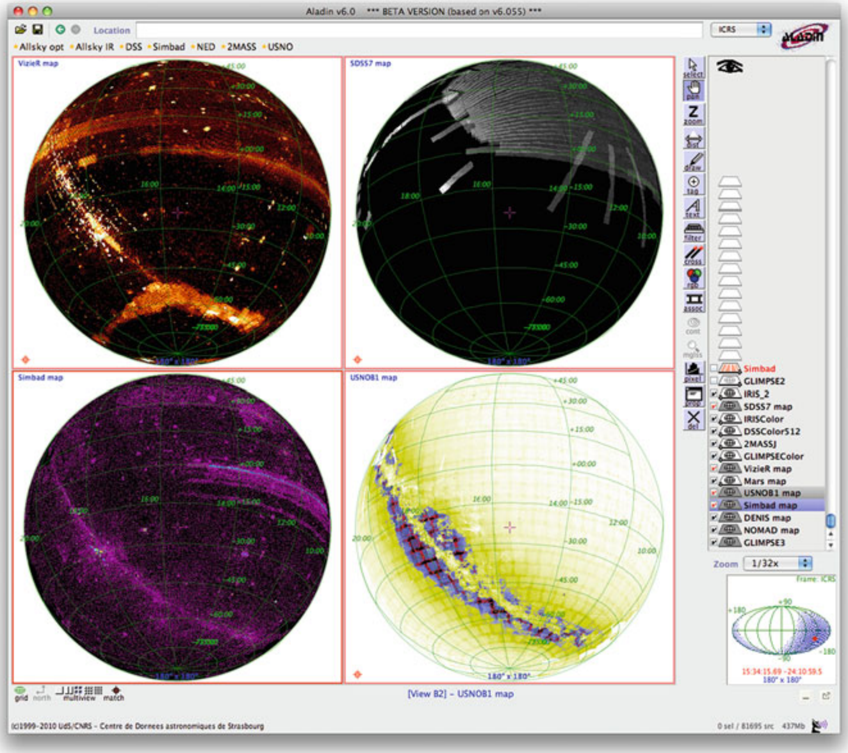


Fig. 14.6 All sky projections of catalogue and survey data in Aladin

the developers of tools beyond the professional astronomy community has the potential to bring new ways of looking at and analysing astronomy data, and is an excellent vehicle for bringing the wonders of astronomy to the wider community.

4.5 The Evolving VObs

This review has only touched on some of the common ways in which the VObs can be used. Going beyond these, there are already many more detailed ways to use the VObs tools and infrastructure. VOExplorer for example allows searching of VObs for data or services based on all the information fields that describe a given VObs resource. The VObs is not just only for observational data, there are services that provide the results of theoretical simulations such as the Millenium simulation results (http://www.g-vo.org/Millennium/Help) via an interface within Topcat. GalMer [7] provides a service for making 'virtual observations' of simulations of interacting galaxies, where many different observables can be computed and sent directly to VObs tools. Furthermore, models of stellar spectra

Look At: Sky

Imagery: SDSS: Sloan Digital Sky Survey

RA : 14h02m23s

Dec : 54:15:17

☑ Plot All

Name_tab1	D_tab1	r_arcmin	_RAJ2000_2	_DEJ2000_2	2XMMi	Flux8	e_Flux8	LEDAS	GroupID	Separatio
NGC 6876	50.9	1.6942206690850075	304.5754	-70.8575	2XMM J201818.1-705126	4.6128E-14	5.3E-15	LEDAS	1	7.181684
NGC 6876	50.9	1.6942206690850075	304.5803	-70.8592	2XMM J201819.2-705133	8.4492E-13	2.09E-14	LEDAS	1	1.463490
NGC 6876	50.9	1.6942206690850075	304.6357	-70.8553	2XMM J201832.5-705119	2.9338E-13	3.13E-14	LEDAS	1	66.95582
NGC 6876	50.9	1.6942206690850075	304.6522	-70.8535	2XMM J201836.5-705112	1.5842E-14	3.29E-15	LEDAS	1	87.35213
NGC 1313	4.13	4.666271811387211	49.4487	-66.5027	2XMM J031747.6-663009	6.5764E-13	3.6E-14	LEDAS	3	170.6890

Fig. 14.7 Microsoft World Wide Telescope with VOTable overlay

and synthetic photometry are provided by services such as the ‘Theoretical spectral Access’ services of the Spanish VO project. These can be accessed for comparison with observed data in tools like VOSpec. Other services provide information about transient events (VOevent), and there is a whole community focused on the possibilities and challenges of time domain astronomy. As a system that is only just now making the transition into an operational phase, the VObs and the methods of using it are sure to evolve with new ideas and innovations.

5 VObs Science

As the overall goal of the VObs is to enable new science, it is important that scientific use of the system be made alongside of the technical development of the infrastructure. This helps to ensure that the system meets the needs of the Astronomy community, and it provides opportunities for astronomers to have early access to new VObs capabilities. A selection of refereed science papers which make use of VObs is maintained on the Euro-VO web pages. There are currently 47 selected papers that make significant use of VObs, and many more with less stringent criteria.

Some of the earlier papers make use of the first interoperability gains provided by the VObs, for example in [16] we were able to identify optically faint, obscured active galactic nuclei (AGN) in the Great Observatories Origins Deep Survey (GOODS) survey fields. By employing publicly available X-ray and optical data and catalogues, 68 type 2 AGN candidates were identified. This work used Aladin to access the data and to filtering the various optical to X-ray catalogues, and then to perform various cross-matches of these catalogues.

In a project supported by the Euro-VO project, [10] derived disk scale lengths for $\sim$30,000 non-interacting disk galaxies in all five SDSS bands. Virtual Observatory methods and tools were used to define, retrieve, and analyse the images for this unprecedentedly large sample classified as disk/spiral galaxies in the LEDA catalogue. Cross correlation of the SDSS sample with the LEDA catalogue allowed us to investigate the variation of the scale lengths for different types of disk/spiral galaxies. The asymmetry, concentration, and central velocity dispersion as indicators of morphological type, and are able to assess how the scale length varies with respect to galaxy type.

Continuing on from the initial study Fathi [9] analyzed the disk scale length and central surface brightness for a sample of 29,955 bright disk galaxies from the Sloan Digital Sky Survey. This allowed investigation of the Freeman law of the relation between the morphology and central surface brightness of disk galaxies, based on a volume-corrected sample of galaxies in the local universe ($z < 0.3$) that is two orders of magnitudes larger than any sample previously studied and deliver statistically significant implications that provide a comprehensive test bed for future theoretical studies and numerical simulations of galaxy formation and evolution.

Chilingarian et al. [6] report their discovery of 21 compact elliptical galaxies with the Virtual Observatory. These galaxies, characterized by small sizes and high stellar densities, are thought to form through tidal stripping of massive progenitors. Only a handful of such galaxies were previously known, preventing understanding of the role played by this mechanism in galaxy evolution. The new objects were uncovered via data mining using high-resolution images and large databases, and were followed-up with spectroscopic observations and numerical simulations. This work shows that all these galaxies exhibit old metal-rich stellar populations different from those of dwarf elliptical galaxies of similar masses but similar to those of more massive early-type galaxies, supporting the tidal stripping scenario.

Another scientific project based on VObs access to catalogues of radio sources is SPECFIND ([19]). The algorithm behind this catalogue makes cross-indentifications of radio sources observed with at least three independent frequencies, with a current version of the catalogue containing 1,07,488 cross-identified objects from over 97 different radio source catalogues. SPECFIND includes a homogenization tool that is used to convert values (flux densities, wavelength/frequencies) from the original catalogues into a common system, and the combined spectral energy distributions are used to detect spectral breaks, to identify different types of radio galaxies.

References

1. M.G. Allen, M. Depretz, Census of European Astronomy data centres, Euro-VO data centre alliance 2008. Data Centre Alliance project, http://www.euro-vo.org
2. ASTRONET, http://www.astronet-eu.org
3. D. Baines, P. Osuna (ed.), Multi-wavelength astronomy and virtual observatory. in *Proceedings of the EURO-VO Workshop, held at the European Space Astronomy Centre of ESA*, Villafranca del Castillo, Spain, 1–3 Dec 2008 (The European Space Agency, Paris, 2008)
4. N.M. Ball, R.J. Brunner, Int. J. Mod. Phys. D **19**, 1049 (2010)
5. T. Budavári, A.S. Szalay, Astrophys. J. **679**, 301 (2008)
6. I. Chilingarian, V. Cayatte, Y. Revaz, S. Dodonov, D. Durand, F. Durret, A. Micol, E. Slezak, Science **326**, 1379 (2009)
7. I.V. Chilingarian, P. Di Matteo, F. Combes, A.-L. Melchior, B. Semelin, Astron. Astrophys. **518**, A61 (2010), http://galmer.obspm.fr
8. Euro-VO Science Software pages, http://www.euro-vo.org/pub/fc/software.html
9. K. Fathi, Astrophys. J. Lett. **722**, L120 (2010)
10. K. Fathi, M. Allen, T. Boch, E. Hatziminaoglou, R.F. Peletier, Mon. Not. R. Astron. Soc. **406**, 1595 (2010)
11. F. Genova, Astron. Soc. P. **376**, 145 (2007)
12. IVOA Newsletter, http://www.ivoa.net/newsletter
13. Microsoft world Wide Telescope, http://www.worldwidetelescope.org
14. Montage, http://montage.ipac.caltech.edu
15. F. Ochsenbein, P. Bauer, J. Marcout, Astron. Astrophys. **143**, 23 (2000)
16. P. Padovani, M.G. Allen, P. Rosati, N.A. Walton, Astron. Astrophys. **424**, 545 (2004)
17. F.-X. Pineau, S. Derriere, L. Michel, C. Motch, Am. Inst. Phys. Conf. Ser. **1082**, 15 (2008)
18. Sky in Google Earth, http://www.earth.google.com
19. B. Vollmer et al. Astron. Astrophys. **511**, A53 (2010)
20. M. Wenger, F. Ochsenbein, F. Bonnarel, S. Lesteven, A. Oberto, Astron. Soc. P. **351**, 662 (2006)
21. M. Wenger et al., Astron. Astrophys. **143**, 9 (2000)

Chapter 15
Doing Astronomy at a Museum

Michael Shara

Abstract Working as an astronomer-curator in a planetarium at a major collections-based museum offers opportunities and challenges very different from those encountered by colleagues at universities and observatories. It may be counterintuitive, but it's true; one can do serious astrophysics research at a museum while connecting millions of people to the wonders of modern science. In this chapter I discuss how the advent of high-resolution digital projectors has brought about "space shows" – 3D immersive journeys through the universe that transcend the earth-bound "sky shows" of the past century. Producing these multi-million dollars movies puts a curator at the intersection of modern astrophysics and Hollywood movie production. Permanent and temporary exhibitions allow curators to present both the basics of our science and cutting edge discoveries to millions of visitors. While we can't interact personally with every child who visits, we can devote significant efforts to upgrading the skills of science teachers. This has a strong multiplicative effect, and is where most of our education efforts are directed. Finally I note that hosting VIPs and donors, both at the museum and on travel is a delightful way to garner support for our science and our museum.

Keywords Planetarium • Sociology of Astronomy

1 Introduction

The invitation to write this piece is irresistible for two reasons. First, I get to make the not-so-obvious point that it's both possible and desirable to work in a museum environment if you want to carry on a very active astrophysics research program.

M. Shara (✉)
Department of Astrophysics, American Museum of Natural History, New York City, USA
e-mail: mshara@amnh.org

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_15,

Second, I get to emphasize the much more obvious point that the opportunities for engaging millions of people in astronomy and in science exist in a museum on a scale unmatched at any other sort of institution.

A large majority of the world's astronomers work in universities, national and private observatories, space agencies, planetaria and, of course, in private industry. The relative fractions of time available for pure research, and for functional work, vary from place to place... as do the inclinations of the astronomers who work there. I suspect that many young scientists for whom research is central would never considering applying to a museum or a planetarium for a permanent position. I hope that this essay prompts some of my younger colleagues to look harder at museums as potentially wonderful places to work. As a sweetener, I note to them that their curatorial colleagues may be world-class experts in the evolution of dinosaurs, the culture of the Maya, or using DNA to establish the family trees of all of the Earth's organisms. I also hope it prompts some of my more established colleagues to partner more with museums as a very cost-effective way of popularizing science and educating the public about the exciting work that we all do.

A curator of astrophysics – equivalent to a university professor – at a major research museum like the American Museum of Natural History (AMNH) has a host of duties that are very different from those of university astrophysics professors. The most important of these tasks are building space shows, permanent exhibition halls, temporary (traveling) exhibitions, and a myriad of efforts aimed at Kindergarten through grade 12 education. In the following sections I will describe how museum curators partner with talented museum graphic artists, exhibition planners and educators to generate the science-rich content and curricula that we produce for millions of visitors.

2 Space Shows

Planetarium shows throughout the twentieth century used analog projectors to show visitors the night sky. The rising and setting of the Sun and stars, the phases of the moon, tails of comets, and the faint glow of the Milky Way were shown with increasing realism by ever-more sophisticated projectors. Even inhabitants of the most light- polluted cities on our planet could be shown the faint zodiacal light with a clarity unmatched at any but the darkest sites on earth. It was obligatory during those shows for northern hemisphere viewers to journey southward (virtually) over the face of our planet and watch the southern constellations rise. The slow precession of the Earth, and even the proper motions of a few of the brightest stars could be shown to illustrate that neither the pole star nor the shapes of the constellations were forever constant. Slide projectors were used to show diagrams and transparencies of telescopic views. I refer to these beautiful space shows as Ptolemaic, because the available technology forced them to be Earth-centered. There was no way to dynamically show what it would be like to journey off the face of the earth and to visit planets, stars and galaxies.

Computer and digital database technologies are the keys to building fully realistic, 3-D immersive space shows that allow us to virtually leave Earth and roam at will through the cosmos. Coupled with views from space-based telescopes and sophisticated numerical simulations from supercomputers, planetaria can finally show visitors what it would be like to journey out into space. With this capability in hand by the 1990s, astronomers designing planetarium shows suddenly faced an embarrassment of riches, in the form of vast numbers of topics that could be portrayed for the first time. A related and very practical problem surfaced simultaneously: money. Modern 3-D space shows take well over a year to produce, involve the collaboration of dozens of scientists, computer graphics experts, writers and musicians, and cost several million dollars apiece.

Both the museum and the curator are responsible for raising the capital needed to turn the curator's dream show into reality. Astrophysics graduate schools usually train their students far more intensively in planetary and stellar physics, and cosmology, than they do in the fine arts of fundraising and movie production. While NASA and the NSF have made essential and generous contributions to our space shows, corporate and private sponsorship are no less important. Grantsmanship and conversations with federal agencies are only part of the equation; successfully curating a space show really has much in common with producing a Hollywood movie. Just as a Hollywood producer assembles the financing to make a movie possible, so does the curator actively participate in the search for capital resources. In practical terms this means working closely with museum development staff, and "pitching" the concept for each new show to corporate and private donors. It also means supporting our Business Development officers who are responsible for leasing our space shows to planetaria around the world. Meeting with visiting planetarium directors and attending the openings of our space shows worldwide is occasionally time-consuming but a fun and important part of ensuring wide distribution of our work.

The creative team that generates a 3-D space show has much in common with a movie production team, and the curator must work seamlessly with all of them. Writers, graphic artists, computer programmers and 3-D visualizers, database and IT specialists, musicians and sound recording specialists each contribute their technical expertise to produce an experience that is "edutainment" – equal parts education and entertainment. Hollywood and Broadway stars donate their time and skill to become the "voice over your shoulder"; it's the curator's job to help generate their script. The curator's principal task is to generate the theme and core ideas of the space show, and then check who in the astronomical community has been doing cutting-edge research and simulations of those core ideas. A dozen or more scientists are thus identified and contacted by the curator about the possibility of using their data sets as part of a space show. Our astronomical colleagues are (not surprisingly) extraordinarily generous with their time and their best ideas. Their reward is seeing their scientific research shown to millions of viewers every year.

During the development phase of a space show, the museum's education department and the curator jointly organize a mini-symposium where all of the chosen scientists present their scientific research to the production team and to each other. It rapidly becomes clear to the attendees of these symposia which images and simulations are the most powerful, the most promising – and the most problematic. Even the most cutting edge astronomical simulations are often of insufficient temporal or spatial resolution for the very high definition imagery essential for a 3-D space show. Curators and production staff frequently collaborate with the scientists to redo or extend their simulations on national facility supercomputers. The results are sometimes the state-of-the-art simulation, offering scientific detail and insight that extends the original research.

The narrative accompanying a space show is essential for each audience member to connect the 3-D imagery and the astrophysics behind it. Astronomers tend to write bone-dry, jargon-rich text, perfect for the Astrophysical Journal. Thus while first drafts of space shows are written by astronomers, professional science writers are always enlisted to translate the technical ideas from Astrophysical Journal-ready style to standard and colloquial English. Many dozens of conversations and meetings accompany the months-long process of producing a script that is engaging, understandable by non-scientists and rigorously correct in its description of the phenomena being presented. Important and supportive critics during every iteration of every script have been the American Museum of Natural History's president (trained as an attorney), provost (a world-renowned paleontologist), and senior vice-presidents for Education, Development and Legal Affairs. None of these highly educated and accomplished individuals has any formal astronomical training, and none of them aspire to be astronomers. As our president once noted in pithy fashion: "If I don't understand a script segment, a 9-year-old is unlikely to understand it". She's right, of course, and the red pens of these erudite non-astronomers have often made the scripts much tighter and easier to understand without dumbing down the science. That doesn't mean that the astronomers and writers accept any and all suggestions without pushback. I've attended heated but respectful discussions where the merits of one five-word phrase versus another occupied half an hour. We've all learned that "if in doubt – don't" is usually good advice, and the simplest way of saying something is almost always the best.

After a script is completed, a well-known Hollywood actor is recruited to record the narration for our audience. We've been incredibly fortunate to have Tom Hanks, Harrison Ford, Robert Redford and Whoopi Goldberg escort millions of our viewers through the universe. Each to them has donated their services gratis in the service of public outreach and education, and their professionalism is remarkable. As an example, I note that Harrison Ford admitted that he knew very little astronomy during his sound studio session. He recorded six to eight versions of each sentence and phrase. Despite intensive, real-time coaching it was clear to the astronomers present that some of Ford's recording wasn't hitting the mark. A few weeks later we invited him to the planetarium to see a first cut of the space show and to hear a rough cut of his script recording. He grimaced through much of the show, denouncing some of his own work in unprintable terms. A few days later he returned to the recording studio and produced an outstanding reading of our script.

About 6 months, and the efforts of many dozens of astronomers and computer scientists are required to generate and assemble the ultrahigh resolution images, diagrams, and simulations that comprise the visually stunning heart of each space show. Rendering and shading the hundreds of billions of triangles that are calculated from the zeros and ones supplied to us by dozens of astronomers consumes vast amounts of supercomputer time – and is an important part of each show's budget. While I've seen each animation segment before on a laptop, the jaw-dropping effect of a much higher resolution video shown on a 30 m dome cannot be underestimated.

Enthusiasm for astronomy and planetarium space shows is worldwide. All of our space shows' scripts are translated into multiple languages by native speaking astrophysics postdoctoral fellows. Local media stars read the scripts, connecting with audiences on six continents (and at sea on a luxury cruise liner).

While I love opera, classical music and jazz, my own talents as a musician and composer are nonexistent, and the same is true of most of my colleagues. Curators choose from an eclectic variety of music generated by our composers. I'm always surprised at how well our composers match and enhance the mood of a scene from a one-paragraph description of it. A powerful score is an essential ingredient of every space show.

The final month before a space show's premiere is extremely trying for the curator. The science, simulations, rendering of scenes, and music are all done. There's nothing you can do! The process is entirely out of your hands as a beautiful, seamless show is assembled and polished by the computer professionals. A few days before the grand opening, AMNH holds an open house for the press. Over 100 reporters show up at 8 AM for bagels, coffee and a brief curatorial lecture about the science behind the new space show. The reporters get a private showing and copious handouts from our communications department describing what they've just seen… some of which ends up verbatim in their published stories. The rest of the day is spent doing newspaper, television and radio interviews which air on opening day in the local and national media. These media flurries are curatorial Andy Warhol episodes of 15 min of fame.

The most important critics of all, of course, are our audiences, especially the schoolchildren. Over 50 million people have seen AMNH space shows all over the world in the past decade. There is no way that I or my curatorial colleagues could individually or collectively have reached even 1% as many viewers with any other platform. I consider it both a privilege and responsibility to produce more space shows in the future… and one of the most rewarding aspects of working in a museum.

3 Collections and Permanent Exhibition Halls

The core message of every museum is contained in its permanent exhibition halls. The American Museum of Natural History is a collections-based institution, with 35 million specimens gathered from every part of planet Earth over the past 150 years. Collections managers catalog, store, preserve, and loan out specimens from our vast collections, just like librarians. The display of one million birds, bats, or fishes would be neither profitable nor feasible, so curators choose the most important and representative specimens – about 1% of our collection – for display in permanent exhibition halls. These carefully chosen specimens may illustrate the evolution of a species, a culture, or the Earth itself. Each hall tells a story, educating and entertaining our visitors simultaneously. A striking example is the Hall of Ocean Life: its eight habitats cover the range of conditions under which life thrives in the sea. Particularly important are the exhibits of vertebrate and invertebrate evolution, based on DNA analyses carried out by our ichthyology curators, and their students and postdoctoral fellows. The point is that curators are not hired by a museum just to assemble and direct exhibitions. They are hired for their scientific and technical expertise, so that exhibitions are informed by up-to-date, cutting-edge science.

Astrophysics is a little bit different… but very much the same. We obviously don't have collections of millions of stars, asteroids, or galaxies preserved in formaldehyde or stuffed. Our collections are digital – terabytes of images and numerical simulations. This is well understood by our senior administrators, who have allocated positions for data collections managers, with position descriptions almost identical to those of the people who manage our millions of birds, bats and fishes. It's the responsibility of astrophysics curators to choose the most important and representative examples of data and simulations to display in the Planetarium, and to work with exhibition planners and designers to create compelling ways of telling the story of the Universe.

Permanent exhibition halls are very expensive, and must have lifetimes of decades. It's a delicate balancing act to summarize the current state of knowledge in a given field in astrophysics in a way that won't be embarrassingly obsolete in 10 years. One wants to include the latest and greatest findings – but what if they're shown to be wrong or incomplete in a year or two? The solution is to "cut metal" only for phenomena and objects that are almost certainly correct and complete. The most cutting-edge results are displayed on high-resolution monitors, driven by hidden computers whose contents can be easily changed. Jupiter and Saturn are unlikely to be displaced as the largest planets in our solar system, and so huge (and expensive) models of them hang on permanent display. In contrast, dark energy and dark matter are regularly and prominently featured on giant monitors whose programs change regularly.

4 Temporary (Special) Exhibitions

Keeping the exhibitions at a museum current, fresh, and cutting edge is nearly impossible in permanent exhibition halls. In addition, topics that may be fascinating to the public, but not core to the mission of the museum should also be presented from time to time. The solution to both of these challenges is the temporary or traveling exhibition. Topics too specialized to be suitable for inclusion in permanent exhibition halls may nonetheless drive exhibitions that draw and challenge hundreds of thousands of visitors. An example is the Einstein exhibition that I was privileged to curate in 2002–2003. A summary is given at http://www.amnh.org/exhibitions/einstein/curator/index.php.

In 2002 I was fortunate enough to be able to borrow many of Albert Einstein's personal papers from the Einstein archives at the Hebrew University in Jerusalem. These priceless documents, including the 1912 Zürich notebook (which many historians of science consider to be the Genesis of General Relativity), were shown to the public for the first time ever. Detailing Einstein's greatest scientific triumphs, especially explanations and interpretations of relativity formed about half the exhibition. Teachers, students of physics, and historians of science gravitated overwhelmingly to this scientific and technical material. A lot of time and effort was expended in making $E = mc^2$ and time dilation understandable and accessible to our visitors. The other half of the exhibition focused on Einstein's life and times. We were particularly fortunate to have been able to borrow both Einstein's August 1939 letter to President Roosevelt, warning of a possible German development of an atomic bomb, and Roosevelt's response to Einstein, noting the very beginnings of the Manhattan Project. These two original documents had never before been shown side-by-side in the same room – and might never be again. No less interesting to me were letters to and from Einstein's wives, mistresses, friends and detractors. The portrait of a complex, brilliant and deeply caring human being emerged. I've gone to many of the openings of the exhibition as it's traveled the globe. Each visit is like catching up with an old friend.

5 VIPs and Donors

It's flattering to have famous and powerful people want to see and understand your work! Every week or two I get a call from our VIP services department, asking if I can give a politician, or captain of industry, or movie store a tour of the planetarium exhibits and the current space show. Many of these VIPs are very bright and deeply interested in astronomy. Black holes and dark matter have as powerful a grip on the imaginations of politicians and movie stars as they do on the imagination of the public. The goal of supporting these visits is goodwill for my Museum. Occasionally, however, there are unexpected benefits of a VIP visit. At the end of his visit several years ago, the late Paul Newman made a generous pledge to enable AMNH to become a partner in the 9-m Southern Africa large telescope. Other generous benefactors have also contributed support to our research and education efforts.

6 Teacher Training

Most astronomers get asked to talk about our profession and current astronomical research at elementary and high schools – I'm no exception. I've always believed that it's incumbent on us to honor at least some of these requests. There're time-consuming, of course, and the long-term effects are hard to measure. Our Education

Department experts maintain that the very best use of curators' time is training and enhancing the skills of science teachers. Every summer my colleagues and I work with teachers at summer institutes held at the American Museum of Natural History. The goals of these institutes are to present to the teachers the latest findings in multiple fields of astrophysics, and to have them turn those findings into practical curriculum materials that they themselves can take back to their classrooms and share with other teachers. Making certain that these materials are consistent with national education guidelines is essential. Only by partnering with professional educators are we able to successfully carry out this mission.

7 Discovery Tours

I'll end with one of the most gratifying and fun parts of a curator's job – traveling as a lecturer to exotic locales with friends of the Museum. About once a year I get to take 20 or 30 deeply interested travelers to the summit of Mauna Kea, to the CERN collider near Geneva, to Machu Picchu or Easter Island or Ayers Rock or the Taj Mahal… well, you get the picture. In return for giving a daily lecture on some topic in astrophysics, I'm privileged to see some of the Earth's most interesting sites in the company of dedicated museum friends and supporters. I couldn't ask for more.

Chapter 16
Being an Astronomer: A Testimony

Bernard Fort

Abstract In this short essay, I examine how and why I became an astronomer and what finally this job has brought to my life. I believe that an irresistible compulsion to better understand the sky has driven my observational work. After my beginnings in solar astronomy, I bet on the rise of solid state technology and, in this way, I was among the first to observe the deep Universe with CCD detectors on a large telescope. This path led me almost naturally to observations of strong and weak lensing of faint distant galaxies by foreground structures. With a short overview of the story of gravitational arcs, I illustrate how astronomy might develop through the opening of new observational windows. To increase our knowledge of the Universe astronomers must be "big builders" and have also a profound expertise in many fields of physics. With large telescopes astounding discoveries have been made, but at the same time these findings have come close to the limits of human logic in trying to understand the true essence of the world and its origin. Thus satisfying a compulsive search for meaning was both for me a source of satisfaction and some disappointment. If being an astronomer brought me moments of happiness it was not always where I initially expected to find them, but rather in friendships created in the collective adventure in search of knowledge. My essay concludes with a personal and probably naive remark concerning how employment in astronomy might change as well as with a few worries if we do not succeed in gaining a better understanding of the role our minds play in constructing our collective human beliefs.

