Astrophysics and Cosmology After Gamow

Edited by

G.S.Bisnovaty-Kogan, S.Silich,
E.Terlevich, R.Terlevich, A.Zhuk

CSP

PROCEEDINGS OF THE GAMOW MEMORIAL INTERNATIONAL CONFERENCE

Astrophysics and Cosmology
After Gamow

Astrophysics and Cosmology After Gamow

Proceedings of the Gamow Memorial International Conference

Edited by

G.S.Bisnovaty-Kogan, IKI, Moscow, Russia
S.Silich, INAOE, Puebla, Mexico
E.Terlevich, INAOE, Puebla, Mexico
R.Terlevich, INAOE, Puebla, Mexico
A.Zhuk, Odessa National University, Ukraine

CAMBRIDGE SCIENTIFIC PUBLISHERS

Printed and bound in the UK by TJ International Ltd, Padstow
Cover illustration: © Prof. Artur Chernin

British Library Cataloguing in Publication Data
A catalogue record for this book has been requested

Library of Congress Cataloging in Publication Data
A catalogue record has been requested

ISBN 978-1-904868-38-5

Cambridge Scientific Publishers Ltd
PO Box 806
Cottenham, Cambridge CB4 8QY
UK
www.cambridgescientificpublishers.com

Contents

Foreword

Dear colleagues,

The scientific conference "Astrophysics and Cosmology after Gamow: Theory and Observations" is devoted to the 100 year Jubilee of Georgij (George) Antonovich Gamow, one of the great physicists and cosmologists of the 20th century.

Gamow has made an important and sometimes decisive contribution to modern physics, cosmology and biology. His work provided solutions to fundamental problems, such as the concept of a Hot Universe and neutron capture processes for production of chemical elements and also has been important for the progress in modern technology. His theory of quantum tunnelling in nuclei is part of the basis of nuclear physics, which is now becoming a powerful industrial branch of the subject. It is not surprising that in the USA Gamow took part in the development of nuclear weapons. While I do not know exactly what his input was, I am sure that he was not a layman there. Gamow himself used to say that his main input in this scientific area was to invite Edward Teller to the USA. He was his friend long before the nuclear race started. The great Russian poet Sergej Yesenin wrote prophetic lines[1]

Litsom k litsu
Litsa ne uvidat'.
Bol'shoye viditsya na rasstoyanii.

It can be translated as

When you are close to something great
You do not recognize its importance.
You should stand at a greater distance to see its essence.

In large measure these words could be applied to Gamow's personality. He was born in Odessa and, even after becoming a refugee in the USA, he retained his character and his sharp Odessian humor. In my opinion, he did not succeed in full to adjust to another culture and medium, different from Russian, Ukrainian and particulary

---

[1] "Pis'mo k zhenshchine" ("A letter to a woman")

from Odessian culture. That can be the reason why he did not receive any special award for his scientific work, which had made an epoch in science. Only UNESCO recognized Gamow's significant contribution to the popularization of science, which he accomplished with the same skill and brightness as he did his scientific work.

During Soviet times, Gamow's name was scarcely mentioned in the land where he was born, where he grew up, studied physics and attained his first scientific achievements. His book entitled "My World Line: an Informal Autobiography" was first translated into Russian only in 1994, just before the first meeting to honour Gamow was held in Odessa.

A century after his birth, enough time has elapsed to recognize and appreciate in full the importance of Gamow's place in science, his contribution and his charming personality. The scientific community is now well aware of his work, and, at last, historians in different countries, not only in his native Russia and Ukraine, study his life and achievements.

G.S. Bisnovatyi-Kogan

I Plenary and Memorial Talks

1

George Gamow – a Giant in 20th Century Science

M.M. Shapiro

University of Maryland, College Park, MD, USA
e-mail: mmshapiro@mailaps.org

I had the good fortune to meet George Gamow sixty-two years ago, and we became friends. At the time he was already known for his ground-breaking work on radioactivity and for his collaborative research with Edward Teller whom he had brought to the physics department of George Washington University in Washington, D.C. soon after Teller immigrated to the USA. Our early encounters occurred at the David Taylor Model Basin, a US Navy laboratory where George served as a consultant, and I was involved in research on the physics of underwater explosions. [It was thanks to this experience that I was invited by J. Robert Oppenheimer to join the wartime laboratory in Los Alamos, NM.]

After a post-war stint in Oak Ridge, I returned to Washington at the invitation of Franz Kurie to join his Nucleonics Division at the Naval Research Laboratory. There I founded the Laboratory for Cosmic-ray Physics, and soon learned of the exciting ideas about the early universe which were being explored by Gamow with his associates Herman and Alpher. They were developing the theory of what has come to be known as the Big Bang – a name somewhat derisively suggested by Sir Fred Hoyle, whose own proposal of a steady-state universe competed with Gamow's theory. [I note parenthetically that at a conference celebrating Hoyle's 60th birthday, he told me privately that, upon the discovery of the cosmic microwave background, he "threw in the towel"].

In 1953 I agreed to be director of the Nucleonics Division at the Naval Research Laboratory, succeeding Dr. Franz Kurie. My first initiative in this position was to telephone George Gamow, inviting him to serve as a consultant to the Division. He agreed, and his regular Friday visit was the highlight of the week. We could always expect to learn something new and exciting. For example, we might hear of the (apparently) colliding galaxies in Cygnus. Or he would tell of John Wheeler's "geons", and their possible role in elementary particle physics. The breezy, humorous tone in which he recounted these developments was typical of the light touch that made his popular books so readable.

George's wide range of interests was legendary. His seminal contributions to cosmology and to nuclear physics would have assured him of a niche among the giants of 20th century science. But his keen probing mind explored the life sciences as well. Thus his insights into the puzzle of the genetic code helped pave the way to its solution. I was privileged to hear his early lectures on this problem at the Cosmos Club in Washington, D.C.

In addition to his work in fundamental science, George was also interested in a variety of applied problems. He was in demand as a consultant in government and industrial laboratories; many areas of defense research benefited from his originality and his masterly grasp of physics. A discussion with Gamow was animated by his appetite for exploring new ideas, and by his intuitive perception of the meaning underlying a new discovery. He always seemed to be in high spirits, fond of telling humorous anecdotes, and dashing off Ogden-Nash-type verses on the spur of the moment.

When first encountered personally, George's irrepressible jocularity evoked pleasant surprise. I remember the puzzled reaction of William G. Penney (later Lord Penney, head of the UK's atomic energy program) during a wartime conference on the physics of underwater explosions – also attended by John von Neumann. Penney had known of Gamow, but had not previously met him. It was a revelation to Penney that this celebrated physicist could be so frolicsome in the midst of a serious technical discussion. Penney turned to me in puzzlement and remarked: "This man Gamow – sort of a buffoon, isn't he!"

Gamow's gift of communication, so evident in his widely appreciated books, made him a charming platform personality. He could amuse and edify and inspire an audience of laymen or students or savants, just as he could hold a coterie of friends spellbound. I remember the impression he made on an audience of adolescents in the Washington area, who had flocked to hear his Christmas lectures on "Big-Bang Cosmology". The Philosophical Society of Washington, which sponsored the lectures, anticipated that the demand for seats would be great, so we arranged for the use of the spacious Lisner Auditorium of George Washington University. On two successive evenings, marked by freezing weather, the hall was filled with youngsters enthralled by his contagious enthusiasm and his instinct for showmanship (he inflated balloons to represent the expanding universe).

The abiding love of George Gamow was the pursuit of astrophysics and cosmology. Predictably, he was in his glory at the Texas Symposia on Relativistic Astrophysics. At the 1964 conference in Austin, when the newly discovered quasars were the mystery of the hour, no one expressed the general mood as succinctly as George, who gaily recited:

"Twinkle, twinkle, little quasar
How I wonder what you are ..."

At a subsequent "Texas Conference" in New York, George basked in the excitement over the discovery of the universal microwave radiation he had predicted many years earlier as a remnant of the primordial "big bang".

Gamow's income was divided into three parts – one part came from teaching, another from consulting, and the third from book writing. His legacy, too, may

be summed up in a trichotomy: his contributions to unraveling the mysteries of nature, the impact of his magnetic personality on his associates, and his superb interpretation of science to the general public.

2

Reflections on George Gamow's Unique Scientific Legacy and Personality

I.B. Pustylnik

Tartu Observatory, Tartu district Tõravere, Estonia
e-mail: izold@aai.ee

Abstract

We discuss the rich scientific legacy of George Gamow, an outstanding figure in XXth century physics and cosmology, whose talent bridged the gap between East and West long before the decline of the totalitarian system. Our analysis is based partly on Gamow's original scientific and popular papers, partly on the reminiscences of his colleagues and contemporaries. G. Gamow's creative style is unique in a sense, representing a peculiar amalgam in which artistic imagination, a sober rational approach of a scholar and psychoanalytic insight are all combined in one creative personality. We discuss how these different facets of Gamow's rare talents are reflected in his transparent physical models, specifically that of a hot primordial Universe and confront some of his predictions with the realities of contemporary extragalactic research and observational cosmology.

Keywords: history of astronomy

1 Introduction

In my contribution I will concentrate on some peculiarities of George Gamow's unique creative style without going into the technical details of his fundamental discoveries. This will be the subject of presentations prepared by cosmologists, specialists in relativity and theoretical physics. Earlier excellent reviews have been presented on the occasion of G. Gamow's 90th birthday anniversary and published in Russian by the well-known physicists A.D. Chernin [2] and V.Ya. Frenkel [3]. My modest analysis is based partly on Gamow's own books – "Thirty Years that Shook Physics" [7], "One, Two, Three... Infinity" [8], "Biography of Physics" [6] and on two memorial volumes "Cosmology, Fusion and other Matters" [1] and the most recent one – "The 1996 George Gamow Symposium" (held in Washington) [17]. It can be regarded as a continuation of my earlier article published eight years ago in *Astronomical and Astrophysical Transactions* (see [13]) and was based on my talk presented almost exactly ten years ago at the first George Gamow international conference held in his home town.

2 Prognostic Power of Gamow's Models

For a historian of science George Gamow is a blessed figure. "Godfather" of the hot primordial universe model, one of the leading 20th century physicists and cosmologists who have shaped our modern views on physical nature of the universe and its constituents, a prolific popular writer, the author of numerous popular books, monographs and several hundreds of scientific papers with their topics ranging from pure mathematics, micro- and macro-physics to astronomy, cosmology and even biology, his unusual charismatic character invariably attracted much attention of both professionals and general public throughout all his scientific career. After his death in 1968 George Gamow's rich legacy is being constantly re-assessed because recent spectacular achievements in space research, observational cosmology have brought about numerous confirmations of his prophetic forecasts. In retrospective it is remarkable to see how the original ideas put forward by G. Gamow have germinated and produced rich fruits 10–20 years after his death. Let us review here just some highlights of the most significant developments. The paramount importance of an amazing isotropy of Cosmic Background Radiation (CMB) has been immediately recognized by both cosmologists and galactic investigators.The attempts to detect small inhomogeneities in the large scale distribution of neutral hydrogen at 21 cm wavelength were undertaken when G. Gamow was still alive. Soon it became clear that the radiation defines an extremely distant frame of reference: that of the matter which last scattered it (see, for example [12]). Thus an attractive idea originated to find an answer to the questions "Where are we going?" and "Where have we been?" by making very precise measurements of CMB radiation both on large scale and small scale respectively. The total velocity of the Earth with respect to this frame should reveal an apparent excess of radiation in the direction of motion due to the Doppler shifts of photons. The idea of a new cosmic "ether" had inspired many at that time to study in great detail the overwhelming influence of inverse Compton scattering in the early history of the universe (for instance, Sunyayev–Zeldovich effect of anisotropy of scattered radiation by the clusters of galaxies [15]). On another front, experimental efforts culminated in 1989 by launching a cosmic platform COBE (Cosmic Background Explorer) which measured the CMB temperature with extremely high precision yielding the value $T = 2.726° \pm 0.01\,K$ and as an additional bonus discovered the quadrupole component in the distribution of CMB [14]. The existence of CMB radiation had been predicted already by G. Gamow in 1946 [4].

Another cosmic experiment, this time staged by nature itself, an explosion in 1987 of a supernova in the Large Magellanic Cloud brought about a brilliant verification of a different prediction made by G. Gamow, according to which a star exploding as a supernova would lose 99% of its energy not in a spectacular optical flash but in a quiet URCA process which is accompanied by the huge neutrino flux release. In reality detection of the signal from this kind of event on the 23rd February of 1987 with the aid of a neutrino telescope has been claimed by the Japanese team of scientists (Kamiokanda II) (for more details see [10]) and later on independently confirmed by an American and Soviet–Italian groups.

The theory of potential barrier penetration and the theory of beta decay developed by G. Gamow jointly with E. Teller serve now in large measure as the foundations of our understanding of why various radioactive elements have their characteristic lifetimes. These, so-called *aeonglasses* constitute now a part of the flourishing branches of both stellar and galactic evolution as well as the early history of the solar system.

3 Different Facets of Gamow's Talent

If one tries to grasp the most essential and at the same time the most intriguing facets in George Gamow's character as a scientist one tends to agree with Stanislav M. Ulam, a mathematician and a close associate of G. Gamow who asserts that "his overwhelming curiosity was directed towards the large lines of the theories which attempt to make us understand the *scheme* of things in the universe – the foundations of physics, the very set-up of the dimensions, and physical variables and the constants from which theoretical physics is built – the stage on which all phenomena take place, that is the nature of space and time in the very small and in the universe at large" [18]. This view is echoed by many others. Thus according to R.A. Alpher and R. Herman (Reflections on Big Bang Cosmology, p. 11) in his paper "Any Physics Tomorrow" Gamow proposed that the main job of theory was to express new empirical constants through existing constants of nature, namely, the velocity of light, Planck's constant, and a fundamental length for microphysics; Boltzmann's constant in a special category; and the gravitational constant for macrophysics. In the last chapter of his famous "collective portrait" of the most outstanding physicists of the 20th century entitled "Thirty Years that Shook Physics" Gamow tries to catch a glimpse of future physics. Starting with a bitter and reproaching remark that "we are now dragging through the lean and infertile years in theoretical physics and looking for better luck in the years to come" and stating that "we have no crystal ball for predicting the future development" he grasps at the straw of cosmonumerology and dimensionless analysis. He selects the velocity of light in vacuum, Planck's constant as the fundamental constants of nature and suggests that the third august quantity should be the classical electron radius. According to his view every physical quantity can be expressed through these three. Even his popular book "Mr Tompkins in Wonderland" whose hero, the little clerk of a big city bank, is transferred in his dreams into the imaginary worlds governed by quantum theory and the principles of relativity carries a subtitle "Stories of c, G, and h".

One of the most amazing features of his talent which brought G. Gamow public recognition is the extreme transparency, visuality of his physical models. According to the testimony of S. Ulam "Gamow always tried to find even in most abstract theories, motivations or similes, i.e., analogies with precisely understood models" [18]. It looks as if this facet of his rare talent has something to do with G. Gamow's artistic abilities. Admittedly he made the most ingenious illustrations in several of his popular books. It may sound a paradox but the more complicated and abstract seemed to be the task he faced, the more eloquently his artistic imagination manifested itself. Let me illustrate this thought with just two examples. In his critical essay reviving Gamow's crucial role in elaborating the drop model of atomic nucleus, R.H. Stewer, American historian of science, offers a straightforward explanation: "Possibly the very shape of the nuclear potential well – which when viewed from above resembles a volcanic cone containing energetic alpha particles inside it – sparkled his thoughts" (The 1996 Gamow symposium, p.36). Here is another, perhaps, even more striking example of what S. Ulam defines as Gamow's urge "to find even in most abstract theories, motivations or similes, i.e. analogies with precisely understood models" (Cosmology, Fusion and other Matters, G. Gamow's memorial volume 1972, p. 60). In his popular book "Biography of Physics", Gamow introduces an uninitiated reader to a rather intricate notion of quantum mechanics – penetration by an α-particle through a high potential barrier surrounding uranium nucleus. He uses a "visible simile", analogy between de Broglie waves and the waves of light. As always, he himself makes the illustration for his book and sketches the slab of glass reminding the reader of a familiar phenomenon in geometric optics of the total internal reflection of light. Next, making a mental experiment he draws another imaginary slab of glass supposedly only several wavelengths away from the original one explaining through Snail's law of refraction the difference between geometric and wave optics thereby vividly illustrating how the photon can "jump" from one glass slab to the other without violating physical laws and thereby facilitating a comprehension of the mechanism leading to nuclear barrier penetration. One cannot help feeling that this rare facet of G. Gamow's numerous talents has something to do with his artistic perception of the world.

4 Oppenheimer's Syndrome?

An intriguing question, which invariably emerges, whenever the scientific legacy of an eminent nuclear physicist is scrutinized, is an issue of the so-called Oppenheimer's syndrome. To what extent G. Gamow was prone to this syndrome, how much reflections on the devastating nature of the forces released from the "bottle" by him and his colleagues engaged in the chain reaction project preoccupied his inquisitive mind? It is well-known that during World War II Gamow worked at the American Navy high explosives laboratory with A. Einstein himself and with J. von Neumann. After the war Gamow was involved in the Bikini bomb test to study the effect of nuclear blast shock waves on the surface structure of ships and on the hydrogen bomb project at Los Alamos jointly with E. Teller (for more details see the reminiscences of F. Saafeld in "The 1996 Gamow Symposium", p. 26).

It seems to us that Gamow's famous cartoons to some extent betray his uneasy thoughts on this delicate issue. It is well-known that during his work in the famous T-division Gamow sketched several mock shields (reminiscent of a medieval coat of arms) dedicated to the project. One of them depicts the leader of T-division Mr. Carson Mark. The shield is encircled on all sides by the motto "For he is a jolly good fellow". Close to a portrait of a hero Gamow places a sinister atomic mushroom. Another corner of the same shield contains a number of figurines, presumably,

the children of "jolly, good fellow" crawling out of something which looks like either the belly of a whale or a horn of plenty. Unwittingly, the whole picture emanates some apocalyptic-sarcastic or even sardonic expressiveness. Here is another example (this time both amusing and ironical) of G. Gamow's evasive style of treating the same ticklish subject. In his book "One, Two, Three... Infinity" he muses: "In respect to nuclear energy we live (or rather lived until quite recently) in a world similar to that of Eskimo, dwelling in a subfreezing temperature for whom the only solid is ice and the only liquid alcohol. Such an Eskimo would never have heard about fire, since one cannot get fire by rubbing two pieces of ice against each other, and would consider alcohol as nothing but a pleasant drink, since he would have no way of raising its temperature above the burning point. And the great perplexity of humanity caused by the recently discovered process of liberating on large scale the energy hidden in the interior of the atom can be compared to the astonishment of our imaginary Eskimo when shown ordinary alcohol burner for the first time" ("One, Two, Three... Infinity", p. 168). Curiously enough, in the same book Gamow refers to the principle of statistic disorder and even the burning question of the difference between living and non-living forms of matter again addressing the meaningful figure of alcohol. "We should have a much closer analogue of a biological process if, for example, the presence of a single alcohol molecule (C_2H_5OH) in a water solution of carbon dioxide gas should start a self-supporting synthesizing process that would unite one by one the H_2O molecules in the dissolved gas forming new molecules of alcohol. Indeed, if one drop of whisky put into a glass of ordinary soda water should begin to turn this soda into pure whisky, we should be forced to consider alcohol as living matter" ("One, Two, Three... Infinity", p. 236). And following this passage Gamow proceeds to a description of recent progress in studying the structure of the simplest living forms – viruses. So why spirits after all? Because it is as contagious as the virus or because G. Gamow himself brooding over the eternal questions of good and evil of his epoch, over the fate of his homeland, from time to time looked into the glass in a desperate search for an answer to the questions that were tormenting him (see an article of D.I. Ivanenko in the supplement to Gamow's biography "My World Line" describing the difficult episodes in Gamow's life)?

Close friends called him affectionately Geo. If we recalls the original meaning of the word for all of us, the inhabitants of this planet, it brings us to Goethe's immortal figure of doctor Faust. The choice of a jocular parody on Goethe's Faust as an epilogue to his book "Thirty Years that Shook Physics" is by no means incidental, it is in a sense symbolic. As G. Gamow explains it, the notion of a massless and chargeless particle with the spin 1/2 was not so easy to swallow even for the outstanding physicists of the thirties. And then they sought refuge in the world of art. But it seems to be only a part of the story. In fact, the book was completed four years before his death. The figure of doctor Faust and his tantalizing question "Where shall I clasp you, infinity of Nature?" was apparently appealing to George Gamow more and more in the epilogue of his life.

Acknowledgements

The author expresses his gratitude to Prof. A. Chernin for stimulating discussions on the scientific legacy of George Gamow. Financial support from the Estonian Science Foundation Grants 4701 and 5760 is gratefully acknowledged.

References

[1] R.A. Alpher, R. Herman, in F. Reines (ed) *Cosmology, Fusion and Other Matters, George Gamow Memorial Volume*, Colorado Associated Univ. Press, p. 304 (1972).

[2] A.D. Chernin, *Uspekhi Fizicheskikh Nauk*, **164**, 867 (1994).

[3] V.Ya. Frenkel, *Uspekhi Fizicheskich Nauk*, **164**, 845 (1994).

[4] G. Gamow, *Nature*, **70**, 572 (1946).

[5] G. Gamow, *Mister Tompkins in Wonderland*, Cambridge Univ. Press, Cambridge (1960).

[6] G. Gamow, *Biography of Physics*, Harper and Brothers, New York, (1961).

[7] G. Gamow, *Thirty Years that Shook Physics, the Story of Quantum Physics*, Heinemann, London (1966).

[8] G. Gamow, *One, Two, Three... Infinity*, The Viking Press, New York, (1961).

[9] G. Gamow, *My World Line, An Informal Autobiography*, Viking Press, New York, (1970).

[10] K. Hirata, T. Kajita, M. Koshiba et al., *Phys. Rev. Letters*, **58**, 1490 (1987).

[11] D.D. Ivanenko, *Gamow's Epoch Seen with the Eyes of His Contemporaries*, in *G. Gamow, My World Line*, Nauka, Moscow, p.278 (1994).

[12] R.B. Partridge, T.D. Wilkinson, *Phys. Rev. Letters*, **18**, 557 (1987).

[13] I. Pustylnik, *Astron. and Astrophys. Trans.*, **10**, 167 (1997).

[14] G. Smoot et al., *Astrophys. J.*, **396**, L1 (1992).

[15] R.A. Sunyaev, Ya.B. Zel'dovich, *Ann. Rev. Astron. Astrophys.*, **18**, 537 (1980).

[16] E. Teller, in F. Reines (ed) *Cosmology, Fusion and Other Matters, George Gamow Memorial Volume*, Colorado Associated Univ. Press, p. 60 (1972).

[17] E. Harper, W.C. Parke, G.D. Anderson (eds) *The George Gamow Symposium: sponsored by the George Washington University and the Carnegie Institution of Washington*, ASP Conf. Ser. **129**, (1997).

[18] S. Ulam, in F. Reines (ed) *Cosmology, Fusion and Other Matters, George Gamow Memorial Volume*, Colorado Associated Univ. Press, p. 272 (1972).

3

The Big Bang Man

A.D. Chernin[1,2]

[1] *Sternberg Astronomical Institute, Moscow, Russia*
e-mail: chernin@sai.msu.ru
[2] *Tuorla Observatory, Turku, Finland*

Abstract

George Gamow was a man with boundless interests and imagination that took him from relativity theory to quantum mechanics and nuclear physics, back to cosmology and then to genetics. He had made seminal contributions to these key areas of modern knowledge which ensured him an enduring place among the giants of modern science. A brief review of his work is given, and new results inspired by Gamow's Big Bang are presented.

Keywords: cosmology

1 Introduction

Gamow wrote twenty-five books, which is plenty for the lifetime of a scientist. The most famous of these is a popular science book entitled "Mr Tompkins in Paperback" printed 19 times since the first publication in 1965. In 1994, it was published again in English in Cambridge, England, and – at last! – in his native Russian for the first time in Moscow. Mr. Tompkins is back (although, in truth, he never really went away).

He enjoyed writing books on popular science, while his major interest was to attack and solve the problems of nature – in physics, cosmology and biology. People often asked him how he wrote books that were so successful. And he replied [1]: "Well, it is a deep secret, so deep that I do not know the answer myself!"

But now we know the answer: it was because he wrote books about what he himself was doing in science. And what he was doing was excellent. His three most significant contributions to science are:

- He discovered the quantum nature of alpha-decay in nuclear physics (1928),
- he proposed the theory of the hot initial state of the Universe based on which he predicted the existence of the cosmic microwave background radiation (1946–1953),
- he found the clue to the genetic code in biology (1954).

And in addition to these contributions, he proposed numerous other ideas in various areas of modern science and technology. Gamow exercised random brilliances that led him to toss off ideas like lawn seed.

Edward Teller, the father of the American H-bomb and Gamow's close friend for more than 35 years, recalled: "Now, Gamow had a fertile imagination. He was an exceedingly nice guy, and furthermore, he was the only one of my friends who really believed I was a mathematician... Now, I'm sorry to say that ninety percent of Gamow's theories were wrong, and it was easy to recognize that they were wrong. But he didn't mind. He was one of those people who had no particular pride in any of his inventions. He would throw out his latest idea and then treat it as a joke. He was a delightful person to work with." (See Blumberg and Panos [2].)

In total, Gamow published about 150 scientific papers. Among these, three papers were on the three ideas mentioned above that brought world recognition and glory to him. He wrote these papers as a single author. Most of the rest of his production was done with coauthors, Nobel Prize winners H. Bethe, F. Bloch, M. Delbruck, S. Chandrasekhar, and L. Landau among them. Not only E. Teller liked to work with him.

If, to be conservative, we assume that only these three solo papers plus those done together with Teller and the five Nobelists were not completely wrong, it would be just ten percent or so of Gamow's work – in good agreement with Teller's mathematical estimate.

2 George Washington University

After a year of work in Europe with M. Curie, P. Langevin, E. Rutherford, and N. Bohr, Gamow received an invitation from George Washington University and soon moved to America. In Washington DC, Gamow and his wife started their American life. He worked at George Washington University for over fifteen years. For the last years of his life Gamow was a professor at Colorado University, and now one of the highest buildings in campus is called the Gamow Tower in his honour.

He died in 1968 having spent 34 years on the American continent and having fully realized the dream of his youth – to travel around the world and to do research in physics.

In 1934, when accepting the professorship in Washington, Gamow managed to get another theoretical physicist of his choice also invited, so that he would have a chance to discuss problems with someone. This man was E. Teller, a Hungarian-born physicist who, at that time, had a temporary position in England. They had first met at Bohr's institute in Copenhagen in 1929 when Teller was only 21. Since then Teller had developed a widespread reputation as an outstanding lecturer and, like Gamow himself, he became a full professor at George Washington University.

Before teaming up with Gamow, Teller had worked primarily on the application of quantum theory to molecules. But their association influenced him to transfer his emphasis to nuclear physics. As a result of their work in Washington, "he stepped onto a path leading to the development of nuclear energy for purposes of war and peace", as Teller's biographers said (Blumberg and Panos [2]), and this may be equally said about Gamow himself.

Probably their most important contribution to pure nuclear physics at that time was the formulation of what is now known as the Gamow–Teller selection rule for beta-decay.

Gamow and Teller were also getting more and more interested in the applications of nuclear physics to astrophysical phenomena. They studied the problem of thermonuclear reactions in stars to explain their energy sources. In 1938, in Washington, Gamow organized a conference on energy production in the sun and other stars. Among the invited physicists and astronomers was Hans Bethe.

As Gamow described in his autobiography [1], Bethe "on his arrival knew nothing about the interior of stars, but everything about the interior of the nucleus". Towards the end of the conference, he came out with a possible scheme of nuclear reactions involving hydrogen and carbon which could produce enough energy to explain the observed radiation of stars – the famous carbon cycle (for this work he was awarded the Nobel Prize in Physics in 1967).

Soon after that, Gamow and his research student Charles Critchfield in collaboration with Bethe developed another scheme of thermonuclear synthesis that was based on direct proton–proton reactions. We know now that proton–proton reactions play the dominant role in the sun and stars fainter than the sun, whereas for brighter stars, such as Sirius, the carbon cycle is dominant.

During World War II, Gamow worked in the Division of High Explosives of the U.S. Navy studying the propagation of shock and detonation waves in various ordinary, not nuclear, highly explosive materials. It would have been, of course, more natural for him to work on nuclear weapons, but he was not cleared for such work until 1948 (after Hiroshima, as he usually emphasized).

The reason was presumably his Russian origin and also his free manner of speech, as he himself suspected. He loved to tell stories and was a great master of this art. One of his lovely stories was about his service as a colonel in the field artillery of the Red Army at an age of about twenty [1]. There was more of art than of reality in his story, but security officers of any nation do not like stories like that.

3 Los Alamos

Gamow joined Los Alamos Scientific Laboratory in July 1948. He was welcomed there by Teller and some other friends, Stanislav Ulam among them. If Teller is known as the "father" of the bomb, then Ulam is recognized as its "mother". They both credited Gamow with initiating the theoretical work in the United States that ultimately led to the biggest manmade explosion. Teller referred to the subject of thermonuclear reactions as "Gamow's game". Teller said that the champion at Gamow's game was Bethe.

We cannot say much about Gamow's work at Los Alamos because we know very little about it. Many years passed, but the information is still mostly classified. So we do not know exactly which "ten percent" of Gamow's ideas were used in the H-bomb production.

Gamow joked afterwards that his major contribution to the development of the American hydrogen bomb was bringing Edward Teller to the US.

However, when Teller testified against Robert Oppenheimer, the father of the A-bomb, before the Atomic Energy Commission (and as a result Oppenheimer was made to leave Los Alamos), Gamow said openly that he considered the accusations against Oppenheimer as false and absurd.

We can only imagine what Gamow would have felt about the "balance of terror", as Winston Churchill called it, between the United States and Russia that began after World War II and intensified late in the summer of 1949, when the Soviet Union exploded its first A-bomb (made, among others, by some of Gamow's close colleagues in Leningrad). In this period of anti-intellectual hysteria inspired by Senator Joseph McCarthy in America, professors, politicians, artists, writers were all challenged to defend their loyalty to the States. We can speculate that this would have reminded Gamow of his personal experience in Russia with the regime of another Joseph.

But what we definitely know is that just at that same time, Gamow developed his idea of the hot Big Bang. Dealing with the greatest manmade explosion, he thought as well about the physics of the greatest natural explosion. He had proved to be mixed up in both.

4 Big Bang

Gamow had not returned to his home country, the country of his brilliant beginnings. But he returned to the science of his youth, to cosmology.

Gamow's Big Bang theory was actually started in 1946, two years before Los Alamos. His initial idea was a bold combination of nuclear physics and cosmology. The basis of the idea is the concept of the expanding Universe developed by Friedmann, Gamow's teacher in Leningrad.

Friedmann's theory has been completely confirmed by direct astronomical observations of the last half-century. It says that the matter of the Universe is expanding and this motion lasts about 15–20 billion years. So it began 15–20 billion years ago with a tremendous explosion that had a universal space-time scale. The theory claimed that the matter of the Universe was extremely dense at the beginning of this explosion.

Gamow added to this that the matter of the Universe was also extremely hot at that stage. The Universe as a whole was even denser and hotter than the interior of stars. If so, thermonuclear reactions could develop in the early Universe, like in the sun and other stars. And these reactions were assumed to produce all the chemical elements observed now in the Universe of which all the cosmic bodies ("and non-cosmic ones, like you and I", Gamow once said) are made.

We will mention here only a few aspects of the theory in its historical retrospective.

First of all, it is remarkable that Gamow believed that the idea of the hot initial state of the Universe was put forward not by him, but by his teacher Friedmann. In his autobiography which was already quoted [1], he wrote: "According to Friedmann's original theory of the expanding Universe, it must have started with a singular state at which the density and temperature of matter were practically infinite."

Note two points in this phrase: original theory and infinite temperature. We studied all of Friedmann's published work on cosmology; their total number is not too large – two articles in international and Russian journals and one popular science book. Our inspection has revealed not a word about the initial temperature of the Universe. A possible (or most probable?) explanation for Gamow's reference to temperature in Friedman's original theory is that this idea was considered as natural or even trivial by Friedmann himself and the people around him. It seems to have been cosmological folklore in Leningrad in the 1920s, so Gamow could not pretend to own this idea.

It is also interesting that Gamow's initial aim to explain the observed abundance of all chemical elements in the Universe was not achieved in his theory. Only helium and a few other light nuclei could be produced during cosmological nucleosynthesis. Other elements are assumed now to be produced in stars, especially when they explode as supernovae.

Finally, the prediction of the cosmic thermal radiation which was considered initially as a simple by-product of Gamow's theory has proved to be its most impressive result.

Let us say something about this radiation and its prediction.

The cosmic microwave background radiation is a wonderful – and at the same time quite natural – phenomenon. It appeared when neither the earth nor the sun and stars existed in the Universe. It may sound like an ancient myth of creation, but so it is that the Universe was full of light from the very beginning, and this light was not radiated by cosmic bodies. It was born together with the Universe and is still filling the cosmic space. Now it exists in the form of radio-waves, mostly in the band from 10 cm to 1 mm, and can be observed with radiotelescopes (or radiometers) from the earth's surface as well as from orbiting space stations.

If one assumes that the matter of the Universe was very hot at a remote past, as Gamow believed, then one would expect that the thermal radiation was in the space alongside with ordinary matter, like protons and electrons. This is a direct consequence of the general laws of physics known since the end of the XIXth century. And this radiation could not disappear with time, but evolved together with other forms of cosmic matter. This evolution led, first of all, to the cooling of matter – both particles and radiation – because of their expansion together with the Universe as a whole.

The whole expansion of the Universe is described by Friedmann's cosmology. His theory covers all about the dynamics and geometry of the Universe. And it may be used to find the law of cooling for matter and radiation during the cosmic evolution. But one needs to know something else to be able to say which temperature the radiation has now. To talk about the present temperature we need to know what temperature the radiation had in at least one moment in the past.

Gamow was the first to recognize that this problem could be solved with the help of nuclear physics applied to the hot and dense matter in the early Universe. In principle, his idea was very simple. He suggested the calculation of the temperature of the cosmic matter at the epoch when thermonuclear reactions took place in this matter. It could be made if we know what the output of these reactions should be. For instance, the reactions should give about 30 percent of helium (by mass), as it is observed in the present Universe. And this is enough to calculate the temperature of the cosmic matter.

He made this suggestion to his young colleagues, Ralph Alpher and Robert Herman. They performed extremely complicated calculations and found the present temperature of the cosmic radiation. This was 5 degrees above absolute zero [3].

Gamow himself hated complicated calculations. He said that he was hopeless at working out simple arithmetic. But he also made his estimation of the radiation temperature – even without any nuclear physics. He used only the simplest formulae of Friedmann's theory and the basic laws of thermodynamics. His result was 7 (or 6, in another version) degrees above the absolute zero of temperature (see [4, 5]).

How did he manage to do it? It was puzzling even for his young colleagues – they knew definitely that something additional was needed, in addition to these simple formulae, to get this result [6]. A closer look at Gamow's work shows that an implicit assumption was made about a relation between cosmological characteristics. It made the problem solvable. And the assumption was so simple, so natural and so implicit that nobody could have any objection, or even an exact realization of it (this is discussed at length in [7]).

It is well-known that the cosmic radiation predicted by Gamow and his colleagues was discovered by American radioastronomers Arno Penzias and Robert Wilson (Nobel Prize of 1972). The measured temperature proved to be about three degrees – in excellent agreement with the calculations mentioned above – both the sophisticated and the simple one. Further observations have confirmed this discovery. The most exact measurements of the temperature of the cosmic microwave radiation were made with the COBE space laboratory which is still in orbit: 2.730 ± 0.060 K. This is a fantastic accuracy for cosmological observations!