Keywords History and philosophy of astronomy • Gravitational lensing • Instrumentation: detectors

---

B. Fort (✉)
Institut d'Astrophysique de Paris, 98 bis boulevard Arago, 75014 Paris, France
e-mail: fort@iap.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_16,

1 Why Does All This Around Me Exist?

Upon being asked to write a testimonial about being an astronomer my first reaction was that I could never succeed in such a project. How could I extract a general lesson from an overview of a personal pursuit of knowledge in a subject that continues to hold such deep mysteries? But in due course, I happily accepted the proposition motivated in great part by remembering the many students with whom I shared this pursuit over the course of many years of research and teaching. In addition to discussing theories and technical problems in astronomy, we talked often of the spirit of science and the search for meaning in life. I found myself anticipating the opportunity to recollect and further analyse some of the questions that we discussed together. But an additional motivation most probably came from the fact that I was at the time of the invitation reading the novel of Julian Barnes [3] “Nothing to be frightened of”. Through a family memoir, Barnes succeeds in projecting a bright overview of the variety of many life attitudes that are held by human beings that are cognizant of the vast knowledge yet to be attained as well as their own mortality that must end in death. I felt myself in a so perfect resonance with this novel that I saw in its coincidence in time with the writing invitation the encouragement I need to set aside my fear of failing to write my own story in a similar sense of inspiration.

As in no doubt true of most people in their youth, I became acutely aware of myself when I formulated for the first time the ancient question “Why does all this around me exist?” After more than sixty years I am still not resigned to the notion that, despite the many brilliant minds that have left us a wealth of knowledge I will never have more than some mere glimmerings of the true answer to the question. But in truth, human history has led us mainly to hypotheses and theoretical constructions which are also “world fictions” of the mind. Neurosciences, for example, are now investigating in detail how such world fictions are equivalent to “world simulations” which arise in each individual brain and how they emerge in a conscious mind [16, 17] to form our beliefs. Such fictions result from the merging of a number of “individual” histories. First is the history of genetic evolution according to which the main driver of life is the collection of our imprinted genes. Following this are cultural histories replete with myth and faith which is determined mainly by where we live and what we learn through our family and professional life. The ultimate result is that our awareness of the surrounding world is permeated at all levels of our mind with a huge number of “fictions” which invariably result from each unique individual experience. However to live in society in harmony with others and to survive as a species we must share common fictions. The natural trend is to adopt our cultural, philosophical or religious fictions in line with our personal trajectory through life. The difficult problem is that many fictions are incompatible and adopting them as absolute truths leads to confrontation, suffering and war. During school, I quickly became skeptical about many fictions which were proposed to me. However, I was rather taken by the beauty of mathematical logic and its ability to solve physical problems. Up until then, I had not discovered philosophers and I was more attracted by people who were constructing a scientific history of the

Universe and humankind over centuries of continuous effort. I became slowly aware that these scientific fictions rely on a mathematical grammar and on physical laws that also have their own limits. For a young student, they especially had the apparent virtue of being comprehensible by everyone throughout the Universe. Indeed I was desirous to know how the World was made, how it operated through experimental facts and what its eventual fate might be.

My career in astronomy has convinced me that scientific theories are cultural products which can be used only within narrow limits and which might be seriously incomplete. But with this caveat, astronomers have contributed to the writing of a new genesis of our world, a theory that is continuously being improved and which seems more acceptable than any other previous theories. Indeed astronomy provides the most reliable history of the formation of planets, stars and galaxies. It is however a discomfort to admit the limits of scientific knowledge, particularly concerning the origin of the Universe, and still choose to rely primarily. For me there was a psychological origin in seeing the logic of mathematics as the best key to understand the world.

I believe that to be a scientist is to be driven by the obsession to better understand the world around us. Astronomers search for answers in the distant Universe. Potential astronomers, in childhood are known to have been motivated to observe the night sky, and this can often drive them toward construction of a first telescope and to observing regularly the planets and stars. But more often, if I believe a quick survey in my laboratory, this attraction emerges from a mind being shaped unconsciously in early childhood, a phenomenon that can be traced later in life. This was my fate. After the end of my engineering studies contemplating many attractive possibilities, I finally chose my first job at the Paris Observatory. Only much later in life was I able to identify the original causes that drove me to make this choice. During my early days, when I was a toddler trying to find my identify and to separate my self from my mother, I was experiencing at the same time the threat of bombs from the sky. I was not really able to understand the war in words: dramatic images were all that I retained. I was nevertheless afraid. I wanted to know what was going on and I started developing an imaginative and anxious personality. At seven, the sudden death of a grandfather who literally fell dead in my arms dealt me another blow, which had a similar impact on my thinking. Therefore, I have followed a path in life determined by a native and compulsive curiosity to understand the world around me. I wanted to make it more predictable and less frightening. My early consciousness had grown primarily around images of the world with a large emotional content. Later, I remained deeply attracted by painting, geometry and astronomical images. From such a cursory analysis of my infancy I suspect that motivation for astronomy does not come at random but rather from the desire to give meaning to things. Unfortunately such an obsessive and compulsive attitude is not sufficient to become an astronomer.

2 First Steps in Solar Astronomy

For my first position in 1965, I became the project engineer of a solar coronagraph project dedicated to coronal observations during long periods of time. The instrument had to be launched with a stratospheric balloon facility of the French National Center for Space Research (CNES) for several trans-European flights. The instrument produced an artificial eclipse of the sun with an occulting disk placed several meters in front of the primary lens of the telescope. Dr Audouin Dolfuss from Paris Observatory loaned me some unpublished laboratory journals of Professor Lyot (period 1938–1950) so I could educate myself in the coronographic arts. He gave me the responsibility of developing the instrument with a tiny team of around five people not much older than me. I will always remember the happy and brilliant team of students I worked with. The main difficulty of the project was to suppress the diffraction of light coming from the external occulting disk [21]. We succeeded and the technical performance was noted by Bob Mac Queen who invited me later to HAO (Boulder Colorado). This achievement brought me to the attention of two radio astronomers, Monique Pick and James Lequeux, and of Raymond Michard then president of the Paris Observatory who encouraged me to take a position of research assistant at the Centre National Recherche Scientifique (CNRS). Today such an astronomical project where a handful of young people can almost without bounds engage funds with no experience but just a lot of enthusiasm is unthinkable. It is worth noting that this post-war pioneer spirit ultimately led to the moon landings. With the development of huge international research funding agencies we have abandoned such a flexible but risky project organization. My generation has lived through these changes which now make possible the construction of large and complex instruments. Sometimes we feel the loss of those times where we were more keenly aware of the intense and pure pleasure of individual discoveries with less constraints arising from the organization of large international consortia with all the attendant political difficulties and pressure for funding.

After the May 1968 student riots in Paris that marked the epoch, at the age of 27, I was offered a position at the CNRS in a laboratory developing photomultiplier devices and electronographic cameras at Paris Observatory. I accepted a work related to the description and understanding of fine structures of the solar corona. The project needed measurements of the intensity and polarization of various emission lines of the low solar corona during total solar eclipses. It culminated in the exceptional solar eclipse of June 30, 1973 which allowed me to defend a “Doctorat d’Etat” on the temperature and density distribution of coronal magnetic loops and their relation with other solar structures such as chromospheric filaments and coronal jets [22]. I was not completely pleased with the results reported in my thesis because I would have preferred them to be more original, a common narcissistic attitude for a young scientist. My thesis work was nevertheless a decisive period of formation. I learned from my supervisors Paul Felenbok and Michel Combes that research means to take risks and how a good group dynamic is important for experimental work. I can say frankly that my subsequent successes in observational astronomy drew their origins from this period.

By the end of the 1960s we knew that observational astronomy would soon undergo a rapid evolution with the emergence of new technologies, with the construction of larger telescopes on good sites and above all with the development of space astronomy. After the moon landing, planetary sciences were expected to be the first to undergo dramatic development. But the solar system appears much too nearby for me and I began to feel an irresistible attraction toward the distant Universe.

3 The CCD Revolution

In 1976, during a one-year visit to Alec Boksenberg's laboratory in London, I discovered that *high-performance* solid-state imaging detectors would follow inevitably from the invention of the charge transfer concept by Boyle and Smith [48] (see also Chap. 7). It was a paradoxical intuition for someone who had worked with electronographic cameras and who was now spending one year in the "temple" of astronomical photon-counting imaging devices. In fact I was influenced by Dr Coleman from the University College in London who suggested that I read more papers on solid state technologies while preparing lectures in imaging science. Willard Boyle and George Smith received the 2009 Nobel prize in physics for their invention, an event astronomers were pleased to hear it because the use of "Charge-coupled devices" (CCDs) has led to a new era for astronomical observations. After my return to Paris, my faith in the future of CCDs led me to develop a camera prototype as soon as the first 100 × 100 Fairchild 202 CCD chips became available. I joined the tiny club of the first users of CCD cameras in astronomy. A few years later with the technical support of Laurent Vigroux's group at the CEA [11] we were able to use a larger thinned RCA CCD chip at the Canada-France-Hawaii Telescope (CFHT), a telescope located on the best observational site in the northern hemisphere, where the *FWHM seeing* is regularly better than 0.8".

In 1982, the CNRS gave Dr Jean-Pierre Picat and myself the possibility to start a new astronomical laboratory at the Toulouse space sciences campus to develop CCD techniques. CCDs were developing much faster than we could have imagined due to military and industrial needs and the development of the Hubble Space Telescope. With our students we observed at CFHT the upper atmosphere of giant red stars with a stellar coronagraph, environments of quasi-stellar objects and elliptical shell galaxies. We developed first multi-object spectrographs operated in real time at CFHT and the European Southern Observatory [20]. We contributed to the use of CCD in medical imaging and to the development of Thomson chips which could be joined together for very wide-field imagery. But most spectacularly of all, we observed distant clusters of galaxies and discovered a strange arc-like structure in Abel 370 [46]. It belonged to a new class of objects [33, 39, 47] that we have never ceased to study: the gravitational arcs. Twenty years later, for the observations of "cosmic shear" [52], almost the same CEA team built the largest CCD camera that had ever been used on a telescope so far, the camera MEGACAM (which has a one degree field of view with a pixel scale of 0.2"/pix, [12]).

Solid-state technologies and more generally the overwhelming intrusion of computer sciences has profoundly affected the work of astronomers. Before illustrating this point, it is important to remember what links all work in Astronomy. Astronomy aims to observe, classify and model the entire chain of events which can be observed in the distant Universe. But when an astronomer observes a remote object, this information is carried by a messenger, a photon or some other particle. The propagation of this object allows one in principle to observe all past events located along this light cone. Since the evolution of the Universe is the same everywhere and the Universe contains billions of galaxies we can observe the evolution of any galaxy types from the time they are born.

When in the 1980s I first became interested in extragalactic astronomy the most distant clusters of galaxies known were at a distance of about three billion light-years [1]. Photographic plates could see no further into the Universe because they were recording only a few percent of the light collected by the telescope. I had in mind that CCDs with a quantum efficiency of around 80% would increase the ability of telescopes to detect faint objects by a factor of twenty, or an immediate increase of a factor of four in effective telescope diameter! A new observing window would be opened on the distant Universe. Actually, CCDs profoundly changed our vision of the Universe. On deep CCD images of the sky, a dense and relatively uniform distribution of faint distant galaxies appeared everywhere (around ten billion for the full sky). They were at great distances. Moreover, it was certain that it would not be long before astronomers observing distant clusters of galaxies with CCD cameras and a telescope like the CFHT would discover gravitational arcs. We were successful because we became curious about a faint and unexpected object on our first CCD images of A370. It might have been an image artifact but I wanted to understand the origin of this strange ring-like structure.

4 From Gravitational Arcs to Cosmic Shear

In deep CCD images background galaxies at an average redshift of $z = 1.$ can be easily detected (something which is obvious today). Some of these galaxies could be magnified by foreground mass condensations like a cluster or a galaxy when a deflecting mass is aligned along the same line of sight. If the central mass concentration of the deflector is large enough it acts like a natural telescope and can even give a mirage with multiple images of the same source [15, 42]. This phenomenon was known for a few distant quasi-stellar objects [53] but surprisingly no observers had imagined it could occur in much greater number with distant galaxies. We observed the first gravitational arc on A370 in 1986 [33, 39, 46]. Gravitational arcs and rings were more easily and accurately studied with the HST in the following years [29, 36]. Now they provide the most spectacular and direct evidence of dark matter distribution in clusters [49] and galaxies [9].

Very soon after the spectroscopic confirmation of the true nature of giant gravitational arcs we rapidly became aware of the distortion of others background

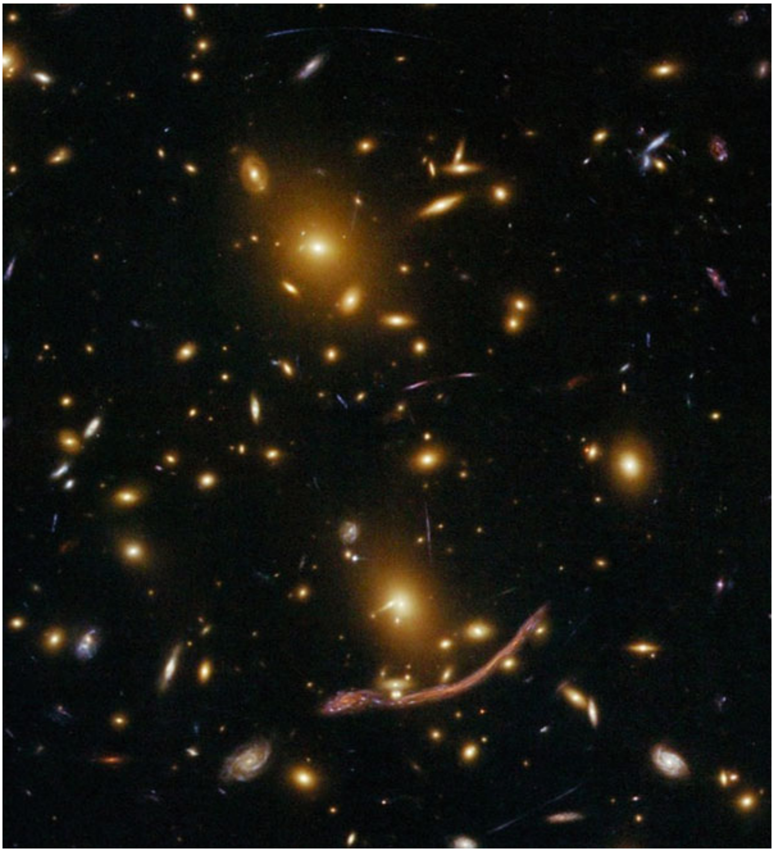


Fig. 16.1 Hubble Space Telescope image of the giant arc (and arclets) in the cluster of galaxies A370

galaxies surrounding the cluster A370 (see Fig. 16.1 for a recent HST image) [23]. At the time I was a visiting fellow at ESO and I remember the excitement I shared with Yannick Mellier and Peter Schneider following this discovery. It would be actually possible to use the distortions of background galaxies to map the dark matter surrounding clusters [10, 18, 50] and later to show that this "cosmic shear" could be detectable [19, 45]. Only in 2000, after an obstinate effort of Yannick Mellier and his colleagues did it become possible to observe and to analyze the cosmic shear [25, 35, 40, 52]. Nowadays the dark matter mapping technique has developed so much that it has become possible to reconstruct the large scale distribution of dark matter in three dimensions along a pencil beam through the Universe [34]. In direct line with all these works I am happy to see that space-based dark energy missions that will use weak lensing techniques are now under consideration at an international level [41].

The story of gravitational arcs is so rich in successive contributions of many observers and theoreticians that it is beyond the scope of this article to report them.

It has undoubtedly benefited from previous works on multiply imaged quasi-stellar objects [7]. Rapidly reviewing this story I see a beautiful confirmation that observational astronomy is a collective and long-term endeavour with all the vagaries of human adventures. I remember how a forgotten idea [51] can reappear and propagate in a more favourable technical environment and the competition between observers as well as decisive conceptual contributions of theoreticians [2, 4–6, 8, 27, 28, 32, 37, 38, 43, 44]. The lensing of background galaxies has become a vast observational field with thousands of publications of the international astronomical community and major advances in our understanding of the DM distribution [9, 13, 14, 24, 26, 30, 31, 35, 49].

When arcs were discovered, the mass concentration at the centre of galaxy clusters was not considered large enough to produce strongly lensed images. The discovery of gravitational arcs has shown that taking risks in developing new observational techniques is a productive way to increase our knowledge in astronomy providing we can set aside our prejudices. In this way with only modest resources a tiny team removed from any major centres of astronomy succeeded in proposing new ideas and concepts which led to the development of a completely new field of observational astronomy. I also learned that new ideas are evanescent entities. Often they can only be incorporated in a body of knowledge after a large amount of rigorous work within a favourable technical environment.

5 Astronomers as Builders

Astronomical observations envision us within space-time with an inconceivable number of objects and complex phenomena which evolve in time. The gravity is at the origin of a matter-energy karma from which all cosmic structures are formed. Stars are born from gas clouds and partially return to gas clouds upon their death. Evolving galaxies collide and merge along cosmic web structures. During the last decade the impact of computer sciences on theoretical astrophysics has been decisive in building a coherent model of such a cosmic karma. Numerical simulations have profoundly changed our investigations and interpretations of the world. I was particularly impressed by the impact and efficiency of computer techniques on our work. But I also see how it is easy to be trapped by numerical models which can subtly replace astrophysical questions with technical ones. They can restrain our imaginative ability to get closer to the essence of things. Towards the end of this chapter I will raise some concerns about the danger that an excess of information could eventually slow down the progress of astronomical knowledge in the future.

Models of the Universe are obviously better adjusted when it is possible to observe all the existing mass distributions of gas, stars and dark matter. Therefore it is necessary to observe the Universe in optical, X, Gamma, IR, mm and radio wavelengths at different scales and to detect particles flows. Such an observational approach demands the construction of powerful telescopes and a research and development program carried out by scientists who have the finest expertise in almost all domains of physics. Therefore modern astronomy concerns a diverse array palette of jobs and engineering expertises. In the light of results obtained during the last twenty years, extending and sharpening our view of the Universe has been quite decisive in making discoveries and in inventing physical concepts which may extend beyond the frontier of current knowledge. Many puzzling questions concerning the formation of planetary systems, the primeval Universe and the apparent existence of dark matter and dark energy have been identified which can probably be solved only with new instruments and extensions of our physical theories.

The astronomical landscape presented in this book demonstrates that astronomers are becoming architects of very large and complex instruments both on Earth or in space to better understand the Universe. Astronomers are among the biggest builders of our scientific era. As a mere example, observing the decoupling of light from the matter 13.4 billion years ago, and deducing a scenario of the galaxy evolution was a large and risky undertaking which required construction and utilisation of several successive satellites (COBE, WMap, Planck) and very large telescopes. To reach back to the formation of the first galaxies, astronomers are now engaged in construction of a new Space Telescope and of huge (radio) telescopes. Even more exciting for humankind, they are working on a space interferometer project (Darwin) which could detect life-signatures on extra-solar earth-like planets. Throughout the history of humanity, astronomers have contributed to the building of pyramids and temples to support a religious understanding of the world. In the future, I hope their observatories will be among the new temples of a society which will produce a scientific interpretation of the world, a knowledge which has also the incomparable virtue to continuously remind us that no-one can pretend to have reached an ultimate truth. Only the confrontation between theoretical models and observations make it possible to perfect our knowledge of the history of the Universe. This confrontation cannot end before we can show that any observation can be coherently interpreted by a single physical model. This challenge gives theoreticians a special place. Less susceptible to onerous technical responsibilities, they are better placed to extrapolate consequences of unexpected discoveries and explore new hypotheses and theories. They are inventing a well-defined structure of physical laws to read how Univers' machinery works. However for theoreticians who become too separated from observations the risk of mis-interpretation is real.

Theoretical studies are essential to check the feasibility of new projects and to estimate the minimum scientific return that might justify a large expenditure for telescope constructions. To illustrate with more strength how theoretical works are valued by astronomers I hazard a guess for gravitational wave detections. Several ambitious projects are aimed at the opening of this new observational window (See Chap. 5). It is exciting because theoreticians are seriously considering Universes with multiple dimensions, with adjacent "branes" where only gravity can leak between them, as well as strong warping of space time in the immediate neighborhood of black hole horizons. Space-time maybe eventually wrinkled in some places. From this strange gravitational effect one might expect that a future space interferometer like LISA could reveal a completely unexpected structure for space-time. Within several tens of years, after an enormous amount of experimental work and hopefully after the detection of first gravitational waves by VIRGO and LIGO I conjecture that we will be at the beginning of one of the biggest scientific adventures of this coming century.

6 Being an Astronomer

Coming back to the initial question "what is being an astronomer?" we easily guess that an answer such as "astronomers observe the Universe to decipher its true nature" is no longer sufficient because in practice they must do much more. They design and build new instruments. They gather and reduce a huge amount of observational data, and structure them into intelligent databases (See Chap. 14). They analyze the data and make numerical models to explain their observations. They formulate hypotheses and theories which could eventually call for new mathematical and logical concepts. Astronomers are also teachers and participate in international conferences and workshops to compare their findings to others interpretations. Indeed some astronomers must also be good politicians to convince their respective governments and public opinion to fund and support their work.

As in any other sciences, knowledge of astronomy grows through a dynamic process where astronomers continuously update a common model of the Universe and a list of unsolved problems, namely the questions of origins. Also many of the technical jobs in astronomy are alike but not identical from those in engineering which contributes to the development of our modern societies. They remain driven by the strong psychological motivation which I tried to identify at the beginning of this article. A motivation which corresponds to an indefatigable desire to find a deeper meaning to the world and which relegates the search for individual profits to secondary importance (this does not prevent successful applications of research, see Chap. 7). For this reason, accomplishments of people who work in astronomy are frequently the source of intense satisfaction. This is why being an astronomer is certainly a privilege in developed societies which imposes to select candidates having this work ethics as well as a high level of expertise and a creative imagination.

7 Final Thoughts

To conclude I would like to return to the future evolution of jobs in astronomy and the feeling of satisfaction that we can achieve in being an astronomer. The standard "Λ-CDM" model of the Universe, our cosmology paradigm, is a construction of the mind of Homo Sapiens which has developed after centuries of trial and error. Up to the end of the last century, a single observer could still hope to be able to add a stone to this edifice. Progress was still a serial process involving successive individuals. During the last decades it has become increasingly apparent that huge international teams are necessary to pursue the exploration of the Universe with giant instruments. As an example, obtaining more observational constraints on the cosmological constant cannot be considered now without a large task force. The contributions of individual astronomers becomes more diluted.

More astonishing is that perhaps the human brain might have insufficient capabilities to probe further into the Universe. The deeper we try to understand the very nature of the Universe, the more our common logic fails and needs to be reconsidered. Only a handful of theoretical physicists can really produce new theories and mathematical concepts which could solve some of the new enigmas of the Universe. Observational astronomers have a superficial view of debates concerning quantum gravity, string theory, the holographic principle and black holes. In contrast, theoreticians are often lost when confronted with complex instrumental devices and the technical details of data processing. Therefore very few (perhaps none) individual astronomers can actually pretend to fully understand and be capable of checking all the stones in the edifice of astronomy. We have reached a point where being of individual consciousness have no choice but to trust the contributions of others and to believe in them. The result is that astronomers are constructing and believing in a cosmological fiction whose larger contours escape them.

Although astronomers have a fragmented view of the Universe today, they are informed almost instantaneously about any advance in their field of research. All the research teams around the world are interconnected like neural networks. Also in astronomy, a virtual super-mind is already emerging from such an information network which seems to be able to assemble all the piece of the cosmic puzzle in a correct way. We are discovering how we feel close to horizon limit of knowledge which might be inherent to the evolutionary state of our brain. Due to the importance of the observer in any world interpretation it appears urgent that we learn how our brain works. Advances in neurosciences are now a real concern for a small but an increasing number of astrophysicists.

There is no doubt that the revolution of information techniques has contributed to extend astronomical knowledge. But if the very essence of the world is more opaque to individuals, how to select information and make sure that our modeling of the world is converging toward the right one? I noted previously that numerical simulations of the Universe that easily capture our awareness on technical problems can sometimes substitute a virtual reality to the real world. I see another risk to our individual creativity with instantaneous access to new observational results and ideas. We may face a stagnation of world models. The continual flood of information astronomers are subjected to is a good way to restrict individual imagination, to miss original ideas and to lose one's research track. Only audacious thinking avoids being trapped in a uniform (non) questioning of the Universe. A colleague, Jean-Philippe Uzan, sums up this concern with this joke: "Minds at the same thermodynamic temperature rule out imagination"!

Our scientific knowledge is telling us more than ever before, that the world around us is built from an unusual logic that we could not comprehend without new mathematical concepts. Can a mere society of information be a new obstacle to the emergence of a better cosmological paradigm? As an observer, I don't really think so because if the Universe seems more creative than us, it has always permitted us to identify its mysteries when we persist in observing it. What I really fear much more is an information society ruled by people with less education in sciences leading to a triumphalist rise of irrational beliefs among many, such as recent creationist pseudo-theories.

In the future our attempts to probe more deeply into the true reality of the Universe will be an opportunity to share more collective adventures than the ones I have encountered. These adventures will bring to more people the same rewards I experienced. I remember moments shared with close colleagues and students when we succeeded in making our observations more intelligible; we experienced emotions similar to ones which seize us when we stand before a great work of art. These moments had the power to bring us together as a social group far beyond any scientific context. After such moments we can be happy just to be together. Being an astronomer has surprisingly led me to learn how to discover others and myself beyond all our respective systems of beliefs. This is obviously true for any field of research and this is why I call passionately for the construction of a society of knowledge and scientific education. It was said that "the Universe is given to us as an enigma to solve", possibly for a constructive way to give meaning to our life.

Finally, I began my explanation of my fate of being an astronomer as arising from a compulsive wish to better understand the world in a rational way. The paradoxical conclusion that I reach at the end of my career is that this attitude has provided me with much pleasure but also much uncertainty. If I succeeded in finding happiness it was in greater part through resonances with this world of senses which I was able to share with others. This is why I take the opportunity in this chapter to warmly thank everyone who made this possible for me, the students, technicians, engineers and scientists who have journeyed with me and who have allowed many friendships to begin and whenever possible to mature.

Acknowledgements Bernard Fort thanks Observatoire de Paris, Observatoire Midi-Pyrénées in Toulouse and Institut d'Astrophysique de Paris for a strong support and a warm hospitality during his career. Many thanks to Henry McCracken, to Hervé Chamley and to the editor for a careful reading of the manuscript which makes this paper more understandable to English speakers, to Jean-Francois Sygnet for comments and technical support and to Jean-Paul Kneib for the use of a recent but spectacular image of the cluster A370.