The discovery of the cosmic microwave background was, and still is, a great triumph of cosmology and of all modern sciences. This way Friedmann's dynamics and geometry together with Gamow's nuclear physics and thermodynamics have given rise to today's cosmology which is called the Big Bang theory.

5 Big Bang Today

Now in 2004, we celebrate Gamow's centennial. The most recent developments in present-day cosmology have been largely due to the Wilkinson Microwave Anisotropy Probe (WMAP) mission which studies the Cosmic Microwave Background. The WMAP results have provided a reliable high-accuracy dataset for a new breakthrough in the studies of important features of the real Universe. Two implications

from the WMAP data seem to be especially impressive: these are the cosmic energy composition [8] and the idea of a finite Universe [9].

The WMAP observations, together with the earlier studies of distant supernovae and in agreement with all the bulk of the concordance data, indicate that the major cosmic energy ingredients are dark energy (70%, in round numbers) and dark matter (25%). The rest is ordinary luminous matter or baryons (4%) and radiation (less than 1%) which is the cosmic microwave background (CMB) photons, neutrinos, gravitons and other possible relic relativistic particles.

The energy ingredients are well measured, but poorly understood. Indeed, the microscopic properties of dark energy and dark matter are unconstrained by cosmological observations and remain quite uncertain. Not only the dark sector, but even baryons are of unknown origin. Actually only CMB photons (0.005%) have a clear physical interpretation. In addition, the energy fractions vary with time because of cosmological expansion. The cosmic energy mix looks preposterously complex, if not strange and absurd.

Alongside with the cosmic energy composition, the WMAP data provide insight into the global structure of the Universe. Evidence for a finite size of the 3D co-moving space [9] comes from the same WMAP dataset and is found in the power spectrum of CMB anisotropy at the lowest harmonics, or the largest spatial scales. In particular, the quadrupole is only about one-seventh as strong as would be expected in an infinite flat space. A similar effect (while not so dramatic) is observed also for the octopole. The lack of power on the largest scales indicates most probably that the space is not big enough to support them. According to Luminet et al. [9], the current size of the 3D co-moving volume is about 15 billion light years. This length is comparable to – actually slightly less than – the current radius of the visual horizon. If confirmed, this is a major discovery about the nature of the Universe.

Surprisingly enough, the idea of a finite Universe suggests a new approach to the problem of the composition of cosmic energy. Indeed, simple regularity can be recognized behind the WMAP figures, if the spatial size of the Universe is considered finite. The new regularity is found in terms of the Friedmann integral which is known since Friedmann's paper [10] on the cosmological expansion and serves as a genuine time-independent characteristic of each of the cosmic energies (see more in [11–13]).

The finite Universe that emerges from the WMAP data reveals new simplicity, symmetry and self-consistency. Simplicity is seen from the time-independent covariant and robust recipe of the cosmic energy composition. The list of the energy ingredients – vacuum (V), dark (D) matter, baryons (B), radiation (R) rewritten in terms of the Friedmann integral A looks very simple indeed:

$$A_V \sim A_D \sim A_B \sim A_R \sim 10^{60\pm1} M_{Pl}^{-1}, \qquad (1)$$

where $M_{Pl} \sim 10^{19}$ GeV is the Planck mass, i.e. the mass at which a particle's Compton wavelength is equal to its Schwarzschild radius; here it is used as a conventional measure of mass/energy, time and length in the units at which the Planck constant and the speed of light are both equal to one. New internal (non-geometrical) time-independent covariant symmetry is behind this relation: the four energy ingredients are members of a set, a quartet, with the same (approximately) values of the

Friedmann integral. The integral is the conservation value appropriate to cosmic internal symmetry.

Cosmic internal symmetry as a remarkable time-independent feature of the evolving Universe must have roots in fundamental physics. A link to theory is identified above under the assumption that the energy ingredients are well described by simple physics. Specifically, it is assumed that dark energy is cosmic vacuum, or the Einstein cosmological constant; and dark matter are weakly interacting stable particles with masses near the electroweak energy scale ~ 1 TeV. If so, the identity of the values of Friedmann integral results from standard freeze-out kinetics at the early epoch of TeV temperatures. The figure for the integral is also given by the theory: it is a power of the hierarchy number $X = \bar{M}_{Pl}/M_{EW} \sim 10^{15}$:

$$A \sim X^4 M_{Pl}^{-1} \sim 10^{60} M_{Pl}^{-1}. \qquad (2)$$

The numerical value of the integral agrees well with the empirical result.

The large dimensionless number X provides a common natural quantitative measure of the cosmological figures and phenomena associated with new symmetry. Among them are

(1) the vacuum density and the present-day non-vacuum densities

$$\rho_V \sim \rho_D \sim \rho_B \sim \rho_R \sim X^{-8} M_{Pl}^4 \sim 10^{-120} M_{Pl}^4; \qquad (3)$$

(2) the total dark matter mass in the finite Universe

$$M_D \sim X^4 M_{Pl} \sim 10^{60} M_{Pl}; \qquad (4)$$

(3) the total number of TeV dark matter particles

$$N_D \sim X^5 \sim 10^{75}; \qquad (5)$$

(4) the total number of CMB photons

$$N_R \sim X^6 \sim 10^{90}; \qquad (6)$$

(5) the cosmic entropy per dark matter particle

$$D \sim X \sim 10^{15}; \qquad (7)$$

(6) the amplitude of the initial perturbations of the gravitational potential (which is time independent and also scale independent for the Harrison–Zeldovich spectrum)

$$\Delta \sim X^{-1/3} \sim 10^{-5}; \qquad (8)$$

(7) the density perturbation amplitude at the epoch of 1 TeV temperature

$$\delta(z_{EW}) \sim X^{-2} \sim 10^{-30}. \qquad (9)$$

The relations show the WMAP Universe as a self-consistent physical system: its basic local (intensive) and total (extensive) quantitative characteristics prove to be in mutual correspondence and harmony.

It is suggested in this paper that the interplay between gravity and electroweak-scale physics is behind new symmetry. Such a conjecture invokes a special significance of the electroweak energy scale in fundamental physics. It assumes that the origin of species and the generation of cosmic perturbations are due to a common physical process at the epoch of the electroweak temperatures. The essential items of the conjecture are as follows:

(1) *The electroweak freeze-out* described by the model of Section 4 which shows how the WIMP dark matter might originate as a thermal relic of the hot Universe; this is a standard approach which may easily be refined in many (secondary) details. The model involves cosmic radiation with the density $\sim M_{EW}^4$ at the electroweak epoch and shows how radiation might come into the internal symmetry relation, $A_R \sim A_D \sim A_V$, with dark matter and vacuum. If electroweak baryogenesis (see below) produces baryons in the desirable quantity, the complete symmetry relation for all the four energies would be guaranteed.

(2) *Electroweak breaking of supersymmetry* which gives rise to dark energy with the observed density. Fundamental theory predicts that there may be various huge contributions to the total energy density of physical vacuum. In particular, it is known that the contributions may be both positive and negative, and at least some of them may be infinite on the absolute value: the vacuum energy density of fermionic particles is $-\infty$, while it is $+\infty$ for bosonic particles. Zeldovich [14] proposed that the contributions might cancel each other, so that the net result would be the zero vacuum density, if there could be an exact symmetry between the fermions and the bosons. That was the first guess on supersymmetry, a few years before the supersymmetry concept was invented. This symmetry proves to be violated, and because of this the expected value for the total vacuum density must be non-zero. It may be assumed that violation is due to electroweak-scale physics. But why is the net sum as it is? Though no answer to the question has been found [15], Zeldovich's idea looks very attractive in the context of the gravity-electroweak interplay.

(3) *Electroweak baryogenesis* which is based on the Sakharov–Kuzmin mechanism. The criteria for baryogenesis include baryon number violation, C and CP violations, and a departure from thermal equilibrium. Electroweak baryogenesis is especially interesting because these criteria can be satisfied in the current Standard Model of particle physics. The quantitative predictions of the theory are favorable, while some additional assumptions are yet needed.

(4) *Electroweak phase transition* which is accompanied by non-causal super-horizon density fluctuations as seed of the observed cosmic structure. The density amplitude of the fluctuation δ might have a natural level at the electroweak epoch. More discussion of this aspect of the gravity-electroweak interplay will be given in a separate paper.

The phenomenological and microscopic understanding of cosmic internal symmetry and gravity-electroweak interplay is a new challenging problem in cosmology stemmed from the impressive new discoveries made in the field after Gamow.

References

[1] G. Gamow, *My World Line: an Informal Autobiography*, Viking Press, N.Y. (1970).

[2] A. Blumberg, L.G. Panos, *Edward Teller: Giant of the Golden Age of Physics*, Scribner, N.Y. (1990).

[3] R.A. Alpher, R. Herman, *Phys. Rev.*, 74, 1737 (1948).

[4] G. Gamow, *Danske Vidensk. Selsk. Mat.-Fys. Medd.*, **27**, nr. 10, 15 (1953).

[5] G. Gamow, *Danske Biologiske Medd.*, **22**, nr. 8, 1 (1955).

[6] R.A. Alpher, R. Herman, in B. Bertotti et al. (eds) *Modern Cosmology in Retrospect*, Cambridge Univ.Press, Cambridge p.129 (1990).

[7] A.D. Chernin, *Physics-Uspekhi*, **37**, Nr.8 (1994).

[8] C.L. Bennet et al., *ApJ Suppl. Ser.*, **148**, 1 (2003).

[9] J.-P. Luminet et al., *Nature*, **425**, 593 (2003).

[10] A.A. Friedmann, *Z. Physik*, **10**, 277 (1922).

[11] A.D. Chernin, *Physics-Uspekhi*, **44**, 1099 (2001).

[12] A.D. Chernin, *New Astron.* **7**, 113 (2002).

[13] A.D. Chernin, *New Astron.*, **8**, 79 (2003).

[14] Ya.B. Zeldovich, *Physics Uspekhi*, **95**, 209 (1968).

[15] A.D. Dolgov, preprint hep-ph/0405089 (2004).

4

George Gamow and the Discovery of Cosmic Microwave Background Radiation

I.D. Novikov[1,2]

[1]*Niels Bohr Institute, NORDITA, Blegdamsvej 17,*
DK-2100 Copenhagen Ø, Denmark
[2]*Astro Space Center of the Lebedev Physical Institute,*
84/32 Profsoyuznaya, Moscow, 117810, Russia

Abstract

In this paper I review some of the scientific contributions to the field of cosmology and especially to the cosmic origin of the chemical elements and prediction of CMB of (by) G. Gamow. I also outline some aspects of the history of the discovery of CMB radiation. Finally I will emphasize the role of CMB as a very powerful tool of modern astronomy and tell about the future PLANCK mission.

It is a great pleasure and honour for me to be a speaker at this Commemoration Meeting to recall Gamow's outstanding contributions to astronomy and cosmology.

For the preparation of this paper I used some material from books by Zeldovich and Novikov [1, 2] and my own review [3]. Many details are in the monograph [4].

George Gamow was the founder of the modern theory of the Hot Universe. It is really the first theory of the physics of the Universe.

Modern cosmology is a vast, rapidly developing field of knowledge. During a rather long period in the second quarter of the XXth century its theoretical basis was the set of cosmological models devised by the Soviet mathematician Alexander Friedmann. Its observational basis was the work that led to the discovery by the American astrophysicist Edwin Hubble of the law describing the redshift in the spectra of galaxies.

Though the kinematics of the evolving Universe became known decades ago, only in more recent years did research into the physics of processes occurring in the expanding Universe receive a reliable observational and theoretical basis. Describing the situation in cosmology at the beginning of the 1960s Steven Hawking [5] wrote: "Cosmology was thought of as a pseudo-science where wild speculation was unconstrained by any possible observations".

Indeed at that period cosmology was only a science about the mechanics of motion of huge mass of matter in the Universe. With very rare exceptions there was not any physics in cosmology at that period. We can compare that period with the period in astronomy at the beginning of the XIXth century, when there was no astrophysics at all. Astronomy itself was only the science about the motion of celestial bodies, the motion of planets and their satellites and so on, but there was nothing about the physical processes. An analogous situation was at the beginning of the second half of last century in cosmology. Mainly it was the science about only two values, the speed of the expansion of the Universe – the Hubble parameter, and the deceleration parameter which characterizes the total average density of matter and the gravity of this matter which leads to the deceleration of its motion.

Nevertheless practically just after the discovery of the expansion of the Universe and during the subsequent period some serious attempts to understand the physics of the Universe were done. Proof that the Universe expands appeared roughly at the same time that modern ideas on nuclear physics arose. (See the work by Friedmann in 1922 and 1924; the announcement in 1919 of the discovery by Rutherford of nuclear reactions; the work of Hubble in 1929; and the discovery of the neutron by Chadwick in 1932). Almost simultaneously there arose, first, the questions of what the Universe and the matter filling it represent near the singularity and, second what the physical preconditions for the answer to these questions are. Already, the most general idea about the singularity, the infinite-density state at the beginning of the expansion, had led to the idea that the original, prestellar matter was more or less homogeneous – in stark contrast with the diverse structure and composition of the stars and of other celestial bodies today. Then, a pleasant-sounding but nebulous formulation appeared: the Universe is a single "primordial atom", or the entire Universe is a single nucleus or drop of nuclear fluid (Lemaitre, the beginning of the 1930s). Lemaitre did the first attempt to understand the physics of the birth of the Universe. It is remarkable how some of his ideas are close to those of modern cosmology. He wrote in 1931 [6]: "I would rather be inclined to think that the present state of quantum theory suggests a beginning of the world very different from the present order of Nature. Thermodynamical principles from the point of view of quantum theory may be stated as follows: (1) Energy of constant total amount is distributed in discrete quanta. (2) The number of distinct quanta is ever increasing. If we go back in the course of time we must find fewer and fewer quanta, until we find all the energy of the universe packed in a few or even in a unique quantum.

Now, in atomic processes, the notions of space and time are no more than classical notions; they fade out when applied to individual phenomena involving but a small number of quanta. If the world has begun with a single quantum, the notions of space and time would altogether fail to have any meaning at the beginning; they would only begin to have a sensible meaning when the original quantum had been divided into a sufficient number of quanta. If this suggestion is correct, the beginning of the world happened a little before the beginning of space and time. I think that such a beginning of the world is far enough from the present order of Nature to be not at all repugnant."

Lemaitre had proposed that the initial state of the matter in the Universe may be thought of as a sea of neutrons. Later it was found that this assumption led to a contradiction with observations.

The situation was as follows. A neutron is an unstable particle. When free, it decays in about 15 minutes into a proton, an electron, and an antineutrino. Hence, as the Universe expands, neutrons inevitably decay, and protons appear. A newborn proton would couple to a surviving neutron and form the nucleus of a deuterium atom (a deuteron). A deuteron would join another deuteron, and so forth. The reaction of forming progressively more complex atomic nuclei would rapidly proceed until alpha particles (nuclei of helium atoms) formed. Calculations show that more complex atomic nuclei would practically never be formed. The entire matter would thus convert to helium. This conclusion sharply contradicts observations. We know that young stars and interstellar gas mostly consist of hydrogen, not of helium.

Observations of the abundance of chemical elements in nature thus completely reject the hypothesis of the cold neutrons beginning of the expansion of the Universe.

In the 1940s, George Gamow [7] published a paper that suggested a "hot scenario" of the initial stage of the expansion; this paper was followed by several more written by him and then by his colleagues Ralph Alpher and Robert Herman (see references and some details in [1]). It was assumed that the temperature of matter at the onset of expansion was extremely high.

The main objective of the authors of the Hot Universe scenario was to analyze nuclear reactions at the start of the cosmological expansion and to obtain the currently observed ratio of abundances of various chemical elements and their isotopes.

Why was it initially assumed that all chemical elements had to be born at the start of the expansion of the Universe? The fact is that it was erroneously decided in the 1940s that the time that had elapsed since the Big Bang was 1–4 billion (10^9) years (instead of the current estimate of 15 billion years). We are aware now that this error was caused by the underestimation of the distances to galaxies and consequent overestimation of the Hubble constant. Comparing this time $(1–4)\times10^9$ years, with the age of the Earth, that is $(4–6)\times10^9$ years, Gamow and his colleagues concluded that even the Earth and other planets, let alone the Sun and other stars, had to grow from the primary matter and that all chemical elements had formed at the early stage of the expansion of the Universe because they simply could not be formed at any later period.

We know now that the time of expansion of the Universe is 15×10^9 years. The Earth has been formed not from primary matter but from matter that went through the stage of nuclear reactions (nucleosynthesis) in stars. The theory of nucleosynthesis in stars successfully explains the main body of observations of the abundance of chemical elements, assuming that primordial stars grew from matter that mostly consisted of a mixture of hydrogen and helium. The matter of old first-generation stars, enriched in heavier elements, was ejected into space. New stars and planets grew out of this matter. As a result, the need to explain the origin of all elements (including such heavy elements as iron, lead, etc.) at the early stage of expansion disappeared. But the main idea of the hot Universe hypothesis proved to be correct.

The theory suggested by George Gamow and his colleagues established that the expanding matter of the Universe turns into a mixture consisting of hydrogen (75 per cent) and helium (25%). This is the material from which stars and galaxies are later formed. What prevents the matter in the Hot Universe scenario from turning into helium, as it did in the scenario that started with cold neutron liquid?

The crucial factor is the hot state of the matter. Hot matter contains numerous energetic photons. It also contains protons and neutrons that tend to merge into deuterons. The photons break down the deuterium nuclei formed by paired protons and neutrons, thus interrupting at the very beginning the chain of reactions leading to the synthesis of helium. When the expanding Universe cools down sufficiently (to a temperature of less than a million degrees), a certain amount of deuterium survives and ultimately gives helium.

The Hot Universe scenario gives definite predictions on the content of helium in prestellar material. We have already mentioned that helium abundance must be about 25% of the total mass.

Gamow's hypothesis was not the last attempt to treat the Big Bang. One attempt, made at the beginning of the 1960s by Ya. Zeldovich [8] was to modernize the cold Universe scenario: the new version predicted the conversion of all matter not to helium (as in the former version) but to pure hydrogen. It was also assumed that all other elements were generated much later, already in stars.

Originally the Hot and Cold Universe models were regarded as attempts to give a detailed explanation of elemental abundances in the prestellar matter. Attempts at determining which of the theories is true were mostly oriented at an analysis of the observational data on abundances of chemical elements. Unfortunately, such observations, and especially their analysis, are extremely complicated and depend on a number of assumptions. If abundances of chemical elements in the Universe were the only test for a theory, the truth could be very hard to unearth. Indeed, it is not easy to decide how much helium and other elements had been synthesized in nuclear processes in stars and how much had been left by the processes occurring in the early Universe.

Fortunately, another method of testing was found. The Hot Universe theory gives a most important observational prediction that is a direct corollary of being hot. This is a prediction of the electromagnetic radiation that has survived in the Universe till our epoch from the time in the past when matter was dense and hot.

The temperature of matter decreases in the process of cosmological expansion, and the radiation also cools off, but nevertheless electromagnetic radiation with a temperature from (in different versions of the theory) a fraction of one degree to 20–30 degrees Kelvin must survive until today.

This radiation that must have been left to cool from the ancient epochs of the evolution of the Universe – if the Universe had really been hot at that time – is known as the Cosmic Microwave Background (CMB) radiation: the Soviet astrophysicist I. Shklovsky suggested calling it the relic radiation. The electromagnetic radiation at such a low temperature consists of radio waves in the centimeter and millimeter wavelength ranges. Correspondingly, the crucial test of whether the Universe had been hot or cold was a search for such a radiation. If found, then the Universe had been hot; otherwise it had been cold.

The prediction of the existence of the CMB radiation was done by G. Gamow and his collaborators. Unfortunately nobody tried to search for CMB. Later we will discuss why it so happened.

The subsequent history is well known. Penzias and Wilson [9] stumbled across an unexplained, regular radio noise at a wavelength of 7.3 cm. At the same time Dicke, Peebles, Roll, and Wilkinson [10] were preparing an apparatus for the measurement of the radio background at a wavelength of 3 cm with the deliberate intention of verifying the theory of the hot Universe and determining the temperature of the CMB radiation. At that time they had made corresponding theoretical estimates. (see the book by Dicke [11]). Then hearing of the results of Penzias and Wilson, Dicke and his group quickly interpreted their results as confirmation of the theory of the Hot Universe and assigned a temperature of about 3 K.

It was in this accidental manner that Penzias and Wilson made the spectacular discovery of the twentieth century, the discovery proving that the Universe had been hot at the beginning of its expansion. In 1978, they won the Nobel Prize for Physics for this discovery.

Let us return now to a problem that belongs to the history of science but that still excites interest in physicists and astrophysicists. The brilliant American scientist Steven Weinberg [12] wrote in his book "The first three minutes: A modern view of the origin of the Universe": "I want especially to grapple here with a historical problem that I find both puzzling and fascinating. The detection of the cosmic microwave radiation background in 1965 was one of the most important scientific discoveries of the twentieth century. Why did it have to be made by accident? Or to put it another way, why was there no systematic search for this radiation, years before 1965?"

The reader is reminded that the prediction of the radiation at a temperature of several degrees Kelvin that fills the Universe was made by Gamow 15 years before the discovery by Penzias and Wilson, at the end of the 1940s and the beginning of the 1950s.

Could it be that sufficiently sensitive radiotelescopes, capable of detecting this radiation, were simply nonexistent? We will see later that this is a rather unlikely reason. Weinberg expresses this opinion too. But the crux of the matter lies elsewhere.

Physics offers numerous examples in which the prediction of a new phenomenon was made long before the discovery could become technically feasible. Nevertheless, if the prediction had solid ground and was important, physicists never let it out of sight. The prediction used to be tested immediately after the technical possibility arose. Weinberg gives an example of the prediction of the antiproton in the 1930s. At that period, physicists could not even dream of the equipment needed for its experimental discovery. In the 1950s, however, it became possible, and a special accelerator was constructed in Berkeley to test this problem.

In the case of the cosmic microwave background, however, radioastronomers knew nothing about it or about the possibilities of detecting it until the middle of 1960s.

Why did it happen?

Weinberg gives three reasons. The first is that the Hot Universe theory was developed by Gamow and his colleagues to explain the abundances of all chemical elements in nature by their synthesis at the very beginning of expansion. As we mentioned above, this hypothesis had to be abandoned since heavy elements were synthesized in stars. Only the lightest elements date back to the earliest moments of expansion. The first version of the theory contained other imperfections as well. Later all shortcomings were corrected for, but at the end of the 1940s and in the 1950s the theory as a whole was not regarded as quite reliable.

In addition during this period Fred Hoyle, Thomas Gold and Herman Bondi created another theory of the Universe: the Steady State theory. This theory became very popular. In the Steady State theory there was no CMB radiation. Fred Hoyle wrote in Frontiers of Astronomy [13]: "It is a suspicious feature of the explosion [big bang] theory that no obvious relics of a superdense state of the Universe can be found".

The second reason was the poor communication between theorists and experimentalists. The former had no idea whether the microwave background was detectable by the available observational equipment; the latter never heard that one would be advised to search for this radiation.

The third reason is psychological. It was extremely difficult for physicists and astrophysicists to accept that calculations dealing with the very first minutes after the beginning of expansion could indeed describe the true picture. Indeed, the contrast of durations was too striking: the first several minutes against the tens of billions of years that separate that epoch from ours.

George Gamow understood this psychological difficulty and tried to clarify the point giving an appropriate comparison. He wrote in *Scientific American*: "To many a reader, the statement that the present chemical constitution of our Universe was decided in half an hour five billion years ago will sound nonsensical. But consider a spot of ground on the atomic proving ground in Nevada where an atomic bomb was exploded.... Within one microsecond the nuclear reactions generated by the bomb produced a variety of fission products. In 1956, 100 million microseconds later, the site was still "hot" with the surviving fission products. The ratio of one microsecond to three years is the same as the ratio of half an hour to five billion years! If we can accept a time ratio of this order in the one case, why not in the other?"

Another reason, which I consider to be the most important, was formulated by Penzias [14] in his Nobel lecture. The point is that neither the first nor subsequent papers by Gamow and his coworkers ever mentioned that the cosmic radiation background could be detected, even in principle. Moreover, it seems that Gamow and the others were of the opinion that the detection was completely unfeasible! Penzias [14] says:

"As for detection, they appear to have considered the radiation to manifest itself primarily as an increased energy incident upon the earth. This contribution to the total energy flux would be masked by cosmic rays and integrated starlight, both of which have comparable energy densities. The view that the effects of three components of approximative equal additive energies could not be separated may be found in a letter by Gamow written in 1948 to Alpher (unpublished and kindly provided to me by R.A. Alpher from his files). "The space temperature of about 5 K is explained by the present radiation of stars (C-cycles). The only thing we can

tell is that the residual temperature from the original heat of the Universe is not higher than 5 K." They do not seem to have recognized that the unique spectral characteristics of the relic radiation would set it apart from the other effects."

It so happened that I played my part at the next stage of this story. I started to work in cosmology at the beginning of the 1960s, shortly before the microwave background was discovered. I had just completed the postgraduate course at Moscow University; my science adviser was Professor Zelmanov. He was mostly interested in the mechanics of motion of masses in cosmological models when no simplifying assumption are made about their uniform distribution. He was less interested in specific physical processes in the expanding Universe. At that time, I knew almost nothing about the Hot Universe model.

Not long before the end of my postgraduate term, I was attracted to the following problem. We know how different types of galaxies radiate in different ranges of wavelength of the electromagnetic radiation. With certain assumptions about the evolution of galaxies in the past and having taken into account the reddening of light from remote galaxies owing to the expansion of the Universe, one can calculate today's distribution of galactic emission over such wavelengths. In this calculation, one has to remember that stars are not the only sources of radiation, and that many galaxies are extremely powerful sources of radio waves in the meter and decimeter wavelength ranges.

I began the necessary calculations. Having completed the postgraduate term, I joined the group of Professor Zeldovich. Our interests focused mostly on the physics of processes in the Universe.

All calculations were carried out jointly with A. Doroshkevich. As a result, we obtained the calculated spectrum of galactic radiation, that is, of the radiation that must fill today's Universe if one takes into account only the radiation produced when galaxies were born and stars began to shine. This spectrum predicted a high radiation intensity in the meter wavelength range (such wavelengths are strongly emitted by radio galaxies) and in the visible light (stars are powerful emitters in the visible range), while the intensity in the centimeter, millimeter and some still shorter wavelength ranges of electromagnetic radiation must be considerably lower.

Since the Hot and Cold Universe scenarios were eagerly discussed in our group (consisting of Zeldovich, Doroshkevich and myself), the paper that Doroshkevich and I prepared for publication added to the total the putative radiation surviving from the early Universe if it indeed had been hot. This hot Universe radiation was expected to lie in the centimeter and millimeter ranges and thus fell into the very interval of wavelengths in which the radiation from galaxies is weak! Hence, the relic radiation (provided the Universe had been hot!) was predicted to be more intensive, by a factor of many thousands or even millions, than the radiation of known sources in the Universe in this range of wavelengths.

This background could, therefore, be observed! Even though the total amount of energy in the microwave background is comparable with the visible light energy emitted by galaxies, the relic radiation would be in a very different range of wavelengths and thus could be observed. Here is what Penzias [14] said about our work with Doroshkevich [15] in his Nobel lecture:

"The first published recognition of the relic radiation as a detectable microwave phenomenon appeared in a brief paper entitled "Mean Density of Radiation in the Metagalaxy and Certain Problems in Relativistic Cosmology" by A.G. Doroshkevich and I.D. Novikov in the spring of 1964. Although the English translation appeared later the same year in the widely circulated *Soviet Physics–Doklady*, it appears to have escaped the notice of other workers in this field. This remarkable paper not only points out the spectrum of the relic radiation as a blackbody microwave phenomenon, but also explicitly focuses upon the Bell Laboratories 20-ft horn reflector at Crawford Hill as the best available instrument for its detection!"

Our paper was not noticed by observers. Neither Penzias and Wilson, nor Dicke and his coworkers were aware of it before their papers were published in 1965; Penzias told me several times that this was very unfortunate.

We are not yet through the chain of missed opportunities that plagued the discovery of the relic radiation.

It is found that the microwave background could have been discovered as early as in 1941 (even before the Hot Universe Theory had been formulated).

At the beginning of the 1940s McKellar was analyzing absorption lines in the spectrum of the star ζ Ophiuchi; the absorption was caused by cyan molecules (compounds of carbon and nitrogen). McKellar ([16, 17]) concluded that these lines (in the visible part of the spectrum) could have arisen only if the light was absorbed by rotating cyanogen molecules. The rotation must be excited by radiation at a temperature of about 2.3 K. Of course, neither McKellar nor anybody else thought at that moment about the possibility that the rotation of molecules is excited by the relic radiation. In fact, the theory of the Hot Universe had not even been developed yet!

It was only after the cosmic microwave background had been discovered, that Shklovsky [18], Field, Hitchcock, Wolf and Thaddeus published their papers, in 1966 and later, which showed that the rotation of interstellar cyanogen molecules observed in the stellar spectra of ζ Ophiuchi and three other stars was indeed excited by the relic radiation.

This shows that one manifestation of the microwave radiation background, even though indirect, was observed in 1941: its effect on the rotational state of interstellar cyan molecules.

But even this is not yet the final chapter of the story. Let us return to the question about the technical feasibility of detecting the cosmic microwave background. At what time did this become possible? Weinberg writes: "It is difficult to be precise about this, but my experimental colleagues tell me that the observation could have been made long before 1965, probably in mid-1950s and perhaps even in mid-1940s." Is this correct?

In the autumn of 1983, Dr. J. Shmaonov of the Institute of General Physics, with whom I was not previously acquainted, telephoned me and said that he would like to talk to me about things relevant to the discovery of the cosmic microwave background. We met the same day and Shmaonov described how, in the middle of the 1950s, he had been doing postgraduate research in the group of well-known Soviet radioastronomers S. Khaikin and N. Kaidanovsky: he was measuring radio waves coming from space at a wavelength of 3.2 cm. Measurements were done with a horn antenna similar to that used many years later by Penzias and Wilson. Shmaonov carefully studied possible sources of noise. Of course, his instrument could not have been as sensitive as those with which the American astronomers worked in the 1960s. Results obtained by Shmaonov were reported in 1957 in his Ph.D. thesis and published in a paper [19] in the Soviet journal *Instruments and Experimental Methods*. The conclusion of the measurements was: "The absolute effective temperature of the radiation background ... appears to be 4 ± 3 K". Shmaonov emphasized the independence of the intensity of radiation on direction and time. Errors in Shmaonov's measurements were high and his 4 K estimate was absolutely unreliable, but nevertheless we now realize that what he recorded was nothing other than the cosmic microwave background. I want to add that Shmaonov's observations were reanalyzed by N. Kaidanovsky and Yu. Parijsky [20]. Their conclusions were: "Thus, one can conclude that the contribution of the relic radiation (CMB) was 2 ± 1 K. Of course the result is rough but unambiguous". Unfortunately, neither Shmaonov himself, nor his scientific advisors, nor other astronomers who saw the results of his measurements, knew anything about the possibility of the existence of the relic radiation and so failed to pay the results the attention that they deserved. They were soon forgotten. When Doroshkevich and I, having completed our calculations, were calling in 1963 and 1964 on several Soviet radioastronomers with the question "Do you know of any measurements of the cosmic background in the centimeter and shorter wavelength ranges?" not one of them remembered Shmaonov's work.

It is rather amusing that even the person who made these measurements failed to appreciate their significance, not only in the 1950s – this is easy to explain – but even after the discovery of the microwave background by Penzias and Wilson in 1965. True, at that time Shmaonov was working in a very different field. His attention turned to old results only in 1983, in response to semiaccidental remarks, and Shmaonov gave a talk on the subject at the Bureau of the Section of General Physics and Astronomy of the USSR Academy of Sciences. This event took place 27 years after the measurements and 18 years after the publication of the results of Penzias and Wilson.

But even this is not the final chapter. Some years ago I was informed that measurements by Japanese radioastronomers at the beginning of the 1950s may also have detected the radiation background. This work, as well as that of Shmaonov, went unnoticed then, was not realized later, and remained practically unknown to the scientific community.

Fate takes unexpected and tortuous turns. Nevertheless, the entire story is very instructive. To hit upon a phenomenon is not yet equivalent to discovering it. One has to realize the significance of the find, and give the correct explanation. A combination of circumstances and sheer luck do play a role here – no doubt about it. Nevertheless, success does not come by accident. Success requires lots and lots of work, vast knowledge, and persistence in the work itself and in bringing the results to the attention and recognition of others.

The discovery of CMB was one of the most unexpected astronomical discoveries of the last century. In contrast to that the existence of the "ripples" (the tiny variations that are seen in the background radiation temperature) in the CMB radiation had been predicted and the corresponding proposal to search for them submitted to NASA in the 1970s. These "ripples" have been discovered by COBE-satellite in

1992, see [21][1]. This project was preceded by the Russian project RELIKT, which was the first space project devoted to the investigation of the CMB anisotropy.

The "ripples" contain the key to the mystery of why cosmic structure finally developed. Observations of the ripples provide data to investigate the process of the birth of the Universe, the origin of primordial small irregularities in the matter distribution in the Universe and they allow us to estimate main cosmological parameters (the Hubble constant, the total amount of matter in the Universe, the amount of baryonic matter and other parameters). All this information can be extracted from the angular power spectrum of the CMB anisotropy. It is the dependence of the intensity of the variations of the CMB temperature on the angular scale (or corresponding multipole number l of spherical harmonics). The theoretical power spectrum has clear peaks. In the standard cosmological model the matter is ionized until a redshift $z_{rec} \approx 1000$. Fluctuations in the hot matter-radiation fluid before z_{rec} oscillate like sound waves. The peaks in the power spectrum are a consequence of these sound waves which we see at the epoch of recombination $z_{rec} \approx 1000$ when the Universe became transparent to radiation. We call these peaks in the power spectrum Sakharov oscillations because a corresponding effect had been pointed out by A. Sakharov [22], see also [1, 23, 24]). The first Sakharov peak (it is called also "Doppler peak") at $\theta \approx 0.5°$ corresponds to the distance that a sound wave can travel at the time of recombination. Other peaks are at smaller angular scales. The COBE satellite mapped the sky with an angular resolution of $\approx 7°$. This was not enough for the measurement of even the first peak. Since then, experiments with finer angular resolution have detected CMB anisotropy on smaller angular scales and have produced evidence for the presence of the first Sakharov peak. Finally two groups [25] (the BOOMERANG experiment) and [26] (the MAXIMA-I experiment) announced the discovery of the first Sakharov peak. During the last decade a series of ground based and balloon borne experiments have been performed. Finally the Wilkinson Microwave Background Probe (WMAP) satellite established many important parameters of the Universe.

The observational data by WMAP opened a new epoch in the investigation of the CMB power spectrum at large multipole numbers l. As we mentioned above, the angular power spectrum of the CMB (in particular the position and amplitudes of the first, second and subsequent Sakharov peaks) provides a unique chance to construct the most likely cosmological model. This model includes information about the Hubble constant, baryonic, dark matter and CMB densities, and dark energy (responsible for the accelerating expansion).

In addition to the anisotropy of the intensity it is possible, though more difficult, to measure the polarization of the relic radiation. Polarization is a secondary effect induced by the scattering of anisotropic radiation on electrons in the cosmic plasma. The importance of the polarization measurements of the relic radiation was pointed out by Rees [27] and this problem has since been discussed in many papers (see for example [28] and [29]). Polarization contains an addition to the anisotropy information about the nature of primordial cosmological perturbations and different types of foregrounds. The polarization field is a combination of two randomly distributed Stokes parameters (Q and U) while the anisotropy is just a scalar.

---

[1] I don't discuss here the large scale dipole anisotropy related to the motion of our Galaxy through the microwave background radiation

In particular, this field is quite sensitive to the presence of tensor perturbations (primordial gravitational waves) and a deviation from zero of the so called pseudo scalar or "magnetic" part of polarization would be an indicator of gravitational waves or vector perturbations. WMAP performed the first measurments of the CMB polarization.