References

1. G.O. Abell, C.E. Seligman, The distribution of clusters of galaxies. Astron. J. **70**, 317−+ (1965). doi:10.1086/109528
2. C.E. Bacon, D.M. Goldberg, B.T.P. Rowe, A.N. Taylor, Weak gravitational flexion. Mon. Not. R. Astron. Soc. **365**, 414–428 (2006). doi:10.1111/j.1365-2966.2005.09624.x
3. J. Barnes, *Nothing To Be Frightened Of* (Knopf, New York, 2008)
4. M. Bartelmann, P. Schneider, Weak gravitational lensing. Phys. Rep. **340**, 291–472 (2001). doi:10.1016/S0370-1573(00)00082-X

5. F. Bernardeau, L. van Waerbeke, Y. Mellier, Weak lensing statistics as a probe of {OMEGA} and power spectrum. Astron. Astrophys. **322**, 1–18 (1997)
6. F. Bernardeau, C. Bonvin, F. Vernizzi, Full-Sky lensing shear at second order. ArXiv e-prints, [arXiv:0911.2244] (2009)
7. R.D. Blandford, C.S. Kochanek, Gravitational lenses. in *Dark Matter in the Universe*, ed. by J.N. Bahcall, T. Piran, S. Weinberg (World Scientific, Singapore, 1987), p. 133
8. R.D. Blandford, A.B. Saust, T.G. Brainerd, J.V. Villumsen, The distortion of distant galaxy images by large-scale structure. Mon. Not. R. Astron. Soc. **251**, 600–627 (1991)
9. J.V. Bolton, S. Burles, L.V.E. Koopmans, T. Treu, R. Gavazzi, L.A. Moustakas, R. Wayth, D.J. Schlegel, The sloan lens ACS survey. V. The full ACS strong-lens sample. Astrophys. J. **682**, 964–984 (2008). doi:10.1086/589327
10. H. Bonnet, Y. Mellier, B. Fort, First detection of a gravitational weak shear at the periphery of CL 0024+1654. Astrophys. J. **427**, L83–L86 (1994). doi:10.1086/187370
11. A. Bouere, J. Cretolle, B. Fort, R. Jouan, M. Gorisse, A. Lecomte, Y. Rio, L. Vigroux, Description and preliminary performance of a charge-coupled device/CCD/camera. Presented at *The Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, vol. 290, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (SPIE, Bellingham, 1981), p. 142
12. O. Boulade, X. Charlot, P. Abbon, S. Aune, P. Borgeaud, P. Carton, M. Carty, D. Desforge, D. Eppele, P. Gallais, L. Gosset, R. Granelli, M. Gros, J. de Kat, D. Loiseau, Y. Mellier, J.L. Ritou, J.Y. Rousse, P. Starzynski, N. Vignal, L.G. Vigroux, Development of MegaCam, the next-generation wide-field imaging camera for the 3.6-m Canada-France-Hawaii telescope. Presented at *The Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, vol. 4008, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. by M. Iye, A. F. Moorwood (SPIE, Bellingham, 2000), pp. 657–668
13. T.G. Brainerd, R.D. Blandford, I. Smail, Weak gravitational lensing by galaxies. Astrophys. J. **466**, 623 (1996). doi:10.1086/177537
14. T. Broadhurst, Gravitational 'convergence' and cluster masses. ArXiv e-prints, [astro-ph/9511150] (1995)
15. W.L. Burke, Multiple gravitational imaging by distributed masses. Astrophys. J. **244**, L1+ (1981). doi:10.1086/183466
16. S. Dehaene, L. Naccache, Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition **79**(1), 1–37 (2001)
17. G. Edelman, G. Tononi, A Universe of Consciousness: how Matter Becomes Imagination (Basic Books, New York, 2001)
18. B. Fort, Y. Mellier, Arc(let)s in clusters of galaxies. Astron. Astrophys. Rev. **5**, 239–292 (1994). doi:10.1007/BF00877691
19. B. Fort, Y. Mellier, M. Dantel-Fort, H. Bonnet, J.P. Kneib, Observations of weak lensing in the fields of luminous radio sources. Astron. Astrophys. **310**, 705–714 (1996)
20. B. Fort, Y. Mellier, J.P. Picat, Y. Rio, G. Lelievre, Multiaperture spectroscopy with rapid mask fabrication and installation. Presented at *The Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, vol. 627, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. by D. L. Crawford (SPIE, Bellingham, 1986), pp. 321–327
21. B. Fort, C. Morel, G. Spaak, The reduction of scattered light in an external occulting disk coronagraph. Astron. Astrophys. **63**, 243–246 (1978)
22. B. Fort, J.P. Picat, M. Dantel, J.L. Leroy, Coronal densities and temperatures derived from monochromatic images in the red and green lines. Astron. Astrophys. **24**, 267 (1973)
23. B. Fort, J.L. Prieur, G. Mathez, Y. Mellier, G. Soucail, Faint distorted structures in the core of A 370 – are they gravitationally lensed galaxies at Z about 1? Astron. Astrophys. **200**, L17–L20 (1988)
24. R. Gavazzi, T. Treu, J.D. Rhodes, L.V.E. Koopmans, A.S. Bolton, S. Burles, R.J. Massey, L.A. Moustakas, The sloan lens ACS survey. IV. The mass density profile of early-type galaxies out to 100 effective radii. Astrophys. J. **667**, 176–190 (2007). doi:10.1086/519237

25. H. Hoekstra, Y. Mellier, L. van Waerbeke, E. Semboloni, L. Fu, M.J. Hudson, L.C. Parker, I. Tereno, K. Benabed, First cosmic shear results from the Canada-France-Hawaii telescope wide synoptic legacy survey. Astrophys. J. **647**, 116–127 (2006). doi:10.1086/503249
26. H. Hoekstra, H.K.C. Yee, M.D. Gladders, Properties of galaxy dark matter Halos from weak lensing. Astrophys. J. **606**, 67–77 (2004). doi:10.1086/382726
27. N. Kaiser, Nonlinear cluster lens reconstruction. Astrophys. J. **439**, L1–L3 (1995). doi: 10. 1086/187730
28. N. Kaiser, G. Squires, Mapping the dark matter with weak gravitational lensing. Astrophys. J. **404**, 441–450 (1993). doi:10.1086/172297
29. J.P. Kneib, Y. Mellier, B. Fort, G. Mathez, The distribution of dark matter in sistant cluster lenses – modelling A:370. Astron. Astrophys. **273**, 367 (1993)
30. C.S. Kochanek, R.D. Blandford, Gravitaional imaging by isolated elliptical potential wells. II. Probability distributions. Astrophys. J. **321**, 676 (1987). doi:10.1086/165661
31. L.V.E. Koopmans, T. Treu, A.S. Bolton, S. Burles, L.A. Moustakas, The sloan lens ACS survey. III. The structure and formation of early-type galaxies and their evolution since z˜1. Astrophys. J. **649**, 599–615 (2006). doi:10.1086/505696
32. I. Kovner, The marginal gravitational lensing. Astrophys. J. **321**, 686–705 (1987). doi:10.1086/165662
33. R. Lynds, V. Petrosian, Luminous arcs in clusters of galaxies. Astrophys. J. **336**, 1–8 (1989). doi:10.1086/166989
34. R. Massey, J. Rhodes, A. Leauthaud, P. Capak, R. Ellis, A. Koekemoer, A. Réfrégier, N. Scoville, J.E. Taylor, J. Albert, J. Bergé, C. Heymans, D. Johnston, J. Kneib, Y. Mellier, B. Mobasher, E. Semboloni, P. Shopbell, L. Tasca, L. Van Waerbeke, COSMOS: three-dimensional weak lensing and the growth of structure. Astrophys. J. Suppl. **172**, 239–253 (2007). doi:10.1086/516599
35. Y. Mellier, Probing the universe with weak lensing. Ann. Rev. Astron. Astrophys. **37**, 127–189 (1999). doi:10.1146/annurev.astro.37.1.127
36. Y. Mellier, B. Fort, J. Kneib, The dark matter distribution in MS 2137-23 from the modeling of the multiple arc systems. Astrophys. J. **407**, 33–45 (1993). doi:10.1086/172490
37. J. Miralda-Escude, Gravitational lensing by clusters of galaxies – constraining the mass distribution. Astrophys. J. **370**, 1–14 (1991). doi:10.1086/169789
38. R. Narayan, M. Bartelmann, Lectures on gravitational lensing. ArXiv e-prints, [astro-ph/9606001] (1996)
39. B. Paczynski, Giant luminous arcs discovered in two clusters of galaxies. Nature **325**, 572–573 (1987). doi:10.1038/325572a0
40. A. Refregier, Weak gravitational lensing by large-scale structure. Ann. Rev. Astron. Astrophys. **41**, 645–668 (2003). doi:10.1146/annurev.astro.41.111302.102207
41. A. Refregier, The dark UNiverse explorer (DUNE): proposal to ESA's cosmic vision. Exp. Astron. **23**, 17–37 (2009). doi:10.1007/s10686-008-9106-9
42. S. Refsdal, The gravitational lens effect. Mon. Not. R. Astron. Soc. **128**, 295 (1964)
43. P. Schneider, J. Ehlers, E.E. Falco, *Gravitational Lenses* (Springer, Berlin/Heidelberg/New York, 1992)
44. P. Schneider, C. Seitz, Steps towards nonlinear cluster inversion through gravitational distortions. 1: basic considerations and circular clusters. Astron. Astrophys. **294**, 411–431 (1995)
45. P. Schneider, L. van Waerbeke, Y. Mellier, B. Jain, S. Seitz, B. Fort, Detection of shear due to weak lensing by large-scale structure. Astron. Astrophys. **333**, 767–778 (1998)
46. G. Soucail, B. Fort, Y. Mellier, J.P. Picat, A blue ring-like structure, in the center of the A 370 cluster of galaxies. Astron. Astrophys. **172**, L14–L16 (1987)
47. G. Soucail, Y. Mellier, B. Fort, G. Mathez, M. Cailloux, The giant arc in A 370 – spectroscopic evidence for gravitational lensing from a source at $Z = 0.724$. Astron. Astrophys. **191**, L19–L21 (1988)
48. M. Tompsett, G. Amelio, W. Bertram, R. Buckley, W. McNamara, J. Mikkelsen, D. Sealer, Charge-coupled imaging devices: experimental results. IEEE T. Electron. Dev. **18**, 992–996 (1971). doi:10.1109/T-ED.1971.17321

49. W.H. Tucker, H. Tananbaum, R.A. Remillard, A search for ‘failed clusters’ of galaxies. Astrophys. J. **444**, 532–547 (1995). doi:10.1086/175627
50. J.A. Tyson, R.A. Wenk, F. Valdes, Detection of systematic gravitational lens galaxy image alignments – mapping dark matter in galaxy clusters. Astrophys. J. **349**, L1–L4 (1990). doi:10.1086/185636
51. F. Valdes, J.F. Jarvis, J.A. Tyson, Alignment of faint galaxy images – cosmological distortion and rotation. Astrophys. J. **271**, 431–441 (1983). doi:10.1086/161210
52. L. Van Waerbeke, Y. Mellier, T. Erben, J.C. Cuillandre, F. Bernardeau, R. Maoli, E. Bertin, H.J. Mc Cracken, O. Le Fèvre, B. Fort, M. Dantel-Fort, B. Jain, P. Schneider, Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. Astron. Astrophys. **358**, 30–44 (2000)
53. D. Walsh, R.F. Carswell, R.J. Weymann, 0957 + 561 A, B – twin quasistellar objects or gravitational lens. Nature **279**, 381–384 (1979). doi:10.1038/279381a0

Part IV
Astronomy at the Frontiers of Knowledge

Chapter 17
Astronomy Versus Astrology

Marek Artur Abramowicz

> *What the populace once learned to believe without reasons,*
> *who could refute it to them by means of reasons?*
>
> Friedrich Nietzsche*, Thus Spake Zarathustra* (3:LXXIII:9)

Abstract Johannes Kepler was a great astronomer and a devoted astrologer. He tried to improve these two disciplines according to his grand mystic vision, based on Pythagorean musical and geometrical harmonies. However, Kepler's enduring analysis of accurate observational data proved that the real nature of planetary motion, summarized in his Three Laws, does not follow from the Pythagorean harmonies. Kepler's discovery of the Three Laws completely reformed astronomy and opened the avenue for modern science. Astrology has experienced no change from Kepler's time. It is still an art of divination, based on groundless, arbitrary and unprecise rules and on embarrassingly loutish mathematics.

Keywords History and philosophy of astronomy

1 Sicut in Caelo et in Terra: Misconceptions About Astrology

There is a widespread acceptance of astrology in all strata of our society. Not only common people, but also celebrities – famed writers, artists, prominent politicians, journalists, successful lawyers, and sport's champions – openly admit their belief in horoscopes. I doubt that this could result *only* from the terror of political correctness that denies the existence of the objective truth and demands that all personal opinions, no matter how naive or absurd, should be equally respected and accepted. Certainly, astrology *is* very politically correct. One cannot belittle this by hoping, that the political correctness represents only an irritating fashion of

---

M.A. Abramowicz (✉)
Department of Physics, University of Gothenburg, 412-96 Gothenburg, Sweden

Copernicus Astronomical Center, Warsaw, Poland

Institute of Astronomy, Czech Academy of Sciences, Prague, Czech Republic
e-mail: marek.abramowicz@physics.gu.se

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_17,

today, and that all fashions are transient. I am afraid that the acceptance of astrology reflects something deeper, and far more disturbing. We may witness here a genuinely confused world-view of many of our contemporaries. In their minds astronomy, astrology, and religious faith are hopelessly mixed up – as was commonplace prior to the Scientific Revolution.

1.1 Kepler the Astronomer

The Scientific Revolution started by labors of Johannes Kepler, Galileo Galilei, Isaac Newton, and others. Kepler discovered his famous laws of heavenly motions of planets. Galileo described terrestrial motions of projectiles and falling bodies. Newton understood that Kepler's heavenly motions and Galileo's terrestrial motions are governed by the same mathematical principles – sicut in caelo et in terra.

Newton's monumental *Philosophiæ Naturalis Principia Mathematica* (1687) created modern science. During the Enlightenment, the change of the paradigm was firmly established. Faith, science and magic have clearly separated, "and never the twain shall meet". Science abandoned arbitrary rules based on superficial illusions, and adopted a method based on testable observations and their objective mathematical analysis.

For Kepler, God was still the key ingredient of the scientific method. Kepler's beautifully Pythagorean model of a heliocentric universe [1] was carefully constructed in the *image and likeness* of the Triune God. In Kepler's mind the universe should be spherical and finite, because only then it could have the perfect shape and be equipped with the three necessary attributes: (*1*) the center occupied by the Sun as an image of the *Father*, (*2*) the spherical boundary of fixed stars as an image of the *Son*, and (*3*) the interior as an image of the *Holly Spirit*. In Kepler's own words [1]:

> Before the universe was created, there were no numbers except the Trinity, which is God himself (...) For, the line and the plane imply no numbers: here infinitude itself reigns. Let us consider, therefore, the solids. We must first eliminate the irregular solids, because we are only concerned with orderly creation. There remain six bodies, the sphere and the five regular polyhedra. To the sphere corresponds the heaven. On the other hand, the dynamic world is represented by the flat-faces solids. Of these there are five: when viewed as boundaries, however, these five determine six distinct things: hence the six planets that revolve about the sun. This is also the reason why there are but six planets.

Although Kepler was very much impressed and influenced by the ideas of Nicolaus Copernicus, he disagreed with Copernicus, with Giordano Bruno, and with several other Copernicans, who clearly opted for an infinite universe.[1] Kepler's trinitarian

---

[1] Kepler accepted Bruno's theory of infinite worlds, but not his infinite universe [39]. The two men never met in person. In 1588 Bruno came to visit the Emperor Rudolph II in Prague, 12 years before Kepler did the same. In 1600, on February 17, Bruno was burned at the stake in Piazza di Campo dei Fiori in Rome, after a log trial and 8 years in prisons in Venice and Rome. Bruno was accused not only of "claiming the existence of a plurality of worlds and their eternity" but also of

argument for a finite universe, based on the triad "center-interior-surface" is not logically complete, for there are finite geometrical objects with spherical symmetry that have no surface or center. According to modern cosmology, the expanding space in our spherically symmetric Universe has only the interior, but no center and no surface, independent on whether it is infinite, or finite. The "Sierpinski carpet", a plane fractal first described by Wacław Sierpiński [40], the famous Polish mathematician, is even stranger: it has no center, no interior, and no boundary.

The concentric spheres corresponding to the six planets, Mercury, Venus, Earth, Mars, Jupiter, and Saturn, have been inscribed and circumscribed in Kepler's model by the five Platonic regular solids, nesting these spheres. Kepler was firmly convinced that there are exactly six planets in Nature *because* there are exactly five regular solids in geometry. The proof that there are exactly five regular solids was given by Euclid in *Elements* (circa 300 BC). Plato in *Timaeus* (360 BC) associated the "four elements", which everything terrestrial consists of, with four of the five regular solids: earth with cube, air with octahedron, water with icosahedron and fire with tetrahedron. Aristotle added the "fifth element" (quintessence, aether) of which heavenly bodies are made, and associated it with dodecahedron. The fifth element was incapable of changes, except that it was experiencing a "natural state of motion" along perfect circles.[2] This was why all astronomers before Kepler were convinced that planets must move along circles, and why they were adding epicycles on epicycles. Only much later in his life, Kepler proved that planets move along "unperfect" ellipses, with the still Sun located exactly in the ellipse's focus.

Kepler found a unique way of ordering the five regular solids in space that gave the observed (relative) sizes of each planet's orbit: octahedron (the innermost), icosahedron, dodecahedron, tetrahedron and cube (the outermost). While sizes of planetary orbits in Kepler's universe were related to a geometrical Platonic harmony, the orbital angular speeds (as seen from the Sun) were given by a Pythagorean harmony of musical tones [14]: the higher the angular speed, the higher the pitch of a tone.

Kepler's grand mystic ideas about the nature of planets, the Solar System and the Universe, were all wrong: it is *not* true that there must be exactly six planets in the Solar System, it is *not* true that the angular speeds of them must agree with musical intervals, it is *not* true that the relative distances between planets must be determined by shapes of the regular polyhedrons. Kepler failed miserably as a visionary mystic.

---

seven other grave charges including immoral conduct and heresy in matters of dogmatic theology, among them "denying the Virginity of Mary".

[2] Medieval scholastics thought, contrary to Aristotle, that the *quintessence* might change its density [29, p. 422]. It is perhaps amusing to note that modern cosmology came back to Aristotle: its quintessence, the Dark Energy, does not change density during the cosmological expansion, which is its "natural" state of motion.

He was very right, however, as a scientist. Paradoxically, his erroneous mystic ideas helped him discover the true physical laws governing planetary motions – the immortal Three Kepler's Laws.[3] *Spiritus flat, ubi vult.*

1.2 Astrology Versus the Church

Kepler's way of thinking about Nature (and God) seems today bizarre, if not absurd.[4] It was already rejected during the Enlightenment, about a century after the Kepler's time. "I had no need of that hypothesis", replied coldly Pierre-Simon de Laplace to Napoleon Bonaparte's question as to why there is no mention of the name of God in his *Mécanique céleste* (1799–1825). Laplace's arrogant bon-mot is of course well remembered and often quoted. Perhaps less known is its remarkable rebuttal by Joseph-Louis Lagrange (to whom Napoleon mentioned Laplace's answer): "Ah, but this is a fine hypothesis! It explains many things". This stinging remark was also very profound. Indeed, *that hypothesis* may not be necessary to explain Laplace's celestial mechanics, but nobody knows whether it is, or it is not, necessary to explain the *"many things"* that Lagrange might have meant. One of them could be, perhaps, the Kantian *moral law within us*. For the moral law, free volition, consciousness, and other *"many things"* stand outside the reach of science as much today, as they stood in the time of Laplace, and earlier of Kepler, and still earlier of Plato – and as maybe they will always stand.

Astrology pretends to form a consistent world-view, linking the patterns in starry heavens with the morality and fate of man. Some see here an analogy with "sicut in caelo et in terra" from the *Lord's Prayer*, and mistakenly assume that astrology and Christian religion act similarly. Following this assumption, they wrongly conclude that astrology and religion must be therefore natural allies in their disapproval of the basic principles and methods of science. This is perhaps the most persistent modern confusion about astrology (and religion). The truth is very different, however. Through its whole history, the Christian Church was strongly opposing astrology, for reasons of uncompromisingly *fundamental* character. For the Catholic Church, astrology is a dangerous delusion that challenges not only the dogma about the free will of man,[5] but even God's omnipotence. Besides, man could *not* know his

---

[3] Kepler's life-long quest for understanding the universe was very dramatic. It is described in several excellent books. My favorites are Arthur Koestler's *Sleepwalkers* [33], and Jerzy Kierul's *Kepler* [32]. Experts value the seminal monograph by Max Caspar [23], who was the editor of Kepler's collected works [18]. Also highly recommended are [41] and [35].

[4] A recent exhaustive (500 pages) monograph on the Holly Trinity [30] ignores Kepler's trinitarian model of the universe. It seems that this particular Kepler's idea embarrasses today not only astronomers but also theologians.

[5] Astrology's conflict with the Christian notion of free will has been pointed out as a serious argument against astrology by several Renaissance humanists, including Giovanni Pico della Mirandola [38].

fate: “Therefore, stay awake, for you know neither the day nor the hour” (Matthew 25:13). The official *Catechism of the Catholic Church*, approved by the Pope John Paul II in 1997, explains plainly the Church’s very negative standing with respect to astrology. Its item 2,116 teaches: “All forms of divination are to be rejected: (…) horoscopes, astrology, palm reading, interpretation of omens and lots, the phenomena of clairvoyance, and recourse to mediums (…). They contradict the honor, respect, and loving fear that we owe to God alone”.

St. Augustine, who lived in the period 354–430 AD and who is one of the most respected Christian philosophers (by the Catholic, Protestant, and Eastern Orthodox Churches alike) wrote against astrology in equally strong terms [20]: “Hence, a devout Christian must avoid astrologers (...) for fear of leading his soul into error by consorting with demons and entangling himself with the bonds of such association”.

A generation earlier, in 321 AD, the Emperor Constantine the Great, who like St. Augustine was a former pagan converted to Christianity, issued the famous edict that condemned all practicing astrologers to death.[6] As a result of its rejection by the Christian Church, astrology disappeared for centuries from medieval Europe. It came back at the dawn of the Renaissance, helped by emerging science. In this second period (after the Antiquity) of astrology blooming in Europe, scholars, artists and many learned aristocrats rediscovered treasures of the old Hellenic, Egyptian and Jewish[7] wisdom. The great intellectual ferment was accompanied by a decline in the Church and imperial powers, and a growing social disorder. In these very confused circumstances, the importance of astrology grew. Court astrologers were employed by Popes and Emperors, as well as by many rulers of a much smaller significance. Cities, towns and provinces also had their official astrologers. The most remarkable of them was Johannes Kepler, the Imperial Mathematician at the Prague court of the Holy Roman Emperor Rudolph II, the Habsburg.

1.3 *Praga Magica*

During Rudolph’s rule (1583–1612), Prague became the leading center of art and science, successfully rivaling Paris, Rome and Vienna [34]. Fascinated by the new Renaissance mood, Rudolph gathered around himself many unusual celebrities.

---

[6] Maria Dzielska in her well-known monograph *Hypatia of Alexandria* [24] wrote that a direct cause for Hypatia’s cruel death could be a false rumor inspired by bishop Cyril, who was Hypatia’s fierce enemy. The rumor accused Hypatia of conducting the forbidden astrological and magical practices. Hypatia, a great philosopher and mathematician, admired by the Alexandrian elite for her knowledge, wisdom, righteousness of character and personal charm, was murdered in 415 AD by a fanatic Christian mob.

[7] Astrological speculations are important in the *Cabbala*. The classical books of Jewish mysticism, *Sefer Zohar* and *Sefer Yetzirah*, give rules of divination often based on astrological calculations and interpretations. However, the medieval Jews distinguished astronomy (the science of stars) from astrology (the art of divination).

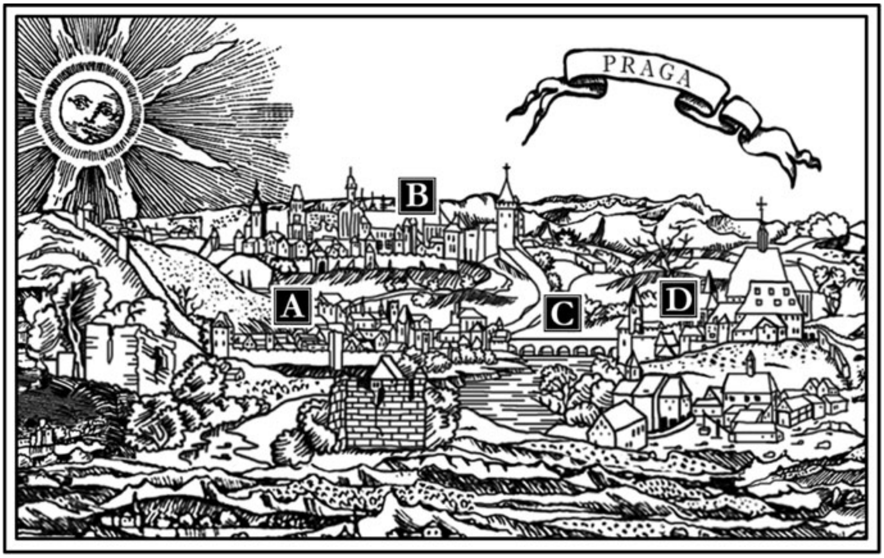


Fig. 17.1 *Praga Magica*: I started writing this essay during my 3 months stay at *Mala Strana* [A] in the heart of the old Prague, at the foot of the *Hradčany Castle* [B], which was the Imperial Seat of Rudolph's II court, and a short walk across the *Charles Bridge* [C] from both *Kepler's* and *Faustus'* houses [D]. This Figure is inspired (not just copied) by a double-page woodcut in the well-known Hartmann Schedel's *Chronicle*, Nuremberg (1493), which was published only a relatively short time after the Johannes Gutenberg's *Bible*, Mainz (1453). This early view of Prague is also used as a detail in Fig. 17.2

The famous painter Giuseppe Arcimboldo was Rudolph's court portraitist. His surrealistic imagination created bizarre illusions in his paintings (Fig. 17.1). He composed fantastically weird portraits of his subjects entirely of fruits, vegetables and flowers, like the well known Rudolph's portrait as *Vertumnus*. Scientists of fame — Kepler, Tycho Brahe, Giordano Bruno, or botanist Charles de l'Ecluse among others — have been residents or frequent visitors to Rudolph's Prague. Rudolph was a very serious practitioner of astrology and alchemy all his life. Nostradamus prepared his horoscope when Rudolph was a prince. Many Europe's best alchemists, including Edward Kelley and John Dee, were working together with him in Prague, trying to find the Philosopher's Stone. Rudolph discussed secrets of Cabbala with rabbi Judah Loew, creator of the legendary Golem. Doctor Faustus also had a Prague connection. Although Christopher Marlowe set his *The Tragical History of Doctor Faustus* (1604) in Wittenburg, there is in Prague a mysterious "Faustus' House" believed by some to be the place where Faustus was performing his diabolical experiments. Rudolph's Renaissance circle, and in different times Wolfgang Amadeus Mozart, Albert Einstein, Franz Kafka and Karel Čapek, as well as Prague's dreamlike castles, palaces and bridges, created her reputation as the magical capital of Europe.

1.4 Kepler the Astrologer

Kepler the astronomer has a sure place in history as one of the greatest creators of the modern science. Kepler the astrologer is perhaps best remembered as a troubled mystic in the dark legend of *Praga Magica*.

I have chosen Kepler to be the main protagonist of my essay, because his conflicting duality illustrates the conflict "astronomy *versus* astrology", which is the main topic here. It should be stressed that Kepler was *firmly convinced* that positions of planets at the moment of birth determine the character, and perhaps also the fate, of the newborn. He *was* an astrologer. He wrote clearly [14, Chapter 7]:

> The soul of the newly born baby is marked for life by the pattern of the stars at the moment it comes into the world, unconsciously remembers it, and remains sensitive to the return of configurations of a similar kind.

This conviction was probably rooted in Kepler's life-long idea, vague but very strongly imprinted in his mind, that the basic physical laws that govern the planets, should be the same as those that govern phenomena on Earth, *sicut in caelo et in terra*.

I would like to recall the well-known polemic between Kepler and Galileo about the nature of tides, in order to stress this point. In 1607 Kepler wrote in a letter to a friend [18, XV:387:29]: "From a German author the following speculation arose in my mind. The seas are attracted by the moon, as all heavy bodies and the seas themselves are attracted by the earth. But the attraction extorted by earth is stronger. Therefore the seas do not leave the earth and do not rise up into the air." As it is well known, Galileo had another (wrong) explanation. He thought that the tides occurred for a purely kinematic reason, caused by a combination of Earth's daily rotation around its axis and its yearly circuit around the Sun. This combination, according to Galileo's misconception, should cause periodic daily differences in rotations of Earth's crust and Earth's body of waters, and therefore also the tides. Galileo strongly criticized Kepler's idea [27]; page 462 in Drake's translation: "Of all great men who have speculated about this marvelous natural phenomenon, I am surprised more by Kepler than by the others. He, whose mind was free and acute, and who had at his disposal the motions ascribed to the earth, nevertheless lent his ear and gave his consent to the moon's domination over the water and to occult qualities and similar childish notions". Galileo criticized Kepler's model for tides because "the Moon's domination over the sea" sounded as a reference to Moon's well-known astrological association to water.[8] Kepler, however, had in mind not this "occult quality" of the Moon, but her gravity – a real physical attribute that affects "all heavy bodies".