Cosmologists need much more information about the Universe which can be extracted from the CMB. The new mission PLANCK will be the third space CMB mission. PLANCK is designed to extract new important information available in the CMB anisotropy and polarization (see details in [4]). We hope that in a few years PLANCK will open a new era in cosmology.

But let us come back to George Gamow who started all this chain of discoveries. I want to conclude with a quotation from the Smoot and Davidson book [30]: Gamow "could ask questions that were ahead of his time", recalls astronomer Vera Rubin, who studied under him.

Acknowledgements

This paper was supported in part by the Danish Natural Science Research Council through its grant No 21-03-0336.

References

[1] Ya.B. Zeldovich, I.D. Novikov, *The Structure and Evolution of the Universe, Relativistic Astrophysics*, **Vol. 2**, The University of Chicago Press, Chicago, IL (1983).

[2] I.D. Novikov, *Black holes and the Universe*, Cambridge University Press, Cambridge (1990).

[3] I.D. Novikov, *Discovery of CMB, Sakharov oscillatons and polarization of the CMB anisotropy*, in V.J. Martinez, V. Trimble, M.J. Pons-Borderia (eds) *Historical Development of Modern Cosmology, ASP Conference Series*, **Vol. 252**, p.43 (2001).

[4] P. Naselsky, D. Novikov, I. Novikov, *Relic Radiation of the Universe*, Nauka, Moscow (2003).

[5] S.W. Hawking, *Hawking on the Big Bang and Black Holes*, World Scientific Publishers, Singapore (1993).

[6] G. Lemaitre, *Nature*, **May 9**, (1931).

[7] G. Gamow, *Phys.Rev. D*, **70**, 572 (1948).

[8] Ya.B. Zeldovich, *Zh. Eksp. Teoret. Fiz.*, **43**, 1561 (1962).

[9] A.A. Penzias, R.W. Wilson, *Astrophys. J.*, **142**, 419 (1965).

[10] R.H. Dicke, P.J.E. Peebles, P.G. Roll, D.T. Wilkinson, Astrophys. J., **142**, 414 (1965).

[11] R.H. Dicke, *Gravitation and the Universe, Jaune Lectures for 1969*, American Philosophical Soc., Philadelphia (1970).

[12] S. Weinberg, *The first Three Minutes*, Basic Books, New York (1977).

[13] F. Hoyle, *Frontiers of Astronomy*, Harper, New York (1955).

[14] A.A. Penzias, *The origin of the elements. Nobel lecture, 8 December 1978, Rev. Mod. Phys.*, **51**, 430 (1979).

[15] A.G. Doroshkevich, I.D. Novikov, *Dokl. Akad. Nauk. USSR*, **154**, 809 (1964).

[16] A. McKellar, *PASP*, **52**, 187 (1940).

[17] A. McKellar, *Publ. Dominion Astrophys. Observ. Victoria, B.C*, **7**, 251 (1941).

[18] I.S. Shklovsky, *Astron. Circular Acad. Sci. USSR*, **364** (1966).

[19] T. Shmaonov, *Pribori i Tekhnika Experimenta*, **Vol. 1**, 83 (1957).

[20] N.L. Kaidanovsky, Yu.N. Parijsky, *Istoriko-Astronomicheskie Issledovaniya*, Nauka, Moscow, p. 59, (1987).

[21] G. Smoot et al. *Astrophys. J.*, **396**, L1 (1992).

[22] A.D. Sakharov, *Zh. Eksp. Teoret. Fiz.*, **49**, 345 (1965).

[23] P. Naselsky, I. Novikov, *Astrophys. J.*, **417**, 14 (1993).

[24] H. Jørgensen. E. Kotok, P.D. Naselsky, I.D. Novikov, *Astron. Astrophys.*, **294**, 639 (1995).

[25] P. De Bernardis et al. *Nature*, **404**, 955 (2000).

[26] S. Hanany et al. *Astrophys. J.*, **545**, L5 (2000).

[27] M. Rees, *Astrophys. J.*, **153**, L1 (1968).

[28] M.A. Basko, A.G. Polnarev, *Sov.Astron.*, **24**, 3 (1979).

[29] A.G. Polnarev, *Sov.Astron.*, **29**, 607 (1985).

[30] G. Smoot, K. Davidson, *Wrinkles in Time*, William Morrow, New York (1993).

5

Gamow Legacy and the Primordial Abundance of Light Elements

E. Terlevich[1], R. Terlevich[1] and V. Luridiana[2]

[1] *Instituto Nacional de Astrofísica Optica y Electrónica, Puebla, México*
e-mail: eterlevi@inaoep.mx, rjt@inaoep.mx
[2] *Instituto de Astrofísica de Andalucía, Granada, Spain*
e-mail: vale@iaa.es

Abstract

The presently accepted theory of the Universe was pioneered 60 years ago by Gamow, Alpher and Herman. As a consequence of the, later dubbed, Hot Big-Bang, matter was neutrons, and after some decay protons, and a history of successive captures built up the elements.

It wasn't until some 15 years later (with the discovery of Cosmic Microwave Background radiation) that Gamow and colleagues theories were validated and present day Standard Big-bang Nucleosynthesis theory was developed.

We will discuss the importance of state of the art observations and modelling in the quest to determine precise values of the primordial abundance of D and ^{4}He, using observations of astrophysical objects and modern day atomic parameters. In particular, we will present the search for understanding and coping with systematic errors in such determinations.

Keywords: cosmology, Big Bang nucleosynthesis, light elements abundance

1 Introduction

The principles of the Standard Big Bang (SBB) model of the universe were laid down by Gamow, Alpher and Herman in work produced between 1945 and 1953 [1–4]. Their work placed Friedmann's and Lemaitre's conception of the universe on a physically observable basis. They had the advantage of knowing already about the expansion of the universe, discovered by Hubble [5]. Their ideas proposed a way to create all the chemical elements.

SBB states that the Big Bang was hot and dense, and that the conditions were ripe for nucleosynthesis during a short time window of a few minutes. Before that moment the radiation was too hard to permit any nucleide to survive, after that the temperature had gone too low. Therefore the first light nucleides were formed only in a time window of approximately 15 minutes, and the relative abundances that result depend on the production and the destruction rates during the nucleosynthesis period, which are a function of the baryon-to-photon ratio η (see, e.g. [6]).

Hayashi in 1950 [7] suggests that at the high density and temperature reached close to the discontinuity, a freezing of the ratio neutrons to protons would have occurred. Considering as well that there are no stable nuclei with atomic weight $A = 5, 8$, led researchers to the conclusion that there was no nucleosynthesis after ^{4}He. Successes in stellar evolution and nucleosynthesis then led to Big Bang Nucleosynthesis theory (BBNS) to be all but abandoned for 10 years, until Penzias and Wilson [8] discovered Cosmic Microwave Background radiation (CMB) that had already been predicted by Gamow and colleagues. People then turned their attention back to Standard Big Bang Nucleosynthesis (SBBN) and a very comprehensive model was developed by Peebles, Wagoner, Fowler, Hoyle, Schramm, Steigman, etc. (see, e.g. [6]).

A schematic illustration of how the first light nucleides were synthesized in a very short time interval following the Big Bang is shown in Figure 1.

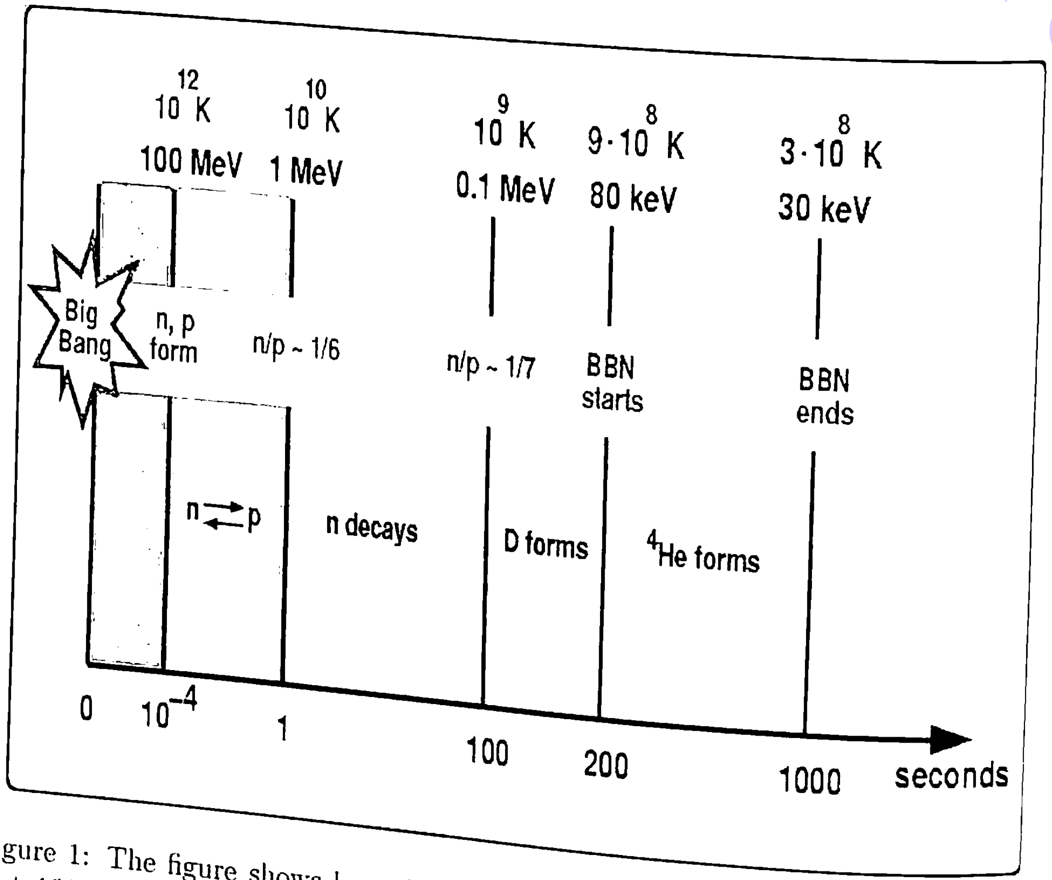


Figure 1: The figure shows how the first light nucleides where synthesized in the first 1000 seconds after the Big Bang. Deuterium, ^{4}He and traces of ^{7}Li and Bo nuclei were formed. After 1000 seconds, and when temperature was 3.10^8 degrees, BBN stopped.

SBBN has been very successful at least in three predictions:

- it very well predicts primordial abundances of Deuterium, Helium 4, Helium 3, Lithium 7 covering 9 orders of magnitude, a commendable achievement.
- it predicts the number of neutrino species.
- it predicts the neutron half life.

There is a very delicate balance between the reaction rates and the expansion of the universe, therefore on the density and temperature, and hence, the primordial abundances of the first elements can be used to determine the baryon-to-photon density η. Figure 2 shows the relationship between the primordial abundances of ^{4}He/H, ^{3}He/H, D/H and ^{7}Li/H and the ratio of the density of baryons-to-photons in units of 10^{-10} (η_{10}). One element alone should be enough to determine η, but consistency with the others would make the theory very robust, while confirming the parameters of atomic physics used in deriving the predictions.

Of all the light elements, ^{4}He is the easiest to measure, but, as Figure 2 shows, it is the least sensitive to η, and therefore, in order to be used as a diagnostic, it has to be measured with very high precision.

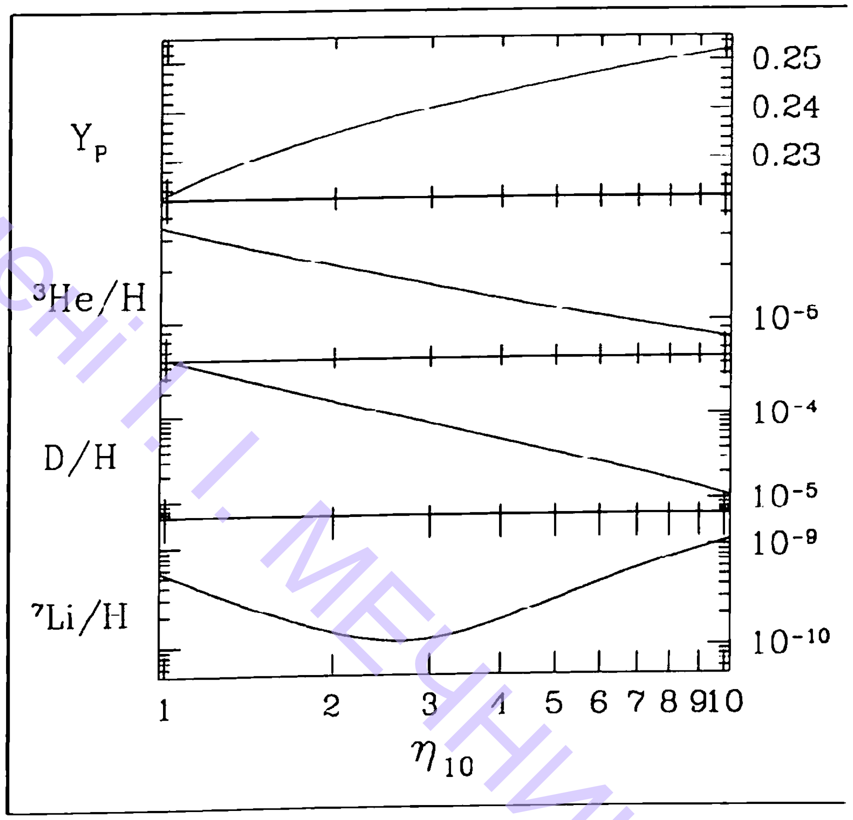


Figure 2: Relationship between the primordial abundances of ^{4}He/H, ^{3}He/H, D/H and ^{7}Li/H and the ratio of the density of baryons-to-photons in units of 10^{-10} (η_{10}). Notice on the right, the 9 orders of magnitude that the abundances cover.

2 Primordial Helium (Y_p)

Abundances of ^{4}He and its primordial value (Y_p), have been inferred for many years by different authors. Analysing their results, it becomes apparent that the different determinations are progressively converging, although a significant scatter remains. In order to see where this scatter comes from, we will review the method used to determine Y_p.

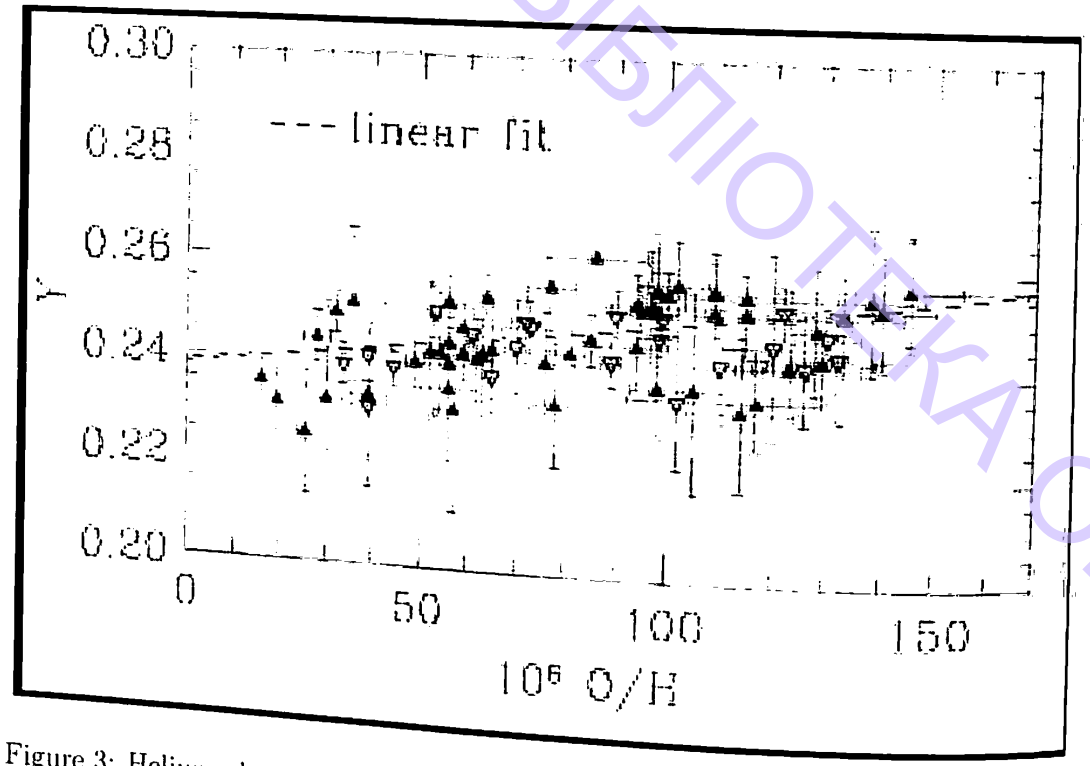


Figure 3: Helium abundance (Y) versus oxygen abundance (O/H) for HII galaxies. The dashed line is a linear fit. Figure taken from Fields & Olive [9].

2.1 The method

The method, essentially devised by Manuel Peimbert and Silvia Torres-Peimbert in 1974 [10] was based on the realization, by Searle and Sargent [11], that the most metal poor known HII galaxy (then and still now, more than 30 years later, but that is another story) – IZw 18 – still has a finite He/H abundance, around 0.24, which therefore had to be primordial. The thought behind [10, 11] is that if the universe was born with zero metallicity ($Z = 0$), the metals we see today were created by nucleosynthesis inside stars, and by the same process, helium abundance grows. Therefore, if we collect a sample of objects, measure their chemical abundances, and extrapolate the relation $Y(Z)$ back to $Z = 0$, we can determine Y_p. An up-to-date representation of this concept can be seen in Figure 3 from a compilation by [9]. It is clear from the figure that a reliable determination of Y_p rests entirely on accurate values of Y and O/H abundances.

2.2 ^{4}He abundance

Many efforts have concentrated over the years on Y_p determinations, and although the value obtained seems to be converging, substantial scatter still remains. Given that the determination of Y_p relies on a good determination of heavier elements abundance, let us review the method commonly used for that purpose.

Chemical abundances in ionized regions of the interstellar medium (HII regions) are determined from their optical spectra (an example of which is represented in Figure 4) applying relations of the type of eq. (1) to the emission-line spectrum as the intensity of a line is linked to the ionic abundance through a function of the electron temperature (T_e).

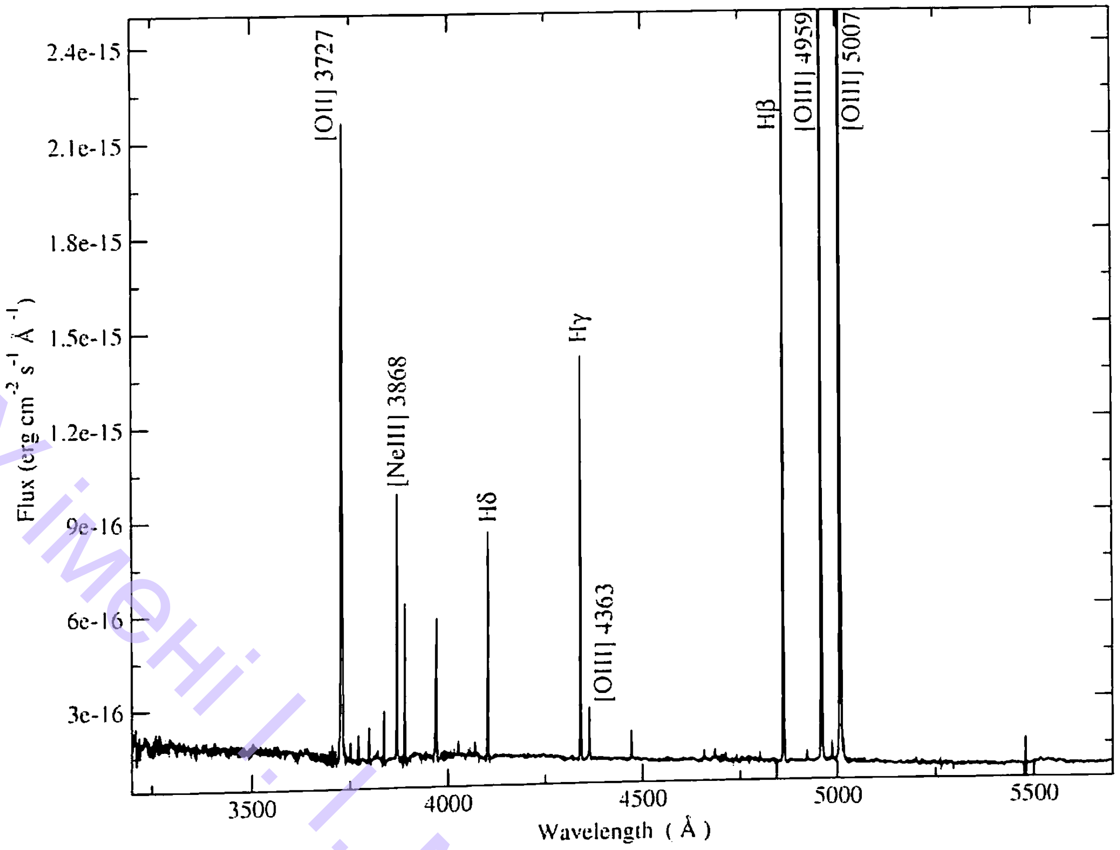


Figure 4: Optical blue spectrum of an HII galaxy. The intense recombination and collisional forbidden lines are indicated. G. Hägele, private communication.

$$\frac{N(X)}{N(H^+)} = f(T_e)\frac{I(\lambda)}{I(H\beta)}. \tag{1}$$

T_e is found by the analysis of suitable temperature sensitive spectral features, e.g. the [OIII]$\lambda\lambda$ 4363/(4959,5007) ratio as it is easy to understand from atomic physics first principles. The ionic abundances are then summed to obtain the chemical abundance of a particular element. For example,

$$\frac{N(He^+)}{N(H^+)} \simeq T^{\alpha_1}\frac{I(\lambda 5876)}{I(H\beta)} \tag{2}$$

$$\frac{N(He^{++})}{N(H^+)} \simeq T^{\alpha_2}\frac{I(\lambda 4686)}{I(H\beta)} \tag{3}$$

$$\frac{N(He)}{N(H)} = \frac{N(He^+)}{N(H^+)} + \frac{N(He^{++})}{N(H^+)} \tag{4}$$

gives the total abundance of He.

2.3 Photoionization models

In cases where the T_e cannot be rigourously determined, we can resort to more empirical methods, in particular, to photoionization models. A schematic view of how they work is presented in Figure 5.

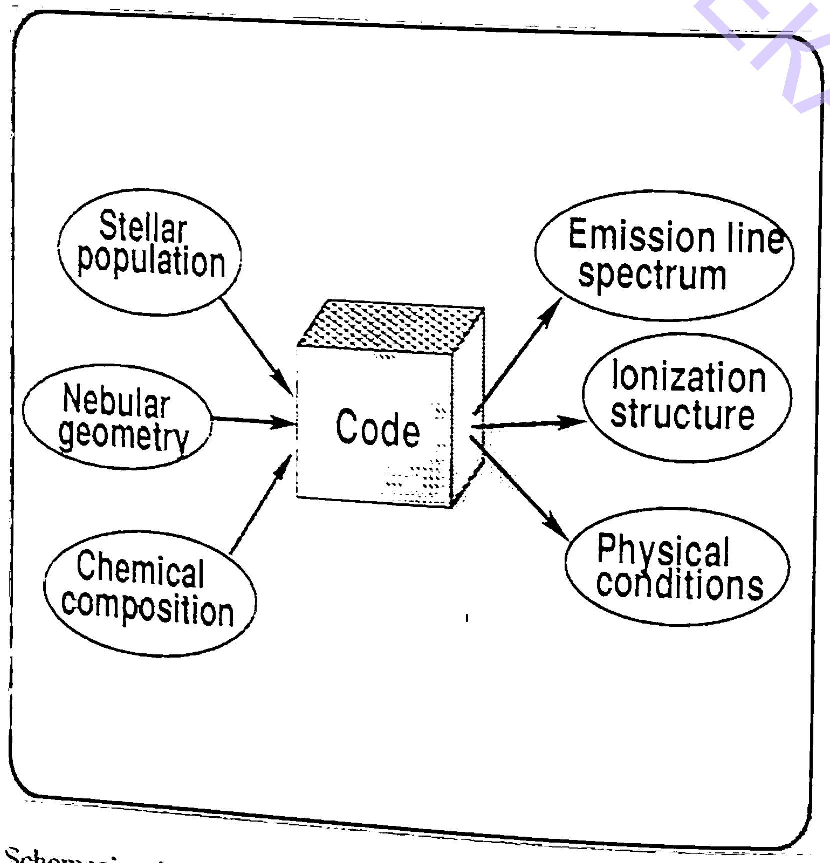


Figure 5: Schematic view of the way in which photoionization models help us to deduce physical conditions of nebular gas.

The inputs to a photoionization code are an ionizing spectrum, a nebular geometry and a chemical composition. The outputs are an emission line spectrum, and the structure of the nebula (ionization structure, electron density and temperature N_e, T_e). If we succeed in reproducing the emission line spectrum of an observed object, we can use the predicted structure to estimate properties such as the ionization structure.

3 Uncertainties in Y_p Determinations

The uncertainties affecting Y_p can be identified in terms of the individual steps that the method includes and can be grouped into three broad categories: the physics, the stellar parameters and the nebular parameters. We can list some of them, but several others exist (e.g., observational errors). In table 1 they are listed in three categories, according to the ingredient of the method where they originate. As it can be seen, some of them are intertwined. We will discuss all of them, except the physics that affect our abundance determinations through atomic parameters. There is nothing we astrophysicists can do about it, but to wait for better laboratory results.

Table 1: Sources of uncertainty

Category	Source of error
physics	atomic parameters
stellar parameters	underlying stellar absorption
nebular parameters	ionization structure temperature structure H I collisional enhancement

3.1 Underlying absorption

The stars that photoionized the nebulae also show their own spectra. They are young hot stars, with H and He lines in absorption, therefore the nebular emission measured is underestimated. One solution is to subtract from the spectra a good estimate (obtained from theoretical stellar models) of the stellar population. We is then left with the pure emission spectrum. Nowadays, state-of-the-art stellar population synthesis models with high spectral resolution allow us to perform a good underlying absorption correction.

3.2 Ionization structure

The simplest assumption to derive the He abundance from the HeI and HeII abundances is that the ionized He and the ionized H spheres coincide. If this is not the case, and the effect is not taken into account, the He abundance is either underestimated or overestimated depending on whether neutral helium is contained in the ionized hydrogen sphere, or H^0 is contained within the He^0 sphere. If HII regions were density-bounded in all directions, the problem would not exist.

There are several ways to deal with the uncertainty associated with the ionization structure:

- applying selection criteria to the object sample,
- building tailored photoionization models,
- using narrow long-slit data.

A first possibility is applying selection criteria to the regions analyzed. Since the ionization structure depends on the shape of the ionizing spectrum, the selection criteria depend on the stellar population. We can therefore exclude those regions ionized by stellar clusters that do not guarantee ionization correction factors ICF(He) ~ 1. This has the disadvantages that precious data points are lost, methods are not so robust as they should be, and selection criteria are VERY model-dependent.

A second possibility is building tailored ionization models and assuming that the ionization structure of the model reflects the one of the real region. Disadvantages of this method are that it is time-consuming, and that the model must really be very constrained to ensure that the predicted ICF is a good assessment.

A third possibility is using long-slit data. In this way, although the problem is not solved, its impact is lessened owed to geometrical reasons (the relative volume of the transition zone is smaller than in the case of a complete sphere, about 2/3).

3.3 Temperature fluctuations

Temperature fluctuations inside the HII regions can bias the abundance determinations. We use eq. (1) to link the abundance of an element with the relative intensities of its lines to a hydrogen recombination line, and the relation depends on a function of the electron temperature. But, does one value of T_e fit all? The answer is no, as each ion is associated with a typical temperature, and adopting one for all introduces a bias in the derived abundance.

The dependence of the line intensity on T_e varies from recombination to collisionally excited lines. Recombination lines (like, e.g. HeI) weight smoothly the T structure according to

$$\frac{I(\lambda 5876)}{I(H\beta)} \simeq T^{-\alpha} \frac{N(He^+)}{N(H^+)} \tag{5}$$

while collisional lines (like, e.g. [OIII]λ 5007Å) are enhanced in T peaks as

$$\frac{I(\lambda 5007)}{I(H\beta)} \simeq T^{-\beta} \times e^{\Delta E/kT} \times \frac{N(O^{++})}{N(H^+)} \,. \tag{6}$$

Therefore, if there are temperature fluctuations in the region and the temperature is determined from the ratio of collisional lines, as is often the case, the inferred temperature will be biased towards the peaks. On the other hand, recombination lines may give an estimate closer to the average values. The problem is that recombination lines are weaker than collisional lines, so they are more difficult to observe.

In any case, this becomes a very hairy problem. The temperature used to find the ionic abundances must be determined with care, otherwise the abundances will be over or underestimated.

3.4 Collisional enhancement of Balmer lines

Another problem is caused by the collisional enhancement of Balmer lines. The expression here

$$\frac{N(He^+)}{N(H^+)} \simeq T^{\alpha} \frac{I(\lambda)}{I(H\beta)} \tag{7}$$

holds under the hypothesis that the intensities derive from recombination only. Helium intensities have an important collisional part, which is routinely computed and subtracted out. Balmer lines have a much smaller collisional part, which is nevertheless important at the present required state of precision. To have an idea of the relative importance of collisions in Balmer lines, let us consider the collisional-to-recombination ratio of Hβ: the ratio depends on the ionic fractions of neutral and ionized hydrogen and on the Boltzmann factor of the transition from the ground state to the upper level of the line.

$$\frac{j(H\beta)_C}{j(H\beta)_R} \propto \frac{N(H^0)}{N(H^+)} e^{-\Delta E/KT} \,. \tag{8}$$

Typical values for the ionic ratio are around 10^{-4}, values for the Boltzmann factor are listed below for the typical range of HII temperatures.

Table 2: Boltzman factor as a function of temperature

T_e	$e^{-\Delta E/KT}$
10000	3.7E-7
12500	7.2E-6
15000	5.2E-5
17500	2.1E-4
20000	6.1E-4

Due to the Boltzmann factor, the collisional contribution to the Balmer lines is particularly important in high-temperature regions. This is a big difficulty because the hottest regions are also the most metal poor regions and are the most important from the point of view of the determination of primordial helium given that the amount of (unseen) He^0 is minimal, and so is the extrapolation to $Z = 0$.

An additional difficulty is that the collisional contribution is higher for Hα than for Hβ, and thus it mimicks (and can be misunderstood for) the reddening due the interstellar attenuation. Therefore, unrecognized collisions have two main effects, one direct and one indirect. The direct one is that the relative hydrogen contribution is overestimated, therefore the He/H ratio is underestimated. The indirect one is that if the Balmer decrement is interpreted in terms of pure extinction, the lines blueward of Hβ are overestimated, and those redward of Hβ are underestimated. The global effect on helium abundance depends on which lines are used, but since the most intense ones (λ 5876 and 6678 Å) are redward of Hβ, the net effect is usually that He is underestimated.

As an example, [12] used tailored photoionization models to study quantitatively the effect of collisional excitation of Balmer lines on the determination of the helium abundance (Y) in individual photoionized regions. The geometry of the model is very complex, simulating a filamentary structure with stellar sources spread everywhere. The revised Y values for the five objects in their sample yield an increase of $+0.0035$ in Y_p, giving $Y_p = 0.2391 \pm 0.0020$.

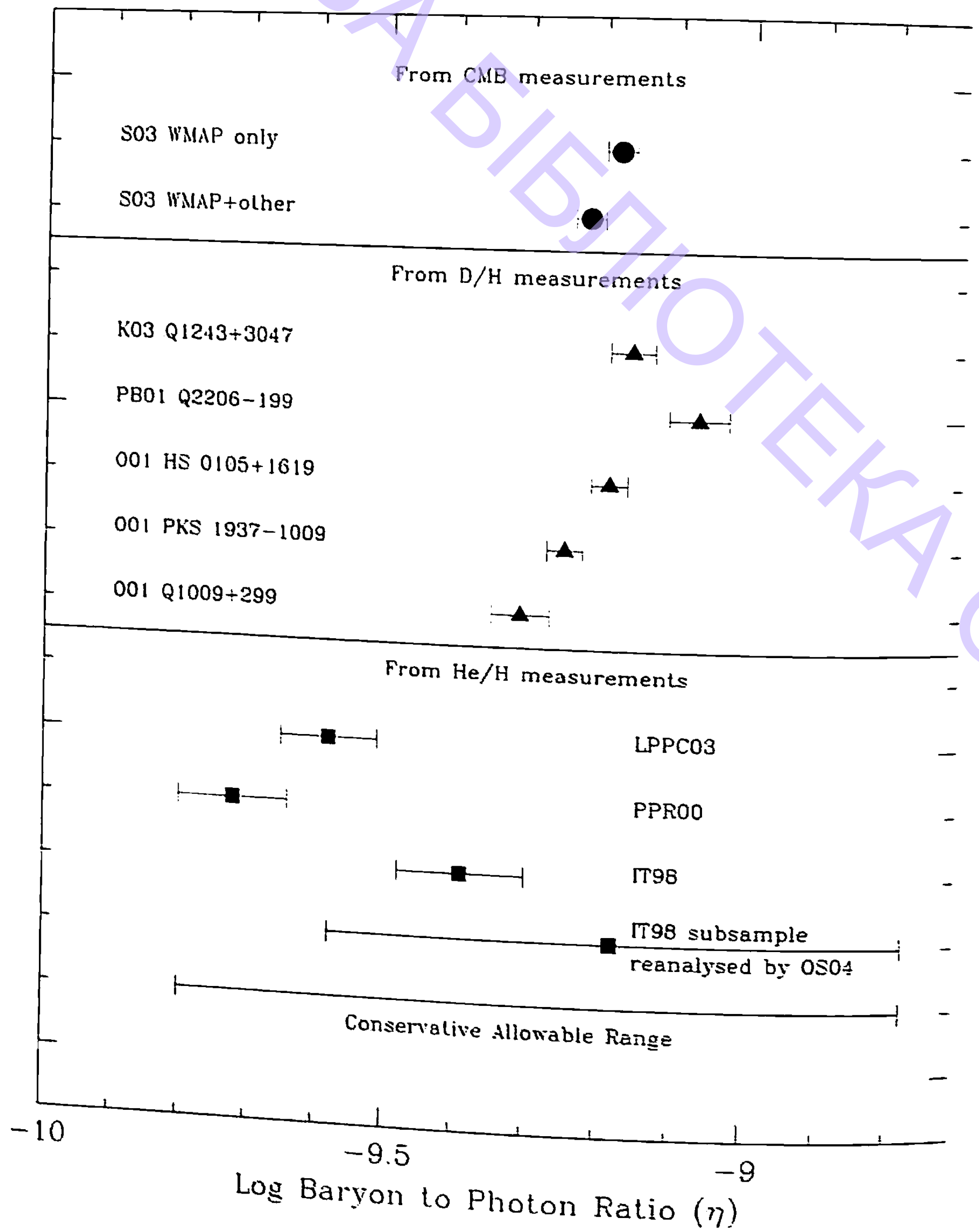


Figure 6: Concordance between WMAP derived light element abundances and observationally determined ones (from [15] and references therein).

Other recent results [13] use a numerous sample of good signal-to-noise spectra in an effort to take into account all the possible systematics. They obtain $Y_p = 0.2421 \pm 0.0021$.