To Kepler, it was very likely that as the Moon affects the ocean on Earth and induces tides, she and other planets might also somehow affect the man,

---

[8] For example, in Shakespeare's *Hamlet* (I,1), Horatio refers to the Moon by saying: "...the moist star upon whose influence Neptune's empire stands..."

by means of a yet undiscovered physical influence. This steamed his profound interest in astrology. The first manifesto of Kepler's astrological research program was described in his book *De Fundamentis Astrologiae Certioribus* [2], entirely devoted to astrology. Kepler all his life was searching for these "more certain foundations of astrology", connected to a physical influence of planets upon man. He was deeply dissatisfied with the foundations and practises of astrology of the time. In particular, Kepler was well aware that the astrology's main concepts, like "houses" and "Zodiac", are superficial and arbitrary. Kepler's suggested remedy for astrology's faults was based on his beloved harmonic Pythagorean framework. But it could not work.

In both, astronomy and astrology, Kepler was guided by erroneous ideas, particular only to him, and based on a strange convolution of the Pythagorean harmonies with the adamant Protestant faith. However, Kepler's mystical vision of the Solar System delivered a very useful byproduct – the correct mathematical description of the planetary motion, the Kepler's Laws. This created a breakthrough in astronomy. Thanks to this development, four centuries later men were walking on the Moon.[9] Nothing like this has happened to Kepler's astrology. The astrologers of today who recall Kepler's interest and involvement in astrology as an argument in astrology's favor, are abusing facts. Undoubtedly, Kepler was a practising astrologer, honestly and truly convinced in astrology's predictive powers.[10] Nobody questions this. However, Kepler's conviction was based on his hope that eventually "more certain foundations of astrology" would replace the arbitrary and superficial principles that existed in his time. Astrology experienced no such progress. Still today it is just an art of divination, based on arbitrary, unprecise and groundless rules.

---

[9] A voyage to the Moon was in fact Kepler's dream. In 1634, 4 years after Johannes Kepler's death, his son Ludwig Kepler published *The Dream (Somnium)* [17], his father's last book (written before 1610), which described adventures of an Icelandic voyager transported to the Moon by aerial demons. In *The Dream* Kepler depicted lunar inhabitants, her fauna and flora, and also the lunar astronomy. I will quote only a fragment most relevant here: "(...) the main features of the entire universe: the twelve celestial signs, solstices, equinoxes, tropical years, sidereal years, equator, colures, tropics, arctic circles, and celestial poles, are all restricted to the very tiny terrestrial globe, and exist only in the imagination of the earth-dwellers. Hence, if we transfer the imagination to another sphere, everything must be understood in an altered form." Kepler went no further. He described no details of the lunar astrology. I think, I may guess why. From the above quoted text it is obvious that Kepler knew that lunar "houses" and lunar "Zodiac" were totally different than those defined on Earth, and therefore they could not have an absolute meaning. Even his beloved "aspects" were to be suspected. Although aspects not involving Moon or Earth are only *slightly* different for terrestrial and lunar astrologers, on different planets they are *very* different. For an astrologer on Saturn, the permanent conjunction of Sun, Mercury and Venus would leave no room for certain horoscope interpretations.

[10] Kepler has casted about 800 horoscopes, not only in order to earn money, but also out of his genuine interest. Most of his adult life, his professional duties included casting horoscopes and publishing yearly astrological prognostic.

2 A Short Review of Astrological Methods and Practices

In the previous Section, I critically discussed the most persistent misconceptions about astrology, all directly connected to astrology's world-view. Now, I wish to discuss astrology's methodology and practices. They are doubtful and shockingly unsophisticated. Even *if* the planetary patterns in the sky *were* entangled with the newborn man's fate, as astrology claims, the existing astrological methods would not be able to pick this up, for they are totaly inadequate for the supposed goal.

A *natal horoscope* gives positions of planets inside the astrological *houses* at the exact moment of birth. Houses result from a division of the whole firmament, seen at this very moment. Usually, the four *cardinal points* are important in these divisions. The first pair of them, *ascendent*, and *descendent*, represents the two points where the ecliptic crosses the horizon. The second pair, *Imum Coeli* and *Medium Coeli*, represents the ecliptic and meridian crossing. Figure 17.2 show the firmament and the four cardinal points. The location of planets in the houses and in the Zodiacal signs, as well as their various *aspects*, are the basis for the horoscope interpretation.

2.1 The Houses

Although the whole firmament is divided into houses, usually only the resulting (projected) division of the ecliptic is taken into account in the astrological practice. There is no consensus about how to divide the firmament into houses. Even the number of them is not fixed. Most often twelve houses are used, but in the Antiquity there was a version of astrology based on the eight-house division. It was not popular in Kepler's time, but still in use. Some practitioners of astrology considered even its sixteen-house modification. The most famous of them was Lord William Brouncker, a noted mathematician who worked on continued fractions and calculated logarithms by infinite series, and became the first president of the Royal Society of London [37]. Even for the standard twelve-house astrology, there is no unique division into houses agreed by astrologers. They use very different methods, all of then arbitrary, which give often *very different* results. I only shortly describe a few of them below (for more details see e.g. an excellent monograph by J.D. North [37]).

1. *The Placidus method*. The cups of the houses are at the four cardinal points, Asc, Dsc, IC, MC, and at the ecliptic crossings with the "hour lines". This is the most widely used methods by modern astrologers in the English speaking countries.
2. *The Standard Method*. The cusps of the houses are again at the four cardinal points, Asc, Dsc, IC, MC, and the ecliptic crossings with the projection circles joining the North and South celestial poles with the uniform division of the cardinal arcs at the equator.

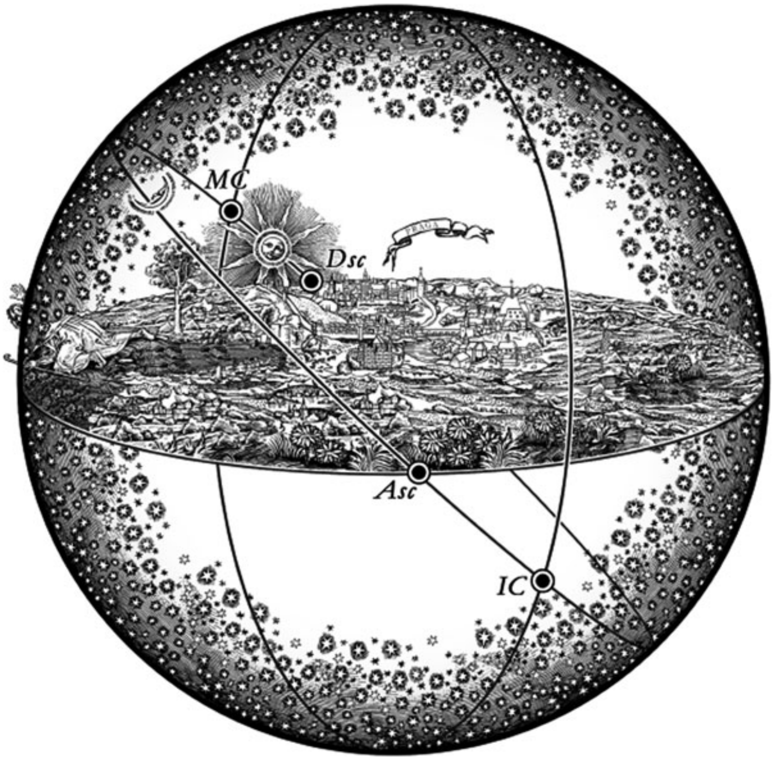


Fig. 17.2 The firmament, horizon (*horizontal circle*), vertical (*vertical circle*), ecliptic (*oblique circle*) and the four cardinal points: 1. ascendent (Asc) called the "rising sign", located in the Eastern part of the firmament, 2. descendent (Dsc), 3. Medium Coeli (MC) called "the midhaeven", and 4. Imum Coeli (IC) called "the bottom of the sky", located below the horizon. For the European, and all other Northern hemisphere observers, IC is in the Northern, and MC is in the Southern, part of the firmament. The illustrative background of this Figure is a composite of a few well-known images, including the famous Camille Flammarion's figure [26] once believed to be a medieval woodcut. It also includes the early view of Prague already mentioned in the Fig. 17.1 caption

3. *The Campanus method Method*. Projection circles join N′, S′ (the North and South points on the horizon) to the uniform (30°) division of the vertical. See Fig. 17.3a.
4. *The Regiomontanus method*. Projection circles join N′, S′ to the uniform (30°) division of the equator. See Fig. 17.3b.
5. *The Equatorial Method*. Projection circles join the North and South celestial poles to the uniformly divided equator.

Nowadays, there are several popular versions of the above methods, for example these proposed by Koch or Krusinski. I will not describe them any further.

Of course, Kepler was using houses, because this was then a standard practice of astrology. Kepler's important (and often learned) clients would probably not accept

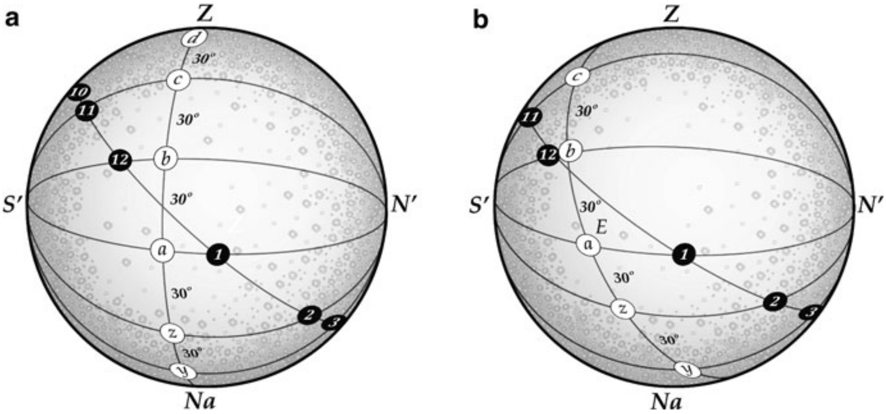


Fig. 17.3 **(a)** Left: The Campanus method of dividing firmament into houses, based on a uniform division of the vertical (Z-a-Na). Z = the Zenith and Na = the Nadir. On the horizon (S′-a-N′): N′ = the North point, S′ = the South point and E′ = the East point. The points of the uniform vertical division are shown by small circles labeled a, b, c, ... The cups of the houses are shown by small black circles, labeled by 1, 2, 3, ... The first house (*horoscopus*) is in between the 1–2 cusps. **(b)** Right: The Regiomontanus method, based on a uniform division of the equator (c-b-a-z-y). Notation the same as for the figure on the left

their horoscopes with no division into houses. However, he considered houses to be an arbitrary construction, with no significance to the real, yet undiscovered, “physical astrology” that he believed would be eventually based on aspects.[11]

2.2 *The Zodiac*

In popular “horoscopes”, published daily in countless newspapers everywhere in the world, the only question that matters is in which Zodiacal sign the Sun was located at the moment of birth of a person. For example, under the tropical zodiac, the Sun is in Pisces roughly from February 18 to March 20, ending on the moment of vernal equinox. People born during this period are called “Pisces” (or sometimes “Pisceans”). A typical newspaper explanation of their character may say, for example,

> The majority of Pisces are kind and gullible. This gentle sign cannot hurt your directly, it is their weaknesses that can scramble your brains. The young Pisces are adventurous, ambitious, impulsive, enthusiastic and full of energy. The Pisces are pioneers both in thought and action, very open to new ideas and a lover of freedom. These people possess a curiously natural understanding, which they do not obtain from books or study. They easily acquire, or rather absorb, knowledge, especially of the history of countries, travel, research, and like subject, etc., etc., etc., ...

---

[11] I am grateful to Jerzy Kierul, the author of a recent, very detailed and competent, best-selling monograph on Kepler [32] who privately explained this point to me.

Fig. 17.4 The twelve signs of Zodiac: Aries (The Ram), Taurus (The Bull), Gemini (The Twins), Cancer (The Crab), Leo (The Lion), Virgo (The Virgin), Libra (The Scales), Scorpio (The Scorpion), Sagittarius (The Archer), Capricorn (The Sea-Goat), Aquarius (The Water Bearer), Pisces (Fish). (*Figure source: wikipedia*)

Professional astrologers consider (and interpret) location of the Zodiacal signs in the houses, and all planets in the Zodiacal signs.

There is an English mnemonic which helps remember the order of the Zodiacal signs (Fig. 17.4),

The Ram, the Bull, the Heavenly Twins,
And next' the Crab, the Lion shines,
The Virgin and the Scales.

The Scorpion, Archer, and the Goat,
The Man who holds the Watering Pot,
And Fish with glittering scales.

2.3 The Planets

In Antiquity, astrology considered the *six wondering lights*: Sun and Moon (the Lights), and Mercury, Venus, Jupiter, Saturn (the Planets). Later Uranus, Neptune and Pluto have been added. Pluto became a very popular planet in astrology, and its recent astronomical degradation out of the planetary status was an astrological embarrassment.[12]

Planetoids (in particular Ceres, Pallas, Vesta, Juno and Hygiea) are included by some astrologers in their horoscopes. In this respect, let me make a remark. I could see a case when a large ($\sim$10 km) planetoid could have a direct and accurately calculable influence on my fate – if it would be discovered on a collision course with Earth.

Comets have no clear status in astrology, probably because they have been considered in the Antiquity as atmospheric (sub-Lunar) phenomena.

Dark planets are hypothetical bodies in the Solar System that some astrologers introduce in order to explain *ex post* failures in their horoscope predictions.[13] The idea behind the astrological dark planets is simple. (1) Suppose, there is an unknown planet in the Solar System, located at F in the firmament, (2) Suppose, it causes an influence X on the event $A = f(X,F)$, (3) Therefore, the horoscope's prediction of the nature of A cannot be correct, because we do not know F. (4) However, we may invert the problem and find $F = g(A,X)$. The logic here is similar to that involved in the astronomy's argument for the presence of the Dark Matter in our Galaxy: theory is correct, but there is something unseen ("dark") that causes the trouble.

2.4 The Aspects

According to the astrological doctrine, aspects are the focal points in horoscopes, and give extra emphasis in the horoscope interpretation. An aspect is the difference $\Delta\lambda$ in the ecliptic longitude of a planet (measured in degrees[14]) with another planet, the ascendant, midheaven, descendant or other important points in the firmament. In the *major* aspects $\Delta\lambda = N \times 30^\circ$, where N is a natural number or zero ($N = 0,...,11$). The particularly important major aspects are: conjunction ($\Delta\lambda \approx 0^\circ$), sextile ($\Delta\lambda \approx 60^\circ$), square ($\Delta\lambda \approx 90^\circ$), trine ($\Delta\lambda \approx 120^\circ$), and opposition ($\Delta\lambda \approx 180^\circ$). Kepler introduced *minor* aspects to astrology. The most commonly used today are: quincunx ($\Delta\lambda \approx 150^\circ$), semisquare ($\Delta\lambda \approx 45^\circ$), sesquiquadrate ($\Delta\lambda \approx 135^\circ$), semisextile ($\Delta\lambda \approx 120^\circ$), quintile ($\Delta\lambda \approx 72^\circ$), biquintile ($\Delta\lambda \approx 144^\circ$). There are (less popular) aspects based on division of the circle by seven segments (septile, biseptile, triseptile), nine segments (novile, binovile, quadnovile), or even 10, 11,

---

[12] "Pluto is the planet of death and transformation. It indicates where negotiation and compromise do not work for the person, or where the person has to change something fundamental in his or her life."

[13] Walter Gorn was the first who introduced (in 1918) a dark planet – the Earth's "dark moon" Lilith. The recent "Uranian" school of astrology claims that there are many unseen planets beyond the orbit of Neptune.

[14] Aspects judged not by difference in angles, but according to the relationship of the Zodiacal signs, are called *platick*.

14, 16, and 24 segments. Some astrologers consider aspects in declination: *parallel*, i.e. when the two planets have the same declinations (with accuracy of about $\pm 1^\circ$), and *antiparallel* i.e. when the two planets have the same declinations, but with opposite signs. Astrologers consider also "near" aspects. For example, if for Jupiter and Venus $\Delta\lambda = 94^\circ$, they say "Jupiter square Venus with the *orb* 4°".

2.5 The Horoscope Interpretation

In the previous sections, I have described the "astronomical" part of the astrological doctrine, which (mostly) gives the rule(s) of how to divide the firmament into the houses, how to place planets, Zodiacal signs, cardinal and other points into the houses, and how to find various aspects. Now its time to say what "Saturn in the 3rd house" should mean for the astrological interpretation of the horoscope. Entering such matters is embarrassing, and I will keep my description of this part of the astrological doctrine as short, as possible – but not shorter.

The houses are related to the newborn's: (1st) life, (2nd) personal property, (3rd) consanguinity, (4th) riches, (5th) children and jewels, (6th) health, (7th) marriage and course of life, (8th) manner of death and inheritance, (9th) intellect, disposition and long journey), (10th) position in life and dignities, (11th) friends and success, (12th) enemies and misfortune. In the Kepler's time, a Latin hexameter was used to help to remember the meaning of the houses:

Vita, lucrum, fratres, genitor, nati, valetudo,
Uxor, mors, sapiens, regnans, benefactaque, daemon.

The Zodiacal signs, in the standard order, (1st) Ram, (2nd) Bull, (3rd) Twins, ..., have similar meanings as the corresponding houses. The reason for meanings of houses and Zodiacal signs is rooted in mythology.

> The Fish, for example, are believed to be Aphrodite and her son Eros. They transformed into the Fish trying to escape the fire god Typhon. For this reason, the Fish are considered to be one of *mutable* signs which are extremely restless. However, they cannot lose each other, because they are firmly tied together with a cord. The Fish are obviously "watery" sign — constantly trying to adapt to mutable feelings.

I will not go further into these matters, except that I will stress the obvious here. The astrological "predictions" are based on mythology – in Europe mostly on the Greek and Roman myths, but in different cultures on different ones.

The six planets known to Kepler rule days of the week, as their Latin names indicate: dies Solis (Sunday), dies Lunae (Monday), dies Martis (Tuesday), dies Mercurii (Wednesday), dies Jovis (Thursday), dies Veneris (Friday), dies Saturni (Saturday). However, in other European languages, names of the weekdays do not match the names of the planets. In Anglo-Saxon and Scandinavian languages same of these names are connected to gods from the forgotten pagan myths – Odin (Wednesday), Thor (Thursday), etc.[15]

In astrology, the Sun, Jupiter, and Mars are masculine, and the Moon and Venus are feminine. Mercury is both masculine and feminine. The Sun, Saturn and Jupiter are the daily planets. The Moon, Mars, and Venus are the nocturnal planets. Mercury belongs to both, day and night. Two planets bring fortune: Jupiter (fortuna major) and Venus (fortuna minor). Two planets bring misfortune: Saturn (infortuna major) and Mars (infortuna minor). The remaining planets – Sun, Moon, Mercury, have a mixed character with respect to fortune.

The interpretation of the horoscope should be based on the rules that I shortly described above. Usually, however, the interpretation is a vague, non-algorithmic procedure. Astrologers, however fine, often tune their horoscopes to needs and expectations of their clients, especially of important clients. Certainly, Kepler was not an exception from this common practice, when in 1608 he casted the horoscope for Albrecht von Wallenstein; it later became one of the most famous horoscopes in history [19, I:338]. Several astronomers pointed out that the Kepler 1608 horoscope, and its rectification done by him in 1625, contain serious inconsistencies [25]. In particular, the horoscopes do not correspond to Wallensteins birth date, and they lead to 4-year shifts in the evaluation of the "directions" used to predict important events in life. It is obvious that Kepler made errors in the 1608 horoscope, and he was covering them in the 1625 follow-up by consciously changing the facts. Kepler's cheating was subtle and in places self-ironic, but one cannot deny that the rules were upset. Contemporary astrologers seem not to be aware of these problems and highly praise Kepler's horoscopes. For example, Ulrike Voltmer compared a detailed year-by-year biography of Wallenstein with her own analysis of the Kepler's 1608 horoscope and decided about its most important features: Aquarius ascendant, Placidius houses, grand conjunction in the 1st house in Pisces, South node of the Moon in Gemini in the 4th house, Mercury in Virgo in opposition to Jupiter/Saturn, Sun in Libra in 7th house, wide opposition to Jupiter/Saturn, Mercury is also in the 7th house.[16]

The Wallenstein horoscope casted by Kepler is a characteristic example of the content and style of the astrological divinations (the fragments of text indicated by boldface should be compared with the same fragments shown in Fig. 17.5):

> **Saturn in ascendency** makes deep, melancholic, always wakeful thoughts, brings inclination for alchemy, magic, sorcery, communion with spirits, scorn and lack of respect of human law and custom, also of all religions, makes everything suspect and to be distrusted which God or humans do, as though it all were pure fraud and there was much more hidden behind it than was generally assumed.

---

[15] In Slavic languages mythology is usually absent in the weekdays names. In Polish, for example, *niedziela* (i.e. Sunday) means the day of rest (no-labor). In Russian the same word, воскресенье sounds as *Resurection*.

[16] http://altairastrology.wordpress.com/2008/05/31/rhythmic-astrology-a-book-review

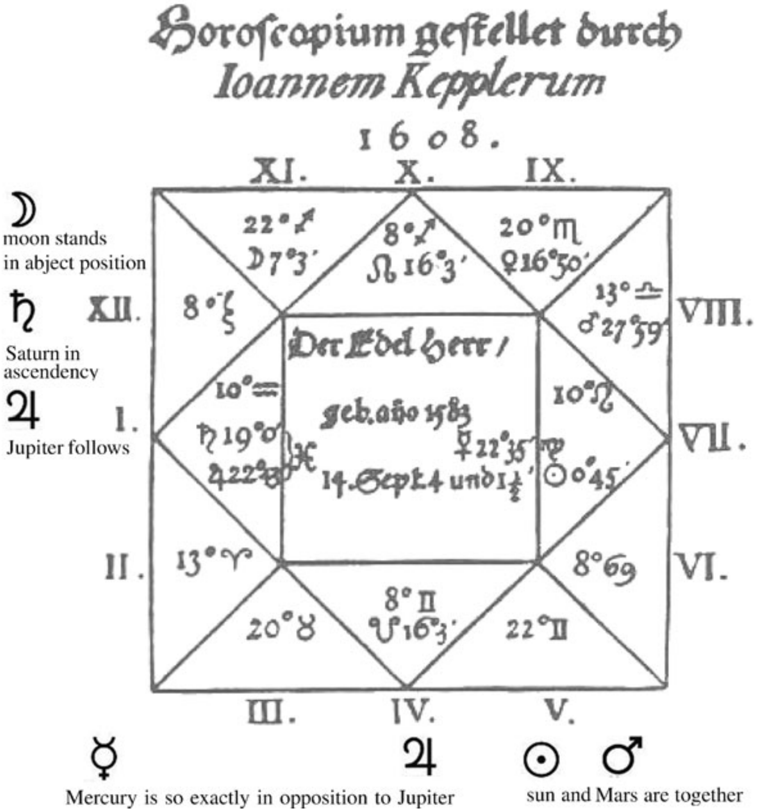


Fig. 17.5 The Wallenstein horoscope. I have superimposed around the original Kepler's woodcut the five fragments (indicated by boldface) of his interpretation of the horoscope

And because the **moon stands in abject position**, its nature would cause considerable disadvantage and contempt among those with whom he has dealings, so that he would also be formed unmerciful, without brotherly or conjugal love, esteeming no one surrendering only to himself and his lusts, hard on his subjects, grasping, avaricious, deceptive, inconsistent in behavior, usually silent, often impetuous, also belligerent, intrepid, because the **sun and Mars are together**, although Saturn spoils his disposition so that he often fears for no good reason.

But the best feature of the positions of the heavenly bodies at his birth is that **Jupiter follows**, bringing hope that with ripe age most of his faults would disappear and thus his unusual nature would become capable of accomplishing important deeds.

For with him can also be seen great thirst for glory and striving for temporal honors and power, by which he would make many great dangerous, public and concealed enemies for himself but also he would mostly overcome and conquer these. It can also be seen that this nativity has much in common with that of the former Polish chancellor, the English queen, and other similar people, who also have many planets standing around the horizon in position of rising and setting, for which reason there is no doubt, provided he only would pay attention to the course of the world, that he would acquire high honors, wealth, and after making a court connection for himself, also a high-ranking lady as his wife.

> And because **Mercury is so exactly in opposition to Jupiter**, it almost looks as though he might yield to wild schemes by means of these attract a great many people to him, or perhaps at some time be raised by a malcontent mob to a leader or ringleader.

Trying to understand astrological reasons for his own fate, character and life, Kepler determined the exact hour of his birth in Weil in Schwaben. It was 2:30pm on 27 December 1571. He also calculated the moment of his conception – 4:37am, on 17 May, 1571, two days after his parents wedding.[17] As a student in Tübingen Kepler casted his own horoscope [19, V:447]. It is very interesting today because it gives Kepler's psychological self-portrait:

> Mercury in the seventh house means speed, and sloth because he is fast. The Sun in sextile with Saturn reveals meticulousness and endurance. The following two characteristics are opposed in the human being: constant regret about lost time and constant waste of time for which one is oneself to blame. Mercury causes devotion to jest and play, also resulting from joy in the more shallow things. For as a boy I devoted myself passionately to playing. As a youth other things delighted my mind, and therefore I devoted myself to other things; thus it is a matter of judgement as to what pleases the human being. Since, however, stinginess frightened me away from play, I often played alone. Here it should be noted that the goal of the stinginess was not the acquisition of wealth, but rather it was based on the fear of poverty – although perhaps all possessiveness may stem from misapplied anxiety... Lust for money imprisons many. I took usefulness and honorableness into consideration. Perhaps much is rooted in the shame of poverty. I am, above all, not arrogant and contemptuous of public opinion, though of course my speech tends to be abrasive.
>
> When, for example, Saturn aspects Mercury, he [Mercury] becomes cold, so that the mind is dulled; when Jupiter aspects Mercury, he [Mercury] makes one moist and hot. In the former case, everything is aimed at the lust for gain, in the latter case, it is the desire for fame and honor. When Mars comes into my view, he frightens. He causes the mind collapse and drags it into anger, play, inconstancy, story-telling, into wars, into excess, into foolhardiness, into busy-ness – all things that adhere to mortal man: he [Mars] incites one to contrariness, fighting, disapproval of all order, criticism of custom. All is conspicuous that I engaged in with my studies: excitement in conversation, dispute, contempt, challenging all immoral habits of some person...

3 The Summary: A Short List of Astrology's Failures

I will now summarize the fundamental defects of astrology in a short list containing six points: (1) astrology has no objective principles, (2) astrology does not use advanced mathematics, (3) astrology offers no physical model to support its claims, (4) astrology fails in testable empirical checks, (5) astrology has no honest methodological standards, (6) there are no great astrology's masters.

---

[17] According to these calculations, his mother was pregnant only 224 days.

3.1 *Astrology Has No Objective Principles*

Astrology's principles are unacceptable for both science and Christian faith. I have discussed this earlier in my essay. Quite in addition, astrology's principles are arbitrary and not objective. Kepler was fully aware of this, as were many other scholars. For example, John Flamsteed, the first Astronomer Royal who died in 1719, criticized astrology on the basis of his solid expertise in the subject – for several years he was casting horoscopes himself, and carefully studied horoscopes casted by others. He wrote[18] [31]:

> Even if we grant the planets some influence, we must still ask how astrologers can be confident of their judgements when they do not agree on which house system to use, nor on how to use fixed stars. They agree that the stars do have an influence, and some pretend to use them when everything else fails, but they never consider their aspects, which may contradict what is promised by planetary aspects. So how can we be certain of the truth of their predictions?