In a novel approach, [14] use the recent determinations of the baryon density by the experiment WMAP (e.g. [15]) and the standard BBNS model to determine quite precise predictions of the primordial light-element abundances. They argue that the discrepancies between the observationally determined Y values and the value favoured by WMAP results are significant if only the statistical errors are considered. They examine in detail some likely sources of systematic uncertainties that may resolve the differences between the determinations and conclude "that the observational determination of the primordial helium abundance is completely limited by systematic errors and that these systematic errors have not been fully accounted for in any published observational determination of the primordial helium abundance". They advocate for a nonparametric approach to the analysis of the observed metal-poor HII region spectra such that the physical conditions and the helium abundance are derived solely from the relative flux ratios of the helium and hydrogen emission lines. The data and the selected objects for the task should be very specific and of inexistent (at present) quality. In practice, the solutions at present are parametric with underestimated error bars. [14] obtain a value of $Y_p = 0.249 \pm 0.009$. They consider as their main result the increase in the size of the uncertainty rather than the shift in the primordial value and go on to claim that most of the spectra analyzed to date do not significantly constrain Y_p. Rather, they argue that a range of allowed values would be $0.232 \leq Y_p \leq 0.258$. Figure 6, taken from their work, shows that this last consideration allows for the eagerly searched concordance between measurements of the baryon-to-photon ratio (η) from WMAP and deuterium and helium abundances. It also shows that individual error bars in D/H do not overlap. As a consequence, averaging the five best absorption system determinations may not be correct.

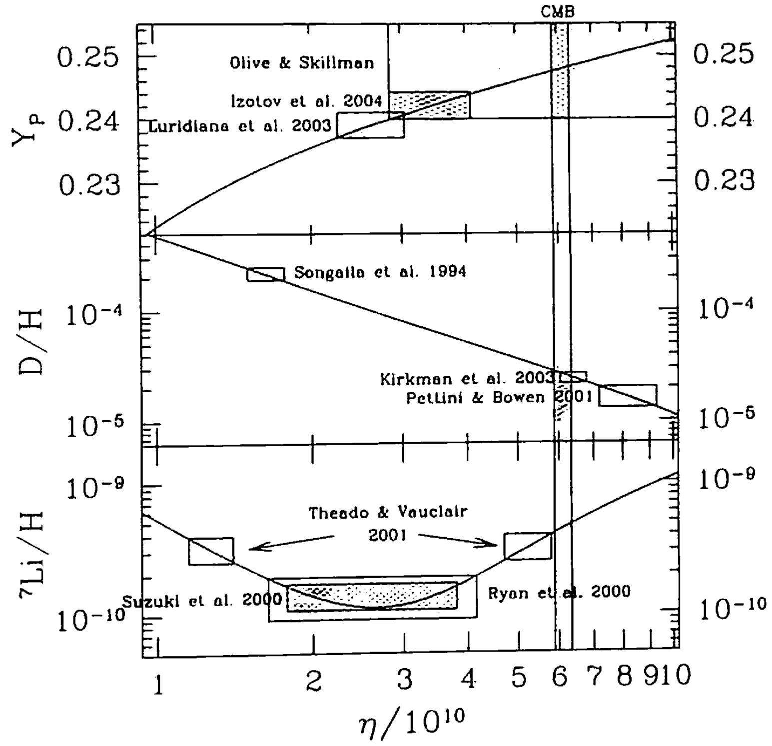


Figure 7: State of the art relation between the cosmological parameter η and the primordial abundances of light elements.

Finally, Figure 7 summarises the state-of-the-art concerning the determination of η vs. the primordial abundances of the light elements.

It is interesting to note that the Songaila 1994 point for D/H has been included just for historical reasons, as it was soon recognised to be wrong.

4 Concluding Remarks

Given the concordance obtained with WMAP D/H determination (e.g. [15]), is the η value secure? The answer is no. Cosmic Microwave Background results constrain combinations of cosmological parameters, therefore independent individual determinations of them are still very important.

Concordance of primordial abundances has been considered a triumph of modern cosmology; regaining it represents an important goal.

In the process, as the community has done in the past, we are sure to learn a lot about chemical evolution of galaxies and about systematic effects in abundance determinations.

Acknowledgments

We are pleased to thank the organizers for their monumental efforts to make this conference such a success. ET is particularly grateful for the opportunity of realizing the long-kept dream of getting to know the land of her ancestors. ET and RT acknowledge the financial support given by the Mexican Research Council (CONACYT) through research grants 40018-A-1 and E32186. VL thanks the hospitality of INAOE during a visit in which this work flourished.

References

[1] G. Gamow, R. Alpher, "The Origin of Chemical Elements", *Physical Review*, April 1 (1948).

[2] G. Gamow, "The Evolution of the Universe," *Nature*, **162**, 680 (1948).

[3] R.A. Alpher, R.C. Herman, "Evolution of the Universe," *Nature*, **162**, 774 (1948).

[4] R.A. Alpher, H.A. Bethe, G. Gamow, "The Origin of Chemical Elements," *Phys. Rev.* **73**, 803 (1948) .

[5] E. Hubble, *MNRAS*, **97**, 506 (1937).

[6] S. Weimberg, *Gravitation and Cosmology*, John Wiley, New York, (1972).

[7] C. Hayashi, *Prog. Theoret. Phys. Japan*, **5**, 224 (1950).

[8] A.A. Penzias, R.W. Wilson, *Astrophys. J.*, **142**, 419 (1965).

[9] B.D. Fields, K.A. Olive, *Astrophys. J.*, **506**, 177 (1998).

[10] M. Peimbert, S. Torres-Peimbert, *Astrophys. J.*, **193**, 327 (1974).

[11] L. Searle, W.L.W. Sargent, *Astrophys. J.*, **173**, 25 (1972).

[12] V. Luridiana, A. Peimbert, M. Peimbert, M. Cerviño, *Astrophys. J.*, **592**, 846 (2003).

[13] Y.I. Izotov, T.X. Thuan, *Astrophys. J.*, **602**, 200 (2004).

[14] K.A. Olive, E.D. Skillman, *Astrophys. J.*, **617**, 290 (2004).

[15] D.N. Spergel et al. *Astrophys. J. Suppl.*, **148**, 175 (2003).

6

Lyman-α Forest – A New Source of Cosmological Information

M. Demiański[1] and A.G. Doroshkevich[2]

[1] *Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland and*
Department of Astronomy, Williams College, Williamstown, MA 01267, USA
e-mail: Marek.Demianski@fuw.edu.pl
[2] *Astro Space Center, Lebedev Physical Institute, Russian Academy of Sciences, 125047 Moscow, Russia*
e-mail: dorr@asc.rssi.ru

Abstract

We present results of an analysis of ~ 4500 Lyα absorption lines found in high resolution spectra of 18 distant quasars. Applying methods developed in our previous papers we notice a possible deviation of the spectrum of primordial density perturbations at small scale from that predicted by the standard ΛCDM model. These data restrict the mass of the dominant component of DM particles to $M_{DM} \approx$ (1.5–5) keV.

Keywords: cosmology, quasars, Lyα absorption lines

1 Introduction

The main goal of observational cosmology is to supply the most accurate values of the basic parameters that are needed to construct a reliable model of the universe. Great technological advances – CCD cameras, fiber optics, adaptive optics, new methods of construction of large mirrors of up to 10 m in diameter and satellite observatories made it possible to observe very distant galaxies and quasars, perform deep galactic surveys and observe the sky with higher precision at wavelengths ranging form microwaves to gamma-rays. The Hubble constant – the fundamental cosmological parameter – has been only quite recently measured with accuracy better than 10% by different methods. The Hubble Space Telescope Key Project led by W. Freedman [19] used Cepheids to measure the Hubble constant, their best value is

$$H_0 = 72 \pm 8 \ \mathrm{km\,s^{-1}Mpc^{-1}} \, .$$

The Hubble constant has been recently independently measured by the WMAP team [5], they find that

$$H_0 = 71^{+4}_{-3} \ \mathrm{km\,s^{-1}Mpc^{-1}} \, .$$

Determinations of velocity curves of spiral galaxies, velocity dispersion of stars in elliptical galaxies and distribution of hot gas in clusters of galaxies as well as gravitational lensing observations consistently show that a large fraction of mass in galaxies and clusters of galaxies is in the form of nonluminous dark matter composed of not yet discovered exotic (nonbaryonic) particles. Observations of distant type Ia supernovae and cosmic microwave background (CMB) radiation lead to a surprising conclusion that the energy-matter content of the universe is now dominated by some kind of dark energy of unknown nature that acts like an effective cosmological constant causing the universe to expand at an accelerated pace. Several other important cosmological parameters have been determined by direct observations or by fitting the observations of the CMB radiation (see the results of the WMAP team [5]).

Confidence in present values of the basic cosmological parameters prompted theoreticians to proclaim that we have already the Standard Cosmological Model [1]. This model incorporates many of the popular ideas of the Big Bang model including however necessary modifications like the very early inflation epoch, when the universe expanded exponentially and primeval density perturbations had been created, and the existence of dark matter and dark energy. Within a 10% accuracy the sum of densities of dark matter and dark energy is equal to the critical density implying that the universe is flat or almost flat. The spectrum of the initial density perturbations generated during inflation as predicted by simple models of inflation is scale invariant which is nicely confirmed by deep galactic surveys and by CMB observations. This important prediction combined with the theory of gravitational instability and the fact that dark matter particles practically do not interact with the baryonic component while before recombination baryons are tightly coupled to the CMB photons provides a generally accepted framework of the theory of structure formation. After recombination of helium and hydrogen the universe entered the so called dark age – at that epoch there were no natural sources of light. Small primordial density perturbations which at the epoch of recombination had amplitudes of order $\delta\varrho/\varrho \sim 10^{-5}$ continuously grew up with time allowing baryons that decoupled from radiation to catch up and condense as well. After some time galaxies and stars were formed and their light illuminated the universe. However not all baryons have been condensed into galaxies and stars and quite a large fraction should be spread out in the intergalactic space. Soon after the dis- covery of quasars when it was realized that they are very distant objects, several people independently noticed that light emitted by a distant quasar is traversing large distances and if along its path the light beam encounters clouds of neutral hydrogen some photons could be absorbed producing absorption Lyα lines. The absence of absorption Lyα lines places strong constraints on the amount of neutral hydrogen in the intergalactic space – this phenomenon is now called the Gunn–Peterson effect [21].

Almost at the same time it has been suggested that the distribution of hydrogen could be clumpy and clouds of neutral hydrogen could produce discrete absorption lines [2]. Such discrete absorption lines have been discovered soon after. It took some time before astronomers realized the nature and potential of the absorption lines appearing in high resolution spectra of distant ($z > 2$) quasars. By now it is generally accepted that the Lyα absorption lines, usually called the Lyα forest, are connected with clouds of intergalactic hydrogen. These clouds form a new class of objects that are more abundant than galaxies and less clustered (for a thorough review see [31]). In the framework of the cold dark matter (CDM) scenario of structure formation Lyα clouds form as a natural outcome of the process of gravitational condensation.

2 Formation and Evolution of Lyman-α Clouds

The main observational characteristics of absorption lines are the redshift, z_{abs}, the column density of neutral hydrogen, N_{HI}, and the Doppler parameter, b, while the theoretical description of structure formation and evolution deals with the mean linear number density of absorbers, $n_{abs}(z)$, temperature, overdensity and entropy of DM and gaseous components, and with the ionization degree of hydrogen. To connect these theoretical and observed parameters, a physical model of formation and evolution of Lyα clouds also called absorbers is required. Many such models were proposed and discussed during the last twenty years (see references in [31], [11] and [12]).

In our considerations we assume that:

1. The DM distribution forms an interconnected structure of sheets (Zeldovich pancakes) and filaments, their main parameters are approximately described by the Zeldovich theory of gravitational instability applied to CDM or WDM initial power spectrum [10, 13]. The majority of DM pancakes are partly relaxed, long-lived, and their properties vary owing to their successive merging and expansion and compression in transverse directions.

2. Gas is trapped in the gravitational potential wells formed by the DM distribution. The gas temperature and the observed Doppler parameter, b, trace the depth of the DM potential wells.

3. For a given temperature, the gas density within the wells is determined by the gas entropy created during the previous evolution. The gas entropy is changing, mainly because of heating by shock waves that develop in the course of merging of pancakes, bulk heating by the UV background and local sources and due to radiative cooling.

4. The gas is ionized by the UV background and for the majority of absorbers ionization equilibrium is assumed.

5. The observed properties of absorbers are changing because of merging, transversal compression and/or expansion and disruption of DM pancakes. The bulk heating and radiative cooling leads to slow changes of the entropy and

density of the trapped gas. Random variations of the intensity and spectrum of the UV background enhance random scatter of the observed properties of absorbers.

6. In this simple model we identify the velocity dispersion of the DM component compressed within pancakes with the temperature of hydrogen and the Doppler parameter b of the absorbers. We consider the possible macroscopic motions within pancakes as subsonic and assume that they cannot essentially distort the measured Doppler parameter.

The formation of DM pancakes as an inevitable first step of evolution of small perturbations was firmly established both by theoretical considerations [35], [32] and by numerical simulations [33]. Theoretical analysis and numerical simulations show the successive transformation of sheet-like elements into filamentary-like elements and, at the same time, the merging of both sheet-like and filamentary elements into richer walls. Such continuous transformation of structure goes on all the time.

In this paper we consider the spatially flat ΛCDM model of the Universe with the Hubble parameter and mean density given by

$$H^2 = H_0^2\Omega_m(1+z)^3[1+\Omega_\Lambda/\Omega_m(1+z)^{-3}]\,, \tag{1}$$

$$\langle\rho_m(z)\rangle = \frac{3H_0^2}{8\pi G}\Omega_m(1+z)^3, \quad H_0 = 100h\,\mathrm{km\,s^{-1}Mpc^{-1}}\,.$$

Here $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ are the dimensionless matter density and the cosmological constant (dark energy), and $h = 0.7$ is the dimensionless Hubble constant.

2.1 Characteristics of the DM component

The fundamental characteristic of DM pancakes is the dimensional, μ, or dimensionless, q, Lagrangian thickness (the dimensionless DM column density):

$$\mu \approx \frac{\langle\rho_m(z)\rangle l_v q}{(1+z)} = \frac{3H_0^2}{8\pi G} l_v\Omega_m(1+z)^2 q\,, \tag{2}$$

$$l_v \approx \frac{6.6}{h\Omega_m}h^{-1}\mathrm{Mpc} = 33.8h^{-1}\mathrm{Mpc}\frac{0.3}{\Omega_m}\frac{0.65}{h}\,,$$

where l_v is the coherent length of the initial velocity field [10, 13]. The Lagrangian thickness of a pancake, $l_v q$, is defined as the unperturbed distance at redshift $z = 0$ between DM particles bounding the pancake. The actual thickness of a pancake is

$$\Delta r = \mu\rho_m^{-1} = l_v q(1+z)^{-1}\delta^{-1}, \quad \delta = \rho_m/\langle\rho_m(z)\rangle\,. \tag{3}$$

Here δ is the mean overdensity of a pancake above the background density.

For Gaussian initial perturbations, the expected probability distribution function for the DM column density is

$$N_q dq \approx \frac{2}{\sqrt{\pi}}e^{-\xi}\,\frac{\mathrm{erf}(\sqrt{\xi})}{\sqrt{\xi}}\left(\frac{q^2}{q^2+q_0^2}\right)^{3/4} d\xi, \quad \xi = \frac{q}{8\tau^2}\,, \tag{4}$$

$$\langle\xi\rangle \approx \frac{1}{2}+\frac{1}{\pi} \approx 0.82, \quad \langle\xi^2\rangle \approx \frac{3}{4}+\frac{2}{\pi}, \quad \langle\xi^3\rangle \approx \frac{3}{16}+\frac{11}{2\pi}\,,$$

[13], where the parameter $q_0 \ll 1$ characterizes the coherent length of the initial density field and the dimensionless "time" $\tau(z)$ describes the evolution of perturbations in the Zeldovich theory. For ΛCDM model (1) and $z \geq 2$ we have

$$\tau(z) \approx \tau_0\left(\frac{1+1.2\Omega_m}{2.2\Omega_m}\right)^{1/3}\frac{1}{1+z} \approx \frac{1.27\tau_0}{1+z}\,, \tag{5}$$

and for the amplitude of perturbations τ_0 we get [17]

$$\tau_0 \approx (0.27 \pm 0.04)\sqrt{\frac{\Omega_m h}{0.2}}, \quad \xi \approx 1.06q(1+z)^2\frac{0.2}{\Omega_m h}\,. \tag{6}$$

2.1.1 Characteristics of gaseous component

The observed column density of neutral hydrogen can be written as an integral over a pancake along the line of sight

$$N_{HI} = \int dx\,\rho_b x_H = 2\langle x_H\rangle\frac{\langle n_b(z)\rangle l_v q}{1+z}\frac{0.5}{\cos\theta}\,. \tag{7}$$

Here $\langle x_H\rangle$ is the mean fraction of neutral hydrogen and $\cos\theta$ takes into account the random orientation of absorbers and the line of sight ($\langle\cos\theta\rangle \approx 0.5$). We assume also that both DM and gaseous components are compressed together and, so, the column density of baryons and the DM component are proportional to each other.

Under the assumption of ionization equilibrium of the gas and neglecting a possible contribution of macroscopic motions to the b-parameter ($T \propto b^2$), for the fraction of neutral hydrogen and its column density we get:

$$\langle x_H\rangle = x_0\kappa_b(z)\delta\beta^{-3/2}(1+z)^3\Theta_x, \quad \beta = b/b_{bg}\,,$$

$$N_{HI} = N_0 q\delta\beta^{-3/2}(1+z)^5, \quad N_0 = 5\cdot 10^{12}cm^{-2}\Theta_H\,, \tag{8}$$

$$\Theta_H = \frac{\kappa_b\Theta_x}{\Gamma_{12}}\frac{0.13}{\Omega_m h^2}\frac{\langle cos\theta\rangle}{cos\theta}\left(\frac{\Omega_b h^2}{0.02}\right)^2\left(\frac{16km/s}{b_{bg}(z)}\right)^{3/2}\,,$$

where $b_{bg} = \sqrt{\frac{2k_B T_{bg}}{m_H}}$, $\Gamma_{12} = \Gamma_\gamma\cdot 10^{12}$, $x_0 \approx \frac{6.7\cdot 10^{-8}}{\Gamma_{12}}\frac{\Omega_b h^2}{0.02}\left(\frac{16km/s}{b_{bg}(z)}\right)^{3/2}$ and Θ_x describes the nonhomogeneous distribution of ionized hydrogen along the line of sight.

2.1.2 Characteristics of absorbers

For the DM column density, q, the average gas entropy, $\Sigma = \ln(F_s)$, and overdensity, δ, we get:

$$q^3 = \frac{N_{HI}}{N_0\delta_0}\frac{\beta^{7/2}}{(1+z)^6}, \quad \delta = \delta_0\frac{q^2}{\beta^2}(1+z)\,, \tag{9}$$

$$\exp(\Sigma) = F_s = \beta^2/\delta_b^{2/3} = \kappa_b^{-2/3}\beta^2/\delta^{2/3}\,.$$

Here we use the standard equation of state $T/T_{bg} = \beta^2(z) = F_s(z)\delta^{2/3}(z)$.

However, comparing the average DM column density of pancakes, $\langle q\rangle$, with expectations (4) we can restrict the possible redshift variations of $\langle\Theta_H\Theta_\delta\rangle$ and estimate the combined uncertainty introduced by unknown factors. As is seen from (9),

$$\xi^3 \approx (1+z)^6\, q^2 = \left(\frac{N_{HI}}{2\cdot 10^{16}\ \mathrm{cm}^{-2}}\right)\left(\frac{b}{16\ \mathrm{km/s}}\right)^{7/2} G_{12}\,,$$

$$G_{12}(z) = \frac{\Gamma_{12}}{\kappa_b\Theta_x\Theta_\Phi\Theta_q^2}\frac{\cos\theta}{0.5}\left(\frac{\Omega_m h^2}{0.13}\frac{\Omega_b h^2}{0.02}\right)^2\,. \tag{10}$$

This relation shows that, for a statistically homogeneous sample of absorbers with $\langle\xi^3\rangle = const$, $\langle\cos\theta\rangle = 0.5$, we can estimate the redshift variations of $\langle G_{12}(z)\rangle$ as follows:

$$\langle G_{12}(z)\rangle \propto \left\langle\frac{\Gamma_{12}}{\kappa_b\Theta_x\Theta_\Phi\Theta_q^2}\right\rangle \propto \langle\xi^3\rangle/\langle N_{HI}b^{7/2}\rangle\,.$$

Precision of these estimates is limited because of strong random scatter of the product $N_{HI}b^{7/2}$.

2.2 Mean number density of absorbers

Following [11] we will characterize the 1D mean number density of absorbers by the dimensionless function

$$n_{abs} = \frac{c}{H_0}\langle l\rangle^{-1} = \frac{H(z)}{H_0}\frac{dN(z)}{dz}\,. \tag{11}$$

Here $dN(z)$ is the mean number of absorbers between z and $z+dz$, $\langle l(z)\rangle$ is the mean free path between absorbers at redshift z and c is the speed of light. When absorbers are identified with the Zeldovich pancakes, this 1D mean number density can be linked with the fundamental characteristics of the cosmological model and moments of the initial power spectrum [13].

As was shown in [13], for Gaussian initial perturbations and for richer DM pancakes with sizes significantly larger than the coherent length of the initial density field, this number density can be approximated by the expression

$$n_{abs} \approx \frac{c}{H_0 l_v}\frac{W_p(\xi_{thr})(1+z)^2}{\langle q(\xi_{thr})\rangle}\,, \tag{12}$$

$$\langle q(\xi_{thr})\rangle = 4\tau^2\left[1+\frac{4\sqrt{\pi\xi_{thr}}\,\mathrm{erf}(\sqrt{\xi_{thr}})+2\exp(-\xi_{thr})}{\pi\exp(\xi_{thr})[1-\mathrm{erf}^2(\sqrt{\xi_{thr}})]}\right]\,,$$

$$W_p(\xi_{thr}) \approx 0.5[1-\mathrm{erf}(\sqrt{\xi_{thr}})]\,,$$

where $\xi_{thr}(z) = q_{thr}/8\tau^2(z)$, $q_{thr}(z)$ is the threshold DM column density of pancakes, $W_p(\xi_{thr})$ and $\langle q(\xi_{thr})\rangle$ are the fraction of matter and the mean DM column density for pancakes with $q \ge q_{thr}$ and the factor $(1+z)^2$ describes the cosmological expansion of the pancakes population.

The formation of pancakes with small q is suppressed due to the small scale cutoff in the initial power spectrum. For pancakes with small q_{thr} the redshift variations of the mean number density is described by the expression

$$n_{abs} \approx \frac{c}{H_0 l_v}\frac{\sqrt{3}(1+z)^2}{16\pi\tau(z)\sqrt{q_0}}\frac{\mathrm{erf}(\sqrt{\xi_{thr}})}{\sqrt{\xi_{thr}}}\exp(-\xi_{thr})\,, \tag{13}$$

where $\xi_{thr}(z) = q_{thr}(z)/8\tau^2$ and the value q_0 characterizes the coherent length of the initial density field [13].

3 The Database

The present analysis is based on the 18 spectra listed in Table 1. The available Lyα lines were arranged into three samples. The richest sample, S^{12}_{14}, includes 4369 absorbers with $10^{15}\ \mathrm{cm}^{-2} \ge N_{HI} \ge 10^{12}\ \mathrm{cm}^{-2}$ from the first 14 high resolution spectra. For comparison, we use the most reliable sample S^{13}_{18} which includes 2643 absorbers from all 18 spectra with $10^{13}\ \mathrm{cm}^{-2} \le N_{HI} \le 10^{15}\ \mathrm{cm}^{-2}$. These lines are more easily identified and they are not so sensitive to outer random influences. The sample W^{12}_{14} contains 2126 weaker lines with $N_{HI} \le 10^{13}\ \mathrm{cm}^{-2}$ from the first 14 QSOs. It is mainly used to characterize possible variations of the UV background. To decrease the scatter of absorbers characteristics we exclude from these samples 45 absorbers with $b \ge 90$ km/s .

4 Statistical Characteristics of Absorbers

In this section, the functions q, ξ, δ and F_s are found for the ΛCDM cosmological model (1) with $b_{bg} = 16$ km/s, $\Omega_b h^2 = 0.02$ and the function $\langle G_{12}(z)\rangle$ with

$$G_0 = 50,\ \Theta_\delta = \langle\Theta_\Phi\Theta_q^2\rangle = 0.5,\ \Theta_H = \left\langle\frac{\kappa_b\Theta_x}{\Gamma_{12}}\right\rangle = 0.1. \tag{14}$$

Some quantitative results of this section depend on the completeness and representativity of the samples and the quite arbitrary choice of Θ_δ and Θ_H. As was noted above, for each absorber these factors change in the course of absorbers evolution and, in fact, the accepted values are averaged over the sample. For the sample S^{12}_{14}, $\sim$ 10–15% of the weaker absorbers could be related to the artificial "noise".

4.1 Redshift variations of the mean observed characteristics

Redshift variations of two observed characteristics of absorbers, $\langle b\rangle$, and $\langle lgN_{HI}\rangle$, are plotted in Figure 1. For both samples S^{12}_{14} and S^{13}_{18}, the mean values $\langle b\rangle$ and $\langle lgN_{HI}\rangle$ surprisingly weakly vary with redshift,

$$\langle b\rangle = (28.6\pm1.6)\ \mathrm{km/s},\quad \langle b\rangle = (31.7\pm1.6)\ \mathrm{km/s}, \tag{15}$$

Table 1: QSO spectra used in our analysis

	z_{em}	z_{min}	z_{max}	Number of HI lines
$0000-260^1$	4.11	3.4	4.1	431
$0055-259^2$	3.66	3.0	3.6	534
$0014+813^3$	3.41	2.7	3.2	262
$0956+122^3$	3.30	2.6	3.1	256
$0302-003^{3,2}$	3.29	2.6	3.1	356
$0636+680^3$	3.17	2.5	3.0	313
$1759+754^4$	3.05	2.4	3.0	307
$1946+766^5$	3.02	2.4	3.0	461
$1347-246^2$	2.63	2.1	2.6	361
$1122-441^2$	2.42	1.9	2.4	353
$2217-282^2$	2.41	1.9	2.3	262
$2233-606^6$	2.24	1.5	2.2	293
$1101-264^2$	2.15	1.6	2.1	277
$0515-441^2$	1.72	1.5	1.7	76
$2126-158^7$	3.26	2.9	3.2	130
$1700+642^8$	2.72	2.1	2.7	85
$1225+317^9$	2.20	1.7	2.2	159
$1331+170^{10}$	2.10	1.7	2.1	69

[1]Lu et al. (1996); [2]unpublished, courtesy of Dr. Kim; [3]Hu et al., (1995); [4]Djorgovski et al. (2001); [5]Kirkman & Tytler (1997); [6]Cristiani & D'Odorico (2000); [7]Giallongo et al. (1993); [8]Rodriguez et al. (1995); [9]Khare et al. (1997); [10]Kulkarni et al. (1996).

$$\langle lgN_{HI}\rangle = (13.1 \pm 0.22), \quad \langle lgN_{HI}\rangle = (13.6 \pm 0.06),$$

respectively. At the same time, the mean Doppler parameter systematically shifts from $\langle b\rangle = (25.6 \pm 2.9)$ km/s for the sample W_{12}^{14} up to $\langle b\rangle = (38.4 \pm 2.7)$ km/s for 660 absorbers with $N_{HI} \geq 10^{14}$ cm^{-2} which indicates a weak correlation of observed properties of absorbers. Detailed discussion of the observed characteristics of absorbers can be found in [24, 25].

The small variations of $\langle b\rangle$ with redshift are accompanied by similar small variations of dispersion $\sigma_b \sim 0.5\langle b\rangle$ which also do not exceed $\sim 8\%$. This fact indicates that the broad distribution function of b is weakly dependent upon the redshift. The same is valid for the depth of potential wells and illustrates a correlation between the DM column density, q, and the overdensity of absorbers, δ described by (9). The nature of such a weak redshift evolution of these statistical characteristics of absorbers remains a mystery.

The function $\langle G_{12}\rangle/G_0$ is plotted in Figure 1. Its weak redshift dependence is a natural consequence of the weak redshift variations of $\langle b\rangle$ and $\langle lgN_{HI}\rangle$.

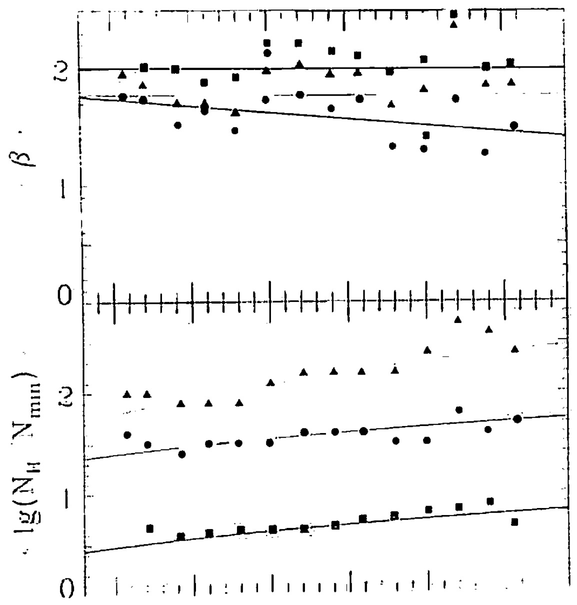


Figure 1: Redshift variations of the Doppler parameter, $\langle b\rangle$ (top panel) and $\langle lg(N_{HI})\rangle$ (bottom panel) for samples S_{18}^{13} (squares) and S_{14}^{12} (triangles). Mean values (15) and fits are plotted by straight lines.

4.2 The mean size of absorbers

Our model of absorbers also allows a rough estimation of their real size along the line of sight, Δr. Due to the strong correlation between q and δ, this size should be obtained by averaging Δr as given by (3). For the model parameters given in (14) and for both samples, S_{14}^{12} and S_{18}^{13}, we have

$$\langle \Delta r\rangle \sim (80 - 100)h^{-1}\text{kpc}\,, \tag{16}$$

and this size increases up to $\sim 150h^{-1}$ kpc for weaker absorbers. For all samples the PDFs of the sizes are similar to the Gaussian function with a dispersion $\sigma_r \sim 0.5\langle \Delta r\rangle$.

The mean size of absorbers in redshift space is associated with the Doppler parameter, b, as follows:

$$\langle \Delta r_b\rangle = 2\langle b\rangle / H(z) \approx 130h^{-1}\text{kpc}\left(\frac{4}{1+z}\right)^{3/2}\sqrt{\frac{0.3}{\Omega_m}}\,.$$

As usual, it is larger than that in the real space (16). This difference agrees with the domination of relaxed gravitationally bound absorbers in the sample.

The expected mean transverse size of absorbers is comparable with their Lagrangian size along the line of sight and was roughly estimated in [13] as

$$\langle \Delta r_t\rangle \approx 5.2\frac{l_v \tau^2}{1+z} \approx \frac{0.6 l_v}{(1+z)^3} \approx 300h^{-1}\text{kpc}\left(\frac{4}{1+z}\right)^3. \tag{17}$$

The difference between estimates $\langle\Delta r\rangle$ and $\langle\Delta r_t\rangle$ agrees with expectations of the Zeldovich theory and indicates that such absorbers could be unstable with respect to the disruption into a system of denser less massive clouds with $\Delta r_t \sim \Delta r$ (see, for example, [16, 34]).

4.3 Characteristics of the initial power spectrum

The amplitude and the shape of large scale initial power spectrum are approximately established by investigations of relic radiation and the structure of the Universe at $z < 1$ detected in large redshift surveys such as the SDSS [15] and 2dF [18]. The shape of small scale initial power spectrum can be tested at high redshifts where it is not so strongly distorted by nonlinear evolution (see, for example, [9]).

Recent discussions on the mass of the dominant fraction of the DM component and the shape of small scale initial power spectrum are focused on the formation of low mass halos in CDM simulations in comparison with observed low mass satellites (see, for example, [6, 7]), and in the simulations of absorbers formation. Thus, [30] estimate the low limit of the mass of DM particles as $M_{DM} \geq 0.75$ keV, while [4] increase this limit up to $M_{DM} \geq 1$–1.25 keV.

Here we can compare these estimates with direct measurements of the redshift distribution of absorbers. In fact, our estimates are related to two spectral moments of the initial power spectrum, m_{-2} and m_0. The first of them depends mainly upon the large scale power spectrum and was independently estimated from the measured characteristics of observed galaxy distribution at small redshifts. Together with cosmological parameters, Ω_m and h, this moment defines the coherent length of initial velocity field, l_v, (2). The second moment, m_0, depends upon the small scale power spectrum and the shape of the transfer function. This means that the estimates of the mass of DM particles are model dependent even though our estimates of q_0 and spectral moments are based on the analysis of observations.

As was shown in [10] and [13], for the CDM and WDM models with the Harrison-Zeldovich asymptotic power spectrum, $p(k) \propto k$, and the transfer function, $T(k)$, given in [3] the coherent length of initial density perturbations, l_ρ, and the parameter q_0 can be written as follows:

$$q_0 = 5\frac{m_{-2}^2}{m_0}, \quad l_\rho = q_0 l_v \approx 0.34h^{-1}\text{Mpc}\left(\frac{q_0}{0.01}\frac{0.2}{\Omega_m h}\right), \tag{18}$$

$$m_{-2} = \int_0^\infty dx\, xT^2(x) \approx 0.023, \quad m_0 = \int_0^\infty dx\, x^3T^2(x)\,,$$

$$x = \frac{k}{k_0}, \quad k_0 = \Omega_m h^2/\text{Mpc}, \quad l_v = \frac{1}{k_0\sqrt{m_{-2}}} = \frac{6.6}{\Omega_m h^2}\text{Mpc}\,,$$

where k is the comoving wave number.

For WDM particles the dimensionless damping scale, R_f, and the damping factor, D_W, are [3]

$$R_f = 0.2\left(\frac{\Omega_m h^2}{M_{DM}}\right)^{4/3},$$

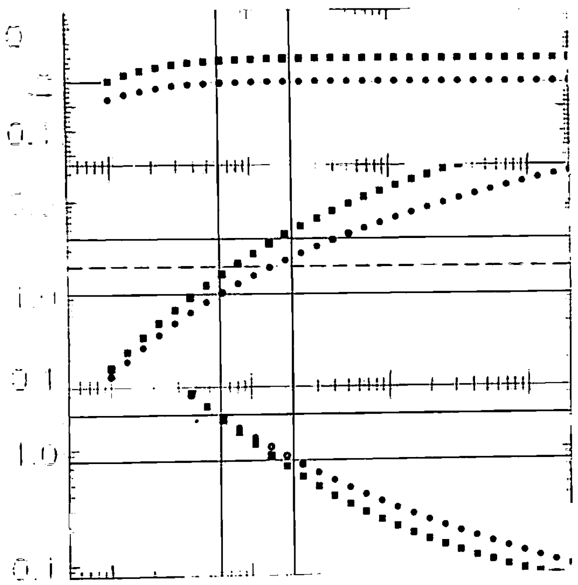


Figure 2: Spectral moments, m_{-2} and m_0, and the parameter q_0 are plotted vs. M_{DM} for the CDM (circles) and WDM (squares) models with the damping factors D_W (19) and with both damping factors, D_W and D_J (20) for $x_J = (2\,\&\,0.5)\cdot 10^{-2}$, (triangles). Two vertical lines show the observational restrictions (21).

$$D_W = \exp[-xR_f - (xR_f)^2]\,, \tag{19}$$

where M_{DM} is measured in keV. The Jeans wave number, k_J, and the damping factor, D_J, can be taken as

$$k_J^{-1} \approx 0.7b_{bg}(1+z)H^{-1}(z), \quad D_J \approx (1 + x_J^2x^2)^{-1}\,, \tag{20}$$

$$x_J = k_0/k_J \approx 2\cdot 10^{-2}\sqrt{\frac{\Omega_m h^2}{0.13}}\sqrt{\frac{4}{1+z}}\,\frac{b_{bg}}{16km/s}\,,$$

(see, for example, [29]).