Astrologers of today seem not to be very concerned about the lack of clear rules for casting a horoscope. The following quote is from a popular astrological blog, where expert astrologers give advise to beginners:

> Q: what house systems should I choose?
> A: This is, unfortunately, a subject where you may get several different subjective opinions. Personally, I use Koch for adults, Equal House for children, and study the shift in house cusps between the two to explore how a child changes as he/she grows up and enters the world. But I know some who are happy with Placidus, or Regiomantus, or Campanus, or Porphry. I know some who do not use houses at all in their work. Oh...and there are more than seven different systems.....I know of at least thirteen.
>
> An interesting technique I learned from [***], a British astrologer: If you don't know the birth time for someone, put their Sun sign on the 5th house cusp. That is, if someone has the Sun in 16 Cancer, put 00 Cancer (not 16) on the 5th house cusp, then put 00 Leo on 6th house cusp, 00 Virgo on 7th house, etc. Then put your natal planets in the houses that correspond to that. To my continuing astonishment, when you work with this chart using transits, they can reveal some mighty interesting stuff. It's a fascinating variation on the whole sign house system. It's clever in that it uses the 5th house, which is the natural house of Leo, ruled by the Sun.

No comments are needed for the above foolishness, it speaks for itself!

3.2 *Astrology Does Not Use Advanced Mathematics*

Astrology employs only embarrassingly simple mathematics, mostly spherical geometry. Its elements have been known already in the Antiquity. Ptolemy, a

---

[18] Flamsteed's intention was to publish his criticism of astrology together with an astronomical almanac (ephemeris) for 1674, but no publisher could be found because the almanac containing a criticism of astrology would not sell. The Flamsteed's text quoted here survived only as a draft manuscript.

famous Alexandrian scholar who lived 90–168 AD, described them in his *Almagest* and in a less known, mostly astrological *Tetrabiblios*. After about Kepler's time astrology contributed nothing to mathematics, not even smallest improvements in calculation routines that would be worth mentioning. This sharply contrasts with astronomy's mathematical connections and achievements. Today, research in astronomy is unthinkable without advanced mathematical methods. On the other hand, astronomy inspired development of important branches of mathematics, most notably of the differential and integral calculus [21].

3.3 Astrology Offers No Physical Model to Support Its Claims

> These failures were one of the reasons that caused me to stop studying astrology and reject it as false. A more important reason was the absence of any way that the planets could influence our actions and thoughts. Thus it was impossible to see how their rays meeting in trine or quartile should be either beneficial or harmful; or how the sun could be more strong in one part of the heavens than another; my experience is that persons with well-placed planets do not attain more than those with ill-placed planets.

In the above quoted argument, Flamsteed [31] refers to what today could be expressed much more strongly: *there is no place for unknown forces that could explain planetary influence on humans assumed in astrology. Planetary forces are accounted for with a great accuracy. We know these forces accurately, because we very accurately measure planetary motions in the Solar Systems.* Already Kepler's *Rudolphine Tables* [16] gave planetary ephemerides with accuracy of about $14''$, i.e. about half of the visible solar disk. The accuracy was therefore sufficient to predict a transits of Mercury and Venus across the Sun. Such transients, predicted by the *Rudolphine Tables* have been observed soon after – of Mercury in 1631 by Pierre Gassendi, and of Venus in 1639 by Jeremiah Horrox. The XIII century *Alphonsine tables* that have been in use before Kepler had accuracy of $\sim 5^\circ$, i.e. more than twenty times magnitude worse, corresponding to about ten solar disks. Today, we check on the theory by measuring the Earth-Moon distance (circa 400,000 km) with the accuracy of about 1 cm by the laser ranging method.

3.4 Astrology Fails in Testable Empirical Checks

Every time when horoscopes have been casted under controlled and unbiased circumstances,[19] the fraction of them that successfully predicted future events or personality of a person, followed from a pure chance. The best known, and

---

[19] Michel Gauquelin's claim [28] that a statistically significant number of sports champions were born just after the planet Mars rises or culminates, attracted attention as the "Mars effect". Other authors proved that the Mars effect was an artifact of a bias in Gauquelin's analysis, see e.g. [36].

fully conclusive, *double-blind test of astrology* was done by Shawn Carlson of the University of California at Berkeley [22]. During the test, astrologers were interpreting natal charts for 116 “clients”, unknown to them. No face-to-face contact with the clients was allowed. For each client’s chart, astrologers were provided three anonymous California Personality Inventory (CPI) profiles. (These profiles measure traits like aggressiveness, dominance, and femininity from a long series of multiple-choice questions.) They are widely used and scientifically accepted. Of the three profiles, one was that of an actual“client”, while the two other have been chosen at random from a large pool containing profiles of other (real people). The astrologers’ task was to find, by comparing the horoscope to the profiles, which of the three was the real profile of the “client”. The result of the experiment, in the author’s own words, clearly refutes the astrological hypothesis:

> We are now in a position to argue a surprisingly strong case against natal astrology as practiced by reputable astrologers. Great pains were taken to insure that the experiment was unbiased and to make sure that astrology was given every reasonable chance to succeed. It failed. Despite the fact that we worked with some of the best astrologers in the country, recommended by the advising astrologers for their expertise in astrology and in their ability to use the CPI, despite the fact that every reasonable suggestion made by advising astrologers was worked into the experiment, despite the fact that the astrologers approved the design and predicted 50% as the “minimum” effect they would expect to see, astrology failed to perform at a level better than chance. Tested using double-blind methods, the astrologers’ predictions proved wrong. Their predicted connection between the positions of the planets and other astronomical objects at the time of birth and the personalities of test subjects did not exist. The experiment clearly refutes the astrological hypothesis.
>
> People believed in astrology for thousands of years and no doubt will continue to do so no matter what scientists discover. They are entitled to their beliefs, but they should know that there is no factual evidence on which to base them.

3.5 Astrology Has No Honest Methodological Standards

Flamsteed [31], whom I already quoted, was surprised and shocked to discover that astrologers of his time were dishonestly abusing their own rules, trying to match events to the horoscope predictions, even if these obviously did not match:

> I soon found that when astrologers found no direction in a corrected chart to match a notable accident, they referred to other indications such as that year’s revolution, seizing on whatever could be made to match the accident despite better arguments to the contrary. And that when they proclaimed the truth of their predictions, they ignored any aspects, directions or transits that failed to show accidents.
>
> Also, if the case could not fairly be proved, they pointed to defects in their ability, or to needing more time to consult their books, rather than acknowledge the least error in astrology. But it is a miracle if the case cannot be proved, because astrologers have so many rules, and so many aspects, transits, directions, revolutions, and progressions to consider,

> and so many ways of considering them, that it is impossible not to find something that matches the event even though it is hard to see why the contrary indications should be overpowered. But if even that approach fails, they say that God has overruled the stars.

We have seen a similar “creative booking” in Kepler’s interpretation of the Wallenstein horoscope. The same abuse of rules is characteristic of the contemporary astrology.

3.6 There Are No Great Astrology’s Masters

Kepler was the last great master of astrology. However, in astrology, he went nowhere, because the “astrological reality” he was trying to discover, existed only as an illusion of his mind – who chases Ghostlight wades through swamps. With Kepler, astrology ended as an acceptable intellectual occupation. The multitude of astrologers who followed Kepler during the last four centuries discovered nothing.

3.7 Final Conclusions

I cannot better conclude this essay than by quoting a few sentences from Flamsteed [31]:

> Since astrology finds no natural grounds to sustain it, and since experience shows us its falsehood, I hope my readers will withdraw any credit they may have given to this imposture. As for astrologers, I have no hope of reforming them because their profession – no matter how foolish and opposite to reason – is too lucrative. My reward for this plain speaking will no doubt be the title of ignorant and peevish.

Acknowledgements Jerzy Kierul, the author of a bestselling book on Kepler [32], and Jarosław Włodarczyk, historian of science and editor of Polish translations of Kepler’s works [9], [17] helped me with information about Kepler’s life, ideas and work. Jiří Bičák, professor at the Charles University in Prague, enriched my knowledge about *Praga Magica*. Maciej Lipowski, a linguist, helped me with Latin and German quotations from old books. I could use these rare books thanks to help received from the staff of the Göteborg University Library. Archbishop Jan Tyrawa, Ordinary of the Bydgoszcz Diocese, explained to me a few theological subtleties relevant for this essay. Astronomers Günther Wuchterl and Klaudia Einhorn told me about their findings of Kepler’s errors and cheats in the Wallenstein horoscope. Małgosia Świentczak drew difficult figures for this essay, guided by my vague, and at times almost capricious, instructions. My wife Henryka and son Tomasz advised me in several editorial matters. I sincerely thank them all. I also thank the Editor of this volume, professor Jean-Pierre Lasota, my friend, for convincing me that I should write an essay on such an unusual subject, and for his constant support and encouragement. Finally, I acknowledge support by Polish grant NN 203 3814 36 and I thank the Czech Academy for very generously supporting my stay in Prague through the 2009 grant *“Program podpory projektů mezinárodní spolupráce AV ČR”*. I thank my Czech colleagues for their hospitality and friendship.

References

I start from the complete chronological list of Kepler's books,[1–17], and his collected works [18, 19]. Not all of these are cited in the text. Then other references, all cited in the text, follow in the alphabetical order. I have read the listed references in English or in Polish (in the later case the Polish source is mentioned).

1. *Mysterium cosmographicum*, (1596), "The Sacred Mystery of the Cosmos"
2. *De Fundamentis Astrologiae Certioribus*, (1601), "Concerning the More Certain Fundamentals of Astrology"
3. *Astronomiae Pars Optica*, (1604), "The Optical Part of Astronomy"
4. *De Stella nova in pede Serpentarii*, (1604), "On the New Star in Ophiuchus's Foot"
5. *Astronomia nova*, (1609), "New Astronomy"
6. *Tertius Interveniens*, (1610)
7. *Dissertatio cum Nuncio Sidereo*, (1610)
8. *Dioptrice*, (1610)
9. *De nive sexangula*, (1611), "On the Six-Cornered Snowflake", Polish translation: Wydawnictwa Uniwersytetu Warszawskiego, (2006)
10. *De vero Anno, quo aeternus Dei Filius humanam naturam in Utero benedictae Virginis Mariae assumpsit*, (1613)
11. *Eclogae Chronicae*, (1615)
12. *Nova stereometria doliorum vinariorum*, (1615), "New Stereometry of Wine Barrels"
13. *Epitome astronomiae Copernicanae*, (1618-1621), "Epitome of Copernican Astronomy, published in three parts,
14. *Harmonice Mundi*, (1619) "Harmony of the Worlds"
15. *Mysterium cosmographicum*, 2nd Edition, (1621), "The Sacred Mystery of the Cosmos"
16. *Tabulae Rudolphinae*, (1627) "Rudolphine Tables"
17. *Somnium*, (1634); "The Dream", Polish translation: Scholar, (2004)
18. *Gesammelte Werke* (ed. M. Caspar)
19. *Joannis Kepleri astronomi opera omnia*, ed. C. Frisch, Heyder & Zimmer, Frankfurt a.M., (1858-1871)
20. Augustine, St.: *De Genesi ad litteram* (2:17:37)
21. C.B. Boyer, *The History of the Calculus and its Conceptual Developments* (Dover, New York, 1949)
22. S. Carlson, A double-blind test of astrology. Nature **318**, 419–425 (1985)
23. M. Caspar, *Kepler*; English translation: C.D. Hellman, Introduction, ed. by O. Gingerich (Dover, New York, 1993)
24. M. Dzielska, *Hypatia of Alexandria* (Harvard University Press, Cambridge, 1995); and the updated 3rd Polish edition: Universitas, Kraków (2010)
25. K. Einhorn, G. Wuchterl, Kepler's astrology and the Wallenstein's horoscopes. Acta Universitatis Carolinae, Mathematica et Physica, **46**(Supplementum), 101 (2004)
26. C. Flammarion, *L'atmosphère: Météorologie Populaire* (Hachette & Co., Paris, 1888), p. 163
27. G. Galilei, *Dialogo sopra i due massimi sistemi del mondo*, (1632); "Dialogue Concerning the Two Chief World Systems", transl. by S. Drake, (University of California Press, Barkely, 1953)
28. M. Gauquelin, *L'influence des astres*, Dauphin, (1955), "The Influence of the Stars"
29. E. Grant, *Planets, Stars, and Orbs: The Medieval Cosmos* (Cambridge University Press, Cambridge, 1994)
30. G. Greshake *Der dreieine Gott: Eine trinitarische Theologie* Herder (2007). Polish translation: abp. Jan Tyrawa, TUM (2009)
31. M. Hunter, *Science and Astrology in Seventeenth Century England: An Unpublished Polemic by John Flamsteed*, in *Astrology Science and Society: Historical Essays*, ed. by P. Curry (Boydell Press, Wolfeboro, 1987)
32. J. Kierul, *Kepler*, PIW, (2007); this book is avaliable only in Polish

33. A. Koestler, *The Sleepwalkers: A History of Man's Changing Vision of the Universe* (Penguin Books, Harmondsworth, 1959)
34. P. Marshall, *The Mercurial Emperor: The Magic Circle of Rudolf II in Renaissance Prague* (Pimilco, London, 2007)
35. R. Martens, *Kepler's Philosophy and the New Astronomy* (Princeton University Press, Princeton, 2000)
36. J.W. Nienhuys, The mars effect in retrospect. Skeptical Inquirer **21**, 24–29 (1997)
37. J.D. North, *Horoscopes and History* (The Warburg Institute, London, 1986)
38. G. Pico della Mirandola, *Disputationes adversus astrologiam divinatricem* (Edizioni Dehoniane Bologna, Bologna, 1493); see *The Occult in Early Modern Europe: A Documentary History*, ed. by P.G. Maxwell-Stuart (St. Martins Press, New York, 1999)
39. A. Saiber, *Giordano Bruno and the Geometry of Language* (Ashgate, Aldershot, 2005)
40. W. Sierpiński, Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. Acad. Sci. Paris **162**, 629–632 (1916)
41. B. Stephenson, *Kepler's Physical Astronomy* (Springer, New York, 1987)

Chapter 18
Fundamental Issues and Problems of Cosmology

George F.R. Ellis

Abstract Cosmology – the study of the origin and evolution of the universe itself – is a unique science because the universe provides the setting and context for all the other sciences. It has made incredible strides in the past century, and particularly in the past two decades, as physical understanding has developed and as vast amounts of new data has come in. We understand the basic evolution of the universe from extremely early times to the present day, as well as the way large scale structures formed in the universe (Dodelson S, Modern Cosmology, Academic, San Diego, 2003; Peter P, Uzan J-P, Primordial cosmology Oxford University Press, Oxford, 2009. However (and partly as a consequence of all this new data) we are inevitably running into a series of limits due to the nature of the subject. This chapter will look at some of these fundamental problems for cosmology, and consider some fundamental issues relating to the nature of the topic (For a discussion of cosmological issues from a philosophical viewpoint, raising many of the issues considered here in the context of the relevant physical theory, see Ellis (Issues in the philosophy of cosmology. In: Butterfield J, Earman J (eds) Handbook in philosophy of physics. Elsevier, Amsterdam, 2006, pp. 1183–1285, http://arxiv.org/abs/astro-ph/0602280).

Keywords Cosmology • Fundamental problems • Laws of physics

1 Limits of Laws of Physics: The Uniqueness of the Universe

Cosmology is a unique science because it is science with only one unique object of study [9, 26]. Now all the historical and geographical sciences describe unique objects (the origin of the Sun and Solar System, the development of the Himalaya

---

G.F.R. Ellis (✉)
Department of Mathematics and Applied Mathematics, Astrophysics Cosmology and Gravitation Center, University of Cape Town, Rondebosch, 7701 Cape Town, South Africa
e-mail: George.Ellis@uct.ac.za

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_18,

Mountains, the evolutionary origin of giraffes on Earth for example). But in each case there are other things to compare with (other stars and planetary systems, other mountain ranges, histories of other species), hence one can propose and test laws for the class of objects in question. Only in the case of cosmology is there no other equivalent physical object to observe or test. In a sense, the issue is that cosmology has one foot in science and the other foot somehow out. The point is that what is actually scientific concerns what is reproducible. In that sense, cosmology is partly in science because of all the phenomena taking place in the universe, but partly out because the universe itself is not reproducible. This is the profound issue of how to handle the generic and the specific in scientific terms: cosmology is at the cusp, because it is an attempt to produce a science of one specific object.

Because of this, the distinction between laws of nature and boundary conditions for the system becomes obscure. What is difficult is to distinguish what is a causal relation (science) and what is an historical connection. Laws of nature are of universal applicability to some class of objects; boundary conditions are contingent features that could in principle have been different, and indeed are different for different members of the class of objects concerned. In the case of the universe the relevant boundary conditions are just given to us as immutable features of our past history and may or may not be determined by some kind of law of boundary conditions. If there is such a law we can never test it and prove it to be so. Hence it is not clear what kinds of ‘laws’ might apply to this context. What we take to be fixed laws may just be particular cases of more general relations that vary from place to place or from time to time, and are realised in a particular way in the universe patch that is accessible to our observation: but they could have been otherwise. They are not immutable laws applicable in all possible universes.

Thus overall there is uncertainty about what is or is not a law of physics in the context of the evolution of the universe itself. One topic where this plays out is as regards the nature of the constants of physics, which are crucial to all physical laws [2, 34]. When we look at them in the cosmological context, involving physical behaviour over vast distances and times, are they in fact constants – a foundational part of the laws themselves – or are they rather more ephemeral quantities that are environmentally dependent? The present day tendency is towards the latter view: that many of the ‘constants’ of physics are in fact contingent, depending on context.

This is in accord with the line of thought developed by Ernst Mach regarding the origin of inertia and Dirac regarding the gravitational constant, where both are seen as being determined by cosmological conditions [10, 15]. Consequently it has been suggested that the gravitational constant G might vary with time [7, 8],[1] and that observations show the fine structure constant α varies with time. This idea of

---

[1] It has also been suggested this might be true for the speed of light c; but the value assigned to this quantity depends on the dimensional units used, and can always be set to unity by appropriate choice of units (measure distance in light years). Thus this is not a good physical proposal [18]. We will of course be in very deep trouble if unity (the quantity represented by the number “1”) ever starts varying with time!

cosmological dependence of the value of constants has been given considerable impetus recently by the development of the idea of the landscape of string theory, where the values of the 'fundamental constants' in fact depend on the vacuum state – and there are a huge number of different such states. Indeed theories such as string theory actually predict that ALL dimensionless constants are dynamical, thus it is possible that they do indeed vary. However, observational evidence that this is indeed the case is not strong [33]. The issue is then why don't we see them vary. There is an issue of scale here. In most multiverse models the constants vary on very large scales but will be constant in each universe in the multiverse. Thus while you can have a universe model with time/space varying constant in this context (if it indeed occurs), it will not occur in our observable universe domain. Furthermore, these theories that predict such variation are not proven, and may or may not be true.

One of the most important areas where this kind of top-down effect from the cosmological to the local scale might be true, is as regards the origin of the arrow of time, in particular as imbedded in the statement of the second law of thermodynamics [6, 35]. Many believe that this crucial feature of physics, which is not determined by the time-reversible equations of fundamental physics, is determined by the boundary conditions at the start of the universe [5, 15, 23].

2 Limits of Laws of Nature: The Start of the Universe

The issue of the limits of laws of physics arises *a fortiori* as regards the origin of the universe – if it indeed had a start, where not only matter but space and time, and perhaps even the laws of physics themselves, came into existence. We do not know if there are laws governing this process, or whether it has happened *sui generis* – a once off affair, not governed by any kind of physical law. It hardly makes sense to say, "there are laws governing the process but it happened only once", for it is the very essence of laws that they apply to multiple events. So perhaps this could be an argument for a multiverse, where the process of universe creation happens many times – if we could show such laws exist. But of course we can't. No possible experiment of any kind could show this is the case. Hence any claims in this regard should be thought of as philosophy rather than science – assuming science is an enterprise with a core value being that one should be able to verify one's hypotheses.

There are two further aspects of this kind of proposal that need to be mentioned. Firstly out of what previous kind of existence should a universe emerge? Various attempts have been made to say it could emerge from "nothing" – but many more envisage emergence as a fluctuation from some kind of previous space time – usually flat spacetime or de Sitter spacetime – or an epoch where time did not have the same nature as it does today. This previous epoch is supposed to just exist: so the problem of existence has not been solved, just transferred back to an earlier epoch or state of existence: why or how did that come into being? In usual mythology, in order to have something, you start from a pre-stage that existed for ever: a sort of nothing that has the potentiality to become not-nothing at some stage. We are now doing the same in cosmology.

This comment applies even to those theories where there is no start to the universe, where it is supposed to be eternal in one way or another (it has no beginning, perhaps being cyclic in its behaviour). The question remains: why does it have the particular form it does, when other forms are presumably possible? [21]. If one could show conclusively that only one form was possible, this issue would be resolved; but we are very far indeed from arriving at this conclusion, indeed the development of the string landscape proposal goes in precisely the opposite direction: there are apparently vast numbers of alternative possibilities for physics, any of which could be realised. But in any case any such explanation would reply on assuming existence of some particular laws of physics or pre-physics: and the same issue would arise in regard to them: why these laws rather than some other?

Underlying all this is the second issue: if we try to present some kind of physical theory for the creation of the universe ("it came into being because of the laws of physics"), then we are assuming the laws of physics pre-existed the universe itself: they somehow were there before space, time, matter, or anything else physical existed. This is certainly not "nothing"! It is the entire apparatus of quantum field theory [25], including symmetry and variational principles, commutation relations, Hilbert spaces, and so on, plus the selection of specific symmetry groups. So what kind of existence is proposed for these laws? Do they live in some kind of eternal Platonic space? Can one be more explicit about this? If not, where or how did they exist? And how do they get their causal power?

These kinds of claims make major philosophical assumptions about what precedes or underlies the existence of the universe, and implicitly about in what way these supposed entities may be said to pre-exist physical existence. This is highly non-trivial: they have some kind of non-physical existence of so potent a nature as to be able to give rise to the coming-into-being of an entire universe (or perhaps many). We are entitled to ask the authors of any such claims to be very clear about such questions. They should of course also give some supporting evidence as to why we should believe their claims be true: which leads to the next issue.

3 Limits to Knowledge of Physical Theories: The Limits of Testing

The previous section related to the nature of what exists – that is, it is concerned with ontology. But the working scientist is concerned with how we can know what exists- that is, she must consider epistemology. And here is one of the major limits to cosmology, as a science: we cannot test the relevant laws of physics that determine what happens in the early universe – or at least we cannot test them under the kinds of conditions that then held. The point is that each of our known experimentally determined laws of physics comes with a domain of applicability: it holds under some conditions and not under others. We cannot reproduce on Earth the conditions that occurred very early on in the hot big bang era, because we can't reach the relevant energies for both technological and economic reasons. Hence in studying these eras, we will always be assuming we know the nature of physics that we cannot test. We certainly cannot test any 'laws' that brought the universe into existence – if they themselves indeed exist.

How do we proceed then? We extrapolate known and tested laws into the unknown, and hope we have made the correct assumptions about which aspects of the laws remain valid in these circumstances – they are the deep nature of reality, often hidden beneath surface appearances. We may or may not make the correct extrapolation, but we can't test the extrapolation, except by their cosmological consequences. The trouble here is that you then have an 'explanation' that applies to only one phenomenon, and cannot be utilised or tested in any other context. It can be adjusted carefully to reproduce one aspect of experience – but if that is all it explains it is not a unifying explanation of different phenomena, it is an ad hoc explanation of the one desired feature. An example is the inflaton field in inflationary theory. It is a theorem that almost any scale factor evolution whatever can be obtained via a scalar field Φ with a suitably chosen potential [13]. Unless some other physical effect of Φ can be predicted and verified, its assumed existence is an *ad hoc* 'cause' of one particular feature (in this case, the desired inflationary evolution) rather than several separate phenomena: hence it does not unify that feature with other experimentally well-established physical results.

This problem of testability of course arises in extreme form as regards theories of creation of the universe itself. There is no possibility whatever of testing such theories. They can be developed and adjusted to give some desired specific result: but this will be their sole application. It is in effect a physical theory applicable only to one historical event and only one existent entity: thus it is an ad hoc theory for a specific instance rather than a generic theory of universal behaviour.

It may be that some specific effect (such as aspects of gravitational wave spectra expected from the early universe) may be predicted from such a theory, but generically this will not be able to uniquely confirm the chain of reasoning that leads to the result: other options may also give the same physical outcome, unless it is solidly proven that no other such option exists. Without such a proof, it will not be a genuinely testable physical theory – hence its scientific status will be questionable (quite apart from all the considerations raised in the previous section). And the whole point of scientific cosmology is to indeed be scientific – to attain the kind of degrees of certainty that scientific study can provide. This is a feature worth defending: for many decades, mainstream scientists regarded cosmology as just a branch of philosophy and hence of dubious scientific quality. That situation has changed through the developments of the last 50 years, so cosmology is recognised as a solid branch of physics (as is confirmed by its inclusion in the annual *Review of Particle Physics* by the Particle Data Group). This achievement should not be lost.

4 Limits of Observation: Tension Between Explanation and Confirmation

Despite these problems, many claims are made for correctness of such theories. Why is this so? It is because they are the outcome of a line of thought that is strongly believed in – even if the outcome is not testable.

In effect what one has is a major tension between explanation and verification in cosmology: one can choose theories largely based on one of them or the other, but not on both, because of observational and testing limits in the cosmological context. Thus the various criteria for good scientific theories will inevitably conflict with each other in the case of cosmology [12].

In the case of possible laws of creation of the universe, one pursues theories because one believes strongly in their explanatory power, even if they are essentially untestable, thus one emphasizes the desirability of explanatory power over against testability. However this is a dangerous road to follow: the key feature that led to the spectacular success of the hard sciences is the way that hypotheses are subjected to experimental or observational test. Many non-scientific theories have (or are claimed to have) great explanatory power.[2] Thus this criterion alone can lead one badly astray. Furthermore the fact that some theory is expressed in mathematical terms does not necessarily mean it is a good scientific theory: testability remains the gold standard for such theories. One should not weaken this requirement of testability lightly; and if one does so, one should clearly label the result as experimentally-based philosophy, as opposed to solidly based scientific fact.

This issue is crucial not only as regards theories of origin of the universe, but also as regards its geometry and structure on the largest scales. Because the universe is very large compared to human size and life span, we can effectively only see the universe on one single past light cone. Because the universe has expanded for a finite time since it first became transparent to light at the time of decoupling of matter and radiation, about 400,000 years after the start of the hot big bang era, what we can see is limited by a visual horizon [16].[3] There is an exceptional case; we could possibly live in a small universe, which is spatially closed on such a small scale that we have seen right round the universe [14]. This is a testable proposition, but so far there is no solid evidence it is indeed the case. Assuming it is not true, what we can determine observationally about the universe is highly restricted: we have no information whatever about what lies beyond the visual horizon, and will never have such information,[4] no matter what technologies we use.

---

[2] Examples are Astrology and Intelligent Design.

[3] This lies inside the particle horizon, which is the causal horizon for all physical effects since the start of the universe.

[4] In a sensible time horizon for science: say in the next 25,000 years.

Consequently any claims for what lies beyond this horizon are purely hypothetical: there is no chance whatever of verifying them. The issue where this comes to prominence is in terms of the claim by many that one or other form of multiverse exists (see e.g. [19, 20, 32]): a universe with disparate expanding domains like the one we see around us but with completely different physical properties in each of them, or perhaps they occur in a completely disconnected form.

This is strongly claimed by some for theoretical reasons: it gives the only scientifically based explanation for why the constants of nature have values that lie in the narrow range that allows intelligent life to come into being [27, 28]. This is particularly advocated as regards the cosmological constant Λ: its very small value (in natural units) has no explanation on the basis of known physical laws; the best current theoretical option (strongly supported by the idea of the 'landscape' of string theory) is to assume Λ takes all sorts of values in a multiverse, and anthropic selection effects (we must be here to do the measurements) explain why we observe that it has a small value. But the proposal is completely untestable: the other supposed domains are far outside any possible observation. Thus the scientific status of this proposal is dubious (see the debate in [4]).

If it was the necessary outcome of tried and tested physics, one could make a strong case for it: but the supposed mechanisms for creating the many other expanding universe domains (usually assumed to occur through Coleman-de Lucia tunnelling), as well as the landscape of string theory, are both hypothetical untested theories. Hence any claims that a multiverse necessarily exists should be treated with extreme caution. This is not a conclusion that necessarily follows from established physics.

5 The Limits of Logic: The Problems of Infinities

A further specific issue in this regard is the often made claim (e.g. [19]) that cosmology – and particularly multiverse proposals – involves infinities of entities: stars, galaxies, even universes. There are two major problems with any such claims.

The first is that it is conceptually problematic. Any claims of actual existence of physical infinities in the real universe should be treated with great caution, as was emphasized by David Hilbert long ago:

> Our principal result is that the infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought… The role that remains for the infinite to play is solely that of an idea… which transcends all experience and which completes the concrete as a totality… ([22], p. 151).

The point is that infinity is not just a very big number, or an ordinary physical entity: rather it is a quantity that can never be attained, that is forever out of reach.

And that leads to the second problem: even if it were to exist, quite apart from the existence of horizons preventing us from seeing all the entities that would then exist, we could never count them even if we could see them – for to do so would take an infinite time, and so would never be completed. Thus there is absolutely no way one can ever prove this claim to be true. Hence is it not a scientific claim – if one believes that science relates to provable statements.[5]

6 The Limits of Scope: The Kinds of Issue Taken to Be Relevant

Now we turn to bigger issues: what is it that cosmology should be concerned with? What is its appropriate topic of study? This determines the nature of the subject: the kinds of questions we want it to answer.

The problem is that, considered in the wider sense, cosmology – the study of the universe as a whole – in principle involves everything there is! How do we trim it down to a manageable size, suitable for scientific research? The standard attitude will be, by omitting almost everything and concentrating on bare bones physics. But there will always be more we want to know: and indeed a good number of papers on cosmology transgress these boundaries – and claim to deal with much more. This occurs specifically when the subject is extended from consideration of purely physical and astrophysical kinds of issues, such as nucleosynthesis in the early universe, the expansion history of the universe, and formation of structures in the universe, to issues such as the origin of life in the universe – surely one of the main issues the public at large would like to hear about.

So do we regard the origin of life as an issue for cosmology, or not? If so, what questions do we want to answer from a cosmological perspective? Two specific kinds of issues arise here. Firstly, the whole business of determining the physical origin of structures such as galaxies, stars, and planets is relevant: the latter being in the domain of astrophysics rather then cosmology, but in the setting provided by cosmology, which determines what is possible. Hence processes for such structure formation are of great interest, as are the conditions that make this possible.

But that leads to an area of real controversy. It has become apparent that if one considers the set of all possible cosmologies, those that admit any complex physical structures, such as living beings, are highly restricted: they are very special within the family of all possible cosmologies [1,3,27,28]). To put it more controversially, on the face of it, it seems that the universe is fine-tuned to allow life to exist. This applies both to the laws of physics that apply in the universe, and to the specific initial conditions that lead to the universe we know, on the basis of those laws. It is well within the bailiwick of cosmology to consider possible universes where the laws of physics are different from those applicable in our neck of the woods; so considering such options is a legitimate activity for a physical cosmologist. But it rapidly leads to areas of much wider concern than just physics.

---

[5] The claim that this infinity necessarily follows from the physics of the situation is not true, even if we accept the unproven underlying physical process: for in reality the situation is that if the inflationary parameters are chosen so that eternal inflation occurs, the implied infinity is an asymptotic state that is never reached in a finite time [17].

The point is not just that it leads firstly to issues to do with planetary formation and secondly to issues of exobiology – an exciting area of study at the present time (see Chaps. 19 and 20). It is that if one pursues this line of argument, it inevitably leads to consideration of issues of ultimate causation: and these are deeply philosophical topics, which are informed by studies in physical cosmology, but cannot be fully resolved by purely physical considerations. Bigger issues are at stake; to assume they can be resolved by purely physical argumentation is a strong philosophical position that can be challenged in many ways, indeed it is highly debatable. But that position is being increasingly taken by a number of cosmologists.

They run the risk of appearing rather naïve to the outside world: being experts in their own area, they are claiming to also be experts in other domains where they not only have no particular expertise, they don't even have the historical and philosophical knowledge required to enter the debate at an adequate level. They are amateurs in these areas – but claim privileged attention because they are scientists. Yet they react very strongly (with good reason) when experts from other areas (such as sociology) make statements about the practice of physics and cosmology. Such asymmetry is rather suspect. They should remember this when they enter these wider discussions – which have a provenance of thousands of years.

So the issue here is, do cosmologists want to try to tackle issues of fundamental causation or not? If they do, they should refrain from proclaiming that philosophy has no value, because they will themselves be practicing philosophy! They should attempt to do so from a sufficiently informed historical and philosophical stance, if they do not want to reinforce the widespread image of the supremely arrogant scientist – expert in his own area but ignorant of wider culture and knowledge. Technical knowledge and ability is not the same as wisdom; but that is what is required to enter these debates adequately.

If one wants to deal with philosophical issues one must do so in an adequately informed way. Any views on the scope and meaning of cosmology – and in particular on their implications for human life – are philosophical ruminations, and should be presented as such, not as scientific conclusions. It is particularly inappropriate to decry philosophy as a topic, while at the same time indulging in philosophical speculation oneself – albeit it of a simple form. Philosophy is philosophy, even if it is low level and uninformed.

7 The Limits of Explanation: What Kind of Causation Will Be Considered?

Overall the underlying theme is what kinds of explanation are regarded as acceptable in cosmology? The current cosmological paradigm is solidly based in theoretical physics, and has achieved enormous success. But we do know that this is not the only kind of causation in the universe: physics has a limited domain of applicability that does not cover all that happens. For example, while physics of course underlies all biology, physics textbooks do not include chapters on Darwinian selection. But that is a central causal feature of biology, demonstrated to be valid by innumerable experiments and observations. It is an emergent higher level principle, in a profound sense based in physics, but representing a kind of causation that is not encompassed by physics. Yet it certainly exists in the universe. This applies also of course to mental events and their causal powers.[6] The implication is crucial: the kinds of causation envisaged by physics do not encompass all the kinds of causation that we know exist in the universe.

So the profound question is, could other kinds of causation be relevant to cosmology than the purely physical, i.e. other than those kinds of effects described in the canon of theoretical physics? One intriguing move in this direction is Smolin's attempt to introduce natural selection into cosmology [30], via a process of rebirth of new universes out of black hole collapse, but with different values of fundamental constants in each case, and a consequent selection process that favours universes with maximal black hole creation. This attempt is not wholly successful in it's own terms [29], but nevertheless is a visionary work in that it moves the concept of causation in cosmology beyond the physical principles described simply by variational and symmetry principles. It proposes other kinds of causation are relevant in the cosmological context.

It is clear that this line of argument can be extended further. We know that intelligence exists in the universe. Is this an accident, or is it a profound feature of the universe? Its emergence is precursored by a possibility space that allows it to exist [11], and therefore in some sense anticipates its coming into being. This is the deep aspect of the anthropic issue: is it plausible that the extraordinary outcome of self-aware intelligent life occurs purely by chance as an unintended by product of inanimate forces at work, or is this outcome so extraordinary in its quality that this possibility must in some sense have been built into the structure of the universe in an intentional way? In brief: is a purely scientific approach adequate as an approach to understanding the deep nature of the universe?

This line of argument will of course be repugnant to many, but it will be very welcome to many others. If one wishes to investigate whether it is a sensible direction to go or not, a key issue is the limits to the kind of data one allows in pursuing the investigation of cosmology: what kinds of data will be taken into account? One will of course have to take into account the experimental data that is the concern of theoretical physics as well as the observations of astronomers. But if one is to investigate the deep nature of all the existence that is permitted by and contained in the physical universe, should one not perhaps admit also as evidence issues of truth and beauty, of good and evil, of justice and injustice, of morality and meaning? They certainly exist in the lives of ordinary people, who exist in the universe and derive their physical being from it: is this to be taken as evidence about

---

[6] Some deny that mental events have causal powers. However the fact that this book exists as a physical object is a disproof of that hypothesis.

the nature of the universe, or not? If not, why not? They are part of what exists in the universe (to prove this, open your morning newspaper and read what is there).

This question is a companion to that considered in the last section: what kinds of issues does one want to answer through cosmological studies? One can of course decide to ignore all except the purely physical, and that is a perfectly valid position to take as a physical cosmologist. It is not a valid position to take however, if one decides to then make pronouncements on the meaning of life, or on ultimate causation in the universe, on the basis of one's cosmological studies.

This is a crucial issue for the study of cosmology in the long term. If one were able to gaze into a crystal ball and try to see, in the mists of future time, the nature of cosmological theory say 50,000 years from now – if humans are still alive on this planet at that time – I would make a guess that this would be the profound change that will have taken place by then: it will have long since been decided that it does not make sense to pursue cosmology as a purely physical subject. The isolation of physics from the rest of human understanding may have by then be come to be seen as a mistake resulting from a great enthusiasm for science soon after its dawn, at the expense of a broader understanding of the nature of things. This may by then have been replaced by a broader and more humanistic perspective which fully takes science seriously, but does not regard it as all that there is to understand in the pursuit of cosmology and our understanding of the universe.

If this is right, the present dominant reductionist viewpoint will by then be looked back on as a short term philosophical mistake. Cosmological theory will try to take cognisance of, and relate to, all the kinds of existence that occur in the universe – including in its ambit the mental and the Platonic as well as the physical [24]. In particular it will try to relate fully to the truly complex as well as the large and the small, and so will deal with the full complexity of all the possibility spaces that underlie existence – which are much more than just the physical [11].

There is much more in the universe than galaxies and stars and planets: theoretical physics as practiced today captures some aspects of reality but omits many others. Its maturity may expand to recognise these limitations, and develop a broader understanding of the nature of existence. The gap between the Two Cultures [31] may by then have been transcended: to the great benefit of human understanding, and also to our cosmological understanding.

Acknowledgments I thank Jean-Philippe Uzan for helpful comments.

References

1. Y.V. Balashov, Resource Letter Ap-1: the anthropic principle. Am. J. Phys. **54**, 1069–1076
2. J.D. Barrow, *The Constants of Nature: From Alpha to Omega – the Numbers That Encode the Deepest Secrets of the Universe* (Vintage Books, London, 2004)
3. J.D. Barrow, F. Tipler, *The Cosmological Anthropic Principle*. (Oxford University Press, Oxford, 1984)
4. B. Carr (ed.), *Universe or Multiverse?* (Cambridge University Press, Cambridge, 2007)

5. S. Carroll, *From Eternity to Here: The Quest for the Ultimate Theory of Time* (Dutton, New York, 2010)
6. P.C.W. Davies, *The Physics of Time Asymmetry* (Surrey University Press, London, 1974)
7. P.A.M. Dirac, The cosmological constants. Nature **139,** 323 (1937)
8. P.A.M. Dirac, New basis for cosmology. Proc. R. Soc. **A165**, 199–208 (1938)
9. S. Dodelson, *Modern Cosmology* (Academic, San Diego, 2003)
10. G.F.R. Ellis, Cosmology and local physics. New Astron. Rev. **46**, 645–658 (2002), http://arxiv.org/abs/gr-qc/0102017
11. G.F.R. Ellis, True complexity and its associated ontology, in *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity*, ed. by J.D. Barrow, P.C.W. Davies, C.L. Harper Jr (Cambridge University Press, Cambridge, 2004)
12. G.F.R. Ellis, *Issues in the Philosophy of Cosmology*, ed. by J. Butterfield, J. Earman. Handbook in Philosophy of Physics (Elsevier, Amsterdam, 2006), pp. 1183–1285, http://arxiv.org/abs/astro-ph/0602280)
13. G.F.R. Ellis, M. Madsen, Exact scalar field cosmologies. Classical Quant. Grav. **8**, 667–676 (1991)
14. G.F.R. Ellis, G. Schreiber, Observational and dynamic properties of small universes. Phys. Lett. **A115**, 97–107 (1986)
15. G.F.R. Ellis, D.W. Sciama, Global and non-global problems in cosmology, in *General Relativity*, ed. by L. O'Raifeartaigh (Oxford University Press, Oxford, 1972), pp. 35–59
16. G.F.R. Ellis, W.R. Stoeger, Horizons in inflationary universes. Class Q. Grav. **5**, 207 (1988)
17. G.F.R. Ellis, W.R. Stoeger, A note on infinities in eternal inflation. GRG J. **41**, 1475–1484 (2009), arXiv:1001.4590
18. G.F.R. Ellis, J-P. Uzan, 'c' is the speed of light, isn't it? Am. J. Phys. **73**, 240–247 (2005), http://arxiv.org/abs/gr-qc/0305099
19. J. Garriga, A. Vilenkin, Many worlds in one. Phys. Rev. **D64**, 043511 (2001), arXiv:gr-qc/0102010
20. A. Guth, Eternal inflation and its implications. J. Phys. **A40**, 6811–6826 (2007), arXiv:hep-th/0702178
21. M. Heller, *Ultimate Explanations of the Universe* (Springer, Berlin/Heidelberg, 2009)
22. D. Hilbert, On the infinite, in *Philosophy of Mathematics*, ed. by P. Benacerraf, H. Putnam (Prentice Hall, Englewood, 1964), pp. 134–151
23. R. Penrose, *The Emperor's New Mind* (Oxford University Press, Oxford, 1989)
24. R. Penrose, *The Road to Reality* (Jonathan Cape, London, 2004)
25. M. Peskin, D. Schroeder, *An Introduction to Quantum Field Theory*(Westview Press, New York, 1995)
26. P. Peter, J.-P. Uzan, *Primordial Cosmology* (Oxford University Press, Oxford, 2009)
27. M.J. Rees, *Just Six Numbers: The Deep Forces that Shape the Universe* (Weidenfeld and Nicholson, London, 1999)
28. M.J. Rees, *Our Cosmic Habitat* (Princeton, Oxford, 2003)
29. T. Rothman, G.F.R. Ellis, Smolin's natural selection hypothesis. Q. J. R. Astron. Soc. **34**, 201–212 (1992)
30. L. Smolin, Did the universe evolve? Class Q. Grav. **9**, 173–191 (1992)
31. C.P. Snow, *The Two Cultures*, 2nd edn. (1993 reissue) (University Press, Cambridge, 1960)
32. M. Tegmark, Is 'the theory of everything' merely the ultimate ensemble theory? Ann. Phys. (N.Y.) **270**, 1–51 (1998)
33. J.-P. Uzan, The fundamental constants and their variation: observational and theoretical status. Rev. Mod. Phys. **75**, 403–455 (2003)
34. J.-P. Uzan, B. Leclercq, *The Natural Laws of Nature* (Praxis, Chichester, 2008)
35. H.D. Zeh, *The Physical Basis of the Direction of Time* (Springer Verlag, Berlin, 1992)

Chapter 19
The Earth and Other Solar-System Bodies

Thérèse Encrenaz

Abstract Our knowledge of solar-system bodies, their formation and evolution, has tremendously improved over the past 50 years. The main reason is the advent of the space era and the in-situ exploration of planets, satellites and comets with flyby probes, orbiters, landers, and even rovers in the case of Mars. In addition, ground-based observations campaigns using large telescopes and an improved instrumentation have led to major discoveries, such as the detection of trans-neptunian objects. Numerical simulations have also been essential for understanding the dynamical history of solar-system bodies. In the future, planetology will face several challenges. The first open question is the formation scenario of the solar system, which looks very different from those of extra solar systems recently discovered. Will we find stellar systems comparable to ours, or is the solar system unique? To better constrain the formation scenario of the solar system, a special emphasis will be given to the study of primitive bodies, such as comets and trans-neptunian objects, which can be seen as remnants of the early stages of solar-system formation. Another challenging question is the search for life in the solar system. We have identified several niches where liquid water might have been probably present (the surface of Mars) or could presently exist (in the interiors of outer satellites). These environments will be favoured targets for space planetary exploration in the forthcoming decades.

Keywords Planetary systems • Planets and satellites: general • Planets and satellites: formation

---

T. Encrenaz (✉)
LESIA, Observatoire de Paris, 92195 Meudon, France
e-mail: therese.encrenaz@obspm.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_19,

1 Introduction

Planetology – or the study of solar-system bodies – is a science which is both old and new. It is an ancient science, as planetary observations, reported since Antiquity, were used in the sixteenth and seventeenth centuries to build the grounds of the Copernican system. The first half of the twentieth century was marked out by the development of stellar astronomy, generated by both instrumental developments (with, in particular, the use of spectroscopy) and theoretical advances. After the 1960s, the situation changed again radically with the beginning of the space era. Our knowledge of the solar system has made a huge progress thanks to the in-situ exploration of planets, satellites and comets. In addition, ground-based observations remain of major importance, as illustrated by the discovery of the Kuiper Belt from telescopic monitoring. In parallel, numerical simulations using supercomputers have allowed scientists to depict the early stages of solar-system formation and the early dynamical history of solar-system bodies [1–6].

2 The Beginning of the Space Era: The Moon Exploration

The space exploration of the Moon, engaged in the middle of the cold war, was obviously not driven by scientific reasons, but science was able to take full benefit of it. The exploration of the Moon was both human – with the Apollo program of NASA and the first man on the Moon on July 21, 1969 – and robotic, with the landing of soviet vehicles. The main scientific result of this adventure was the collection of lunar samples and solar wind particles which were analysed on Earth in laboratories. From these studies, the absolute age of the Moon was estimated, and used as a calibrator for the datation of solar-system solid-surface bodies. It is interesting to note that the human exploration of the Moon was not necessary for this research, as lunar samples were also collected by the soviet vehicles (although, however, in a much smaller amount than with the Apollo program).

3 The Space Exploration of Mars

Space missions toward Mars were also designed in the 1960s, but the first ones encountered many failures. In the 1970s, two space missions, designed by NASA, have been especially successful. The first one was the Mariner 9 orbiter which was launched in 1969 (its twin Mariner 8 was lost), and monitored the atmosphere and surface of Mars with, in particular, the first images of the volcanoes in 1972–1973. A few years later, the ambitious Viking mission was an outstanding success (Fig. 19.1), both from a scientific and a technical point of view. Two orbiters and two landers monitored the Martian atmosphere and surface, exploring the seasonal cycles of carbon dioxide, water vapor and dust. Compiled over several years, the

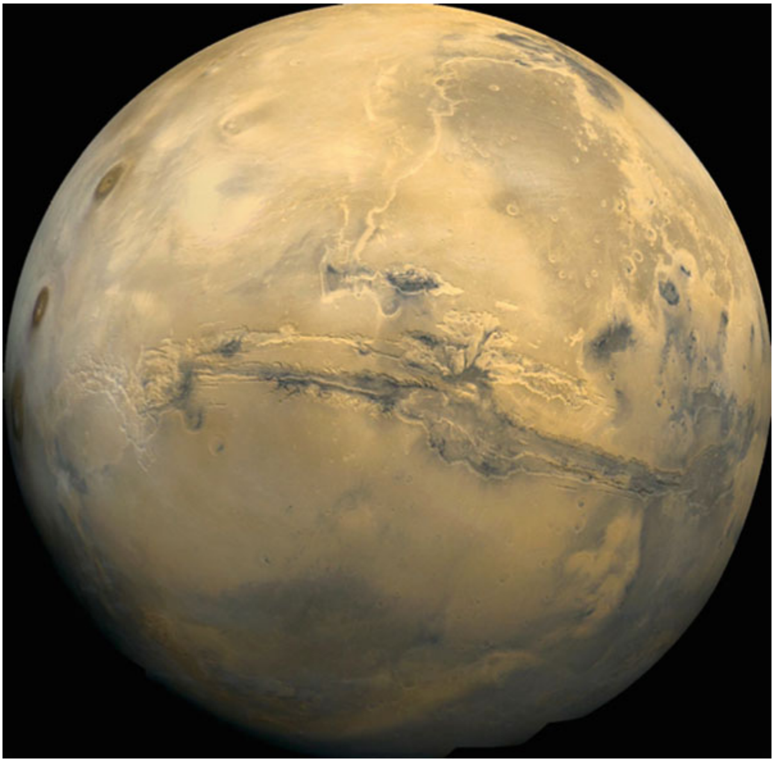


Fig. 19.1 Planet Mars, as observed by one of the Viking orbiters. The great fracture in the middle of the disk is Valles Marineris, a huge canyon 4,500 km long, 150 km wide and 8 km deep ([7] ©NASA)

Viking dataset is still used as a reference together with more recent space observations. Viking however was unsuccessful in finding any trace of life at the surface of Mars. As this question was at the origin of the mission's justification, the interest for the red planet fainted in the US, and the program was stopped for two decades.

In 1989, the Soviet Union launched the Phobos mission, which included two orbiters around Mars and descent modules which should have landed on Mars' satellite Phobos. The mission however failed, apart from a 2-month monitoring of Mars from one of the orbiters. At the end of the 1990s, new probes were sent by NASA, with again an important rate of failures. The next successful space missions were two orbiters (Mars Global Surveyor launched in 1996 and Mars Odyssey launched in 1999) and a small rover (Mars Pathfinder launched in 1997). In 2003, the European Space Agency successfully launched its first planetary orbiter Mars Express while NASA sent two identical rovers, Spirit and Opportunity, with an equal success.

The space exploration of Mars has allowed us to draw a new image of Mars, and to learn about its early history. With its obliquity and rotation period close to the terrestrial ones, Mars shows interesting similarities with the Earth, but it also exhibits two main differences: at 1.5 AU from the Sun, the planet is colder than the Earth, and with a mass of one tenth of the terrestrial one, Mars is also smaller. As a result, its atmosphere is much more tenuous and cold. About a third of its main atmospheric compound, carbon dioxide, is transported from pole to pole where it condenses on polar caps, together with water, along the seasonal cycle. These pressure variations induce strong winds and dust storms.

There are several indices which support the presence of liquid water on Mars in its early history (dried valley networks, outflow channels, presence of water below the poles and clays in old terrains) but the question of when and for how long this water was present is still open. The enrichment of heavy water (HDO) by a factor of 5 with respect to the terrestrial value, suggests that water escaped massively at the beginning of Mars' history, which could imply that the early atmosphere of Mars was warmer and thicker than today. Did this episode last long enough for life to appear and develop? If so, can we look for traces of ancient life? And why and how did the Martian atmosphere escape? All these questions will drive the future exploration of Mars over the coming decades.

4 The Space Exploration of Venus

Venus has been a favored target of planetary exploration for the Soviet Union. Venus, our neighbouring planet, hosts especially hostile conditions with a surface pressure close to 100 bars, a surface temperature of 730 K (460°C) and a permanent cloud deck of sulfuric acid. The Venera missions, designed in the 1980s, consisted in orbiters and landers which, once at the surface, had a very short lifetime. Still, the first images of Venus's surface were sent back to Earth in 1982 (Fig. 19.2). In 1986, an atmospheric balloon was sent in Venus' atmosphere as the Vega spacecraft, still from Soviet Union, continued its journey to encounter comet Halley. After the Pioneer Venus mission in 1978 and the Galileo flyby in 1990, NASA launched the Magellan mission which achieved the radar mapping of the surface. It was then discovered that the surface of Venus is covered with relatively young craters, with an age of 1 Gy at most. In 2005, the European Space Agency has launched an orbiter, Venus Express, still in operation in Venus' orbit.

As in the case of Mars, the atmosphere of Venus is mostly composed of carbon dioxide, with a few percent of molecular nitrogen. As on Mars, water vapor and carbon monoxide are also present. A noticeable difference with Mars is the presence of sulfur and the sulfur cycle, which could be fed by volcanism. Due to the H_2SO_4 cloud layer, the lower atmosphere is opaque to optical radiations, but it can be probed in the near infrared range at some specific wavelengths, on the dark side of the planet: the surface is warm enough for thermal emission to be detectable. Observations, performed from the ground and from space, have led to the determination of the deuterium abundance in water. The D/H ratio measured in

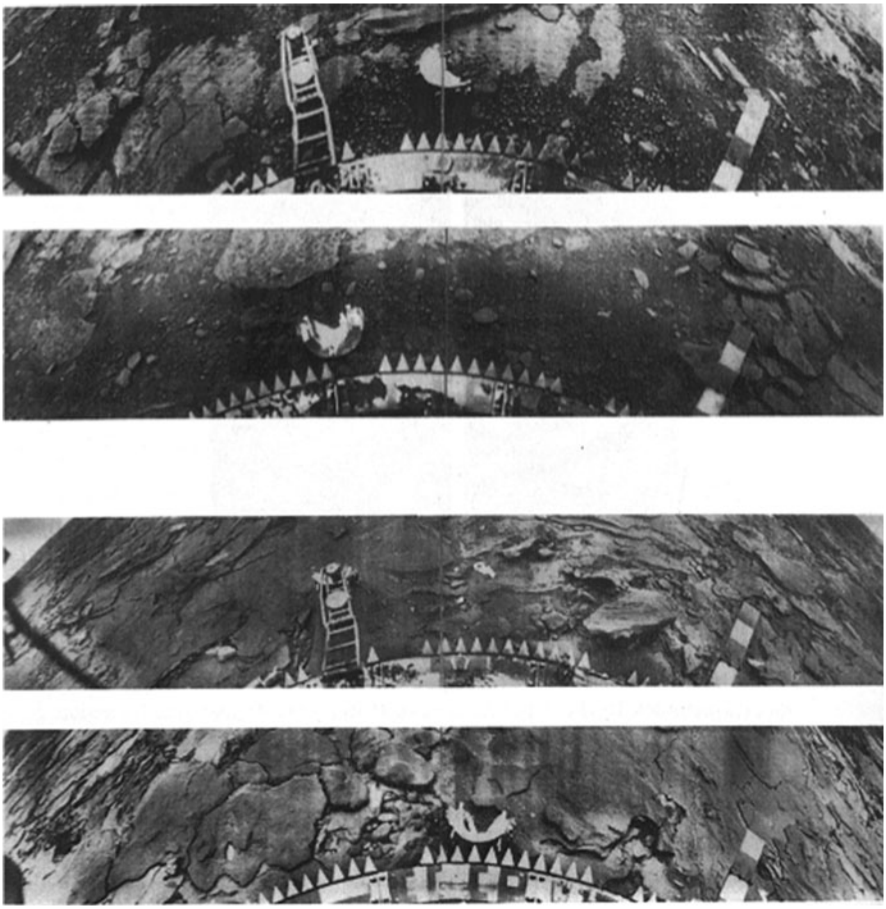


Fig. 19.2 The surface of Venus, observed by Venera 13 (*top*) and Venera 14 (*bottom*) in March 1982. These images are unique, as the surface of Venus, hidden behind a thick cloud deck of sulfuric acid, cannot be observed from outside in visible light ([7] ©Academy of Sciences of Soviet Union)

water has been found to be enriched by a factor 120 which implied, as on Mars but to a much greater extent, a massive outgassing of water in the early history of Venus.

5 Comparative Evolutions of the Terrestrial Planets

In the light of these results, it is interesting to compare the evolutions of the atmospheres of Venus, Mars and the Earth. Mercury, the innermost telluric planet, could not retain a stable atmosphere because of its low mass which implies a very low escape velocity (4 km/s, as compared with 11 km/s for the Earth), and its high surface temperature. The three other terrestrial planets probably started with an early atmosphere of comparable composition, with large relative fractions of CO_2 and H_2O and a minor contribution from N_2. How did these atmospheres evolve?

In the case of Venus, a liquid water ocean may have existed when the young Sun was fainter than today. As the solar flux increased with time, the water probably turned into vapor and, together with carbon dioxide, contributed to feed a runaway greenhouse effect, responsible for the high surface temperature observed today. Water disappeared probably by photodissociation, although this mechanism is not fully understood.