Variations of the spectral moments and q_0 with the mass of DM particle obtained by integration of the power spectrum with these damping factors are plotted in Figure 2 for the CDM and WDM transfer functions and for $x_J = 2\cdot 10^{-2}$ and $0.5\cdot 10^{-2}$. As is seen from this figure, the low limit $q_0 \geq 0.07\sqrt{x_J}$ caused by the Jeans damping (20), restricts the range of masses of DM particles detectable with this approach to $M_{DM} \leq (3$–$5)$ keV.

Our 1σ estimates

$$q_0 \approx (0.6 - 1.2)\cdot 10^{-2}, \quad m_0 \approx 0.15 - 0.5,$$

$$M_{DM} \approx (1.5 - 5)\text{keV}\,, \tag{21}$$

are close to those of [30] and [4]. They are based on absorbers with the redshifts $z \leq 3.5$ where the scatter of measured n_{abs} is minimal and, so, they weakly depend upon both the Doppler parameter and the velocities of absorbers.

5 Summary and Discussion

Here we investigated a more representative sample of $\sim 4\,500$ absorbers that allows us to improve the physical model of absorbers introduced in [11] and [12] and to obtain a reasonable description of the physical characteristics of absorbers. However, the representativity of this sample is not sufficient for the more detailed study of structure evolution. Further progress can be achieved with a richer sample of observed absorbers.

The main results of our analysis can be summarized as follows:

1. The approach used in this paper allows us to link the observed and other physical characteristics of Lyα absorbers such as the overdensity and entropy of the gaseous component and the column density of the DM component accumulated by absorbers.
2. The basic observed properties of absorbers are quite successfully described by the statistical model of DM confined structure elements (Zeldovich pancakes). Comparison of independent estimates of the DM characteristics of pancakes confirms the self consistency of the physical model for richer absorbers. However, some characteristics of pancakes formation and evolution remain uncertain.
3. In the framework of this approach, all characteristics of absorbers can be expressed through two functions, ξ and v, describing the systematic and random variations of the absorber properties. They are directly expressed through the observed parameters, z, N_{HI} and b.
4. The main stages of structure evolution are illustrated by the separation of three subpopulations of absorbers with high and low entropy and low overdensity.
5. The absorbers with high Doppler parameter, $b \geq 90$ km/s, can be naturally identified with the embryos of wall-like structure elements observed in the spatial distribution of galaxies at small redshifts.
6. The strong suppression of the mean number density of absorbers at $z \leq 1.8$ can be naturally explained by the influence of the observational threshold $N_{HI} \geq 10^{12}$ cm^{-2}.
7. The redshift distribution of weaker absorbers indicates the probable systematic redshift variations of the UV background. They can be caused by the reionization of $HeII$.
8. The PDF of the DM column density and the redshift distribution of absorbers are consistent with Gaussian initial perturbations.
9. We estimate the spectral moment m_0 which in turn is linked to the cutoff of the power spectrum caused by the mass of the dominant fraction of DM particles and the Jeans damping.

Acknowledgements

This work was supported in part by the Polish Ministry of Science, grant 1-P03D-014-26 and by EC network HPRN-CT-2000-00124.

References

[1] N. Bahcall, J.P. Ostriker, S. Perlmutter, P.J. Steinhardt, *Science*, **284**, 1481 (1999).

[2] J.N. Bahcall, E.E. Salpeter, *Astrophys. J.*, **142**, 1677 (1965).

[3] J.M. Bardeen, J.R. Bond, N. Kaiser, A. Szalay, *Astrophys. J.*, **304**, 15 (1986).

[4] R. Barkana, Z. Haiman, J.P. Ostriker, *Astrophys. J.*, **558**, 482 (2001).

[5] C.L. Bennett et al., *Astrophys. J. Supp. Ser.*, **148**, 1 (2003).

[6] P. Bode, J.P. Ostriker, N. Turok, *Astrophys. J.*, **556**, 93 (2001).

[7] P. Colin, V. Avila-Reese, O. Valenzuela, *Astrophys. J.*, **542**, 622 (2001).

[8] S. Cristiani, V. D'Odorico, *Astronom. J.*, **120**, 1648 (2000).

[9] R.A.C. Croft et al., *Astrophys. J.*, **581**, 20 (2002).

[10] M. Demiański, A. Doroshkevich, *MNRAS.*, **306**, 779 (1999).

[11] M. Demiański, A.G. Doroshkevich, V.I. Turchaninov, *MNRAS*, **318**, 1177 (2001).

[12] M. Demiański, A.G. Doroshkevich, V.I. Turchaninov, *MNRAS*, **318**, 1189 (2001).

[13] M. Demiański, A. Doroshkevich, preprint astro-ph/0206282 (2002).

[14] S.G. Djorgovski, S. Castro, D. Stern, A.A. Mahabal, *Astrophys. J.*, **560**, L5 (2001).

[15] S. Dodelson et al., *Astrophys. J.*, **572**, 140 (2002).

[16] A.G. Doroshkevich, *Sov. Astronom.*, **24**, 152 (1980).

[17] A.G. Doroshkevich, D.L. Tucker & S. Allam, preprint astro-ph/0206301 (2002).

[18] G. Efstathiou et al., *MNRAS*, **330**, L29 (2002).

[19] W.L. Freedman et al., *Astrophys. J.*, **553**, 47 (2001).

[20] E. Giallongo, S. Cristiani, A. Fontana, D. Trevese, *Astrophys. J.*, **416**, 137 (1993).

[21] J.E. Gunn, B.A. Peterson, Astrophys. J., **142**, 1633 (1965).

[22] E.M. Hu, K. Tae-Sun, L. Cowie, A. Songaila, *Astronom. J.*, **110**, 1526 (1995).

[23] P. Khare et al., *MNRAS*, **285**, 167 (1997).

[24] T.S. Kim, R.F. Carswell, S. Cristiani, S. D'Odorico, E. Giallongo, *MNRAS*, **335**, 555 (2002).

[25] T.S. Kim, S. Cristiani, S. D'Odorico, *Astron. Astrophys.*, **383**, 747 (2002).

[26] D. Kirkman, D. Tytler, *Astrophys. J.*, **484**, 672 (1997).

[27] V.P. Kulkarni et al., *MNRAS*, **279**, 197 (1996).

[28] L. Lu, W.L.W. Sargent, D.S. Womble, M. Takada-Hidai, *Astrophys. J.*, **472**, 509 (1996).

[29] S. Matarrese, R. Mohayaee, *MNRAS*, **329**, 37 (2002).

[30] V. Narayanan et al., *Astrophys. J.*, **543**, L103 (2000).

[31] M. Rauch, *Ann. Rev. Astron. Astrophys.*, **36**, 267 (1998).

[32] S. Shandarin, Ya.B. Zeldovich, *Rev.Mod.Phys.*, **61**, 185 (1989).

[33] S. Shandarin et al., *Phys.Rev.Let.*, **75**, 7 (1995).

[34] E.T. Vishniac, *Astrophys. J.*, **274**, 152 (1983).

[35] Ya.B. Zeldovich, *Astrophysica*, **5**, 20 (1970).

7

George Gamow and the Genetic Code

M. Demiański

Institute of Theoretical Physics, University of Warsaw,
Warsaw, Poland and
Department of Astronomy, Williams College,
Williamstown, MA 01267, USA
e-mail: Marek.Demianski@fuw.edu.pl

Abstract

A brief review of Gamow's work on the genetic code is presented.

Keywords: DNA, RNA, genetic code

George Gamow was an extraordinary man. After developing in the early 1930s the theory of nuclear barrier penetration – the quantum mechanical effect that made it possible to understand the thermonuclear energy generation processes that power our Sun and other stars – and in the mid 1940s inventing the Big Bang model of the early evolution of the universe; in 1950s he turned his attention to, we would say now, molecular biology.

In April and May 1953, James Watson and Francis Crick published two short papers in Nature [7, 8] in which they suggested a structure of the deoxyribonucleic acid (DNA) molecule and a way this molecule can self-replicate. In their model, the DNA molecule is a helix made of a pair of very long chains of phosphate–sugar rings. In this chain, to each sugar is attached to one of the four possible bases: adenine, thymine, guanine and cytosine. In one chain, adenine pairs only with thymine of the other chain while guanine pairs only with cytosine.

By the time Watson and Crick proposed a structure of the DNA molecule it was generally accepted that DNA carries all the genetic information. In the Watson–Crick model of DNA the four letter alphabet A–C–G–T encodes information on all the proteins needed in the intricate chemical machinery of a cell. In their Nature paper of May 1953 Watson and Crick speculated on the possible mechanism of reading off the structure of proteins from a DNA molecule. They write: "Our structure, as described in [7], is an open one. There is room between the pair of polynucleotide chains ... for a polypeptide chain to wind around the same helical axis. It may be significant that the distance between adjacent phosphorus atoms, 7.1 Å, is close to the repeat of a fully extended polypeptide chain."

Gamow must have read this paper as soon as this issue of Nature became available in the United States. Already on July 8, 1953 he wrote a four page long letter to Watson and Crick asking for clarification on the nature and location of genes [6]. Gamow had seriously considered the idea that sequences of nucleotides in the DNA molecule could directly control the process of arranging sequences of amino acids into proteins. Gamow assumed that all proteins are built out of 20 amino acids. The main problem was to find a prescription on how to associate a sequence of nucleotides with one specific amino acid. Gamow correctly realized that using a four letter alphabet, a three letter long codons (words) will provide $4^3 = 64$ different combinations. To get the necessary 20 codons Gamow assumed that the required correspondence between codons and amino acids can be "established by considering the 'key-and-lock' relation between various amino acids, and the rhomb-shaped holes formed by various nucleotides in the deoxyribonucleic acid chain. Figure 1 shows schematically the structure of the deoxyribonucleic acid molecule as derived by Watson and Crick. We see that each hole is defined by only three of four nucleotides forming it since, indeed, two nucleotides located across the axis of the cylinder are related by 1 ↔ 2 and 3 ↔ 4 binding conditions. It can easily be seen that there are twenty different types of such holes, as shown in Figure 2." [2].

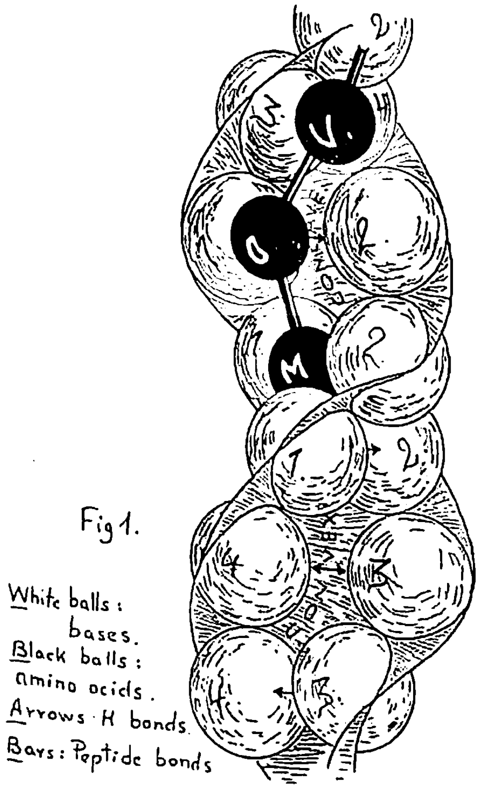


Figure 1: Gamow original drawing of a polypeptide strand being synthesized on a DNA molecule.

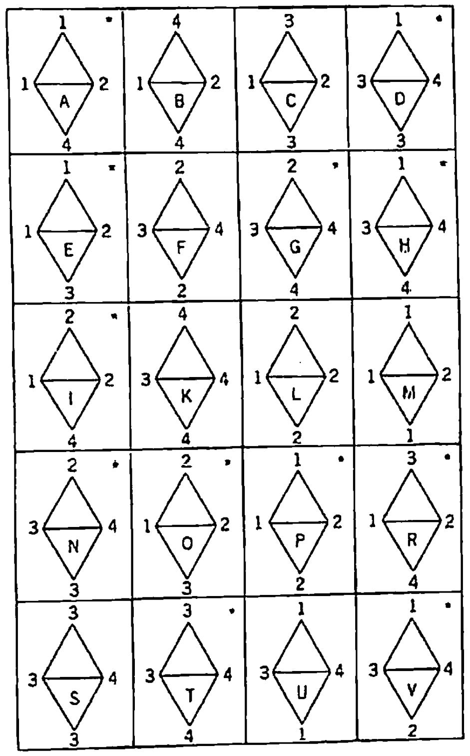


Figure 2: Table 1 with Gamow's assignment of different letters to diamonds. The * denotes diamonds with internal symmetry.

Using his original idea of the genetic code Gamow realized that, because in his model a sequence of three bases determines uniquely the fourth one, the sequence of amino acids in a protein can not be random and there should be correlations. These correlations could provide hints on how different "diamonds" are associated with different amino acids. From a purely combinatorial point of view there are 20! = 2.43×10^{18} possibilities of different assignments. Correlations inherent in Gamow's model could dramatically reduce this number. It was even more interesting to compare the known sequences of amino acids in proteins which became available. Gamow assigned a letter to each of the 20 diamonds and noticed that eight diamonds (A, D, E, G, H, I, N, O) have much higher affinity for other diamonds than the remaining twelve. The proposed assignments also show that the six letters (B, C, F, R, U, V) can occur only singly, the twelve letters (A, D, E, G, H, I, K, L, M, N, O, S) can occur in pairs and the remaining two (P and T) can repeat consecutively an unlimited number of times. At that time the sequence of amino acids was known only for the insulin but already this very limited information showed that Gamow's diamond code is in trouble. The insulin A is a chain of 21 amino acids long containing 11 different amino acids. Gamow was able to decode the first eleven places in insulin A but he encountered a contradiction, the next amino acid in the chain is Ser and according to his rules it can not follow Cys as it does in insulin [3]. It was necessary to give up the idea of diamonds and concentrate on the concept that codons are linear sequences of three letters formed by sequences of nucleotides.

Gamow initially assumed that proteins are synthesized directly on the DNA molecule and this process does not require participation of intermediary molecules.

In this model amino acids would bind to specific holes in the DNA molecule and after forming a polypeptide chain they will spontaneously detach and drift from the nucleus to the cytoplasm. At that time it was known however that proteins in cells are synthesized outside the nucleus in the cell region where there are ribosomes. Gamow was not dogmatic about this part of his hypothesis and immediately realized that the problem of genetic code became more complicated and it was necessary to consider all 64 possible combinations of three letter codons. To pursue discussions and free flow of new ideas Gamow, stimulated by Watson, organized the RNA Tie Club limited to 20 members who were interested in the problems of coding. There were also four honorary members, one for each of the four nucleotide bases. Gamow had a very fortunate hand in picking members of the RNA Tie Club, over the years 12 out of the 20 members were awarded Nobel Prizes. The first meeting of this club took place in August 1954 in the Marine Biology Laboratory at Woods Hole on Cape Cod.

The involvement of the RNA molecule in the process of protein synthesis complicated the picture of information transfer between the DNA molecule and the final polypeptide chain. The RNA molecule is a one strand of nucleotides very similar to either helix of the DNA molecule but with one important difference, it contains uracil instead of thymine. The problem of how the information stored in the DNA molecule is translated into a sequence of amino acids in a polypeptide chain forming a protein remained open. Gamow was mostly interested in the theoretical aspects of the code and he became quite sceptical about the possibility of finding a realistic model of the correspondence between the 64 codons and 20 amino acids, there were just too many different possibilities. To get some hints on the structure of the genetic code Gamow pursued statistical studies of distribution of amino acids in proteins. In this analysis to reduce the number of codons to 20 Gamow and Yčas [4] assumed that triplets of nucleotides that differ only by order, code the same amino acid; in other words, they assumed that the code is degenerate so the same amino acid is coded by different codons but the code is not overlapping so that means that codons do not share bases with adjacent codons. Analysis performed on a limited sample of proteins and RNA molecules with known sequences seemed to support the hypothesis that amino acids in polypeptide chains are selected by triplets of nucleotides taken without regard to order.

The code proposed by Gamow has not explained how it is actually read. The sequence of nucleotides in the DNA molecule and also in the RNA molecule should contain punctuation marks indicating the beginning of a sequence and its end. To solve this problem Crick and his collaborators [1] proposed a so called comma-free triplet code. In this model codons are divided into two groups: genuine codons which are related to amino acids and nonsense codons. The genuine codons when placed next to each other form such a sequence that the overlapping triplets of letters are always nonsense codons. From this assumption it follows that all codons of the type AAA should be nonsense codons since when repeated AAAAAA they could be read ambiguously. For the same reason only one of the three codons ABC, BCA and CAB should be genuine and the other two should be nonsense codons. In this way the 64 possibilities of triplet codons contained 20 genuine ones and 44 nonsense codons. Such a code is non degenerate. Soon a new restriction on the general property of the genetic code was imposed. It was connected with the more complicated information transfer from the DNA molecule to the polypeptide chain. The actual information is contained in only one helical strand of the DNA molecule, but since the DNA molecule has no intrinsic sense of direction and its two helices run in opposite directions if one helix contains genuine codons the other complementary part should contain only nonsense codons. Gamow became more and more frustrated and in November 1955 in a letter to Watson he declared "I think I will go back to Cosmology!". Well, not quite, for the next two years he actively participated in discussions concerning the genetic code and presided over the RNA Tie Club.

The real break through in the study of the genetic code came quite unexpectedly when it became possible to synthesize strands of RNA molecule in cell-free laboratory experiments. Using this possibility Nirenberg and Matthaei [5] synthesized a strand of RNA build out of polyuridylic acid (poly-U) and showed that the codon UUU represents the amino acid phenylalanine. This created a surge of activity and during the next five years the genetic code was cracked. The history of the early attempts to crack the genetic code is nicely described by Yčas [10].

	Second Position								
First Position (5' End)	U		C		A		G		Third Position (3' End)
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA*	Stop	UGA*	Stop	A
	UUG		UCG		UAG*	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 3: The genetic code.

Let us briefly summarize what is now known about the genetic code. First of all the process of information transfer from the DNA molecule to the final polypeptide chain turned out to be more complicated than initially supposed and it involves three different types of the RNA molecule: the messenger RNA (mRNA) which is just a copy of a portion of one helical strand of a DNA molecule (with thymine replaced by uracil), the transfer RNA (tRNA) which plays a role of supplier of appropriate amino acids to a specific place on the mRNA chain and the ribosomal RNA (rRNA) which plays a role of assembly plant where a polypeptide chain (protein) is being synthesized. The genetic code that operates in cells (see Figure 3) is degenerate, 61 codons represent 20 amino acids and three codons play a role of markers indicating the end point of the chain, and it is comma-free. Two codons AUG and GUG play the role of initiation codons but they also code amino acids. The code is almost universal and it operates in all living cells (for more information see [9]). Cracking the genetic code was a great achievement and it opened up the era of genetic manipulations.

There are fundamental questions connected with the genetic code that are still open. It is not known, for example, how the genetic code has developed in nature. Somehow during the period of only about 4 billion years nature produced a stable and optimal code. This problem waits for another Gamow.

Acknowledgements

This work was supported in part by the Polish Ministry of Science grant 1-P03D-014-26 and by EC network HPRN-CT-2000-00124.

References

[1] F.H.C. Crick, J.S. Griffith, L.E. Orgell, *Proc. Nat. Acad. Sci. U.S.*, **43**, 416 (1957).

[2] G. Gamow, *Nature*, **173**, 318 (1954).

[3] G. Gamow, *Kgl. Dansk. Videnskab. Selskab. Biol. Medd.*, **22**, 1 (1954).

[4] G. Gamow, M. Yčas, *Proc. Nat. Acad. Sci. U.S.*, **41**, 1011 (1955).

[5] M.W. Nirenberg, J.H. Matthaei, *Proc. Nat. Acad. Sci. U.S.*, **47**, 1588 (1961).

[6] J.D. Watson, *Genes, Girls and Gamow*, Alfred A. Knopf, New York (2002).

[7] J.D. Watson, F.H.C. Crick, *Nature*, **171**, 737 (1953).

[8] J.D. Watson, F.H.C. Crick, *Nature*, **171**, 967 (1953).

[9] J.D. Watson, N.H. Hopkins, J.W. Roberts, J.A. Steitz, A.M. Weiner, *Molecular Biology of the Gene*, The Benjamin/Cummings Publishing Company, Inc., Menlo Park (1987).

[10] M. Yčas, *The Biological Code*, North-Holland Publishing Company, Amsterdam (1969).

8

George Gamow and Nuclear Physics in Ukraine

Yu. Ranyuk[1], O. Shevchenko[1] and P. Josephson[2]

[1] *National Scientific Center "Kharkiv Institute of Physics and Technology", Kharkiv, Ukraine*
e-mail: ranyuk@kipt.kharkov.ua
[2] *Technology and Society Colby College, Waterville, ME 04901 USA*

Abstract

This paper is devoted to the beginning of nuclear Physics in Ukraine and Gamow's role in the organization of the first nuclear experiments. On the 16th of May 1928 the Government of Ukraine accepted a decision concerning the formation of the Ukrainian Physics-Technological Institute (UPTI) and already in May of 1929 the new institute resonantly announced the first All-Union conference on theoretical physics in Kharkiv. Gamow's presentation at the conference was the only one that touched upon the issues of nuclear physics. He presented his already famous work about α-particle disintegration, for which he received worldwide recognition.

There should be no doubt that it was Gamow's initiative and assertiveness what led to the beginning of atomic nuclei fission research both in Cambridge and in Kharkiv. Ioffe stated this: "Gamow's theories opened the path to the penetration of the nucleus".

Keywords: Gamow, nuclear physics

In 1911, Ernest Rutherford discovered the atomic nucleus and in 1919 first disintegrated it with the aid of natural radium α-particles. Research on atomic nuclei with the aid of natural α-particles had significant drawbacks, among them low radiation intensity and lack of possibility to regulate the energy of the particles. Niels Bohr wrote: "Although the research produced much important data, it became much more readily apparent that insufficient natural α-particles existed for substantive nuclear research and man made accelerated ions was desirable." This was all the more reason for an alternative new methodology as experimentalists presently better understand. An insurmountable obstacle surfaced with α-particle production in a laboratory – the imperative necessity of the availability of a source of high electrical tension of several million volts, as at that time particle acceleration was only achieved with the aid of an electrical field. The magnitude of the electrical tension was defined by the energy of nuclear α-particles, and also by the performance calculations based on classical physics.

Figure 1: Group of Petrograd University students. On the top is G. Gamow, on the right top is D. Ivanenko, right down is L. Landau.

This is how it was before George Gamow, who, as a 24 year old student at Leningrad University, applied the principles of the new wave mechanics to α-decay phenomena and to the calculation of the probability of nuclei penetration by charged particles. According to his calculations the necessary energy levels are not so large and therefore completely accessible for the experiment. This occurred during 1928 in Göttingen, Germany. The young scientist rapidly informed N. Bohr of his sensational result. Bohr immediately referred him to Rutherford so that he could present him his results. This is how T. Allibone, one of Rutherford's pupils, describes the events [1]: "Right at the time in the winter of 1928–29, Russian scientist Gamow arrived in Cambridge, and lectured on a new concept in quantum mechanics – the existence of an energy barrier around the nuclei. I remember when after the lecture

Figure 2: George Gamow in 1932.

we, together with Ernest Thomas Sinton Walton, descended down the stairs of the laboratory and approached Sir John Douglas Cockcroft, who was working in the room. Cockcroft was just setting forth into formulae Gamow's calculations that were able to be obtained at that time for ionic currents – that is 1 mkA protons, accelerated with energy of 0.5 MeV so as to explain the probable penetration of protons through energy barriers of boron nuclei. Even after factoring in possible losses these calculations stood out as completely acceptable, and within some time he provided Rutherford with an explanatory notation with the proposal of reaching an energy current of 0.5 MeV for proton acceleration, similar to mine, but that existed in a constant current (I worked with the variable current). Walton ceased his attempts at electron acceleration and joined Cockcroft's approach. As is known, in three years they attained success". Here is how Gamow himself remembers the events [2]:

Just plainly – Rutherford stated with amazement "I had guessed it will lead you to fill up reams of paper with damned formulas". "In no case", I replied.

Rutherford invited John D. Cockcroft and Ernest T.S. Walton, with whom he previously discussed the experiment's possibilities. "Build me an accelerator of one million electron volts; then without any problems we will smash the lithium atom", Rutherford stated. And so they did. Later, when I was staying in Leningrad and was notified of their success, I sent Cockcroft the following telegram: "Good impact, Johnny; nice protons for golf".

Now let us move to Kharkiv. On 16th of May 1928, the Government of Ukraine accepted a decision concerning the organization of the Ukrainian Physics-Technological Institute (UPTI) and already in May of 1929, the new institute resonantly

Figure 3: Cirill Sinelnikov and George Gamow travel throughout England. 1929.

announced that the first All-Union conference on theoretical physics was to take place in Kharkiv [3]. Gamow's presentation at the conference was the only one that touched on nuclear physics. He presented his already famous work about α-decay, for which he received worldwide recognition.

Figure 4: Soviet arrival in Cambridge. From left to right: George Gamow, Peter Kapitza, Ivan Obreimov, Cirill Sinelnikov. 1929.

At the time of the formation of the UPTI in the Soviet Union, there was no institute, not even a laboratory, that was entirely dedicated to nuclear physics research. This research was not planned for the UPTI which was created as a purely cryogenic laboratory. By the end of the summer of 1931, preparations began at the institute for the atomic nuclei disintegration experiment applying artificially accelerated protons. Similar preparations were already underway in Cambridge.

One of the participants in the future experiments, Cirill Sinelnikov arrived in Kharkiv on the 2nd of June 1930, after spending two years at the Cavendish Laboratory. There is no doubt that he carefully followed the preparatory research steps of Cockcroft and Walton, initiated by his friend Gamow, who within a year also relocated to Cambridge.

Figure 5: Left to right: Cirill Sinelnikov, his Bride Edna Cooper, George Gamow, at the Sinelnikovs flat in Cambridge. 1929–1930.

Why was it that this particular direction of research was taken at the UPTI in August 1931? Unfortunately the institute's pre-war archives were destroyed by fire during the Second World War. Nevertheless, we have a possibility to learn about events of that time through the letters that Sinelnikov's wife Edna Cooper wrote to her sister in England in August 1931. I reproduce below the relevant passages:

2nd August 1931 Darlingest, We are having Dr. Allibone, from Cambridge, for two days on the sixth. And Gamow wired to say he is coming next week from Leningrad. Nobody knows why. Then Cocroft and Webster come from Cambridge later, so we shall be very gay until September.

Figure 6: E. Rutherford and C. Sinelnikov.

11 August 1931 Darlingest, Johnny Gamov has returned to Russia and Kharkov, and, being almost a resident in our flat, he's presented me with this piece of hand-

made paper which I immediately dedicate to you. He's just the same, so bored if we don't do things. The first days were rather jolly and we rushed about to kinos and theatres and for ices, but our pockets don't run to it for ever, and now I don't want to spend any more until Cockroft comes, fortunately Johnny is already bored with Kharkov, and is telegraphing wild messages to Dimus in Leningrad to buy him tickets for a trip down the Volga. As he hopes to go abroad again almost immediately, I don't see how he can afford it, but that's his affair. He came home simply desperately in love and planning a home for his future bride, who has almost been persuaded to come to Russia, but he's been here four or five days and is already definitely cooling off. It's a great pity, because she looks charming, a Swedish dancer, did I tell you all this before, and Johnny might become a more responsible person if married. As far as I can see he's come to Kharkov hoping to get all the advantages of the Institute without working for 'em.

I've just been correcting the thesis he intend to deliver in English at the Conference in Rome, 'orribly learned, didn't understand a word, but if this doesn't sound like Marie Stopes I'am a Dutchman "As the fundamental proton level in radioactive nuclei is very deep, a proton would have to be very highly excited before it could be ejected".

12 August 1931 Darlingest, Our scientific friend, who must be nameless, has just gone off to Leningrad to meet Dimus before going for a trip on the Volga, but incidentally he has decided Kharkov is too dull a place to live in, and in spite of being given a flat, that was prepared for someone else, and being considerably helped with his new foreign passport, by our Institute, he just says he can't stand it and that he's going to Leningrad. I do not know how Kira and Ivan Vasilitch feel about it, but it seems to me a pretty dirty trick, to come and get money and all he can for nothing, and then show his heels.

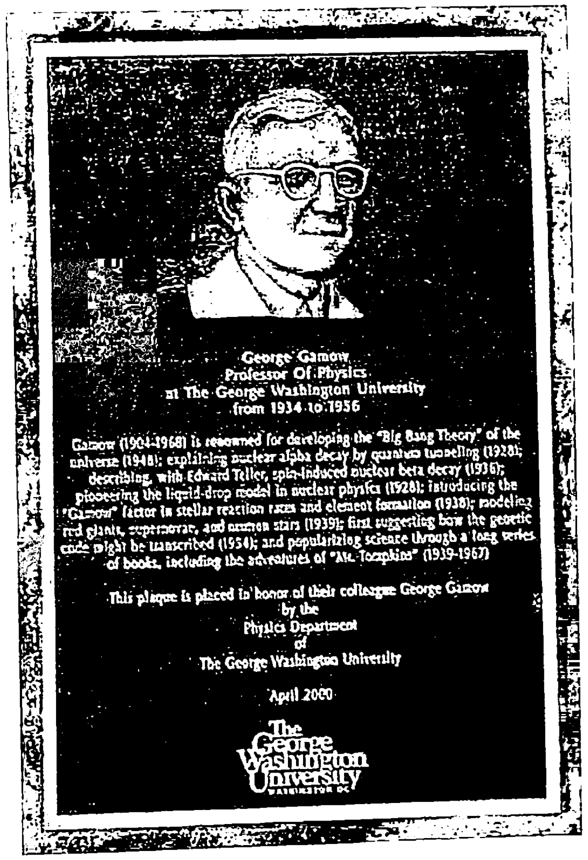
20 August 1931 Dearest Old Thing, I'm busy because Cockroft and Webster are dining here to-morrow. Kira is happy they are here. Though neither of us ever want to live in England, we did enjoy Cambridge days and it is great to renew our Cambridge friendships like this.

26 August 1931 Darling, Our English visitors have been and are going away to-day, Cockroft and Webster and two other physicists and seven or eight other scientific people with two wiwes, not two each! My room could hardly hold them all. But after the first "Reception" they didn't all come at once, only three or four at a time. They have been very kind. They took us to dinner at their hotel, and to Dinamo for ices, in between rushing round "doing" Kharkov. I may say everybody is very impressed with our Institute and says it's wonderfully clean and well organized, better than Cavendish. So you see it can't be so bad.

Bunny

Edna's letters that relate to Gamow cannot be taken at face value. Not because it was her fabrication, but because George was a big comedian, practical joker, and skillful at playful shenanigans.

Edna Alfredovna's letters may leave the impression that Gamow came to Kharkiv only to cheer up the scholarly "UPTI" ladies that for him was indisputably and luckily successful. One speculates whether he really only went there for the sake of money. And there should be no doubt that Gamow's initiative and assertiveness is what led to the beginning of atomic nuclei fission research both in Cambridge and in Kharkiv. Joffe stated this at the time [4]: "Gamow's theories opened the path to the penetration of the nucleus".



George Gamow
Professor Of Physics
at The George Washington University
from 1934 to 1956

Gamow (1904-1968) is renowned for developing the "Big Bang Theory" of the universe (1948); explaining nuclear alpha decay by quantum tunneling (1928); describing, with Edward Teller, spin-induced nuclear beta decay (1936); pioneering the liquid-drop model in nuclear physics (1928); introducing the "Gamow" factor in stellar reaction rates and element formation (1938); modeling red giants, supernovae, and neutron stars (1939); first suggesting how the genetic code might be transcribed (1954); and popularizing science through a long series of books, including the adventures of "Mr. Tompkins" (1939-1967)

This plaque is placed in honor of their colleague George Gamow
by the
Physics Department
of
The George Washington University

April 2000

Figure 7: Plaque erected in honor of Gamow at George Washington University.

We probably should not doubt that he met and spoke with physicists, whom he knew well after Leningrad and Sinelnikov after Cambridge, and the topic of discussion was research relating to atomic nuclei fission. Truthfully a question arises as to the activities of the "Cambridgites who landed" en masse at the institute in August 1931. And why would Gamow arrange his visit, so as to be physically present in Kharkiv, while at the same time not meeting anyone of them. We guess that these very visits by Gamow to Kharkiv were in the capacity of a consultant to UPTI and that could explain why Edna Alfredovna in her letters writes frequently about money.

This contribution was never mentioned at the institute. There is nothing strange about this however, as Gamow was "not returning'. To recall and mention any contacts with him, even mentioning his name, was dangerous. And when the time finally came there was no one left to mention him. Strangely there was not even a mention of his role in any newspaper publication on the occasion of the successful atomic nuclei fission in 1932.

Sinelnikov's daughter Jill recalls how her mother always mentioned Gamow with great warmth. She talked about Gamow's and Sinelnikov's close friendship and how they both traveled by motorcycle across almost the entire English countryside.

George very much facilitated Edna's and Kira's musical evenings in Cambridge - when she played the violin, and he accompanied her - which in the end culminated in matrimony.

In 1951, Cockcroft and Walton received the Nobel Prize "for the transmutation of elements through man made particle acceleration". Sadly, our great compatriot received no recognition in this award although he genuinely and particularly deserved the award.

In 2000, a monument was erected in honor of Gamow on the grounds of George Washington University in Washington, DC. Inscribed on it are his fundamental scientific achievements. A similar monument is expected to be erected in Kharkiv in 2004 on the occasion of the 100th anniversary of the birth of our distinguished compatriot.

References

[1] T. Allibone, in *Rutherford – scholar and professor*, From the publication of academic Peter L. Kapitza. Science. Moscow (1973).

[2] G. Gamow, *My World Line: an Informal Autobiography*, Viking Press, N.Y. (1970).

[3] *Ukrainian physics papers*, **2**, 116 (1930).

[4] A.F. Joffe, *Atomic nuclei today. Lessons, lecture presented 19 February 1934 in the 1 MGU club*, Moscow, p. 31, (1934).

9

Phenomenology of Brane-world Cosmological Models

U. Günther[1,3*] and A. Zhuk[2,3]

[1] *Potsdam University, Potsdam, Germany*
e-mail: u.guenther@fz-rossendorf.de
[2] *Odessa University, Odessa, Ukraine*
e-mail: zhuk@paco.net
[3] *ICTP, Trieste, Italy*

Abstract

We present a brief review of brane-world models – models in which our observable Universe with its standard matter fields is assumed as localized on a domain wall (three-brane) in a higher dimensional surrounding (bulk) spacetime. Models of this type arise naturally in M-theory and have been intensively studied during the last years. We pay particular attention to the covariant projection approach, the Cardassian scenario, to induced gravity models, self-tuning models and the ekpyrotic scenario. A brief discussion is given of their basic properties and their connection with conventional FRW cosmology.

Keywords: Brane-world models, induced gravity, self-tuning, ekpyrotic scenario

1 Introduction

In the brane-world picture, it is assumed that ordinary matter (consisting of the fields and particles of the Standard Model) is trapped to a three-dimensional submanifold (a three-brane) in a higher dimensional (bulk) spacetime. This world-brane is identified with the currently observable Universe. These types of scenarios are inspired by string theory/M-theory. One of the first examples of this type was the Hořava–Witten setup [1, 2] of 11-dimensional M-theory compactified on an S_1/Z_2 orbifold (i.e. a circle folded on itself across a diameter). In this setup, E_8 gauge fields are confined to two $(1+9)$-dimensional planes/branes located at the two fixed

---

*Present address: Research Center Rossendorf, P.O. Box 510119, D-01314 Dresden, Germany points of the S_1/Z_2 orbifold. In the low-energy limit, the E_8 symmetry is broken-down to the $SU(3) \times SU(2) \times U(1)$ group of the Standard Model. Simultaneously, the 9-brane is endowed with the structure of a product manifold $M_3 \times M_6$ consisting of a noncompact real 3D manifold M_3 and a compact real 6D manifold M_6 with the complex structure of a Calabi–Yau threefold $M_6 \sim CY_3$ and a characteristic compactification scale r_{CY_3}. When the distance between the planes is much larger than the compactification scale, $R >> r_{CY_3}$, then the CY_3 degrees of freedom can be integrated out (see e.g. [3, 4]) and we arrive at an effective 5-D heterotic M-theory with two $(1+3)$-dimensional planes. One of these planes describes our brane/Universe and the other one corresponds to a hidden brane (see Figure 1). In the formal limit $R \to \infty$, we obtain a model with a single brane.