In the case of the Earth, its heliocentric distance was such that water could stay in the liquid phase when the Sun acquired its present flux (the planet may have encountered ice ball episodes in its early history; these episodes may have been stopped by volcanism outgassing of CO_2, H_2O and possibly CH_4 which were able to warm up the planet). The presence of liquid oceans allowed the carbon dioxide to be trapped in the form of calcium carbonates. As a result, the greenhouse effect remained moderate and the surface temperature stayed relatively constant over the planet's history. Life appeared presumably in the oceans, leading to the formation of molecular oxygen and to the present atmospheric composition, dominated by N_2 and O_2. The formation of an ozone layer allowed life to spread over the continents.

The case of Mars is different in two ways: the planet is colder and smaller than the Earth. Being smaller, its internal energy (which comes from the radiogenic elements of its interior) and its gravity field were smaller, which implied a more tenuous atmosphere. After a period of volcanic activity, the internal energy of the planet decreased, its greenhouse effect slowed down, and the planet became cold enough for water to be in the form of ice or permafrost, as we see it today. As mentioned above, there are still major open questions related to the evolution of the Martian atmosphere, the loss of its early atmosphere, the loss of water, and the possible presence of traces of an extinct life. The future space exploration of Mars will require an international cooperation between NASA, ESA and other partners. The first steps of this program are the ExoMars mission, planned for two launches in 2016 and 2018, and later an ambitious program of Martian sample collection, called Mars Sample Return, which should take place in different steps along the forthcoming decades.

6 The Exploration of the Giant Planets

Giant planets have been targets of planetary space exploration for about 40 years. The first program was issued by NASA with the two identical probes Pioneer 10 and 11. The probes flew by Jupiter in 1973 and 1974, and Pioneer 11 flew by Saturn in 1979. At the same time, the much more ambitious mission Voyager, also launched by NASA, flew by Jupiter, with two identical spacecraft Voyager 1 and Voyager 2.

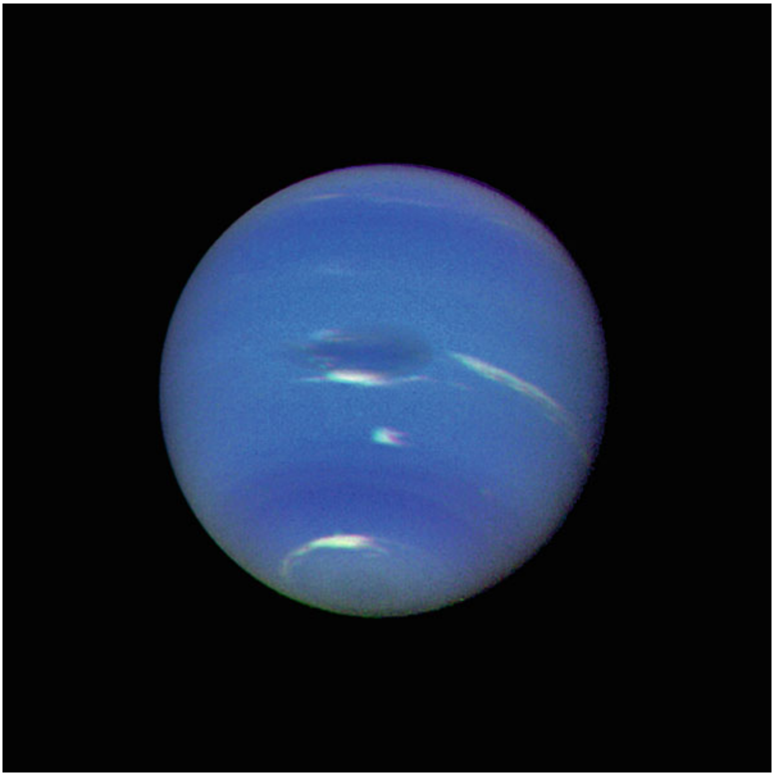


Fig. 19.3 Planet Neptune observed by Voyager 2 during its flyby of the planet in August 1989. The deep blue colour might be due to the presence of large amounts of methane. The white clouds are believed to be high-altitude cirrus of methane ([7] ©NASA)

Voyager 1 flew by Saturn and Titan in 1980; Voyager 2 encountered Saturn in 1981, then Uranus in 1986 and Neptune in 1989 (Fig. 19.3). The Voyager mission was a historical success, which brought lots of new discoveries; the Voyager dataset is still a reference today, especially for Uranus and Neptune which have not been explored in-situ by other spacecraft. Among the highlights of Voyager mission, one should quote the complexity of Jupiter's cloud structure (including the Great Red Spot); the discovery of active volcanism on Io; the possible evidence of a liquid water ocean below the surface of Europa; the complexity of Saturn's rings; the presence of many hydrocarbons and nitriles in Titan's atmosphere; the complex tectonics features of Uranus' satellite Miranda; the strong dynamics of Neptune's atmosphere and the discovery of cryovolcanism on Neptune's satellite Triton.

In the history of planetary exploration, the first step is the flyby; then come the orbiters for a long-term monitoring of the planets, and the probes for an in-situ analyses of their atmosphere and surface. After the success of Voyager and its exciting discoveries about Io and Europa, NASA designed the Galileo mission to pursue an in-depth exploration of the Jupiter system. Launched in 1989, the Galileo mission approached Jupiter in 1995 and sent a probe inside its atmosphere. Then the orbiter approached the four Galilean satellites with successive orbits and ended its mission in 2003. The Galileo mission encountered several problems. First its launch was delayed by several years after the Challenger disaster in January 1986; in addition, its big antenna could not deploy properly and a low-gain antenna had to be used for all telecommunications, which reduced by a great factor the volume of data sent to Earth.

In spite of these difficulties, the Galileo mission was a tremendous success. Among the many discoveries of Galileo, one should quote the unique results transmitted by the descent probe on December 7, 1995. The probe was able to sent data down to a pressure of 22 bars, a pressure much higher than anticipated by the engineers. Information was obtained about the temperature profile of Jupiter, its cloud structure, its wind profiles, and the chemical composition of its atmosphere. These last results turned out to be a key diagnostic to constrain the formation model of the giant planets, as will be discussed below.

Another highlight of the Voyager mission was the exploration of Titans' atmosphere and its possible analogy with the primitive Earth, illustrated by its nitrogen-rich composition, its surface pressure of 1.5 bars and the presence of a complex prebiotic chemistry. These results drove an enhanced interest for Saturn's largest moon and led to the definition of a new ambitious mission. Launched by NASA in October 1997, the Cassini mission was composed of an orbiter, under NASA's responsibility, and the Huygens probe, designed by ESA. This joint mission was a perfect example of cooperation between Europe and the United States, with scientists of both sides of the Atlantic involved in the different instruments of the two vehicles. After a flyby of Jupiter in December 2000 (Fig. 19.4), the Cassini spacecraft approached Saturn in 2004 (Fig. 19.5), and the Huygens probe successfully landed on Titan's surface on January 14, 2005, revealing the first images of Titan's surface, hidden below a permanent haze of aerosols. The images showed a flat surface covered with dried hydrocarbon deposits and eroded boulders, probably made of water ice. Information was also retrieved on Titan's thermal profile, winds and atmospheric composition. The images showed no evidence for any liquid surface. However, in 2007, lakes of hydrocarbons were discovered by the radar of the Cassini spacecraft, mostly in the northern hemisphere of the satellite, which is presently in winter.

Another major result of the Cassini mission was the discovery of active cryovolcanism at the surface of Saturn's small satellite Enceladus. This activity takes place near the south pole of Enceladus and the ejected material (mostly water) is believed to feed the tiny E-ring of Saturn, located near the satellite's orbit. The composition of the E-ring, as well as the observed plumes ejected from Enceladus, suggests, as in the case of Jupiter's satellite Europa, the presence of a salted liquid ocean below the surface, which has important potential implications for astrobiology.

Fig. 19.4 Planet Jupiter, observed by the Cassini spacecraft during its flyby of the planet in December 2000. The image illustrates the complexity of the dynamical atmospheric structure, in particular the Great Red Spot in the middle of the image ([7] ©NASA)

7 The Formation of Planets

As in the case of the terrestrial planets, the exploration of the giant planets has allowed us to study them in a comparative way, to better understand their formation and evolution processes. A first question to answer is the origin of the two classes of planets (Table 19.1). Close to the Sun, the four terrestrial planets are relatively small and dense; they have only a few satellites, if not none. At larger heliocentric distances, the giant planets are large with a low density, a ring system and a large number of satellites. How can we explain these two classes of planets?

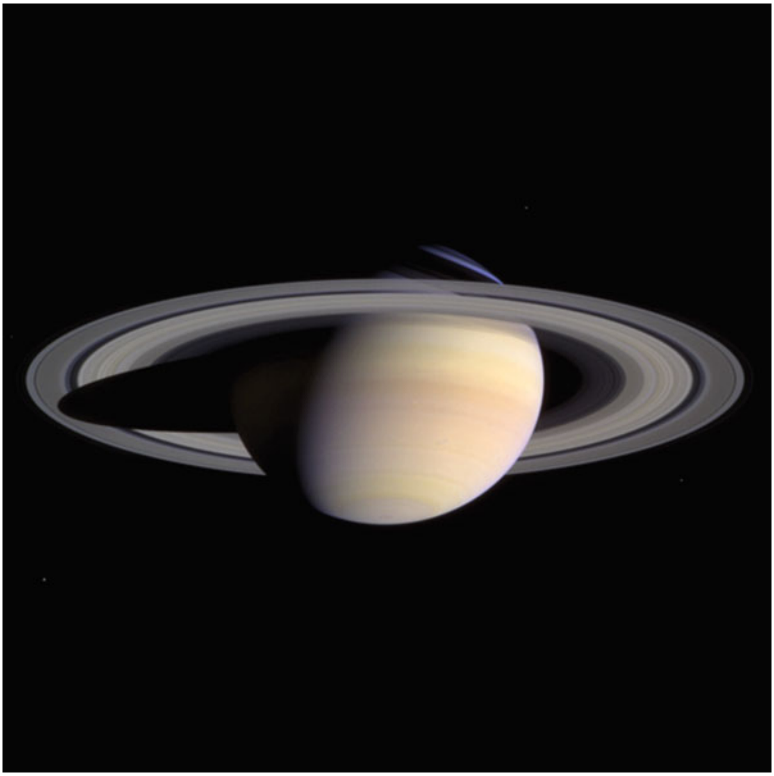


Fig. 19.5 Saturn and its ring system, as observed by the Cassini spacecraft during its approach of the planet in 2004. The shadow of the planet projects on the ring system on the left side of the image ([7] ©NASA)

Table 19.1 Orbital and physical properties of solar-system planets

Planet	SMA(AU)	$R(R_E)$	$MM_{(E)}$	$dg/(cm^3)$	P(y)
Mercury	0.4	0.38	0.05	5.44	0.24
Venus	0.7	0.95	0.81	5.25	0.61
Earth	1.0	1.00	1.00	5.52	1.00
Mars	1.5	0.53	0.11	3.94	1.88
Jupiter	5.2	11.2	317.8	1.24	11.85
Saturn	9.5	9.4	95.1	0.63	29.42
Uranus	19.9	4.0	14.6	1.21	83.75
Neptune	30.1	3.8	17.2	1.67	163.7

This question can be simply answered if we consider the scenario of the solar-system formation. Since the seventeenth century, it has been suggested that solar-system bodies were formed within a disk, resulting from the collapse of a rotating interstellar cloud. This "solar nebula" model was based on a few simple observations: all planets rotate around the Sun in the same direction (also the one of the Sun's rotation), on almost circular, concentric and coplanar orbits. Over the past decades, this model has received further support from astronomical observations which show that more than half of young stars are surrounded by a protoplanetary disk.

The protosolar disk was composed of gas and dust, and its temperature decreased with increasing distance to the center. The relative abundances of the chemical elements followed the cosmic values, with hydrogen being by far the most abundant atom, and the heavier elements being less and less abundant. Within the protosolar disk, planets formed from the accretion of solid particles, as an effect of coagulation and mutual collisions. The largest embryos were later able to grow by gravity, attracting the nearby particles. In this scenario, the difference between terrestrial and giant planets is a direct consequence of the condensation sequence. Near the Sun, at temperatures of a few hundred K, the only elements or molecules in solid form were silicates and metallic compounds, which are relatively rare in the Universe; as a result, the planets formed close to the Sun were relatively small and dense. In contrast, at larger heliocentric distances, where the temperature fell below about 200 K, the simple molecules (H_2O, NH_3, CH_4, H_2S, CO_2...) were in the form of ice and could be incorporated into planetary embryos. As these molecules were more abundant than the heavier ones, they could form nuclei as large as ten to fifteen terrestrial masses. When this stage is reached, models show that the gravity is sufficient for the nucleus to attract the surrounding material (mostly hydrogen and helium) by gravitational collapse. This explains the formation of the giant planets, with a big volume and a low density. The collapse of the surrounding subnebula also explains the presence of several satellites in the equatorial plane of the giant planets, as well as their ring systems.

Do we have an observational confirmation of this scenario? This answer is yes, based of the measurements (both from ground and space) of the chemical composition of the giant planets. Following this scenario, the giant planets must be enriched in heavy elements (i.e. all elements heavier than helium) with respect to the protosolar composition. In contrast, if the giant planets had formed directly from the contraction of a subnebula of protosolar (cosmic) composition, no enrichment in heavy elements would be expected. Actually, the measurements of the Galileo probe in the case of Jupiter, and the determination of the C/H ratio (derived from CH_4/H_2) in the three other giant planets, have provided evidence for the enrichment in heavy elements, which gives full support to the nucleation model of the giant planets.

In the light of this simple scenario, it would be tempting to extrapolate it to other planetary systems. In the early 1990s, before the discovery of the first exoplanets, one could reasonably expect that if extrasolar planetary systems did exist, they would show the same classification between small and dense planets near their star and giant planets at further distances. Surprisingly, it turned out to be the opposite. The first exoplanet, 51 Peg b, was a giant planet in the immediate vicinity of its host star, and most of the exoplanets discovered later shared the same strange property. Actually, the discovery of exoplanets was a true revolution for astronomers: it showed that the formation scenario of the solar system was not universal. The question was raised: how can a gaseous giant planet be formed in the immediate vicinity of its star? The nucleation model cannot account for it. To solve the paradox, the most popular mechanism currently invoked is migration: giant exoplanets are formed far from their stars and then migrate toward it as an effect of interactions between the planet and the disk. This scenario leads to another question: why was this migration absent in the case of the solar system? Actually, recent numerical simulations have shown that, within the solar system, giant planets might have encountered a moderate migration in their early history, with Jupiter moving slightly inward and the three other giants migrating outward. Numerical simulations of the dynamical evolution of the outer solar system strongly suggest that, before the end of the first Gy after the planets' formation, the Jupiter:Saturn system crossed the 2:1 resonance, which strongly perturbed the inclinations and eccentricities of all small bodies and modified their trajectories, in particular in the Kuiper Belt, as will be discussed below. As a result, the collision rates increased dramatically, which led to the Late Heavy Bombardment (LHB) observed on the surface of all old bare surfaces, which show the signature of an intense bombardment rate some 3.8 Gy ago.

8 Comets, Early Remnants of the Solar-System History

Comets have been known since Antiquity. For centuries, the unexpected apparition of these small objects which travel around the Sun on very eccentric orbits led to fears and superstition. Now the nature of comets is well understood. The study of comets has known a renewed interest over the past decades, as comets now appear as primitive remnants of the solar-system history, and also provide a link toward the study of the interstellar medium.

Comet Halley, known since Antiquity, is the most famous of all comets. Its period is 76 years, and its numerous passages have been used by Edmund Halley to understand the nature of its orbit and successfully predict its return in 1758. In 1910, the apparition of comet Halley was quite spectacular and could be monitored with photographic plates and spectra in visible light. The next apparition, 1986, was not favorable in terms of its geometrical configuration (the comet being behind the Sun at the time of perihelion) but an internal campaign using all possible ground-based and space means was designed to monitor the event. Five probes (the European probe Giotto, the two Vega probes from Soviet Union and two Japanese probes) approached the comet in March 1986. The Giotto probe sent to Earth the first images of a comet nucleus (Fig. 19.6). Data were recorded about the cometary composition, showing evidence for water (as predicted by F. Whipple as early as 1950), carbon dioxide and carbonaceous compounds. The cometary composition was found to show close analogies with the interstellar matter. In addition to space exploration, the study of comet Halley and subsequent comets largely benefited from the developed of ground-based infrared and millimeter spectroscopy, which led to the identification of many parent molecules outgassed from the nucleus.

Another important milestone of cometary research came 10 years later with the apparition of a new comet, especially big and bright, Hale-Bopp. With a diameter

Fig. 19.6 The nucleus of comet Halley, as observed by the camera of the Giotto probe during its flyby of the comet on March 13, 1986. This image was the first picture of a cometary nucleus. It shows that the nucleus is far from spherical (its dimensions are 15× 7 ×8 km) and very dark (its albedo is 0.04). Water is outgassed in localized jets and most of the nucleus surface is inactive ([7] ©ESA)

above 50 km (while most of comets have diameters smaller than 10 km), comet Hale-Bopp was an exceptional object. About twenty parent molecules were detected from ground-based millimeter and submillimeter spectroscopy. In addition, the comet could be observed by the Infrared Space Observatory (ISO), launched by ESA in November1995, which was operating in Earth orbit at the time of the comet's apparition (April 1997). In addition to the discovery of parent molecules, the nature of the cometary dust could be identified from its infrared spectrum. Its closest analog was forsterite, a magnesium-rich olivine, also identified in the circumstellar disk of some young stars. This discovery again illustrated the close connection between cometary and interstellar matter.

What is the origin of comets? The answer comes from their dynamical history. Some comets, such as Hale-Bopp, come for the first time, are usually very bright and have a high eccentricity. Their orbital history shows that they come from the Oort cloud, a large shell located at the edge of the solar system, at a distance of about 40 000 AU (or about a fifth of the distance of the closest star). They were initially formed in the vicinity of Uranus and Neptune, at 20 or 30 AU from the Sun, but have been ejected outside as a result of planetary perturbations. Occasionally, a comet is ejected from this reservoir and is sent back to the inner solar system. Then, its orbit may change again toward a shorter period, due to other planetary perturbations; this is the case of comet Halley. Another class of comets is characterized by a low inclination and a short period. They are believed to come from another reservoir located between 40 and 100 AU, the Kuiper Belt. The existence of this reservoir was suspected for decades on dynamical theoretical grounds, but was identified only recently, as discussed below.

9 The Kuiper Belt

The results mentioned above have mostly emphasized the major contribution of space research in our knowledge of solar-system bodies. Here is an example which illustrates the decisive role of both numerical simulations and ground-based observations for our understanding of the outer solar system.

About half a century ago, two astronomers, K. Edgeworth and G. Kuiper, independently suggested the presence of a population of small bodies outside the orbit of Neptune. In the early 1980s, this idea was again proposed with the hypothesis that this population could be the reservoir for short-period comets. Still, the observational means were not sensitive enough to allow the detection of these objects. The first discovery, by D. Jewitt and J. Luu, was made in 1992, and was soon followed by many others.

We know today more than 1,300 trans-neptunian objects, which can be classified in different categories according to their dynamical properties and their heliocentric distances. The "classical" objects (about 2/3 of the total), located between 42 and 47 AU from the Sun, have low eccentricities and inclinations. The "resonant" objects (about 12%), also called "Plutinos" have the same semi-major axis as Pluto, and are trapped with Neptune in a 3:2 resonance. Further out from the Sun, the "scattered" objects have a perihelion close to Neptune's orbit and large eccentricities. As an example, Sedna's aphelion distance is close to 1,000 UA. Numerical simulations have shown that the structure of the Kuiper Belt, in particular the classical and resonant objects, is a direct consequence of Neptune's outward migration.

10 Conclusions and Perspectives

Our knowledge of the solar system has made a tremendous progress over the past 50 years, by far more important that during the first half of the twentieth century. There are several reasons to this success. First, the space exploration has given us large observational datasets on planets, their main satellites, and a few small bodies. The space missions have carried both remote sensing and in-situ instruments. One should stress, in particular, the uniqueness of close-by images of surfaces (Venus, giant planets, Titan, Io, Europa, comet Halley), the mass-spectrometry measurements of the atmospheric compositions (Jupiter, Titan), and the in-situ analysis of giant planets' magnetospheres.

In addition, the use of large telescopes, associated with more and more performing imagers, photometers and spectrometers, has provided several highlights: the discovery of the trans-neptunian objects and the detection of new comets (Hale-Bopp in particular) which requires continuous monitoring and could not have been achieved from space; the identification of parent molecules in comets using millimetre heterodyne spectroscopy; the detection of the ring systems of Uranus (in 1977) and Neptune (in 1984) using stellar occultation experiments. More than ever, ground-based and space research are complementary and must be pursued in parallel for the exploration of the solar system.

We have seen the major role of numerical simulations for our understanding of the dynamical history of solar-system bodies. Such calculations allow us to constrain the formation scenario of the planets, but also to follow the dynamical history of specific objects. They allow to describe the evolution of Mars' obliquity over the past tens of My, and offer possible solutions to explain the peculiar obliquities of Venus and Uranus.

Another aspect which should not be underestimated is the analysis of extraterrestrial samples. We have mentioned their importance in the case of the lunar samples. The laboratory study of meteoritic and micrometeoritic samples (which come from Mars, asteroids and comets) is the basis for datation analyses of solar-system bodies. Samples of cometary dust have been collected by the StarDust mission and are currently under analysis. In the next decades, this research will develop further with the expected collection of extraterrestrial samples from Mars, nearby asteroids and possibly comets.

What are the main challenges of tomorrow's planetology? Among many others, two main questions appear. First, the discovery of exoplanets and their unexpected properties raises the question of the uniqueness of the solar-system formation scenario. We already know that it is not ubiquitous, as all extrasolar systems discovered so far are very different from ours. Is this an observational bias? In this case, other stellar systems, more like ours, remain to be discovered. The improving techniques of direct and indirect exoplanet detection should allow us to answer this question within the next decade. In any case, we have to understand the reasons which made our system different from the other planetary systems. In particular, why was migration moderate in the case of the solar system? Is our Kuiper Belt equivalent to the debris disk observed around other young stars? If so, why is the Kuiper Belt so much less massive? Future studies will concentrate on the early history of the solar system, both in its physical properties (composition of comets and trans-neptunian objects) and its dynamical evolution (numerical simulations). The Rosetta mission, launched by ESA in 2004, will perform an in-situ exploration of a comet, Churyumov-Gerasimenko. This mission should be a new milestone in our knowledge of the nature of cometary matter.

A second key question is the search for life in the solar system. Over the past decades, we have learned that liquid water was once present at the surface of Mars (it may have been also present on Venus during its early history but, because of Venus' recent volcanism, we cannot find any trace of it). We have also learned that a salted liquid water ocean might exist under the surface of Europa and Enceladus, and also possibly in the interiors of other outer satellites. Could life have ever appeared and developed in these environments? If so, could we look for traces of it? We have identified in Titan a possible laboratory for prebiotic chemistry, and even an equivalent of the primitive Earth. To address all these questions, a major challenge of the next decades will be the exploration of Mars' ancient sites where life might have existed, as well as an in-depth exploration of the Jovian and Saturnian systems. Space missions are under study at ESA and NASA, in the frame of a large international cooperation, and exciting discoveries have to be expected in the next coming decades.

References

1. I. de Pater, J. Lissauer, *Planetary Physics*, 2nd edn. (Cambridge University Press, Cambridge, 2010)
2. T. Encrenaz, J.-P. Bibring, M. Blanc, M.-A. Barucci, F. Roques, P. Zarka, *The Solar System*, 3rd edn. (Springer, Berlin, 2004)
3. J.S. Lewis, *Physics and Chemistry of the Solar System*, Revised edn. (Academic, San Diego, 1997)
4. P. Murdin, *Encyclopedia of Astronomy and Astrophysics* (Institute of Physics Publishing/Nature Publishing Group, Bristol, 2001)
5. S.T. Taylor, *Solar System Evolution*, 2nd edn. (Cambridge University Press, Cambridge, 2001)
6. P. Weissman, L. McFadden, T. Johnson, *Encyclopedia of the Solar System*, 2nd edn. (Academic, New York, 2007)
7. http://nssdc.gsfc.nasa.gov/photo_gallery/ (Accessed 13 Feb 2011)

Chapter 20
Exobiology: An Example of Interdisciplinarity at Work

Muriel Gargaud and Stéphane Tirard

Abstract Exobiology is an interdisciplinary field that deals with the origins of life on Earth, its evolution and its possible distribution elsewhere in the Universe. This new field appeared around the 1960s, generated by NASA's Apollo missions. Since then it has experienced a very rapid expansion, far beyond its original purpose, due to new spatial missions (Mars, Titan), associated advances in comparative planetology, and new knowledge provided by chemists, biologists and geologists. However, more immediate than looking for other life forms elsewhere in the Universe, one of the field's main goals is to understand how life, as we know it now on Earth, was able to arise. For that, we need to know from the geologists what comprised the environment of early Earth, from the chemists how a prebiotic chemistry could have been able to develop, and from the biologists how it has evolved through the last 3.5 billion years. This understanding can only take place through an interdisciplinary community, which has a need to formulate and agree to some common rules before representatives of these disciplines can truly work together.

Keywords Astrobiology • Astrochemistry

M. Gargaud (✉)
Laboratoire d'Astrophysique de Bordeaux, Université Bordeaux1-CNRS,
2 Rue de l'Observatoire, 33271 Floirac, France
e-mail: gargaud@obs.u-bordeaux1.fr

S. Tirard
Centre François Viète d'Epistémologie et d'Histoire des Sciences et des Techniques,
Université de Nantes, 2 rue de la Houssinière, 44 000 Nantes, France
e-mail: stephane.tirard@univ-nantes.fr

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2_20,

1 Why Can't We Approach Exobiology Problems as We Approach Other Scientific Questions?

Exobiology – or bioastronomy or astrobiology – whatever the name one gives to the study of the origins of life on Earth, its evolution and its possible distribution somewhere in the Universe – is indeed one of the most interdisciplinary domains of twenty-first century science. Fields like astronomy, planetology, geology, geochemistry, chemistry, biology and history of sciences are all concerned with exobiology. Sometimes questions can be completely solved by knowledge from one or two fields, sometimes questions are addressed to all fields; and even so, there is sometimes no unique answer or, even worse, there appears to be no answer at all – at least no completely satisfactory answer.

Let us take the example of a "real" interdisciplinary question, which in fact is a triple one: where, how and when did life appear on Earth?

This triplet-question is a real interdisciplinary one in the sense that it does admit some limited monodisciplinary answers, but it must certainly admit a global one able to reconcile all the monodisciplinary ones, and in doing so, to eliminate those that are mutually incompatible.

The first step one has to take when faced with such a difficult question is to identify specialists able to bring a real building block on which to construct the whole edifice. In the present case one needs:

a geologist, to reconstruct the story of the Earth since its very beginning 4.46 Ga ago ($1\,\mathrm{Ga} = 1\mathrm{giga\ annum} = 10^9$ years),

a planetologist, to definitively place the Earth in the Solar system and to determine to some degree of plausibility if it is really an exceptional planet which "by chance" has been able to harbour life, contrasted with those that do not seem to have,

an astronomer, able to reconstruct the story of Earth since the formation of the Solar system and able, for example, to discuss the origin of water, this so "simple molecule" (in terms of number of atoms) which seems to be indispensable to life, as least as we know it.

All these scientists have the opportunity to hold in their hands some (otherwise silent) witnesses (rocks, meteorites, minerals) that can be dated by geochemistry with a very high precision, and which constitute, in a sense, the "Rosetta stone" of Early Earth. However, astronomy, planetology, geology and geochemistry are able to answer only partially the questions "when" and "where", and certainly not "how". "How" is the domain of chemistry and biochemistry, which, as the opposite of astronomy and geology have no chronological clues, but only hypotheses and models, which have to be compatible with the geochemistry witnesses (meteorites, rocks, minerals) on the one hand, and with the first traces of life, identified by the geo/biologists, on the other.