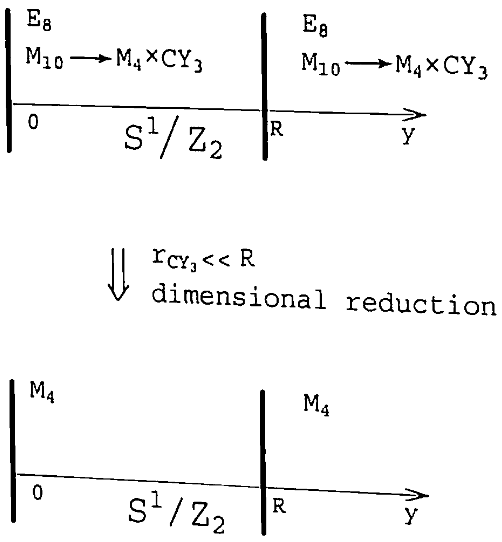


Figure 1: Schematic of the dimensional reduction of a heterotic $S^1/\mathbb{Z}_2$ orbifold setup.

The described scenario is one of the possible scenarios for an embedding of branes into a higher-dimensional bulk manifold. Because up to now only certain aspects of the underlying fundamental higher dimensional theory are understood, the model building process is still going on and an uncovering of new aspects of brane nucleation/embedding mechanisms also can be expected for the future.

The presently known brane-world models can be distinguished by the number of branes involved, their co-dimensions, as well as the topological structure of the branes and the higher dimensional bulk spacetime. One of the basic tests that any brane-world model has to pass, is the test on its compatibility with observable cosmological data and their implications. From the large number of currently present scenarios we select in the present mini-review only a few examples – to illustrate some of the possible deviations from usual FRW cosmology. The aim of the brief discussion is twofold. Firstly, for each of the selected models we briefly review the modifications of the Friedmann-like cosmological equations and demonstrate the regimes where the corresponding physics coincides with conventional FRW cosmology; and, secondly, we briefly indicate the type of observable deviations from it.

2 4D Projected Equations along the Brane

Let us consider a single 3-brane embedded in a five-dimensional bulk spacetime. Denoting the Gaussian normal (fifth) coordinate orthogonal to the brane by y we assume the brane position at some fixed position $y = \text{const}$ which without loss of generality can be set as $y = 0$. Furthermore, we assume that there is bulk matter present in 5D and 4D matter on the brane. Λ_5 denotes the bulk cosmological constant and λ the cosmological constant on the brane (the brane tension). The 5D Einstein equations for this system read

$$^{(5)}G_{AB} = -\Lambda_5 {}^{(5)}g_{AB} + \kappa_5^2 \left[{}^{(5)}T_{AB} + (T_{AB} - \lambda\, g_{AB})\, \delta(y) \right] . \qquad (1)$$

The components along the brane can be easily derived by projecting out the normal components. With the ansatz $n_A dX^A = dy$, where n^A is the unit normal to the surface (brane) at $y = 0$, the metric can be split as

$$^{(5)}g_{AB} = g_{AB} + n_A n_B \implies {}^{(5)}g = g_{\mu\nu}(x^\alpha, y)\, dx^\mu \otimes dx^\nu + dy \otimes dy \,. \qquad (2)$$

The Gauss–Codazzi equations yield then the effective 4D Einstein equations along the brane (for the details we refer to [5] and to the reviews [6, 7])

$$\begin{aligned} G_{\mu\nu} &= -\frac{1}{2}\Lambda_5 g_{\mu\nu} + \frac{2}{3}\kappa_5^2 \mathcal{F}_{\mu\nu}({}^{(5)}T) + K K_{\mu\nu} - K_\mu{}^\alpha K_{\alpha\nu} \\ &\quad + \frac{1}{2}\left[K^{\alpha\beta}K_{\alpha\beta} - K^2\right] g_{\mu\nu} - \mathcal{E}_{\mu\nu} \,, \end{aligned} \qquad (3)$$

where

$$\mathcal{F}_{\mu\nu}({}^{(5)}T) \equiv {}^{(5)}T_{AB}\, g_\mu{}^A g_\nu{}^B + \left[{}^{(5)}T_{AB} n^A n^B - \frac{1}{4}{}^{(5)}T \right] g_{\mu\nu} \qquad (4)$$

is the contribution of the bulk matter,

$$\mathcal{E}_{\mu\nu} \equiv {}^{(5)}C_{ACBD} n^C n^D g_\mu{}^A g_\nu{}^B \qquad (5)$$

is the projection of the bulk Weyl tensor (orthogonal to n^A) and

$$K_{AB} = g_A{}^{C\,(5)}\nabla_C n_B \qquad (6)$$

is the extrinsic curvature of the surface $y = 0$.

The on-brane matter can be taken into account via the Israel junction condition

$$K_{\mu\nu}|_{brane} = -\frac{1}{2}\kappa_5^2 \left[T_{\mu\nu} + \frac{1}{3}(\lambda - T) g_{\mu\nu}) \right] . \qquad (7)$$

Substitution of this condition into (3) yields the induced field equation on the brane:

$$G_{\mu\nu} = -\Lambda g_{\mu\nu} + \kappa_4^2 T_{\mu\nu} + \underline{\kappa_5^4 S_{\mu\nu}(T^2) - \mathcal{E}_{\mu\nu}({}^{(5)}C) + \frac{2}{3}\kappa_5^2 \mathcal{F}_{\mu\nu}({}^{(5)}T)} \,. \qquad (8)$$

The underlined terms are the specific brane-world contributions which are not present in usual general relativity. The first of these terms,

$$S_{\mu\nu}(T^2) \equiv \frac{1}{12} T T_{\mu\nu} - \frac{1}{4} T_{\mu\alpha} T^\alpha{}_\nu + \frac{1}{24} g_{\mu\nu} \left[3 T_{\alpha\beta} T^{\alpha\beta} - T^2 \right] , \qquad (9)$$

describes local bulk effects (local, because it depends only on the matter on the brane). The corresponding quadratic corrections follow from the quadratic extrinsic curvature terms in (3) and the Israel junction condition (7). The second and third underlined terms represent non-local bulk effects, with the traceless contribution $\mathcal{E}_{\mu\nu}$ playing the role of a *dark radiation* on the brane. The 4D effective cosmological and gravitational constants are defined as

$$\Lambda := \frac{1}{2}\left[\Lambda_5 + \kappa_4^2\lambda\right] \quad \text{and} \quad \kappa_4^2 := \frac{1}{6}\lambda\kappa_5^4 . \tag{10}$$

From the latter relation we are led (for this concrete type of models) to the following three conclusions. Firstly, only on positive-tension branes ($\lambda > 0$) gravity has the standard sign $\kappa_4^2 > 0$. A negative tension brane with $\lambda < 0$ would correspond to an anti-gravity world. Secondly, the on-brane matter gives a linear contribution to the generalized gravity equations (8) only in the case of a non-vanishing brane tension. A vanishing brane tension $\lambda = 0$ implies via vanishing effective gravitational constant $\kappa_4^2 = 0$ a vanishing linear energy-momentum contribution, $\kappa_4^2 T_{\mu\nu} = 0$, so that only quadratic terms will remain – which would spoil a transition to conventional cosmology in this limit. Thirdly, (8) reproduces standard Einstein gravity in the simultaneous limit $\kappa_5^2 \to 0$, $\lambda \to \infty$ for κ_4^2 kept fixed and $\mathcal{E}_{\mu\nu}({}^{(5)}C) \to 0$.

Furthermore, it can be shown that for absent bulk matter (${}^{(5)}T_{AB} = 0$) there will be no leakage of usual matter from the brane into the bulk,

$${}^{(4)}\nabla^{\nu} T_{\mu\nu} = 0 , \tag{11}$$

and that in this case the on-brane Bianchi identities set the following constraint on the dark radiation:

$${}^{(4)}\nabla^{\mu}\mathcal{E}_{\mu\nu} = \frac{6\kappa_4^2}{\lambda}\,{}^{(4)}\nabla^{\mu} S_{\mu\nu} . \tag{12}$$

Cosmological implications of the model are most easily studied with the help of a perfect fluid ansatz for the on-brane energy momentum tensor (EMT)

$$T_{\mu\nu} = (\rho + P)u_{\mu}u_{\nu} + P g_{\mu\nu} . \tag{13}$$

Assuming, furthermore, a natural equation of state for the dark radiation, $P^* = (1/3)\rho^*$, the corresponding energy momentum contribution can be approximated as

$$-\frac{1}{\kappa_4^2}\mathcal{E}_{\mu\nu} = \frac{1}{3}\rho^*(4u_{\mu}u_{\nu} + g_{\mu\nu}) . \tag{14}$$

Substitution of (13) and (14) into (8) shows that it is convenient to introduce an effective perfect fluid with total energy density and pressure

$$\rho_{tot} := \rho\left(1 + \frac{\rho}{2\lambda} + \frac{\rho^*}{\rho}\right) , \qquad P_{tot} := P + \frac{\rho}{2\lambda}(2P + \rho) + \frac{\rho^*}{3} \tag{15}$$

and an effective equation of state

$$\omega_{tot} = \frac{P_{tot}}{\rho_{tot}} = \frac{\omega + (1 + 2\omega)\rho/2\lambda + \rho^*/3\rho}{1 + \rho/2\lambda + \rho^*/\rho} , \tag{16}$$

where $\omega = P/\rho$. The second and the third terms in (15) correspond to local and non-local bulk corrections. These corrections will be small (and (16) will tend to the ordinary equation of state $\omega_{tot} \approx \omega$) in the limit of a large brane tension λ and a small energy density ρ^* of the dark radiation: $\rho^* \ll \rho \ll \lambda$. We expect this to happen, e.g., during late-time stages of the evolution of the Universe. The considered brane-world model predicts strong deviations from conventional cosmology in the high-energy limit. Assuming that the phenomenological description via (8), (15), (16) is still applicable to energy densities $\rho \gg \lambda, \rho^*$ at early evolution stages of the Universe, we obtain an equation of state parameters $\omega_{tot} \approx 2\omega + 1$ (i.e. $\omega_{tot} = 5/3$ for $\omega = 1/3$ and $\omega_{tot} = 1$ for $\omega = 0$) which would imply inflation regimes quite different from those in usual 4D FRW models. We will demonstrate this fact explicitly below.

First, endowing the brane with a FRW metric, we express the conservation equations (11) and (12) in terms of the energy densities:

$$\dot{\rho} + 3H(\rho + P) = 0 \quad \Longrightarrow \quad \rho \sim \frac{1}{a^{3(\omega+1)}} \tag{17}$$

and

$$\dot{\rho}^* + 4H\rho^* = 0 \quad \Longrightarrow \quad \rho^* \sim \frac{1}{a^4} , \tag{18}$$

where, as usual, the Hubble constant is given as $H = \dot{a}/a$, and overdots denote derivatives with respect to the synchronous/cosmic time on the brane.

Subsequently, we restrict our attention to the simplest case of a model without bulk matter ${}^{(5)}T_{AB} = 0$ and dark radiation $\mathcal{E}_{\mu\nu} = 0 \Rightarrow \rho^* = 0$ (a model with vanishing projection of the bulk Weyl tensor onto the brane). Then the generalized Friedmann equation is of the type

$$H^2 = \frac{1}{3}\kappa_4^2\rho\left(1 + \underline{\frac{\rho}{2\lambda}}\right) - \frac{1}{6}{}^{(3)}R + \frac{1}{3}\Lambda . \tag{19}$$

As above, we underlined the specific local bulk-induced contribution which is responsible for the deviation from the standard Friedmann equation[1]. Again, we observe that this deviation becomes dominant in the high-energy limit $\rho \gg \lambda$, whereas conventional cosmology is reproduced at late times when $\rho \ll \lambda$.

In the simple case of a flat brane ${}^{(3)}R = 0$ with vanishing effective cosmological constant $\Lambda = 0$ (compensation of brane tension λ and on-brane effects of the bulk cosmological constant Λ_5 according to (10)) the solution of (19) is easily found explicitly (see [8]) as:

$$a(t) = (t - t_0)^{1/[3(\omega+1)]}(t + t_0)^{1/[3(\omega+1)]} , \qquad t_0 := \frac{1}{3(\omega + 1)}\left(\frac{\kappa_5}{\kappa_4}\right)^2 . \tag{20}$$

At early times $t \to t_0$ the scale factor behaves as $a \sim (t - t_0)^{1/[3(\omega+1)]}$, whereas for late times $t \gg t_0$ the conventional FRW behavior is restored: $a \sim t^{2/[3(\omega+1)]}$. Here, the parameters of the model should be chosen in such a way that the transition to conventional cosmology occurs before BBN.

---

[1] Here, the standard Friedmann equation follows from general relativity with Λ-term.

We will briefly analyze next the influence of the additional contributions of local bulk effects in (8) and (19) on inflation. We demonstrate the basic effect with the help of the same simple flat-brane model as above ($^{(3)}R = \Lambda = 0$) – assuming that the energy density and pressure in the state equation are those of a homogeneous minimally coupled scalar field which lives on the brane: $\rho = (1/2)\dot\phi^2 + V(\phi)$, $P = (1/2)\dot\phi^2 - V(\phi)$. The conservation equation (11) $^{(4)}\nabla^\nu T_{\mu\nu} = 0 \Rightarrow \dot\rho + 3H(\rho + P) = 0$ is then equivalent to the field equation $\ddot\phi + 3H\dot\phi + V' = 0$. For the Friedmann equation we get

$$\begin{aligned} H^2 &= \left[\frac{8\pi}{3M_{Pl}^2}\right]\frac{\rho^2}{2\lambda} \quad\Rightarrow\quad H(1-\frac{1}{3}\epsilon) = \sqrt{\frac{4\pi}{3M_{Pl}^2\lambda}}V(\phi) \equiv AV(\phi)\,, \qquad (21)\\ \epsilon &\equiv \frac{1}{6A}\frac{1}{H}\left(\frac{H'}{H}\right)^2 \approx \frac{2\lambda}{V}\frac{M_{Pl}^2}{16\pi}\left(\frac{V'}{V}\right)^2 \end{aligned}$$

in the high energy limit $\rho \gg \lambda$ and

$$\begin{aligned} H^2 &= \left[\frac{8\pi}{3M_{Pl}^2}\right]\rho \quad\Rightarrow\quad H^2(1-\frac{1}{3}\epsilon_c) = \frac{8\pi}{3M_{Pl}^2}V(\phi)\,, \qquad (22)\\ \epsilon_c &\equiv \frac{M_{Pl}^2}{4\pi}\left(\frac{H'}{H}\right)^2 \approx \frac{M_{Pl}^2}{16\pi}\left(\frac{V'}{V}\right)^2 \end{aligned}$$

in the limit of a conventional FRW setup (i.e. in the formal limit $\lambda \to \infty$). In these equations, the gravitational constant has been expressed in terms of the Planck mass, $\kappa_4^2 = 8\pi/M_{Pl}^2 = 8\pi G_N$, with G_N denoting the Newton constant. The important point is the relation between the slow-roll parameters ϵ and ϵ_c corresponding to these two limiting cases

$$\epsilon \approx \frac{2\lambda}{V(\phi)}\epsilon_c\,. \qquad (23)$$

During inflation the high-energy limit with $\rho \approx V(\phi) \gg \lambda$ holds so that $\epsilon \ll \epsilon_c$ and the local bulk corrections lead to a strong relaxation of the slow-roll conditions [9]. Additionally we note that the deviation of the cosmological equation (21) from the conventional one (22) will necessarily lead to a measurable imprint in the CMB anisotropies (corresponding details are discussed, for example, in [7]).

3 Induced Gravity Brane-world Models

In the previous section, the 4D on-brane gravitational equations were obtained via the projection of the 5D curvature along the brane. Another approach is based on the assumption that 4D effective gravity for matter fields confined to the brane can be induced on the brane by interactions with 5D gravity in the bulk [10–17]. This results in an effective 4D scalar curvature term $^{(4)}R$ (as well as higher order curvature corrections) in the effective on-brane action functional. For models with vanishing bulk cosmological constant ($\Lambda_5 = 0$) and brane tension ($\lambda = 0$) it was shown by Dvali, Gabadadze and Porrati (DGP) [15] how quantum interactions can induce effective 4D on-brane gravity. In their setup, the values of the 5D and 4D gravitational constants do not depend on each other, rather they define a characteristic length scale at which the gravitational attraction experienced by on-brane matter changes from 4D Newton's law to that in 5D (see below). Cosmological implications of induced gravity models were considered in [12, 14, 16, 18–21]. Subsequently, we describe specific issues of induced gravity cosmology along the lines of [20, 21].

Let us consider a model with the following action:

$$\begin{aligned} S = &\frac{1}{2\kappa_5^2}\left[\int_{M_5} d^5x\sqrt{|^{(5)}g|}\left(^{(5)}R - 2\Lambda_5\right) - 2\int_{M_4} d^4x\sqrt{|^{(4)}g|}K\right]\\ &+ \int_{M_5} d^5x\sqrt{|^{(5)}g|}L_5\\ &+ \frac{1}{2\kappa_4^2}\int_{M_4} d^4x\sqrt{|^{(4)}g|}\,^{(4)}R + \int_{M_4} d^4x\sqrt{|^{(4)}g|}\left(L_4 - \lambda\right)\,, \qquad (24) \end{aligned}$$

where M_4 denotes the 4D brane[2], and n_a is the vector field of the inner normal to the brane. The gravitational constants $\kappa_{5,4}^2$ are given in energy units as: $\kappa_5^2 = 8\pi/M_*^3$ and $\kappa_4^2 = 8\pi/M_{Pl}^2$. Variation of the action (24) with respect to the 5D bulk metric and the 4D induced metric yields:

$$^{(5)}G_{AB} + \Lambda_5\,^{(5)}g_{AB} = \frac{8\pi}{M_*^3}\,^{(5)}T_{AB} \qquad (25)$$

and

$$^{(4)}G_{AB} = \frac{8\pi}{M_{Pl}^2}\left(^{(4)}T_{AB} - \lambda g_{AB}\right) + \underline{\frac{M_*^3}{M_{Pl}^2}S_{AB}}\,, \qquad (26)$$

where g_{AB} denotes the induced metric (2) on the brane and

$$S_{AB} := K_{AB} - g_{AB}K\,. \qquad (27)$$

Equation (26) describes gravity on the brane and the underlined term defines the bulk correction which is new compared to general relativity. Let us roughly estimate the length scale at which this term becomes important. Clearly, it can be dropped when $(M_*^3 S_{AB})/(M_{Pl}^2\,^{(4)}G_{AB}) \ll 1$. Introducing the characteristic scales $r_{1,2}$ for $^{(4)}G_{AB}, S_{AB}$, i.e. $^{(4)}G_{AB} \sim r_1^{-2}$, $S_{AB} \sim r_2^{-1}$, and assuming as usual for cosmological applications $r_1 \sim r_2$, we find for this ratio $(M_*^3 r_1^2)/(M_{Pl}^2 r_2) << 1 \Longrightarrow r << M_{Pl}^2/M_*^3$. This means that the correction plays an essential role at length scales $r \gtrsim l := 2M_{Pl}^2/M_*^3$. Explicitly, we have the following estimates:

$$l \sim \left\{\begin{array}{l} 10^{15}\ \mathrm{cm} \sim 10^{-3}\ \mathrm{pc} \\ 10^{24}\ \mathrm{cm} \sim 1\ \mathrm{Mpc} \\ 10^{60}\ \mathrm{cm} \gg H_0^{-1} \end{array}\right. \quad \mathrm{for} \quad M_* \sim \begin{array}{c} 1\ \mathrm{TeV}\,, \\ 1\ \mathrm{GeV}\,, \\ 10^{-3}\ \mathrm{eV}\,. \end{array} \qquad (28)$$

For energies $M_* \sim 10^{-4/3}$ GeV the characteristic length scale is of the order of the present horizon scale $l \sim H_0^{-1}$, for higher energies it is smaller, and for the

---

[2] In this section, the brane is assumed as a boundary of a M_5 manifold. The theory can be easily extended to the case where M_4 is embedded into M_5.

dark energy scale $M_* \sim 10^{-3}$ eV it is much larger than H_0^{-1}. Hence, in the latter case and, in general, for $M_* < 10^{-4/3}$ GeV the corrections will be unobservable at present time. The crucial role of the distance l can be demonstrated on the behavior of the Newtonian potential. In [15, 19], it was approximated as

$$V(r) \approx -\frac{G_N}{r}\frac{2}{\pi}\left\{\frac{\pi}{2} + \left[-1 + \gamma + \ln\frac{4r}{l}\right]\frac{4r}{l} + \mathcal{O}(r^2)\right\}, \quad r \ll l\,, \tag{29}$$

(effective 4D Newtonian scaling) and as

$$V(r) \approx -\frac{G_N}{r}\frac{2}{\pi}\left\{\frac{l}{4r} + \mathcal{O}(\frac{1}{r^2})\right\}, \quad r \gg l\,, \tag{30}$$

i.e. at large distances $r \gg l$ the gravitational potential scales as $1/r^2$ in accordance with the laws of a 5D theory.

Let us now illustrate the cosmological implication of this scenario. For this purpose we consider the simplest model – a setup without bulk matter: ${}^{(5)}T_{AB} = 0$. The Codazzi relation guarantees then the conservation of matter on the brane:

$${}^{(4)}\nabla^{\mu}T_{\mu\nu} = 0\,. \tag{31}$$

Here, ${}^{(4)}\nabla$ is the covariant derivative associated with the induced metric ${}^{(4)}g_{\mu\nu} \equiv g_{\mu\nu}$ (in normal Gaussian coordinates (2)) and we use the notations ${}^{(4)}T_{\mu\nu} \equiv T_{\mu\nu}$, ${}^{(4)}G_{\mu\nu} \equiv G_{\mu\nu}$. The complete trace of the Gauss relation yields

$${}^{(4)}R - 2\Lambda_5 + S_{AB}S^{AB} - \frac{1}{3}S^2 = 0\,, \tag{32}$$

and plugging S_{AB} from (26) into this relation gives the following on-brane equations system:

$$\begin{aligned}&\left[M_{Pl}^2 G_{\mu\nu} + 8\pi\left(\lambda g_{\mu\nu} - T_{\mu\nu}\right)\right]\left[M_{Pl}^2 G^{\mu\nu} + 8\pi\left(\lambda g^{\mu\nu} - T^{\mu\nu}\right)\right]\\ &= \frac{1}{3}\left[M_{Pl}^2\,{}^{(4)}R - 8\pi\left(4\lambda - T\right)\right]^2 - M_*^6\left({}^{(4)}R - 2\Lambda_5\right)\,.\end{aligned} \tag{33}$$

Although these equations seem to form a closed system (they contain only on-brane quantities), the higher order derivatives allow for additional degrees of freedom which should be fixed via integration. One such term which reappears in the corresponding cosmological equations is the dark radiation. With the help of a perfect fluid ansatz (similar to that in the previous section), it was shown in [20] that integration of (33) yields:

$$\frac{1}{\kappa_4^4}\left[H^2 + \frac{k}{a^2} - \frac{1}{3}\kappa_4^2\left(\rho + \lambda\right)\right]^2 = \frac{1}{\kappa_5^4}\left(H^2 + \frac{k}{a^2} - \frac{\Lambda_5}{6} - \frac{\mathcal{C}}{a^4}\right), \tag{34}$$

where $k = \pm 1, 0$ and $\mathcal{C}$ is a constant of integration[3]. Obviously, the term $\mathcal{C}/a^4$ can be identified as dark radiation. Assuming as before that the 5D and 4D gravitational

constants $\kappa_5^2 = 8\pi/M_*^3$ and $\kappa_4^2 = 8\pi/M_{Pl}^2$ can be fixed independently, the solution of (34) can be resolved for the Hubble parameter,

$$H^2 + \frac{k}{a^2} = \frac{1}{6}\Lambda_5 + \frac{\mathcal{C}}{a^4} + \frac{1}{l^2}\left[\sqrt{1 + l^2\left(\frac{1}{3}\kappa_4^2\left(\rho + \lambda\right) - \frac{1}{6}\Lambda_5 - \frac{\mathcal{C}}{a^4}\right)} \pm 1\right]^2, \tag{35}$$

where $l = 2M_{Pl}^2/M_*^3 = 2\kappa_5^2/\kappa_4^2$.

Again, the dynamics of on-brane matter depends on the characteristic length scale l. At short distances $r \ll l$, what corresponds to high energy densities, the bulk corrections play no important role. Equation (35) shows that in the (UV) limit $\rho \to \infty$ we restore the equation of conventional cosmology

$$H^2 + \frac{k}{a^2} \approx \frac{1}{3}\kappa_4^2\rho\,. \tag{36}$$

In the particular case $\lambda = \Lambda_5 = \mathcal{C} = 0$ this happens already at[4] $\rho \gtrsim \rho_c \equiv 1/(\kappa_4 l)^2 \sim M_*^6/M_{Pl}^2$.

In the opposite case of the IR limit with low energy densities (i.e. $\rho \to 0$) the bulk corrections become dominant. Depending on the sign in (35) two regimes can be distinguished. (Again, we perform the corresponding sketchy estimates for the simplest model with $\lambda = \Lambda_5 = \mathcal{C} = 0$ where the low energy density regime corresponds to $\rho < \rho_c$.) In the first regime with "minus" sign in (35) we find

$$H^2 + \frac{k}{a^2} \approx \frac{1}{9}\kappa_5^4\rho^2\,, \tag{37}$$

which formally coincides with the high energy limit $\rho \gg \lambda$ in (19) of the previous section. The different numerical prefactors 1/9 in (37) and 1/36 in (19) are the result of different setups: in the previous section the brane was embedded into the bulk, whereas here it is a boundary of the bulk. For a flat brane, $k = 0$, the scale factor behaves non-conventionally: $a \sim t^{1/[3(\omega+1)]}$. We have to conclude that a Universe (brane-world) which is filled with ordinary matter (with $\omega > 0$) undergoes no late time acceleration in this regime. The situation is completely different for models described by (35) with "plus" sign. Here, the low energy density regime $\rho << \rho_c$ is necessarily connected with a dynamics of the type

$$H^2 + \frac{k}{a^2} \approx \frac{\kappa_4^4}{\kappa_5^4}\,. \tag{38}$$

Thus, in this scenario a phase of matter or radiation dominated cosmology is followed by a late phase of accelerated expansion [18].

A more detailed analysis of the cosmological behavior in models with induced

---

[3] It was shown in [20], that the generalized on-brane Friedmann equation (19) (obtained by the projection method of the previous section 2) can be recovered from (34) by the formal limit $\kappa_4^2 \to \infty$. For a model with $\mathbb{Z}_2$–symmetry and a brane embedded in the bulk (and not located on the bulk boundary as considered in the present section) this formal $\kappa_4^2 \to \infty$ limit gives

$$H^2 + \frac{k}{a^2} = \frac{1}{36}\kappa_5^4\left(\rho + \lambda\right)^2 + \frac{1}{6}\Lambda_5 + \frac{\mathcal{C}}{a^4}$$

what via relations (10) exactly reproduces (19) with an additional dark radiation term.

[4] This characteristic energy density changes within the limits 10^{-93} g/cm$^3 \le \rho_c \le 10^{-3}$ g/cm^3 depending on the concrete value of M_* from the interval 10^{-3} eV $\le M_* \lesssim 1$ TeV. In the special case $M_* \sim 10^{-4/3}$GeV it holds $\rho_c \sim \rho_0 \sim 10^{-29}$ g/cm^3.

gravity can be found, for example, in [19, 21][5]. An interesting generalization of DGP-type models to branes with co-dimensions $D \geq 2$ was proposed in [25]. In the case of a co-dimension 2 model, the characteristic scale at which gravity becomes six-dimensional is $l \sim M_{Pl}/M_*^2$, which is of the order of the Hubble radius H_0^{-1} for the dark energy mass scale $M_* \sim 10^{-3}$ eV. However, it was pointed out [26] that such models are not free of ghosts and will lead to violations of the equivalence principle.

4 Cardassian Expansion

The discovery of the present late-time acceleration of the Universe posed a great puzzle to modern theoretical cosmology. Up to now several scenarios have been proposed as possible solutions, such as a cosmological constant, a decaying vacuum energy and quintessence. However, none of them is fully satisfactory. On the other hand, as we have seen above, a possible explanation could consist of an IR modification of gravity and with it a modification of the effective Friedmann equation (see e.g. the low energy limit of (35)). A natural phenomenological ansatz for an effective Friedmann equation could consist, e.g., of adding a nonlinear term in the energy density,

$$H^2 = \frac{1}{3}\kappa_4^2\rho + B\rho^n\,, \tag{39}$$

where B is a constant prefactor. Assuming the matter as dust with $P = 0\,,\ \rho \sim a^{-3}$ this would lead, for a dominating nonlinear energy density term, to

$$H^2 \approx B\rho^n \quad \Longrightarrow \quad a \sim t^{2/(3n)}\,. \tag{40}$$

Thus, for $n < 2/3$ the Universe would undergo an accelerated expansion. Therefore, for large energy densities the main contribution would come from the first term in (39) and we would recover conventional cosmology. In contrast to this, for small energy densities the second term would dominate so that starting from a certain evolution stage the Universe would undergo a late time acceleration. Such a scenario was dubbed "Cardassian expansion"[6] [27]. The appealing feature of this ansatz is the fact that the acceleration is explained without involving vacuum energy. The question is how to derive the specific (phenomenological) nonlinear contributions in the energy density from a more fundamental theory (such as string theory/M-theory)?

One possible way could consist of taking advantage of bulk corrections in brane-world models. For the special case with $n = 2$ this has been demonstrated in the

---

[5] It was shown in [21] that conventional on-brane dark matter can lead to a cosmology with equation of state parameter $\omega < -1$ but without phantom-like future singularity as well as with a transient acceleration with subsequent matter dominated regime. Additionally, it should be noted that in spite of the appealing features of the DGP model, it predicts [22, 23] intensive and strong interactions at the energy scale $(M_{Pl}/l^2)^{1/3}$, which for $l \sim H_0^{-1}$ corresponds to distances ~ 1000 km [22]. Obviously, this would conflict with experimental data. However, it was noted in [24] that this problem can be cured with the help of a proper UV completion of the theory and it may not occur for certain configurations of the general type (24) (with $\Lambda_5, \lambda \neq 0$) [22].

[6] The name Cardassian refers to a humanoid race in "Star Trek" whose goal was to take over the Universe, i.e. an accelerated expansion.

previous two sections. In [28] it was shown that, in general, it is possible for such models to get a modified Friedmann equation $H^2 \sim \rho^n$ with any n. The main idea consists of defining the bulk moduli N, a and b in the metric

$$ds^2 = -N^2(t,y)dt^2 + a^2(t,y)d\vec{x}^2 + b^2(t,y)dy^2 \tag{41}$$

in such a way that the energy density of on-brane dust behaves (via the Israel junction conditions) in accordance with (40). On the other hand, these moduli specify the bulk EMT via the 5D Einstein equations (see also [29]). The main drawback of the approach consists of a very clumsy and inelegant form of this bulk EMT [27, 28]. It is very difficult to guess what kind of matter could produce such an EMT. Furthermore, the combination $\rho + \rho^n$ with arbitrary n is not yet obtained by this method. But precisely this combination is required to induce a matter dominated cosmology which is followed by a late phase of accelerated expansion (for $n < 2/3$).

5 Self-tuning Universe

One of the most natural candidates for dark energy causing the present late time acceleration of the Universe is a positive vacuum energy (in other words, a cosmological constant). The energy density of vacuum fluctuations (with natural cut-off $k_c \sim M_{Pl}$) can be estimated as

$$< \rho_v > = \int_0^{k_c} \frac{d^3k}{(2\pi)^3}\frac{1}{2}\sqrt{k^2+m^2} \sim \frac{k_c^4}{16\pi^2} \sim M_{Pl}^4 = \rho_{Pl} \sim \left(10^{19}\ \mathrm{GeV}\right)^4 \tag{42}$$

and is 123 orders of magnitude greater than the observable dark energy density $\rho \sim (10^{-3}\ \mathrm{eV})^4 \sim 10^{-123}\rho_{Pl}$. Usually, this discrepancy is resolved by an extra ordinary parameter fine tuning and for this reason it is well known as the fine tuning problem. However, there seems to exist a possibility to avoid this problem. The idea again consists of a modification of the Friedmann equation. Let us suppose that the standard (flat space) Friedmann equation $H^2 = (1/3)8\pi G_N\rho$ is modified in such a way [30] that it holds

$$H^2 = f(\rho, P)(\rho + P) + \dots\,, \tag{43}$$

where $f(\rho, P)$ is a smooth function of the energy density ρ and the pressure P, and dots denote subdominant terms of other type. Then the state equation of the vacuum energy, $P = -\rho$, shows that its contribution will exactly cancel and that it will not affect the evolution of the Universe, i.e. its energy density can be arbitrarily large. In general, such a modification can be in agreement with modern observational data. It is easily demonstrated on the simplest model with

$$H^2 = 2\pi G_N(\rho + P)\,. \tag{44}$$

During the matter dominated stage with $P = 0$ this equation reads $H^2 = 2\pi G_N\rho = (3/4)(1/3)8\pi G_N\rho$. The additional factor $3/4$ is rather close to 1 and such a deviation from the standard Friedmann equation is very difficult to extract from observational data. During the radiation dominated stage with $P = (1/3)\rho$ the standard equation $H^2 = (1/3)8\pi G_N \rho$ is restored, which is very important for the BBN. Scenarios with a modified Friedmann equation of the type (43) were dubbed "self-tuning Universe". They were first proposed in [31–33] for specific brane-world models. An excellent explanation of the basic mechanisms underlying this setup was given in [30]. In our subsequent brief outline, we mainly follow this work.

The starting point for the scenarios is a 5D brane-world model with action

$$S = \int d^5x \sqrt{|{}^{(5)}g|} \left[\frac{1}{2\kappa_5^2} {}^{(5)}R - \alpha \left({}^{(5)}\nabla\varphi \right)^2 \right] + S_4 \left[\varphi, {}^{(4)}g, \Psi_i \right] , \tag{45}$$

where φ is a dilatonic bulk field and Ψ_i are matter fields living on the brane. The 5D metric ${}^{(5)}g_{AB}$ in the Einstein frame is connected with the 5D metric ${}^{(5)}\tilde{g}_{AB}$ in Brans–Dicke (BD) frame as

$${}^{(5)}\tilde{g}_{AB} = \left(e^{\beta\varphi}\right) {}^{(5)}g_{AB} \quad \Longrightarrow \quad {}^{(4)}\tilde{g}_{\mu\nu} = \left(e^{\beta\varphi}\right) {}^{(4)}g_{\mu\nu} , \tag{46}$$

with the induced metrics ${}^{(4)}\tilde{g}_{\mu\nu}$ and ${}^{(4)}g_{\mu\nu}$ connected by a similar relation. The setup allows for two possible types of brane-matter coupling in the effective 4D action on the brane: either the matter is coupled to the induced 4D metric in the BD frame,

$$S_{(1)4} = \int d^4x \sqrt{|{}^{(4)}\tilde{g}|}\, L_4 \left(\Psi_i, {}^{(4)}\tilde{g} \right) , \tag{47}$$

or it is coupled to the induced metric in the Einstein frame,

$$S_{(2)4} = \int d^4x \sqrt{|{}^{(4)}\tilde{g}|}\, L_4 \left(\Psi_i, {}^{(4)}g \right) = \int d^4x \sqrt{|{}^{(4)}g|}\, \left(e^{2\beta\varphi}\right) L_4 \left(\Psi_i, {}^{(4)}g \right) . \tag{48}$$

Using the ansatz (41) for the 5D metric (in the Einstein frame) and the special parameter tuning

$$\frac{\beta^2}{\alpha} = \frac{2}{3}\kappa_5^2 \tag{49}$$

we find the following on-brane equations

$$\frac{\ddot{a}_0}{a_0} + \frac{\dot{a}_0^2}{a_0^2} = \dot{H} + 2H^2 = \begin{cases} -\frac{1}{32}\kappa_5^4 e^{4\beta\varphi_0} \left(\tilde{\rho} + \tilde{P} \right)^2 & - \quad \frac{1}{2}\beta^2\dot{\varphi}_0^2 , \quad \text{(I)} \\ -\frac{1}{36}\kappa_5^4 \left(\rho + P\right)\left(\rho + 2P\right) & - \quad \frac{1}{2}\beta^2\dot{\varphi}_0^2 , \quad \text{(II)} \end{cases} \tag{50}$$

which contain energy-density–pressure combinations of a form similar to the self-tuning combinations of the generalized Friedmann equation (43). In (50), equation (I) corresponds to the BD frame coupling and equation (II) to the Einstein frame coupling. The tilded and non-tilded energy densities and pressures $(\tilde{\rho}, \tilde{P})$ and (ρ, P) are defined from (47) and (48) with respect to metrics ${}^{(4)}\tilde{g}_{\mu\nu}$ and ${}^{(4)}g_{\mu\nu}$, respectively. 0-subscripts refer to on-brane quantities.

Comparing equations (I) and (II) in (50) with the self-tuning equations (43) and (44), we immediately realize that the quadratic energy-density–pressure term in equation (I) would yield a self-tuning but it would spoil the possibility for a limit to conventional cosmological behavior. In contrast to this, equation (II) can provide the necessary linear dependence. Splitting, for example, the on-brane energy density and pressure into a static background component (which can be identified with a vacuum contribution[7]) and a dynamical component of ordinary (non-vacuum) matter,

$$\rho = e^{2\beta\varphi_0}\lambda + \rho_m , \quad P = -e^{2\beta\varphi_0}\lambda + P_m , \tag{51}$$

(with λ as brane tension) one gets from (II) for constant on-brane values of the dilaton[8] ($\dot{\varphi}_0 = 0$) an equation suitable for self-tuning

$$\dot{H} + 2H^2 = \frac{\kappa_5^4}{36} \left[e^{2\beta\varphi_0}\lambda \left(\rho_m + P_m\right) - \left(\rho_m^2 + 3\rho_m P_m + 2P_m^2\right) \right] . \tag{52}$$

In a late-time regime, the quadratic term in the dynamical energy density and pressure would be subdominant and it remains to check whether the resulting linear model will be capable to reproduce conventional cosmology. The analysis can be made explicit, e.g., by considering the simplest case of a flat-brane ansatz for the metric (41)

$$ds^2 = \omega(y) \left[-dt^2 + a_0^2(t) d\vec{x}^2 \right] + dy^2 . \tag{53}$$

The concrete form of $\omega(y)$ was found in [31, 32, 33] as

$$\omega(y) = \sqrt{1 - y/y_c} , \quad y_c = \frac{3\kappa_5^{-2}}{2\lambda} e^{-2\beta\varphi_0} , \tag{54}$$

where y_c corresponds to a boundary (opposite to the world-brane) at which the metric experiences a singularity, $\omega(y_c) = 0$. The effective 4D gravitational constant on the brane can be obtained via dimensional reduction as

$$\frac{1}{8\pi G_N} = \frac{1}{\kappa_4^2} = \frac{1}{\kappa_5^2} \int_0^{y_c} dy \omega(y) = \frac{1}{\lambda\kappa_5^4} e^{-2\beta\varphi_0} . \tag{55}$$

Thus, up to quadratic terms, (52) reads

$$\dot{H} + 2H^2 \approx \frac{2\pi G_N}{9} \left(\rho_m + P_m\right) \tag{56}$$

and yields after integration

$$H^2 = \frac{4\pi G_N}{9} a^{-4} \int a^3 \left(\rho_m + P_m\right) da . \tag{57}$$

For a Universe which is dominated by a combination of dust ($\rho_d \sim 1/a^3, P_d = 0$) and radiation ($\rho_r \sim 1/a^4, P_r = (1/3)\rho_r$) this leads to a Hubble parameter of the type

$$H^2 = \frac{1}{3} 8\pi G_N \left[\frac{1}{6}\rho_d + \frac{2}{9}\rho_r \ln\left(\frac{a}{a_*}\right) \right] \tag{58}$$

with a_* as constant of integration.

---

[7] The self-tuning scenario was designed to cancel the huge vacuum contributions, but in its current form it is not yet capable to elegantly induce accelerated expansion – like inflation or late-time acceleration (see also footnote 9).

[8] Here, the stabilization problem for the on-brane value of the dilaton is left aside. It would require a separate investigation.

Obviously, the considered toy model does not exactly reproduce the dynamics of conventional cosmology. During the matter dominated stage the additional prefactor 1/6 induces a deviation from conventional cosmology which probably will leave an imprint in the CMB anisotropy. Also the dynamics of the radiation dominated stage is governed by a different equation – which can affect the BBN. This may require a special tuning mechanism[9] to produce the correct light-element abundances. Nevertheless, the simple toy model allowed us to illuminate the main ideas of the self-tuning mechanism. Further details of the scenario have been worked out, for example, in [34–37].

6 Ekpyrotic Universe

A very interesting scenario which is based on a different general setup – compared with those in the previous sections – was proposed in [38]. It is based on the assumption that the evolution of our Universe (in accordance with known observational data) can be explained with the help of a relatively simple brany M-theory setup with S^1/Z_2 orbifold symmetry. The starting point is the Hořava-Witten model (briefly mentioned in the introduction section 1) which after dimensional reduction leads to an effective 5D action of the type

$$S = \frac{1}{2\kappa_5^2}\int_{M_5} d^5x\sqrt{|^{(5)}g|}\left[{}^{(5)}R - \frac{1}{2}(\partial\varphi)^2 - \frac{3}{2}\frac{e^{2\varphi}\mathcal{F}^2}{5!}\right] + \text{brane terms}\,. \tag{59}$$

The scalar field φ defines the volume scale (modulus) of the Calabi-Yau three-fold and $\mathcal{F}$ is the field strength of a four-form gauge field. The key point of this model is a static (nearly) Bogomol'nyi–Prasad–Sommerfeld (BPS) solution [38] which describes a setup consisting of three 3-branes:

$$\begin{aligned} ds^2 &= D(y)\left(-N^2d\tau^2 + A^2d\vec{x}^2\right) + B^2D^4(y)dy^2\,, \\ e^{\varphi} &= BD^3(y)\,, \qquad\qquad (60) \\ \mathcal{F}_{0123y} &= -\alpha A^3NB^{-1}D^{-2}(y)\,, \quad (y<Y)\,, \\ &= -(\alpha-\beta)A^3NB^{-1}D^{-2}(y)\,, \quad (y>Y)\,, \qquad (61) \end{aligned}$$

where τ denotes the conformal time, $D(y)$ is defined as

$$\begin{aligned} D(y) &= \alpha y + C\,, \quad (y<Y)\,, \\ &= (\alpha-\beta)y + C + \beta Y\,, \quad (y>Y) \end{aligned} \tag{62}$$

and A, B, C, N, Y are constants (with $C>0$). Two of the three branes are located at the orbifold fixed points $y=0$ and $y=R$, whereas the third (bulk) brane is positioned at $0<y=Y<R$. Our visible brane (at $y=0$) is assumed to have negative tension $\alpha_1 \equiv \alpha < 0$, the hidden brane (at $y=R$) $\alpha_2 \equiv \alpha - \beta > 0$ and

---

[9] Additionally, it should be noted that up to now no convincing mechanism has been found for an elegant matching of the self-tuning scenarios with inflation. The point is that for conventional cosmology a vacuum equation of state with $P=-\rho$ yields $H=$const. In contrast to this, setting in (52) $\rho_m = 0 = P_m$ (i.e., a vacuum of the dynamical contributions) gives $H \sim 1/(2t)$ and H behaves in the same way as during the RD stage of conventional cosmology.

the bulk brane $\alpha_3 \equiv \beta > 0$ to have a positive one. Furthermore, it is assumed that the bulk brane is nucleated near the positive tension (hidden) brane at $y=R$. It is supposed that the bulk brane is very light: $\beta << \alpha$. A very important property of the (nearly) BPS solution is that all these three branes are flat and parallel to each other.

Additionally, it is assumed that a non-perturbative interaction between the bulk and boundary branes leads to an effective potential

$$V(Y) \approx -ve^{-m\alpha Y}\,, \tag{63}$$

which for small Y suddenly becomes zero (v and m are positive dimensionless parameters of the model). Via the potential the light bulk brane is attracted to the visible brane and moves adiabatically slowly towards the visible brane (see the schematic illustration in Figure 2).

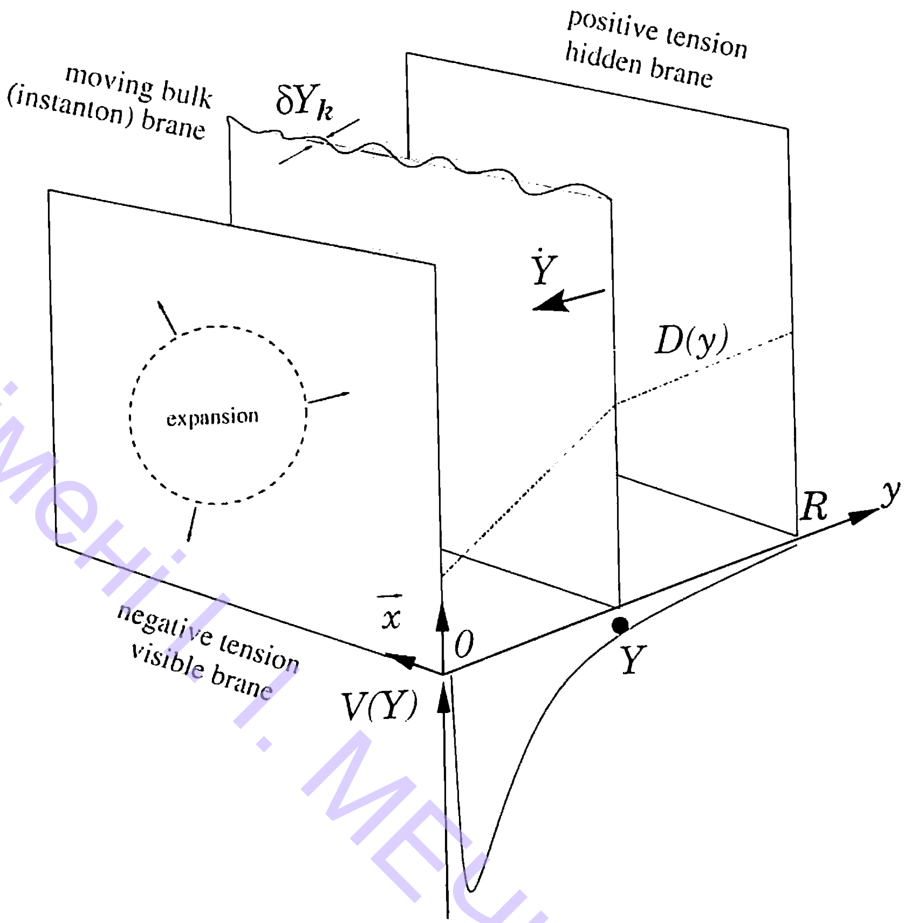


Figure 2: Schematic of the ekpyrotic scenario (from [39]).

In this scenario, branes and bulk are assumed to start out from a cold, symmetric state with (60) which is nearly BPS. The beginning of our Universe, i.e. the Big Bang, is identified in this model with the impact of the bulk brane on the visible brane. After this collision, a part of the kinetic energy of the bulk brane transforms into radiation which is deposited in the three dimensional space of the visible brane.

The scenario was called “Ekpyrotic Universe” [38][10]. An early detailed analysis of its features was performed in [38–44].

It was argued that this model could solve the main cosmological problems without invoking inflation and as alternative to it[11]:

- It is assumed that the *monopole problem* can be avoided if the parameters of the model are chosen in such a way that the temperature on the visible brane after collision is not high enough for a production of primordial monopoles.
- The *flatness and homogeneity problems* are automatically solved[12] because the nearly BPS branes are assumed as sufficiently flat and homogeneous (see however the critical comments in [39, 40] concerning the naturalness of the required high parameter tuning).
- The *horizon problem* is solved by the extremely slow motion of the bulk brane so that, before collision, particles on the visible brane can travel an exponentially long distance. As a result, the horizon distance d_{hor} can be much bigger than the Hubble radius at collision $H_c^{-1} \sim T_c^{-2}$, where T_c is the temperature of the visible brane at collision.
- It is assumed that the *large-scale structure* on the visible brane is induced by quantum fluctuations (ripples) in the position of the bulk brane. These fluctuations would lead to a space-dependent time delay in the brane collision which on its turn would result in density fluctuations on the visible brane. Because small ripples can generate a large time delay, the collision could induce a spectrum of density fluctuations which would extend to exponentially large, super-horizon scales. In the considered scenario, the spectrum of perturbations is approximately scale invariant. However, in the simplest variant of the ekpyrotic model the spectrum of adiabatic fluctuations is blue [42, 43].

Summarizing, the M-theory based scenario of an Ekpyrotic Universe has a large number of interesting features which could present an alternative to inflation. Future investigations will show which of the scenarios is more robust. Further developments in this direction can be found in [46, 47].

7 Conclusions

New observational data directly indicate that our knowledge about the Universe and its evolution is limited to a great extent. In this situation new ideas and new theoretical models are required to explain the observational data. In turn, these new models will predict new observable phenomena which can be used as test tools to single out those scenarios which are most close to nature. One possible direction

---

[10] This term was drawn from the Stoic model of cosmic evolution in which the Universe is consumed by fire at regular intervals and reconstituted out of this fire – a conflagration called ekpyrosis. In the present scenario, the Universe is made through a conflagration ignited by a brane collision.

[11] See, however, the critical comments in [39, 40] and the reply to these comments in [41].

[12] For the solution of these problems in other brane world models, e.g. such with large extra dimensions, different smoothing mechanisms have been proposed [45].

to modify the currently accepted standard theory of cosmology are brane-world models – M-theory inspired setups in which our observable Universe is interpreted as a 3D submanifold embedded in a higher dimensional bulk spacetime. The aim of the present mini-review was to give a very brief description of some of the basic features of these models – illustrating them with the help of a few concrete toy model setups.

Acknowledgements

We thank the High Energy, Cosmology and Astroparticle Physics Section of the ICTP (Trieste) for their warm hospitality during the preparation of this mini-review as well as for financial support from the Associate Scheme (A.Z.) and from HECAP (U.G.). Additionally, U.G. acknowledges support from DFG grant KON/1806/2004/GU/522.

References

[1] P. Hořava, E. Witten, *Nucl. Phys.*, **B460**, 506 (1996).

[2] P. Hořava, E. Witten, *Nucl. Phys.*, **B475**, 94 (1996).

[3] A. Lukas, B.A. Ovrut, D. Waldram, *Nucl. Phys.*, **B532**, 43 (1998).

[4] A. Lukas, B.A. Ovrut, K.S. Stelle, D. Waldram, *Phys. Rev.*, **D59**, 086001 (1999).

[5] T. Shiromizu, K. Maeda, M. Sasaki, *Phys. Rev.*, **D62**, 024012 (2000).

[6] R. Maartens, *Phys. Rev.*, **D62**, 084023 (2000).

[7] R. Maartens, *Living Rev. Rel.*, **7**, 1 (2004).

[8] A. Campos, C.F. Sopuerta, *Phys. Rev.*, **D63**, 104012 (2001).

[9] G. Huey, J.E. Lidsey, *Phys. Lett.*, **B514**, 217 (2001).

[10] M. Pavšič, *Grav. Cosmol.*, **2**, 1 (1996).

[11] S. Nojiri, S.D. Odintsov, S. Zerbini, *Phys. Rev.*, **D62**, 064006 (2000).

[12] S.W. Hawking, T. Hertog, H.S. Reall, *Phys. Rev.*, **D62**, 043501 (2000).

[13] H. Collins, B. Holdom, *Phys. Rev.*, **D62**, 105009 (2000).

[14] S. Nojiri, S.D. Odintsov, *Phys. Lett.*, **B484**, 119 (2000).

[15] G.R. Dvali, G. Gabadadze, M. Porrati, *Phys. Lett.*, **B485**, 208 (2000).

[16] Yu.V. Shtanov, preprint hep-th/0005193 (2000).

[17] G.R. Dvali, G. Gabadadze, *Phys. Rev.*, **D63**, 065007 (2001).

[18] C. Deffayet, *Phys. Lett.*, **B502**, 199 (2001).

[19] C. Deffayet, G.R. Dvali, G. Gabadadze, *Phys. Rev.*, **D65**, 044023 (2002).

[20] Y.V. Shtanov, *Phys. Lett.*, **B541**, 177 (2002).

[21] V. Sahni, Yu. Shtanov, *JCAP*, **0311**, 014 (2003).

[22] M.A. Luty, M. Porrati, R. Rattazzi, *JHEP*, **0309**, 029 (2003).

[23] V.A. Rubakov, preprint hep-th/0303125 (2003).

[24] A. Nicolis and R. Rattazzi, *JHEP*, **0406**, 059 (2004).

[25] G. Dvali, G. Gabadadze, X. Hou, E. Sefusatti, *Phys. Rev.*, **D67**, 044019 (2003).

[26] S.L. Dubovsky, V.A. Rubakov, *Phys. Rev.*, **D67**, 104014 (2003).

[27] K. Freese, M. Lewis, *Phys. Lett.*, **B540**, 1 (2002).

[28] D.J.H. Chung, K. Freese, *Phys. Rev.*, **D61**, 023511 (2000).

[29] P. Binetruy, C. Deffayet, D. Langlois, *Nucl. Phys.*, **B565**, 269 (2000).

[30] S.M. Carroll, L. Mersini, *Phys. Rev.*, **D64**, 124008 (2001).

[31] N. Arkani-Hamed, S. Dimopoulos, N. Kaloper, R. Sundrum, *Phys. Lett.*, **B480**, 193 (2000).

[32] S. Kachru, M.B. Schulz, E. Silverstein, *Phys. Rev.*, **D62**, 045021 (2000).

[33] S. Kachru, M.B. Schulz, E. Silverstein, *Phys. Rev.*, **D62**, 085003 (2000).

[34] S.M. Carroll, M.M. Guica, preprint hep-th/0302067 (2003).

[35] H.-P. Nilles, A. Papazoglou, G. Tasinato, *Nucl. Phys.*, **B677**, 405 (2004).

[36] J.E. Kim, H.M. Lee, *Phys. Lett.*, **B590**, 1 (2004).

[37] H.M. Lee, *Phys. Lett.*, **B587**, 117 (2004).

[38] J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, *Phys. Rev.*, **D64**, 123522 (2001).

[39] R. Kallosh, L. Kofman, A.D. Linde, *Phys. Rev.*, **D64**, 123523 (2001).

[40] R. Kallosh, L. Kofman, A. Linde, A. Tseytlin, *Phys. Rev.*, **D64**, 123524 (2001).

[41] J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, preprint hep-th/0105212.

[42] D.H. Lyth, *Phys. Lett.*, **B524**, 1 (2002).

[43] R. Brandenberger, F. Finelli, *JHEP*, **0111**, 056 (2001).

[44] J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, *Phys. Rev.*, **D66**, 046005 (2002).

[45] G.D. Starkman, D. Stojkovic, M. Trodden, *Phys. Rev. Lett.*, **87**, 231303 (2001).

[46] J. Khoury, B.A. Ovrut, N. Seiberg, P.J. Steinhardt, N. Turok, *Phys. Rev.*, **D65**, 086007 (2002).

[47] P.J. Steinhardt, N. Turok, *Phys. Rev.*, **D65**, 126003 (2002).

10

Braneworld Cosmology

J. Soda and S. Kanno

Department of Physics, Kyoto University, Kyoto 606-8501, Japan
e-mail: jiro@tap.scphys.kyoto-u.ac.jp,
sugumi@tap.scphys.kyoto-u.ac.jp

Abstract

We present a general formalism for studying an inflationary scenario in the two-brane system. We explain the gradient expansion method and obtain the 4-dimensional effective action. Based on this effective action, we give the basic equations for the background homogeneous universe and the cosmological perturbations. As an application, we propose a born-again braneworld scenario.

Keywords: cosmology, low-energy, gradient expansion, born-again braneworld

1 Introduction

It is widely believed that the initial singularity problem in the conventional standard cosmology can be resolved by the quantum theory of gravity. It is also known that the most promising candidate for the quantum theory of gravity is the superstring theory. Remarkably, the superstring theory precisely predicts the existence of extra-dimensions. As our universe can be recognized as the 4-dimensional Lorentzian manifold by experience, the necessity for reconciling the predictions of the superstring theory with our experience is apparent.

The solution to this problem was proposed by Kaluza and Klein a long time ago. Their idea is simply that the extra-dimensions are too tiny to see them. In fact, in order to fit our real world, the sizes of the internal dimensions are constrained to be the Planck length. This beautiful answer has been considered to be a unique one for a long time. However, the recent discovery of Dp-brane solutions has changed the situation. The Dp-brane is a kind of soliton solution with the p-dimensional spatial and 1-dimensional temporal extension on which the boundary of open strings can be attached. Closed strings can freely propagate in the whole 10-dimensional spacetime. Since the open string contains the standard matter and the closed string contains gravitons, the D-brane solution yields the idea of the braneworld where the standard matter is confined to the 4-dimensional braneworld, while gravitons can propagate in the 10-dimensional bulk spacetime. Interestingly, in the case of the braneworld, the constraint on the sizes of the internal dimensions is not so stringent compared with the case of Kaluza–Klein mechanism. Indeed, accelerator experiments in the laboratory cannot give strong constraints because the standard matter is confined to the 4-dimensional braneworld. Moreover, the experiment of Newtonian force is rather poor at scales smaller than 0.1 mm. Hence, the sizes of the extra-dimensions could be 0.1 mm in this braneworld picture.

Because of the possibility of the large extra-dimensions in the braneworld models, the early universe must be studied with careful consideration of the effects of the bulk geometry. In this sense, the extra-dimensions can be regarded as the new elements of the universe.

To be more precise, what we want to explore is the effect of the bulk geometry on the generation and evolution of the cosmological fluctuations. In the conventional 4-dimensional cosmological models, curvature fluctuations with a flat spectrum are generated from the quantum fluctuations during the inflationary expansion. Then, its amplitude is frozen until the wavelength crosses the horizon again. Now, these curvature fluctuations can be observed through the cosmic microwave background radiation. There are various possible sources to alter this standard scenario in the braneworld models. The most obvious possibility is the collision between different braneworlds. Related to this case, there may be the effect of the radion, namely the distance between two braneworlds. Moreover, there are effects due to gravitational waves propagating in the bulk, these are often called Kaluza–Klein (KK) effects.

There are many braneworld models. Here we will concentrate on the simplest Randall–Sundrum two-brane model [1]. In this paper, using the gradient expansion method, we formulate the general scheme to investigate effects of the bulk geometry. As an application, we propose the born-again braneworld scenario and discuss its resemblance to the pre-big-bang model [2]. We also make observational predictions on the spectrum of the primordial gravitational waves which turn out to be very blue.

2 Inflation on Braneworld

Let us start with the effective Friedmann equation

$$H^2 = \frac{\kappa^4\mu^2}{12} - \frac{3}{\ell^2} + \frac{\kappa^2}{\ell}\rho + \kappa^4\rho^2 + \frac{\mathcal{C}}{a^4} , \qquad (1)$$

where H, a and ρ are, respectively, the Hubble parameter, the scale factor and the total energy density of each brane. The Newton's constant can be identified as $8\pi G_N = \kappa^2/\ell$ and $\mathcal{C}$ is a constant of integration associated with the mass of a black hole in the bulk. This constant $\mathcal{C}$ is referred to as the dark radiation in which the effect of the bulk is encoded. The effective cosmological constant for the braneworld becomes

$$\Lambda_{\rm eff} = \frac{\kappa^4\mu^2}{12} - \frac{3}{\ell^2} , \qquad (2)$$

where μ and ℓ are the tension of the brane and the curvature scale in the bulk which is determined by the bulk vacuum energy, respectively. For $\kappa^2\mu = 6/\ell$, we have Minkowski spacetime as a solution. In order to obtain the inflationary universe, we need the positive effective cosmological constant. In the brane world model, there are two possibilities. One is to increase the brane tension and the other is to increase ℓ. The brane tension can be controlled by the scalar field on the brane. The bulk curvature scale ℓ can be controlled by the bulk scalar field. The former case is a natural extension of the 4-dimensional inflationary scenario [3]. The latter possibility is a novel one peculiar to the brane model. Recall that, in the superstring theory, scalar fields are ubiquitous. Indeed, the dilaton and moduli exist in the bulk generically, because they arise as the modes associated with the closed string. Moreover, when the supersymmetry is spontaneously broken, they may have the non-trivial potential. Hence, it is natural to consider the inflationary scenario driven by these fields [4]. Now, we develop the formalism to investigate these inflating braneworlds.

3 Gradient Expansion Method

In this section, for simplicity, we explain the gradient expansion method using the single-brane model. We use the Gaussian normal coordinate system

$$ds^2 = dy^2 + g_{\mu\nu}(y,x)dx^\mu dx^\nu \qquad (3)$$

to describe the geometry of the brane world. Note that the brane is located at $y = 0$ in this coordinate system. Decomposing the extrinsic curvature $K_{\mu\nu} = -g_{\mu\nu,y}/2$ into the traceless and the trace part $K_{\mu\nu} = \Sigma_{\mu\nu} + \frac{1}{4}g_{\mu\nu}K$, we obtain the basic equations which hold in the bulk:

$$\Sigma^\mu{}_{\nu,y} - K\Sigma^\mu{}_\nu = -\left[\overset{(4)}{R}{}^\mu{}_\nu - \frac{1}{4}\delta^\mu_\nu \overset{(4)}{R}\right] , \qquad (4)$$

$$\frac{3}{4}K^2 - \Sigma^\alpha{}_\beta\Sigma^\beta{}_\alpha = \left[\overset{(4)}{R}\right] + \frac{12}{\ell^2} , \qquad (5)$$

$$\nabla_\lambda \Sigma_\mu{}^\lambda - \frac{3}{4}\nabla_\mu K = 0 , \qquad (6)$$

where ∇_μ denotes the covariant derivative with respect to the metric $g_{\mu\nu}$ and $\overset{(4)}{R}{}^\mu{}_\nu$ is the curvature associated with it. We also have the junction condition

$$\left[K^\mu{}_\nu - \delta^\mu_\nu K\right]\Big|_{y=0} = \frac{\kappa^2}{2}\left(-\mu\delta^\mu_\nu + T^\mu{}_\nu\right) , \qquad (7)$$

where $T_{\mu\nu}$ is the energy momentum tensor of the matter. Recall that we are considering the Z_2 symmetric spacetime.

The problem now is separated into two parts. First, we will solve the bulk equations of motion with the Dirichlet boundary condition at the brane, $g_{\mu\nu}(y = 0, x^\mu) = h_{\mu\nu}(x^\mu)$. After that, the junction condition will be imposed at the brane. As it is the condition for the induced metric $h_{\mu\nu}$, it is naturally interpreted as the effective equations of motion for gravity on the brane.

In this paper, we will consider the low energy regime in the sense that the energy density of matter on the brane, ρ, is smaller than the brane tension, i.e., $\epsilon \equiv \rho/\mu \ll 1$. In this regime, we can use the gradient expansion method where the above ratio ϵ is the expansion parameter [6].

At zero order, we can neglect the curvature term. The anisotropic term must vanish in order to satisfy the junction condition. Now, it is easy to solve the remaining equations. The result is $\overset{(0)}{K} = \frac{4}{\ell}$. Using the definition of the extrinsic curvature, we get the zero order metric as

$$ds^2 = dy^2 + a^2(y) h_{\mu\nu}(x^\mu) dx^\mu dx^\nu \ , \quad a(y) = e^{-2\frac{y}{\ell}} \ , \tag{8}$$

where the tensor $h_{\mu\nu}$ is the induced metric on the brane. From the zero order junction condition, we obtain the well-known relation $\kappa^2 \mu = 6/\ell$. Hereafter, we will assume that this relation holds exactly.

The iteration scheme consists in writing the metric $g_{\mu\nu}$ as a sum of local tensors built out of the induced metric on the brane, the number of gradients increasing with the order. Hence, we will seek the metric as a perturbative series

$$g_{\mu\nu}(y, x^\mu) = a^2(y) \left[h_{\mu\nu}(x^\mu) + \overset{(1)}{g}_{\mu\nu}(y, x^\mu) + \overset{(2)}{g}_{\mu\nu}(y, x^\mu) + \cdots \right] \ , \tag{9}$$

where $a^2(y)$ is extracted and we put the Dirichlet boundary condition $\overset{(i)}{g}_{\mu\nu}(y = 0, x^\mu) = 0$, so that $g_{\mu\nu}(y = 0, x) = h_{\mu\nu}(x)$ holds at the brane. Other quantities can be also expanded as

$$\begin{aligned} K^\mu{}_\nu &= \frac{1}{\ell}\delta^\mu_\nu + \overset{(1)}{K}{}^\mu{}_\nu + \overset{(2)}{K}{}^\mu{}_\nu + \cdots \\ \Sigma^\mu{}_\nu &= \qquad + \overset{(1)}{\Sigma}{}^\mu{}_\nu + \overset{(2)}{\Sigma}{}^\mu{}_\nu + \cdots \ . \end{aligned} \tag{10}$$

The next order solutions are obtained by taking into account the terms neglected at zero order. Substituting the zero order metric into $\overset{(4)}{R}(g)$, we obtain

$$\overset{(1)}{K} = \frac{\ell}{6a^2} R(h) \ . \tag{11}$$

Hereafter, we omit the argument of the curvature for simplicity. A simple integration of equation (4) also gives the traceless part of the extrinsic curvature:

$$\overset{(1)}{\Sigma}{}^\mu{}_\nu = \frac{\ell}{2a^2}(R^\mu{}_\nu - \frac{1}{4}\delta^\mu_\nu R) + \frac{\chi^\mu{}_\nu(x)}{a^4} \ , \tag{12}$$

where the homogeneous solution satisfies the constraints $\chi^\mu{}_\mu = 0$, $\chi^\mu{}_{\nu|\mu} = 0$. This term corresponds to dark radiation at this order.

At this order, the junction condition can be written as

$$\left[\overset{(1)}{K}{}^\mu{}_\nu - \delta^\mu_\nu \overset{(1)}{K} \right] \Bigg|_{y=0} = \frac{\ell}{2}\left(R^\mu{}_\nu - \frac{1}{2}\delta^\mu_\nu R \right) + \chi^\mu{}_\nu = \frac{\kappa^2}{2} T^\mu{}_\nu \ . \tag{13}$$

Thus, we have obtained the effective equation on the brane.

4 Effective Action for Two-brane System

We consider the two-brane system in this section. Without matter on the branes, we have the relation between the induced metrics, $g^{\ominus\text{-brane}}_{\mu\nu} = e^{-2d/\ell} g^{\oplus\text{-brane}} \equiv \Omega^2 g^{\oplus\text{-brane}}$ where d is the distance between the two branes. Here, $\oplus$ and $\ominus$ represent the positive and the negative tension branes, respectively. Although Ω is constant for vacuum branes, it becomes a function of the 4-dimensional coordinates if we put the matter on the brane.

Adding the energy momentum tensor to each of the two branes, and allowing deviations from the pure AdS$_5$ bulk, the effective (non-local) Einstein equations on the branes at low energies take the form [10],

$$\begin{aligned} G^\mu{}_\nu(h) &= \frac{\kappa^2}{\ell}\overset{\oplus}{T}{}^\mu{}_\nu - \frac{2}{\ell}\chi^\mu{}_\nu \ , && (14) \\ G^\mu{}_\nu(f) &= -\frac{\kappa^2}{\ell}\overset{\ominus}{T}{}^\mu{}_\nu - \frac{2}{\ell}\frac{\chi^\mu{}_\nu}{\Omega^4} \ . && (15) \end{aligned}$$

where $h_{\mu\nu} = g^{\oplus\text{-brane}}_{\mu\nu}$, $f_{\mu\nu} = g^{\ominus\text{-brane}}_{\mu\nu} = \Omega^2 h_{\mu\nu}$ and the terms proportional to $\chi_{\mu\nu}$ are 5-dimensional Weyl tensor contributions which describe the non-local 5-dimensional effect. Although equations (14) and (15) are non-local individually, with undetermined $\chi_{\mu\nu}$, we can combine both equations to reduce them to local equations for each brane. Since $\chi_{\mu\nu}$ appears only algebraically, we can easily eliminate $\chi_{\mu\nu}$ from equations (14) and (15). Defining a new field $\Psi = 1 - \Omega^2$, we find

$$\begin{aligned} G^\mu{}_\nu(h) = {} & \frac{\kappa^2}{\ell\Psi}\overset{\oplus}{T}{}^\mu{}_\nu + \frac{\kappa^2(1-\Psi)^2}{\ell\Psi}\overset{\ominus}{T}{}^\mu{}_\nu \\ & + \frac{1}{\Psi}\left(\Psi^{|\mu}{}_{|\nu} - \delta^\mu_\nu \Psi^{|\alpha}{}_{|\alpha} \right) + \frac{3}{2\Psi(1-\Psi)}\left(\Psi^{|\mu}\Psi_{|\nu} - \frac{1}{2}\delta^\mu_\nu \Psi^{|\alpha}\Psi_{|\alpha} \right) \ , \end{aligned} \tag{16}$$

where $|$ denotes the covariant derivative with respect to the metric $h_{\mu\nu}$. Since Ω (or equivalently Ψ) contains the information of the distance between the two branes, we call Ω (or Ψ) the radion. We can also determine $\chi^\mu{}_\nu$ by eliminating $G^\mu{}_\nu$ from equations (14) and (15). Then,

$$\begin{aligned} \chi^\mu{}_\nu = {} & -\frac{\kappa^2(1-\Psi)}{2\Psi}\left(\overset{\oplus}{T}{}^\mu{}_\nu + (1-\Psi)\overset{\ominus}{T}{}^\mu{}_\nu \right) \\ & -\frac{\ell}{2\Psi}\left[\left(\Psi^{|\mu}{}_{|\nu} - \delta^\mu_\nu \Psi^{|\alpha}{}_{|\alpha} \right) + \frac{3}{2(1-\Psi)}\left(\Psi^{|\mu}\Psi_{|\nu} - \frac{1}{2}\delta^\mu_\nu \Psi^{|\alpha}\Psi_{|\alpha} \right) \right] \end{aligned} \tag{17}$$

Note that the index of $\overset{\ominus}{T}{}^\mu{}_\nu$ is to be raised or lowered by the induced metric on the $\ominus$-brane, $f_{\mu\nu}$. The condition $\chi^\mu{}_\mu = 0$ yields

$$\Box\Psi = \frac{\kappa^2}{3\ell}(1-\Psi)\left\{ \overset{\oplus}{T} + (1-\Psi)\overset{\ominus}{T} \right\} - \frac{1}{2(1-\Psi)}\Psi^{|\mu}\Psi_{|\mu} \ . \tag{18}$$

The effective action for the $\oplus$-brane which gives equations (16) and (18) is

$$S_\oplus = \frac{\ell}{2\kappa^2}\int d^4x\sqrt{-h}\left[\Psi R - \frac{3}{2(1-\Psi)}\Psi^{|\alpha}\Psi_{|\alpha}\right] + \int d^4x\sqrt{-h}\mathcal{L}^\oplus + \int d^4x\sqrt{-h}\,(1-\Psi)^2\,\mathcal{L}^\ominus\,. \tag{19}$$

We now extend the analysis to the two-brane system with the bulk scalar field ϕ coupled to the brane tension $\sigma(\phi)$ but not to the matter $\mathcal{L}_{\rm matter}$ on the brane. The action reads

$$S = \frac{1}{2\kappa^2}\int d^5x\sqrt{-g}\mathcal{R} - \int d^5x\left[\frac{1}{2}g^{AB}\partial_A\phi\partial_B\phi + U(\phi)\right] - \sum_{i=\oplus,\ominus}\int d^4x\sqrt{-g^{i-\text{brane}}}\mu_i(\phi) + \sum_{i=\oplus,\ominus}\int d^4x\sqrt{-g^{i-\text{brane}}}\mathcal{L}^i_{\rm matter}, \tag{20}$$

where κ^2, $\mathcal{R}$ and $h_{\mu\nu}$ are the gravitational constant, the scalar curvature in the 5-dimensional space constructed from the metric g_{AB} and the induced metric on the brane, respectively. We assume the potential $U(\phi)$ for the bulk scalar field takes the form

$$U(\phi) = -\frac{6}{\kappa^2\ell^2} + V(\phi)\,, \tag{21}$$

where the first term is regarded as a 5-dimensional cosmological constant and the second term is an arbitrary potential function. The brane tension is also assumed to take the form $\mu_i(\phi) = \mu_{i0} + \tilde{\mu}_i(\phi)$. The constant part of the brane tension, μ_{i0} is tuned so that the effective cosmological constant on the brane vanishes. The above setup realizes a flat braneworld after inflation ends and the field ϕ reaches the minimum of its potential.