When one has understood that, a big step has been taken, but one has also understood that the task undertaken is probably impossible. Indeed, an astronomer (or a geologist, or a chemist or...) has probably had no contact with geologists (or astronomers, or biologists, or. . .) since the time they were all undergraduates. Not only do they not know the other scientific communities but also, above all, they are unable to understand a paper written by a geologist (or a chemist or. . .). The question that appears more crucial is then: Why is it that I, as an astronomer (or a geologist, or a chemist. . .), am unable to understand what a geologist (or a chemist or a biologist.) says about the emergence of life on Earth? Is it simply that the "subject" is too far from my field or do I "only" feel like, say, a French person and a Chinese person who would like to say to each other that "the weather is fine today" but are only able to smile at each other. To overcome this difficulty, there is only one solution: to "learn" the language of the other.

Science, whatever the field is, must be above all "rigorous". As in the case of the French-Chinese encounter, one can build a new language (for example the "frenchine" or the "chinefrench") which will allow one to speak more or less about the weather but certainly not to read Diderot or Wu Cheng'en directly in the original version. However, to be a good scientist, one must be able to read science "in the original version", and if one considers that "exobiology" must be a scientific domain equal to all the others, the only solution one has is to become as bi-tri-quadrilingual as possible. And that will take time. Perhaps about five years will suffice to be able to speak about the "scientific weather" and probably ten before being able to write something really original about it.

Indeed, and if we believe our own experience, the different steps to acquire this bi-tri-quadri-linguality are the following:

to try to be as pedagogical as possible in teaching my scientific mother tongue to the others,

to show real good will and perseverance, and to learn the scientific mother languages of all fields linked to exobiology. This is not easy: as a well-known researcher in my field I must accept my role as a young student and to restart from the very beginning in another field and to probably appear sometimes stupid with very naive questions. (But these are the rules of the game and if all participants accept them, this is no problem).

Let us admit now that – thanks to several summer schools regularly organised by this new emerging community – we are now more or less bi-tri-quadrilingual.

Again, if we come back to the French-Chinese example, the next step is not only to speak more or less the same language as the other (fluency in several scientific fields is probably impossible but let us admit we have reached a not-so-bad level) but "to think" as the other. Perhaps the reader of this paper has had the opportunity to spend several months abroad, far from home: there is no need to insist on and to say that, when you settle in a new country, you find that most of what you had in your own home town was better (weather, food, etc.) and when you leave, as it's just the opposite, you understand how much you have adapted to this (no more) new country. . .

Exobiology is more or less the same as an exotic trip. To prepare for this trip you have learned the language of the land in which you will stay, now you have to think like your host. Let us come back to the previous questions "where", "when" and "how" did life appear on Earth? As an astronomer (or a geologist or a chemist or…) you know why you are interested in such questions and which particular aspects are interesting in your field. But the questions you really cannot answer are:

- Why is this question of "where", "when", "how" important for the others' fields?
- How does each field express its interest?
- Which monodisciplinary questions must be undertaken?
- What are the means in the hands of your colleagues to solve these questions (models, theories, observations, experiments…)?
- What are the limits and uncertainties of the results obtained with these tools?

And little by little, you discover that if you really want to discover the answers to all these questions, you are led to revisit (but also to defend!) your own monodisciplinary model or observations or results in the light of all the others' arguments. And sometimes (unfortunately not always!), after some rather hard debates, new interpretations force themselves upon all the others and interdisciplinarity wins the impossible bet. But as we will see now, it is not so new a problem…

2 Exobiology and Interdisciplinarity: An Historical Relation

2.1 Interdisciplinarity in Origins of Life Study: A Necessity

Exobiology and astrobiology are new words, but the scientific preoccupations they describe are ancient and have always called for an interdisciplinary answer.

In 1748, a posthumous book was published anonymously under the title: *Telliamed*. The author, Benoit de Maillet (1636–1738) (Telliamed was an anagram) presented a theory of panspermia and explained how life evolved on Earth. To present his thesis, he used the rhetorical method of a dialogue between an Indian man and a missionary. The Indian man who disagreed with the missionary and his classical and religious view of the occidental society of the moment pronounced Maillet's thesis: life came from space and developed later on in the oceans. Maillet's was one of the first panspermia theories, and symbolically recalls us that every theory of origins of life depends on a large spectrum of knowledge of space, Earth and life and may be the basis of interdisplinarity.

Around the middle of the nineteenth century, the understanding of biology was drastically changed by two greats discoveries. Firstly, Darwin's evolutionary theory [3] of variation and natural selection, which claimed that life had begun with very simple living organisms. Secondly, the definitive abandon of spontaneous generations after Louis Pasteur's works, which demonstrated that life could not emerge in Nature's current state. On the other hand, and as far as origins of life on Earth were concerned, the second part of the nineteenth century also saw the development of two new theories: the panspermia one and the evolutive abiogenesis conceptions. Let us come back now to some aspects of these new ideas and see how interdisciplinarity played a major role in their development.

The panspermia theories claimed that life did not emerge on Earth, but was present in the Universe for eternity, as all matter is. In 1871, Lord Kelvin (William Thomson) proposed that germs of life came to Earth with meteorites, linking for the first time biology and astronomy. Later on, Svante Arrhenius went further by saying that microscopical germs of life were present in space and driven by the motion of radiation pressure. The decline of panspermia theory occurred in 1910 with the work of the young French biologist Paul Becquerel, a specialist in latent life and the resistance of organisms under drastic conditions. He claimed, on the basis of his own experimentation, that terrestrial resistant forms of living beings (seeds, spores…) cannot survive spatial conditions (ultraviolet radiation). Therefore panspermia, as an astronomical theory of origins of life, was definitively ruled out by biology. Interdisciplinarity played then a double role: it first intervened in the elaboration of theories with complementary data, but also in the scientific debate and in the construction of contradictory arguments.

The evolutive abiogenesis was based on the idea of evolution of matter. During the second part of the nineteenth century, some chemists and biologists (H. Spencer, T. Huxley, E. Haeckel, W. Pflüger…) proposed that the complex organic matter present in living beings could be progressively synthesized and could have produced the first and simplest organisms, at the bottom of the evolution tree. Darwin himself in a famous letter to Hooker (1871) suggested that a very primitive synthesis was able to exist at the beginning of evolution.

> It is often said that all the conditions for the first production of a living organism are now present, which could have been present. But if (and oh what a big if) we could conceive in some warm little pond with all sort of ammonia and phosphoric salts, - light, heat, electricity &c present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured, or absorbed, which would not have been the case before living creatures were formed. (Darwin 1871 in [2])

What is important in that well-known text is that, not only is the first step of evolution a chemical process that can lead to life, but also the presence of life in the environment can stop current formation of life! Therefore, and as far as the origins of life are concerned, Darwin showed indubitably that chemistry, biology, evolution and historicity of life were all interdependent.

The end of the nineteenth century corresponds also to the development of biological chemistry, a field that studied the matter constituting living beings and explained the chemical mechanisms in organisms [14]. This field gave a lot of data to evolutive abiogenesis and opened the way to evolutionary explanations of origins of life. Several scientists (Darwin, as we have seen, but also Herbert Spencer, Thomas Huxley or Ernst Haeckel…) described then how matter could evolve from mineral matter to complex organic matter.

During the 1920s, two great scientists, the Soviet biologist A.I. Oparin [12] and the British biochemist and geneticist J.B.S. Haldane [10], published independently two important texts describing different scenarios for the origins of life on Earth. An important innovation of these two texts came from the fact that they were not limited to description of chemical mechanisms of evolution of matter. Indeed, they contained also important data on the geological evolution of Earth before the emergence of life and on the early Earth when life did emerge. These complementary data came from astronomy, geology, and geochemistry… Indubitably, these authors had understood the necessity of including an evolutionary chemistry hypothesis in a general context where the knowledge of initial conditions is absolutely necessary. Interdisciplinarity appeared then as a necessity and Oparin concluded his text with these words underlining the respective roles of the different disciplines:

> What we do not know today we shall know tomorrow. A whole army of biologists is studying the structure and organisation of living matter, while a no less number of physicist and chemists are daily revealing to us new properties of dead things. Like two parties of workers boring from the two opposite ends of a tunnel, they are working towards the same goals. The work has already gone a long way and very, very soon the last barriers between the living and the dead will crumble under the attack of patient work and powerful scientific thought.

Therefore at this moment, the study of the origins of life became the problem of a broad diversity of scientists. For his part, Haldane showed another aspect of the complexity of this interdisciplinarity and insisted on the consequences of historicity for the origins of life questions:

> The question at issue is “How did the first such system on this planet originate?” This is a historical problem to which I have given a very tentative answer on the not unreasonable hypothesis that a thousand million years ago matter obeyed the same laws that it does today.
>
> […] The biochemist knows no more, and no less, about this question than anyone else. His ignorance disqualifies him no more than the historian or the geologist from attempting to solve a historical problem.

Regarding interdisciplinarity, this quotation is important for two main reasons. Firstly, Haldane did not only speak of the necessity to use an interdisciplinary approach in the study of origins of life, but he spoke also of the equal legitimacy for a panel of disciplines to study this topic. This point has to be underlined because Haldane explained here, perhaps for the first time, that several disciplines are authorized to share the same problem. Secondly, we have to notice that Haldane claimed that origins of life are an historical problem and that all the disciplines have tools to solve this problem.

During the 1930s and 1940s, several important texts were engaged in an interdisciplinary way. The first of all was Oparin’s book, *Origins of life*, published in 1936, in Russian, and later on in English [13]. This book presented a scenario very close to the 1924 one, but more complete with its 200 pages and a lot of scientific references in a very broad panel of disciplines. Oparin included origins of life as a step in evolution of Earth that he accurately described, and he insisted on the primitive conditions in which chemical evolution of matter could lead to primitive living beings. This book confirmed the interest, and also the necessity, of thinking of origins of life in terms of interdisciplinarity. Oparin clearly showed how a broad variety of data, coming from a lot of scientific disciplines, are indispensable for construction of scientific hypotheses of origins of life. This view was confirmed by other scientists, in particular by J. D. Bernal and A. Dauvillier, who proposed very complete and interdisciplinary scenarios [1, 4]. However, each of them added some interrogations on the difficulties of becoming competent in such a diversity of specialities!

To conclude these first historical aspects, we can say that the question of the origins of life imposes the necessity of an interdisciplinary approach. The first part of the twentieth century was characterized by formulations of complete scenarios. However if they were complex compilations of scientific data, they conserved a very theoretical nature.

2.2 From Prebiotic Chemistry to Exobiology

During the second part of the twentieth century, the rise of prebiotic chemistry opened a new epistemological field. The problem of origins of life on Earth was then studied for the first time with a specific experimental method consisting in reproducing the conditions of the early Earth's atmosphere which could have allowed the emergence of life on Earth: this is the famous Miller's experiments [11], on the basis of Urey's hypothesis. In an apparatus built to circulate a mixture of CH_4, NH_3, H_2O and H_2 (mimicking the composition of the atmosphere of the Early Earth as planetologists imagined it at that time) and exposed during one week to electric discharges (mimicking flashes of lighting through the atmosphere), the young American chemist observed the production of amino acids (glycine, α alanine and β alanine).

This experiment showed, among many other more important conclusions, that prebiotic chemistry was strongly dependent upon an environment that needs to be described by other sciences such as planetology. On the other hand, if we look at the development of prebiotic chemistry during its first two decades, we observe that the first molecules that chemists tried to synthesize were amino acids and sugars and, later (end of the 1950s), bases of nucleic acids. In between, prebiotic chemistry was influenced by the development of molecular biology concepts that revealed the importance of nucleic acids.

These two examples show that prebiotic chemistry was, and undoubtedly is currently, limited but also led by two groups of disciplines. The first group is composed of disciplines describing primitive conditions on Earth (comparative planetology, geochemistry,...), and therefore giving the initial conditions of experiments. The second group is composed of biological disciplines (biochemistry, molecular biology, cellular biology) that study and try to synthesise molecules of early life interest.

With the founding of prebiotic chemistry, the interdisciplinary nature of studies on the origins of life changed. It was no longer an occupation of isolated scientists assembling data, but the result of the activity of a complex community of scientists, belonging to a variety of disciplines, who started to meet regularly (in workshops, summer schools and congresses). Since the 1970s they have regrouped themselves around different societies such as ISSOL – The International Astrobiology Society – and, later on, Bioastronomy, Commission 51 of the International Astronomical Union.

On the other hand, the spatial conquest of the early 1960s definitively gave birth to the new field we now call exobiology or astrobiology. Indeed with development of a spatial challenge, some questions were formulated at the Solar system scale: Is life present elsewhere in the Solar system? Do specific studies of spatial environments (research of water, nature of atmospheres…) help toward comprehension of the origin of life on Earth?

One of the first steps of this new scale of exploration concerned radioastronomy and the detection of interstellar organic molecules, but also the SETI program with a search for extraterrestrial intelligence through radio signals. But the most important event which indeed gave birth to the term exobiology (in its original sense: exo-biology, life outside the Earth) is due to spatial conquest and particularly to the first mission to the Moon. Indeed the main reason why NASA started to be interested at that time in the origins of life and its possible distribution elsewhere in the Universe came from the risk of contamination the astronauts could encounter by landing on the Moon. Later, development of the Apollo program gave the opportunity of searching for life on the Moon.

It is not the point in this chapter to list all spatial missions, all of which have played a great role in development of the exobiology field, but let us only recall two or three of them (see also Chap. 19):

- In 1976, the probe called Viking inaugurated interplanetary missions and was launched to Mars. It mapped the surface, studied the atmosphere and searched for the possibility of biological activity on the surface of the planet. The term "bioastronomy" was invented then.
- In 1996, the Global Surveyor mission mapped and transmitted a complete map of the planet and Mars Pathfinder, the first rover, explored the surface and sent famous photographs. The word "astrobiology" was invented then, but it would take too long here to try to explain the differences between "bioastronomy" and "astrobiology", and even with "exobiology", which, in Europe, is generally used instead of the two previous ones.
- In 2003, the mission Mars Explorating Rovers studied the possibility of an ancient presence of water on Mars which as of now, constitutes one of the conditions of the development of life, at least as we know it.
- In 2004, the probe Cassini-Huyghens, composed of the spacecraft Cassini and of the Huygens probe, went in orbit around Saturn. In 2005, the Huygens probe successfully landed on Titan's surface and in 2007 the radar of the Cassini spacecraft discovered a lake of hydrocarbons (see Chap. 19). But probably one of the most important results of this mission, from an astrobiology point of view, was the observation of ejected plumes from Enceladus (Saturn's satellite), which suggest the presence of a salted liquid ocean.

After this fast review of some historical aspects of the search for the origins of life on Earth and its possible detection elsewhere in the Universe, let us come back now to the "exobio daily life" of an interdisciplinary group trying to answer these questions.

3 The Question of "When" as Seen by an Astronomer, a Geologist, a Chemist and a Biologist

Among the numerous questions scientists are interested in, those related to time are probably the most fascinating. For physics, time is "just" the fourth dimension but, on the other hand, the time dimension, as we perceive it, is qualitatively different from the space dimensions. Time "flows"; time is irreversible and associated with past, present and future.

Of course, not all sciences are historical, i.e. not all sciences focus their study on similarities or differences occurring between past, present and future events. For example, a chemist studying the evolution of a reaction as a function of time certainly knows that if pressure, temperature and all other experimental conditions are kept the same, the reaction will evolve tomorrow in a deterministic way, exactly as it does today. Even if the reaction under study is irreversible, chemists (as all scientists) know that the physical laws do not change with time.

The situation is completely different for an astronomer or a geologist or a biologist interested in evolutionary problems. They must take into account the historical time and, therefore, the irreversible flow of time, the so-called "arrow of time". These scientists need to measure time with respect to a conventional reference time. Their situation can be compared to that of a historian who, in Western countries, uses as reference time the birth of Christ, even though its date is still debated among historians, being uncertain by several years. All of them, historians, astronomers, geologists or biologists have in common their need to apply to a time reference chosen by convention.

As our interdisciplinary group is interested in the history of life on Earth, the "logical" reference time t_0 could have been the time when Earth was completely formed. On the other hand, as Earth is part of the Solar system, and as we have in hand some meteorites (accurately dated), which are witnesses to Solar system formation, it seems "logical" to take as reference time the one corresponding to the oldest dated solids formed in the proto-solar nebula. That is $t_0 = 4568.5 \times 10^6$ years. However, it took probably some millions of years before this meteorite was formed, and some astronomers involved in exobiology are extremely interested by this preceding period, starting with the collapse of a molecular core cloud and ending with the first solid formed. We can then define an absolute (non-datable) time t_0^* (corresponding to the beginning of this collapse) which allows astronomers to describe the first million years by reference to t_0^* and to introduce the different stages of a protostar and a T-Tauri star phase, necessary to form the Sun. This period, which precedes t_0 by some millions of years, cannot be identified (nor named) by the International Stratigraphic Chart, since there are absolutely no geological witnesses to that period. The time between t_0 and the dating of the first rock is (informally) identified as "Hadean" by the ISC, the Archaean starting at 4.0 Ga with the dating of the first rocks dated up to now, the Acasta Gneiss in Canada. Note however that: (i) the lower limit of the Hadean is a moving one. It is dated by the Acasta gneisses but could be replaced for example by the rock recently discovered in Nuvvuagittuq if the age of 4.28 Ga can be confirmed. (ii) t_0 is the age of Calcium Aluminium Inclusions in the Allende meteorite, not the age of the Solar system for which we will never know "exactly" when it started to form.

All this discussion is to contend that working with an interdisciplinary group is not much more complicated than with a monodisciplinary one, except that in a monodisciplinary setting, all participants have been brought up with the same rules, and so a base exists from which to proceed. In an interdisciplinary setting, we all have to come back to basic principles, and to agree on them, whatever they are. (It is no more stupid or intelligent to drive on the left-hand side of the road, but depending on whether you were born British or not, it appears more or less evident. And when it has been – arbitrarily – decided that the best choice is the right one or the left one, you have to be sure that all will accept and respect this choice!) In the present case, our group could have chosen different relative times (t_1 = impact of Theia on Earth, t_2 = formation of proto-ocean, t_3 = end of the Late Heavy Bombardment, etc...) and described the different following events by reference to these relative times. For practical reasons we finally agreed to take by convention a unique t_0 (whatever it could be) and to sometimes introduce time elapsed since t_0. Indeed it's by far easier to remember that the Moon formed between 30 and 150 millions years after the formation of the first solid in the Solar system than remembering that it formed between 4.537 and 4.417 billion years ago.

After some very enriching discussions on the determination of t_0 (which we stress again is without real importance but is the visible part of the time-problem iceberg), came the problem of the representation of the "arrow of time". Indeed, let us accept that this group finally agrees on a t_0 and wants now to go further and to analyse which events are dated, with which accuracy and how relevant they are for the origins of life, etc. The best way to identify and compare them is to put them on a time scale coming from 4.567 Ga ago and ending today, or on the contrary starting from today and going back to 4.567 Ga. At first sight one will say that the choice is without importance, but again when working together, one has to use the same rules. Did the Solar system form at $t = -4.567$ Ga (that means 4.567 Ga before the present), or is the Solar system formation the reference time and is the Solar system 4.567 Ga old? Of course it's the same, except when we want to represent them on a scale. The first proposition is the "way of thinking" of geologists: they use a time scale that takes as reference time the "present time" defined as 1950 AD (AD = Anno Domini = after Jesus Christ), which is the reference used for ^{14}C

dating. Any time is thus expressed in "years before present" (year BP). On this scale, the accretion of the Solar system took place approximately 4.567 Ga BP (notated also $t = -4.567$ Ga). On the contrary, astronomers will consider that the reference time is the one of the Solar system formation, and that indeed the age of the Solar system is 4.567 Ga but that time is going from t_0 forwards to the present. For various reasons these different communities have to use different time scales, both forward and backwards, even for events that are a common subject of study. The point is just to use the same conventions, otherwise the −4.4 Ga event of the geologist will never fit with the 4.4 Gyr of the astrophysicist. (At this point we can notice also that one billion years is expressed as one giga year (Gyr) for astrophysicists and one giga annum (Ga) for geologists).

These conventions being fixed and accepted by all, the next difficult step for a participant in an interdisciplinary project is, probably, to be as critical with his "non-mother scientific field" as he could be with his own and, in the present case of the "when" question, to understand perfectly the meaning of dates associated to events. That supposes in particular a more or less good understanding of the power and limits of the chronometers each field is using.

Astronomers can collect a lot of information but the chronometers they have in hand are indirect and only of statistical nature (for example they can use the Herzsprung-Russel diagram to classify stars and understand their evolution). On the contrary, geochemists have very efficient radioactivity chronometers but the difficulty for them is to determine what is exactly dated and what they can infer from these data (the latest measurements of Calcium Aluminium Inclusion in the Allende meteorite gives a very precise age of 4.5685 ± 0.0004 Ga, but on what reliable hypotheses can we deduce that this gives also the age of the Solar system with an error bar of less than 1Ma?). For chemists, the situation is clear: chemistry is not an "historical science", (this is of course not the case of geochemistry) and they cannot use any absolute chronometer. The best they can hope for is to identify a series of very important events (appearance of membranes, of genetic information and translation, of metabolism, etc.) and to propose more or less plausible scenarios based on the present knowledge of the primitive Earth and on what is known about chemical reactivity in such environments. One should keep in mind that, with reference to what happened between 4.4 Ga (where oceans were probably formed on Earth and have made this planet "potentially" habitable) and 2.7 Ga (where we have the absolute proof of life but traces around 3.5 Ga are now nearly confirmed), the chemists have not only not a clue, but it is for them just "mission impossible" to reconstruct the prebiotic chemistry period (which lasted 1.7 Ga in the worst case, but still 900 Ma in the best). Hopefully the situation will improve in biology where molecular clocks are invaluable tools for reconstructing evolutionary timescales. However, biochemical and biological problems are so complex that of course the reliability of chronometers helps but doesn't solve everything in a definitive way.

4 A Possible Interdisciplinary Scenario of the Emergence of Life on Earth

Let us now have a look at some clues, observations and models given by astrophysics, geochemistry and chemistry and to the conclusions we could reach.

The discovery at Jack Hills (Australia) of a zircon dated 4.4 Ga by U-Pb radioactivity methods has allowed (*geochemists*) to infer, from an isotopic ratio $\delta^{17}O/\delta^{16}O$, that water was delivered on Earth at a very early stage of its formation. On the other hand *astrophysical and geochemical models* show that the formation of oceans could have taken no more than 150 Ma, so that the early Earth may have been "habitable" from 4.3 Ga.

Observations of lunar craters (*by planetologists*) suggest that the Moon (and so the Earth, even if impact craters on Earth have disappeared due to tectonics) were submitted around 4.0 Ga to an intense meteoritic bombardment for which the last sterilizing impact seems to have occurred around 3.8 Ga. That means in particular that if life had already appeared on Earth between 4.3 Ga and 3.8 Ga, probably all traces of it have been erased by this bombardment, so that we'll probably never find witnesses to that period.

Consequently, and as far as traces of life are concerned, this period between 4.3 Ga and 3.8 Ga relies nearly exclusively upon models. In particular, the problem is: can *planetologists* model the early atmosphere to learn if its composition could have been comparable to that proposed by the Miller experiment and could have led to the emergence of life or, on the contrary, could life have appeared with a non-reducing atmosphere (*chemist relevance*)?

It was not the purpose of this paper to follow in detail any scenario (for that see for example [5–9],) but on the contrary to show how exobiology is dependent upon the expertise of each participant in the quest and how careful one must be before accepting a scenario.

Let us come back to the short scenario mentioned above and list carefully all questions linked to some assertions:

- What is the reliability of the U-Pb measurement (*geochemistry*)? What do geochemists really measure?
- How, from one mineral dated to 4.4 Ga, are we sure that water was delivered on Earth at that time (*geochemistry, geology*) and that oceans have been able to form within a short time (*astrophysics, geology*)? What is a "short" time for a geologist compared to a biologist? Where did this water come from (*astrophysics, geology*)?
- How do we know how long the last intense meteoritic bombardment lasted? (*astrophysics, geology*)? What was the origin of this intense bombardment? (*planetology*)
- What was the nature of the early atmosphere? (*planetology, chemistry*)
- What is a biosignature or a clue of life? (*geochemistry chemistry, biology*)

- How are our models dependent upon geochemistry dating? What would happen to this scenario if Jack Hills's zircon were dated to 4.5 Ga or if we found a meteorite older than Allende's?
- What exoplanets discovery will tell us more about uniqeness or banality of our planet Earth?

Again the purpose of this paper was not to answer all these questions (it would take up the entire book – and some of the answers can be found in the references below) but rather to show how all these questions are inter-dependent and how the answer to one can completely change the whole landscape and open new roads for other fields.

As a conclusion we could say that whatever the methods used by the different communities involved in the origins of life (bottom-up: from the bricks of life to life as we know it; top-down: from present life to the origins of it), exobiology (and interdisciplinarity in general) have indeed a beautiful future.

Epilogue and Acknowledgments The interdisciplinary results obtained by the French exobiology community has been possible thanks to the GDR Exobio which gave recently birth to the Société Française d'Exobiologie http://www.exobiologie.fr/, the CNRS (with the Program OPV (Origines des Planètes et de la vie) and EPOV (Environnement Planétaire et Origines de la Vie)) and the CNES (Centre National d'Etudes Spatiales). All of them have supported us in the organization of Exobio'N Propriano Summer schools, Rencontres Exobio pour Doctorants (RED'N), and several workshops (Chronology of the history of life on Earth, Biosignatures, Evolution, Defining Life). These summer schools and workshops have given rise to several college textbooks, special issues of reviews and even to an Encyclopaedia of Astrobiology.

References

1. J.D. Bernal, *The Physical Basis of Life* (Routledge and Kegan Paul, London, 1951)
2. M. Calvin, *Chemical Evolution: Molecular Evolution Towards the Origin of Living Systems on the Earth and Elsewhere* (Oxford University Press, New York, 1969)
3. C. Darwin, *The Origins of Species* (1859) (Penguin Books, London, 1985)
4. A. Dauvillier, E. Desguins, La genèse de la vie, Phase de l'évolution chimique. *Actualité Scientifique et Industrielle*, Biologie Générale **917** (Herman, Paris, 1942), 127 p
5. M. Gargaud, P. Lopez-Garcia, H. Martin, T. Montmerle, R. Pascal, P. Claeys, J. Reisse, D. Despois, From suns to life: a chronological approach to the history of life on Earth. Earth Moon Planets **98**(1–4), 1–312 (2006)
6. M. Gargaud, P. Lopez-Garcia, H. Martin, T. Montmerle, R. Pascal, *Le Soleil, la Terre... la vie* (Belin Pour la Science, Paris, 2009)
7. M. Gargaud, J. Reisse, C. Mustin, Traces of past or present life: biosignature and potential life indicators. Palevol **8**(7), 593–691 (2009)
8. M. Gargaud, P. Lopez-Garcia, H. Martin, *Origins of Life: An Astrobiology Perspective* (Cambridge University Press, Cambridge, 2010), 450 pp
9. M.A.R. Gargaud, J. Cernicharo, J. Cleaves, W. Irvine, D. Pinti, M. Viso, *Encyclopedia of Astrobiology* (Springer Reference 2011), 1851 pp
10. J.B.S. Haldane, The origin of life *Rationnalist Annual*, in: (1991) *On Being the Right Size and Other Essays* (Oxford University, Oxford, 1929)

11. L. Miller Stanley, A production of amino acids under possible primitive Earth conditions. Science **117**, 528–529 (1953)
12. A. I. Oparin, *The Origin of Life, Proiskhozhdenie zhizny*. Moscow, in *The Origin of Life*. Trad. Ann Synge, ed. by J.D. Bernal (1967) (Weidenfeld and Nicholson, London, 1924)
13. A. I. Oparin, *The Origin of Life* (The Macmillan Company, New York, 1936), 1938
14. S. Tirard, L'histoire du commencement de la vie à la fin du XIX^{e} siècle. *Les sciences historiques*. Cahiers François Viète **9**, 105–118 (2005)

Index

J.-P. Lasota (ed.), *Astronomy at the Frontiers of Science*, Integrated Science & Technology Program 1, DOI 10.1007/978-94-007-1658-2,