Most interesting phenomena occur at low energy in the sense that the additional energy due to the bulk scalar field is small, $\kappa^2\ell^2V(\phi) \ll 1$, and the curvature on the brane R is also small, $R\ell^2 \ll 1$. Here, again, we can use the gradient expansion method explained above. The effective action can be deduced as [6]

$$S = \frac{\ell}{2\kappa^2}\int d^4x\sqrt{-h}\left[\Psi R - \frac{3}{2(1-\Psi)}\partial^\alpha\Psi\partial_\alpha\Psi - \kappa^2\Psi\left(\partial^\alpha\eta\partial_\alpha\eta + 2V_{\rm eff}\right)\right] + \int d^4x\sqrt{-h}\,\overset{\oplus}{\mathcal{L}} + \int d^4x\sqrt{-h}\,(1-\Psi)^2\overset{\ominus}{\mathcal{L}}\,, \tag{22}$$

where the boundary field $\eta = \phi(y=0,x)$ is defined and the effective potential takes the form

$$V_{\rm eff} = \frac{1}{\ell}\left[\frac{1}{\Psi}\frac{d\tilde{\mu}_\oplus}{d\eta} + \frac{(1-\Psi)^2}{\Psi}\frac{d\tilde{\mu}_\ominus}{d\eta}\right] + \frac{2-\Psi}{2}V(\eta)\,. \tag{23}$$

As this is a closed system, we can analyze a primordial spectrum to predict the cosmic background fluctuation spectrum.

In the Einstein frame $g_{\mu\nu} = \Psi h_{\mu\nu}$ with $\Psi = 1/\cosh^2\frac{\kappa}{\sqrt{6\ell}}\rho$, the above action becomes

$$S = \frac{\ell}{2\kappa^2}\int d^4x\sqrt{-g}R - \int d^4x\sqrt{-g}\left[\frac{1}{2}\partial^\alpha\varphi\partial_\alpha\varphi + \frac{1}{2}\partial^\alpha\eta\partial_\alpha\eta + V^{\rm E}_{\rm eff}\right] + \int d^4x\sqrt{-g}\cosh^4\frac{\kappa}{\sqrt{6\ell}}\varphi\,\overset{\oplus}{\mathcal{L}} + \int d^4x\sqrt{-g}\sinh^4\frac{\kappa}{\sqrt{6\ell}}\varphi\,\overset{\ominus}{\mathcal{L}}\,, \tag{24}$$

where

$$V^{\rm E}_{\rm eff} = \frac{d\tilde{\mu}_\oplus}{d\eta}\cosh^4\frac{\kappa}{\sqrt{6\ell}}\varphi + \frac{d\tilde{\mu}_\ominus}{d\eta}\sinh^4\frac{\kappa}{\sqrt{6\ell}}\varphi + \ell\left(\cosh^2\frac{\kappa}{\sqrt{6\ell}}\varphi - \frac{1}{2}\right)V(\eta)\,. \tag{25}$$

5 Cosmological Evolution

In order to study the cosmological evolution of the background spacetime and the fluctuations, we need to analyze the following action

$$S = \frac{1}{2\kappa^2}\int d^4x\sqrt{-g}R + \int d^4x\sqrt{-g}\left[-\frac{1}{2}\partial^\mu\varphi\partial_\mu\varphi - \frac{1}{2}e^{2b(\varphi)}\partial^\mu\chi\partial_\mu\chi - V(\varphi,\chi)\right], \tag{26}$$

where e^b is chosen as 1 for the bulk inflation model and $\cosh^2\frac{\kappa}{\sqrt{6\ell}}\varphi$ for the brane inflation model. Taking the metric

$$ds^2 = a^2(\tau)\left[-d\tau^2 + \delta_{ij}dx^idx^j\right]\,, \tag{27}$$

we have the equations of motion for the background fields

$$\varphi'' + 2\mathcal{H}\varphi' + a^2V_\varphi = b_\varphi e^{2b}\chi'^2 \tag{28}$$

$$\chi'' + 2\left(\mathcal{H} + b_\varphi\varphi'\right)\chi' + e^{-2b}a^2V_\chi = 0 \tag{29}$$

$$\mathcal{H}^2 = \frac{\kappa^2}{3}\left[\frac{1}{2}\varphi'^2 + \frac{1}{2}e^{2b}\chi'^2 + a^2V\right] \tag{30}$$

$$\mathcal{H}' - \mathcal{H}^2 = -\frac{\kappa^2}{2}\left[\varphi'^2 + e^{2b}\chi'^2\right]\,, \tag{31}$$

where $\mathcal{H} = a'/a$ and the prime represents the derivative with respect to the conformal time τ. Here, the suffix of the functions b, V mean the derivatives. The trajectory of the classical solution is parametrized by the proper length σ in the superspace (φ,χ), namely

$$\sigma' = \cos\theta\varphi' + \sin\theta e^b\chi'\,. \tag{32}$$

Here, θ is defined by

$$\cos\theta = \frac{\varphi'}{\sqrt{\varphi'^2 + e^{2b}\chi'^2}}\,,\qquad \sin\theta = \frac{e^b\chi'}{\sqrt{\varphi'^2 + e^{2b}\chi'^2}}\,. \tag{33}$$

In terms of this variable, we obtain

$$\sigma'' + 2\mathcal{H}\sigma' + a^2V_\sigma = 0 \tag{34}$$

where $V_\sigma = \cos\theta V_\varphi + e^{-b}\sin\theta V_\chi$. This is nothing but the equations of motion for the single scalar field if V_σ depends only on σ. Thus, the perturbation of the field σ must be adiabatic and that of the orthogonal field s must be the entropy perturbation. Note that the following formula is useful for later calculation:

$$\theta' = -\frac{a^2 V_s}{\sigma'} - \sigma' b_\varphi \sin\theta \ , \tag{35}$$

where $V_s = e^{-b}\cos\theta V_\varphi - \sin\theta V_\chi$.

6 Cosmological Perturbations

Now we consider the cosmological perturbations. As the action must be written by the gauge invariant quantities, we calculate the action in a completely gauge fixed manner. It is convenient to start with the longitudinal gauge

$$ds^2 = a^2(\eta)\left[-(1+2A)d\eta^2 + (1-2\psi)\delta_{ij}dx^i dx^j\right] \ . \tag{36}$$

We also consider the fluctuation of the scalar fields

$$\varphi \to \varphi + \delta\varphi \ , \quad \chi \to \chi + \delta\chi \ . \tag{37}$$

By the simple substitution of the metric and the scalar fields into the action (26), we obtain the second order action

$$\begin{aligned} S \ = \ & \frac{1}{2\kappa^2}\int d^4x a^2 \Big[-6\psi'^2 - 4A_{|i}\psi^{|i} + 2\psi_{|i}\psi^{|i} \\ & \quad +6\mathcal{H}(A+\psi)(A'-\psi') + 3(\mathcal{H}-\mathcal{H}^2)(A+\psi)^2\Big] \\ & + \int d^4x a^2 \Big[\frac{1}{2}\left(\delta\varphi'^2 - \delta\varphi_{|i}\delta\varphi^{|i}\right) + \frac{1}{2}e^{2b}\left(\delta\chi'^2 - \delta\chi_{|i}\delta\chi^{|i}\right) \\ & \quad -\frac{a^2}{2}\left(V_{\varphi\varphi}\delta\varphi^2 + 2V_{\varphi\chi}\delta\varphi\delta\chi + V_{\chi\chi}\delta\chi^2\right) - (A+3\psi)\left(\varphi'\delta\varphi' + e^{2b}\chi'\delta\chi'\right) \\ & \quad -(A-3\psi)\, a^2\left(V_\varphi\delta\varphi + V_\chi\delta\chi\right) + A\,(A+3\psi)\left(\varphi'^2 + e^{2b}\chi'^2\right) \\ & \quad +\left(\frac{1}{2}\varphi'^2 + \frac{1}{2}e^{2b}\chi'^2\right)\left(-\frac{1}{2}A^2 - 3A\psi + \frac{3}{2}\psi^2\right) + 2b_\varphi e^{2b}\delta\varphi\chi'\delta\chi' \\ & \quad -b_\varphi e^{2b}\chi'^2\delta\varphi\,(A+3\psi) + \frac{1}{2}e^{2b}\chi'^2\left(b_{\varphi\varphi} + 2b_\varphi^2\right)\delta\varphi^2\Big] \ . \end{aligned} \tag{38}$$

To simplify the action, we introduce the adiabatic and the entropy perturbations defined respectively as

$$\delta\sigma = \cos\theta\delta\varphi + e^b\sin\theta\delta\chi \ , \quad \delta s = e^b\cos\theta\delta\chi - \sin\theta\delta\varphi \tag{39}$$

respectively. Using these variables, we can rewrite the action as

$$\begin{aligned} 2\kappa^2 S \ = \ & \int d^4x a^2 \Big[-6\psi'^2 - 4A_{|i}\psi^{|i} + 2\psi_{|i}\psi^{|i} - 12\mathcal{H}A\psi' - 2\left(\mathcal{H}' + 2\mathcal{H}^2\right)A^2 \\ & +\kappa^2\delta\sigma'^2 + \kappa^2\delta s'^2 + \kappa^2\theta'^2\delta s^2 + \kappa^2\theta'^2\delta\sigma^2 \\ & -\kappa^2\left(\delta\sigma_{|i}\delta\sigma^{|i} + \delta s_{|i}\delta s^{|i} + a^2V_{\sigma\sigma}\delta\sigma^2 + a^2V_{ss}\delta s^2\right) - 4\kappa^2a^2V_sA\delta s \\ & + \left\{\kappa^2\sigma'^2\left(1+2\sin^2\theta\right)\cos^2\theta + \kappa^2\sigma'^2b_{\varphi\varphi} + 2\kappa^2\sigma'b_\varphi\cos^2\theta\sin\theta\right. \\ & \left. -\kappa^2b_\varphi\left(1+\sin^2\theta\right)a^2\left(V_\sigma\cos\theta + V_s\sin\theta\right)\right\}\delta s^2 \\ & -2\kappa^2a^2V_\sigma A\delta s - 2\kappa^2\sigma'A\delta\sigma' + 6\kappa^2\sigma'\psi'\delta\sigma + 4\kappa^2\frac{a^2V_s}{\sigma'}\delta\sigma'\delta s \\ & +12\kappa^2\mathcal{H}\frac{a^2V_s}{\sigma'}\delta\sigma\delta s + 4\kappa^2\frac{a^2V_\sigma V_s}{\sigma'^2}\delta\sigma\delta s - \kappa^2\sigma'^2b_\varphi^2\sin^2\theta\delta\sigma^2 \\ & +\kappa^2b_\varphi a^2V_\sigma\sin^2\theta\cos\theta\delta\sigma^2 - \kappa^2b_\varphi\sin^3\theta a^2V_s\delta\sigma^2\Big] \ , \end{aligned} \tag{40}$$

where

$$V_{\sigma\sigma} \ = \ V_\varphi\cos^2\theta + e^{-b}V_{\varphi\chi}\sin 2\theta + V_{\chi\chi}\sin^2\theta \ , \tag{41}$$

$$V_{ss} \ = \ V_\varphi\sin^2\theta - e^{-b}V_{\varphi\chi}\sin 2\theta + V_{\chi\chi}\cos^2\theta \ . \tag{42}$$

However, to simplify the action, it is necessary to move onto the uniform adiabatic gauge. Hence, we perform the gauge transformation

$$A_L = A_U + B'_U + \mathcal{H}B_U \ , \quad \psi_L = \psi_U - \mathcal{H}B_U \ , \quad \delta\sigma_L = \sigma' B_U \ . \tag{43}$$

Notice that the entropy field is gauge invariant. Using the relation

$$V'_\sigma = \theta' V_s - b_\varphi\sigma'\cos\theta\sin\theta e^{-b}V_\chi + \sigma' V_{\sigma\sigma} \tag{44}$$

we get

$$\begin{aligned} 2\kappa^2 S \ = \ & \int d^4x a^2 \Big[-6\psi'^2 - 12\mathcal{H}A\psi' - 2\left(\mathcal{H} + 2\mathcal{H}^2\right)A^2 \\ & -4\kappa^2a^2V_sA\delta s - 2\psi_{|i}\left(2A - \psi\right)^{|i} - 4\triangle B\left(\psi' + \mathcal{H}A\right) \\ & +\kappa^2\delta s'^2 - \kappa^2\delta s_{|i}\delta s^{|i} + \kappa^2\theta'^2\delta s^2 - \kappa^2a^2V_{ss}\delta s^2 \\ & + \left\{\kappa^2\sigma'^2\left(1+2\sin^2\theta\right)\cos^2\theta + \kappa^2\sigma'^2b_{\varphi\varphi} + 2\kappa^2\sigma'b_\varphi\cos^2\theta\sin\theta\right. \\ & \left. -\kappa^2b_\varphi\left(1+\sin^2\theta\right)a^2\left(V_\sigma\cos\theta + V_s\sin\theta\right)\right\}\delta s^2\Big] \ , \end{aligned} \tag{45}$$

where we omitted the suffix for simplicity. Now, we can obtain the reduced action by eliminating the unphysical degrees of freedom. Taking the variation with respect to B, we have

$$A = -\frac{\psi'}{\mathcal{H}} \ . \tag{46}$$

Substituting this into equation (45), we obtain our main result:

$$\begin{aligned} 2\kappa^2 S \ = \ & \int d^4x a^2 \Big[2\left(1 - \frac{\mathcal{H}'}{\mathcal{H}^2}\right)\left(\psi'^2 - \psi_{|i}\psi^{|i}\right) + 4\kappa^2\frac{a^2V_s}{\mathcal{H}}\psi'\delta s \\ & +\kappa^2\delta s'^2 - \kappa^2\delta s_{|i}\delta s^{|i} - \kappa^2M_s^2\delta s^2\Big] \ , \end{aligned} \tag{47}$$

where

$$\mathcal{M}_s^2 = a^2 V_{ss} - \theta'^2 - \sigma'^2\left(1+2\sin^2\theta\right)\cos^2\theta - \sigma'^2 b_{\varphi\varphi} - 2\sigma' b_\varphi \cos^2\theta\sin\theta + b_\varphi\left(1+\sin^2\theta\right)a^2\left(V_\sigma\cos\theta + V_s\sin\theta\right) \ . \qquad (48)$$

Defining the new variable

$$\mathcal{U} = z\psi \ , \quad z = \frac{a\sigma'}{\mathcal{H}} \ , \qquad (49)$$

we finally obtain the canonically normalized action

$$S = \int d^4x \left[\frac{1}{2}\left(\mathcal{U}'^2 - \mathcal{U}_{|i}\mathcal{U}^{|i} + \frac{z''}{z}\mathcal{U}^2\right) + 2\frac{a^2 V_s}{\mathcal{H}}\left(\frac{\mathcal{U}}{z}\right)'\delta s + \frac{1}{2}a^2\delta s'^2 - \frac{1}{2}a^2\delta s_{|i}\delta s^{|i} - \frac{1}{2}\mathcal{M}_s^2 a^2\delta s^2\right] \ . \qquad (50)$$

The above action clearly shows the non-trivial coupling between the adiabatic and the entropy perturbations.

7 Born-again Braneworld

In this section, we use our results to present the born-again braneworld scenario [14]. For simplicity, we do not consider the bulk scalar field. When we consider only the vacuum energy on each brane, the effective potential for the radion takes the form depicted in Figure 1. The unstable extrema correspond to the de Sitter two-brane system.

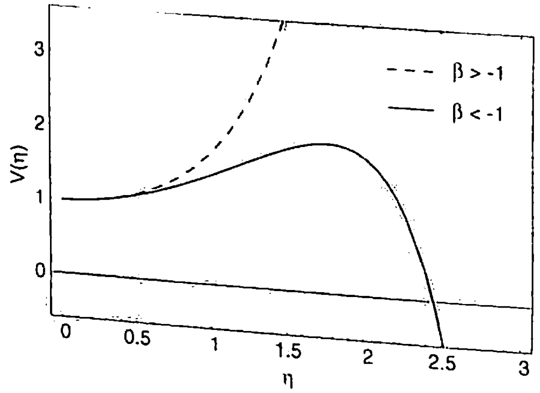


Figure 1: Effective potential for the radion.

Our model has the features of both inflationary and pre-Big-Bang scenarios (see Figure 2). In the original frame, which we call the Jordan frame since gravity on the brane is described by a scalar-tensor type theory, the brane universe is assumed to be inflating due to an inflaton potential. The radion, which represents the distance between the branes and acts as a gravitational scalar on the branes, has non-trivial dynamics and vacuum branes can collide and pass through smoothly. After collision the positive tension and the negative tension branes exchange their role. Then, they move away from each other, and the radion becomes trivial after a sufficient lapse of time. The gravity on the originally negative tension brane (whose tension becomes positive after collision) will then approach the conventional Einstein theory except for tiny Kaluza–Klein corrections.

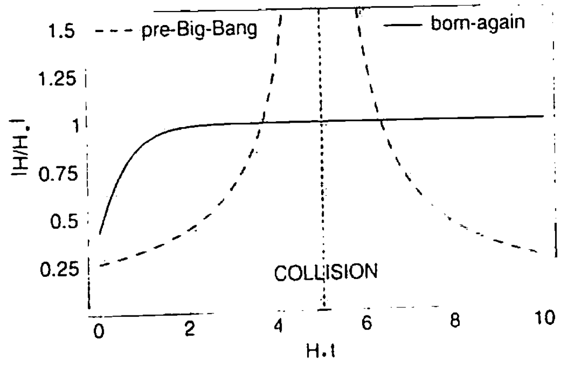


Figure 2: The evolution of the Hubble constant in the Jordan frame. The de-Sitter spacetime is an attractor of the solution. We also plotted the pre-Big-Bang solution in the Einstein frame.

We can also consider the cosmological evolution of the branes in the Einstein frame. Note that the two frames are indistinguishable at present if our universe is on the positive tension brane after collision. In the Einstein frame, the brane universe is contracting before the collision and we encounter a singularity at the collision point. This resembles the pre-Big-Bang scenario. Thus our scenario may be regarded as a non-singular realization of the pre-Big-Bang scenario in the braneworld context.

7.1 Gravitational waves

In this section, we shall discuss the observational consequences of our scenario. The universe rapidly converges to the de-Sitter regime in the Jordan frame. As the metric couples with the radion, the non-trivial evolution of the radion field affects the tensor perturbations. This possibility discriminates our model from the usual inflationary scenario. Let us consider the tensor perturbations in the Einstein frame:

$$ds^2 = \alpha^2(\tau)\left[-d\tau^2 + (\delta_{ij} + h_{ij})dx^i dx^j\right] \qquad (51)$$

where α is the scale factor and h_{ij} satisfy the transverse-traceless conditions, $h^j_{ij} = h^i_i = 0$. As for the gravitational tensor perturbations, we have

$$h_k'' + 2\mathcal{H}h_k' + k^2 h_k = 0 \ , \qquad (52)$$

where h_k is the amplitude of h_{ij} and $\mathcal{H} = \alpha'/\alpha$. Since $\mathcal{H} \sim (2\tau)^{-1}$ near the collision point, h_k has the spectral index $n = 4$ for the gravitational waves (with the spectral index defined by $P_h(k) \propto k^{n-4}$ as in the case of scalar curvature perturbation; for the tensor perturbation, the conventional definition is $n_T = n - 1$). Provided that inflation ends right after collision, this gives a sufficiently blue spectrum that can amplify the density parameter Ω_g by several orders of magnitude or more, on small scales as compared to conventional inflation models. Thus, there arises a possibility that it may be detected by a space laser interferometer for low frequency gravitational waves such as LISA.

7.2 Inflaton perturbation

On the other hand, the inflaton does not couple directly with the radion field. Hence, the inflaton fluctuations are expected to give adiabatic fluctuations with a flat spectrum. To calculate the inflaton perturbation rigorously, we need to introduce an inflaton field explicitly and consider a system of equations fully coupled with the radion and the metric perturbation. It can be demonstrated [14] that the inflaton fluctuations are not much affected by the radion fluctuations [14]. Thus, the inflaton fluctuations will have a standard scale-invariant spectrum.

8 Conclusion

The general formalism for studying an inflationary scenario in the two-brane system is presented. In this approach, the gradient expansion method plays a central role. In particular, the second order action for the cosmological perturbations has been derived. This result should be useful when we discuss the quantization of the cosmological fluctuations. As an application, we have proposed a born-again braneworld scenario. As our braneworld is inflating and the inflaton has essentially no coupling with the radion field, adiabatic density perturbations with a flat spectrum are naturally realized. As the collision of branes mimics the pre-Big-Bang scenario [2], primordial background gravitational waves with a very blue spectrum may be produced. This implies that we may be able to see the collision epoch by a future gravitational wave experiment such as LISA.

Acknowledgements

This work was supported in part by Grant-in-Aid for Scientific Research Fund of the Ministry of Education, Science and Culture of Japan No. 155476 (SK) and also by a Grant-in-Aid for the 21st Century COE "Center for Diversity and Universality in Physics".

References

[1] L. Randall, R. Sundrum, *Phys. Rev. Lett.*, **83**, 3370 (1999).

[2] M. Gasperini, G. Veneziano, *Astropart. Phys.*, **1**, 317 (1993).

[3] R. Maartens, D. Wands, B.A. Bassett, I. Heard, *Phys. Rev.*, **D62**, 041301 (2000).

[4] S. Kobayashi, K. Koyama, J. Soda, *Phys. Lett.*, **B501**, 157 (2001).

[5] Y. Himemoto, M. Sasaki, *Phys. Rev.*, **D63**, 044015 (2001).

[6] S. Kanno, J. Soda, *Phys. Rev.*, **D66**, 043526 (2002).

[7] S. Kanno, J. Soda, *Phys. Rev.*, **D66**, 083506 (2002).

[8] S. Kanno, J. Soda, *Gen. Rel. Grav.*, **36**, 689 (2004).

[9] S. Kanno, J. Soda, *Phys. Lett.*, **B588**, 203 (2004).

[10] S. Kanno, J. Soda, *Astrophys. Space Sci.*, **283**, 481 (2003).

[11] J. Soda, S. Kanno, *Astrophys. Space Sci.*, **283**, 639 (2003).

[12] T. Wiseman, *Class. Quant. Grav.*, **19**, 3083 (2002).

[13] T. Shiromizu, K. Koyama, *Phys. Rev.*, **D67**, 084022 (2003).

[14] S. Kanno, M. Sasaki, J. Soda, *Prog. Theor. Phys.*, **109**, 357 (2003).

11

Quantized Black Holes, Correspondence Principle and Holographic Bound

I.B. Khriplovich

Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia,
and Novosibirsk University, Russia
e-mail: khriplovich@inp.nsk.su

Abstract

An equidistant spectrum of the horizon area of a quantized black hole does not follow from the correspondence principle or from general statistical arguments. Such a spectrum obtained earlier in loop quantum gravity (LQG) does not comply with the holographic bound. This bound fixes the Barbero–Immirzi parameter of LQG, and thus leads to a unique result for the spectrum of the horizon area.

Keywords: quantum gravity, horizon area, holographic bound

1 Is the Spectrum of a Quantized Black Hole Equidistant?

The idea of quantizing the horizon area of black holes was put forward many years ago by Bekenstein in a pioneering article [1]. It was based on the intriguing observation, made by Christodoulou and Ruffini [2, 3] that the horizon area of a nonextremal black hole behaves as an adiabatic invariant. Of course, the quantization of an adiabatic invariant is perfectly natural, in accordance with the correspondence principle.

One more conjecture made in [1] is that the spectrum of a quantized horizon area is equidistant. The argument therein was that a periodic system is quantized by equating its adiabatic invariant to $2\pi\hbar n$, $n = 0, 1, 2, , \dots$.

It was pointed out later by Bekenstein [4] that the classical adiabatic invariance does not guarantee by itself the equidistance of the spectrum, at least because any function of an adiabatic invariant is itself an adiabatic invariant. However, up to now articles on the subject abound in assertions that the form

$$A = \beta\, l_p^2 n, \quad n = 1,\, 2,\, \ldots \tag{1}$$

for the horizon area spectrum[1] is dictated by the respectable correspondence principle. The list of such references is too lengthy to be presented here.

Let us consider an instructive example of the situation when a nonequidistant spectrum arises in spite of the classical adiabatic invariance. We start with a classical spherical top of angular momentum **J**. Of course, the projection J_z of **J** is an adiabatic invariant. If the z-axis is chosen along **J**, the value of J_z is maximum, J, or $\hbar j$ in the quantum case. The classical angular momentum squared (J^2) is also an adiabatic invariant, with eigenvalues $\hbar^2 j(j+1)$ when quantized. Let us try now to use the operator $\hat{j}^2$ for the area quantization in quite natural units of l_p^2. For the horizon area A to be finite in the classical limit, the power of the quantum number j (for $j \gg 1$) should be the same as that of $\hbar$ in l_p^2 [5]. For $l_p^2 \sim \hbar$, we obtain

$$A \sim l_p^2 \sqrt{j(j+1)}. \tag{2}$$

Since $\sqrt{j(j+1)} \to j+1/2$ for $j \gg 1$, we have come back to the equidistant spectrum in the classical limit. However, the equidistance can be avoided in the following way. Let us assume that the horizon area consists of sites with area of the order of l_p^2, and ascribe to each site i its own quantum number j_i with contribution to the area equals $\sqrt{j_i(j_i+1)}$. Then eq. (2) changes to

$$A \sim l_p^2 \sum_i \sqrt{j_i(j_i+1)} \tag{3}$$

(in fact, this formula for a quantized area arises as a special case in loop quantum gravity, see below). Of course, to retain a finite classical limit for A, we should require that $\sum_i \sqrt{j_i(j_i+1)} \gg 1$. However, any of the j_i can be comparable to unity. So, in spite of the adiabatic invariance of A, its quantum spectrum (3) is not equidistant, though it is of course discrete.

One more argument in favour of the equidistant spectrum (1) has been presented in [4, 6]. On the one hand, the entropy S of a horizon is related to its area A through the Bekenstein–Hawking relation

$$S = \frac{A}{4l_p^2}\,. \tag{4}$$

On the other hand, the entropy is nothing but $\ln g(n)$ where the statistical weight $g(n)$ of any state n is an integer. In [4, 6] the requirement of integer $g(n)$ is taken literally, and results after simple reasoning not only in an equidistant spectrum (1) but also in the following allowed values for the numerical factor β in this spectrum:

$$\beta = 4 \ln k, \quad k = 2,\, 3,\, \ldots.$$

---

[1] Here and below $l_p^2 = \hbar k/c^3$ is the Planck length squared, $l_p = 1.6 \cdot 10^{-33}$ cm, k is the Newton gravitational constant; β is a numerical factor.

Let us imagine however that with some model for S we obtain for $g(n)$, instead of an integer value K, a noninteger one $K+\delta$, $0 < \delta < 1$. Then, the entropy will be

$$S = \ln(K+\delta) = \ln K + \delta/K.$$

Now, the typical value for the black hole entropy $S = \ln K = A/4l_p^2$ is huge, something like 10^{76}. So, the correction δ/K is absolutely negligible as compared to $S = \ln K$. Moreover, it is far below any conceivable accuracy of a description of entropy in general, and of the semiclassical Bekenstein–Hawking formula (4) which relates the black hole entropy to the quantized horizon area in particular. Therefore, this correction can be safely omitted and forgotten. As it is usual for macroscopic objects, the fact that $g(n)$ is an integer has no consequences for entropy.

Thus, contrary to the popular belief, the equidistance of the spectrum for the horizon area does not follow from the correspondence principle and/or from general statistical arguments [7].

2 “It from Bit”, Loop Quantum Gravity and Holographic Bound

What we discussed above does not mean however that any model leading to an equidistant spectrum for the quantized horizon area should be automatically abandoned. A rather simple and elegant version of such a model, so called “it from bit”, for a Schwarzschild black hole was formulated by Wheeler [8]. The assumption is that the horizon surface consists of ν sites, each of them supplied with an “angular momentum” quantum number j with two possible projections $\pm 1/2$. The total number K of degenerate quantum states of this system is obviously

$$K = 2^{\nu}. \tag{5}$$

Then the entropy of the black hole is

$$S_{1/2} = \ln K = \nu \ln 2. \tag{6}$$

And finally, with the Bekenstein–Hawking relation (4) we obtain for the area spectrum the following equidistant formula:

$$A_{1/2} = 4 \ln 2\; l_p^2\, \nu. \tag{7}$$

This model of the quantized Schwarzschild black hole looks quite right by itself.

This result was also obtained in [9] in the framework of loop quantum gravity (LQG) [11–15]. We discuss below whether the “it from bit” picture, if considered as a special case of the area quantization in LQG, can be reconciled with the holographic bound [16–18].

More generally, a quantized surface in LQG is described as follows. We ascribe to it a set of punctures. Each puncture is supplied with two integer or half-integer “angular momenta” j^u and j^d:

$$j^u,\, j^d = 0, 1/2, 1, 3/2, \ldots\,; \tag{8}$$

at least one of them should not vanish. j^u and j^d are related to edges directed up and down the normal to the surface, respectively, and add up into an angular momentum j^{ud}:

$$\mathbf{j}^{ud} = \mathbf{j}^u + \mathbf{j}^d; \quad |j^u - j^d| \le j^{ud} \le j^u + j^d. \tag{9}$$

The area of a surface is

$$A = 8\pi\gamma\, l_p^2 \sum_i \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}\,. \tag{10}$$

The numerical factor γ in (10) cannot be determined without an additional physical input. This ambiguity originates from a free (so called Barbero–Immirzi) parameter [17, 18] which corresponds to a family of inequivalent quantum theories, all of them being viable without such an input. Once the general structure of the horizon area in LQG is fixed, the only problem left is to determine this overall factor, i. e., the Barbero–Immirzi parameter.

To make the discussion more clear and concrete, we restrict our discussion from now on to the simplified version (quite popular) of the general equation (10) when $j^{d(u)} = 0$, so that this formula for a surface area reduces to

$$A = 8\pi\gamma\, l_p^2 \sum_i \sqrt{j_i(j_i+1)}\,, \quad j = j^{u(d)}. \tag{11}$$

It is worth mentioning that this particular case of eq. (10) coincides with the naive model (3).

The result (7) was obtained in [9] under an additional condition that the gravitational field on the horizon is described by the $U(1)$ Chern–Simons theory[2]. Formula (7) is a special case of the general one (11) when all $j = 1/2$. As to the overall factor γ, its value in this case is

$$\gamma = \frac{\ln 2}{\pi\sqrt{3}}. \tag{12}$$

Let us turn now to the holographic bound [16–18]. According to it, the entropy S of any spherical nonrotating system confined inside a sphere of area A is bound as follows:

$$S \le \frac{A}{4l_p^2}, \tag{13}$$

with the equality attained only for a system that is a black hole.

A simple intuitive argument confirming this bound was formulated in [16] and goes as follows. Let us allow the discussed system to collapse into a black hole. During the collapse the entropy increases from S to S_{bh}, and the resulting horizon area A_{bh} is certainly smaller than the initial confining one, A. Now, taking into account the Bekenstein-Hawking relation (4) for a black hole, we arrive through the obvious chain of (in)equalities to

$$S \le S_{bh} = \frac{A_{bh}}{4l_p^2} \le \frac{A}{4l_p^2}$$

---

[2] Earlier attempts to calculate the black hole entropy in LQG go back at least to [19–22]. They did not lead however to concrete quantitative results.

at the discussed bound (13).

The result (13) can be formulated differently: among the spherical surfaces of a given area, it is the surface of a black hole horizon that has the largest entropy.

The last statement was used as an assumption by Vaz and Witten [23] in a model of the quantum black hole as originating from dust collapse. Then this assumption was adopted by us [24] in the problem of quantizing the horizon of a black hole in LQG. In particular, the coefficient γ was calculated in [24] in the case when the area of a black hole horizon is given by the general formula (10) of LQG, as well as under some more special assumptions on the values of j^u, j^d, j^{ud}. Moreover, it was demonstrated in [24] for a rather general class of the horizon quantization schemes that it is the maximum entropy of a quantized surface that is proportional to its area.

3 Problem of Distinguishability of Edges

We consider in fact the "microcanonical" entropy S of a quantized surface defined as the logarithm of the number of states of this surface for a fixed area A (instead of the fixed energy in common problems). Obviously, this number of states K depends on the assumption made on the distinguishability of the edges. So, let us discuss first of all which of *a priori* possible assumptions is reasonable from the physical point of view [7].

It is instructive to start the discussion with a somewhat more detailed analysis of the "it from bit" model. The natural result (5) for the total number of states in this model corresponds in fact to the assumption that among all the sites with $j = 1/2$, those with the same j_z, $+1/2$ or $-1/2$ (their numbers are ν_+ and ν_-, respectively), are indistinguishable, and those with different projections are distinguishable. Indeed, in this case the number of states with given ν_+, ν_- is obviously $\nu!/(\nu_+!\,\nu_-!)$, and the total number of states and entropy:

$$K = \sum_{\nu_+,\,\nu_-} \frac{\nu!}{\nu_+!\,\nu_-!} = \sum_{\nu_+=0}^{\nu} \frac{\nu!}{\nu_+!\,(\nu-\nu_+)!} = 2^\nu \quad \text{and} \quad S = \nu \ln 2\,, \tag{14}$$

coincides with eqs. (5) and (6).

It should be noted that the configuration with $\nu_+ = \nu_- = \nu/2$ not only gives here the maximum contribution to K and S but for $\nu \gg 1$ both the "it from bit" statistical weight and entropy are strongly dominated by this configuration. Indeed,

$$\ln\frac{\nu!}{(\nu/2!)^2} = \nu\ln 2 - \frac{1}{2}\ln\nu. \tag{15}$$

Let us also note in passing that if we confine our study to the states with total angular momentum $J = 0$, the entropy is still about the same:

$$\nu\ln 2 - \frac{3}{2}\ln\nu.$$

Let us analyse now the more general case with various j's. One possibility that might look quite appealing here, is that of complete indistinguishability of edges. It means that no permutation of edges results in new states. To analyze this situation, let us rewrite eq. (11) as follows:

$$A = 8\pi\gamma\, l_p^2 \sum_{jm} \sqrt{j(j+1)}\; \nu_{jm}\,. \tag{16}$$

Here the projections m of an angular momentum j run as usual from $-j$ to j; ν_{jm} is the number of edges with given j and m. Under the assumption of complete indistinguishability, the total number of angular momentum states created by $\nu_j = \sum_m \nu_{jm}$ edges of a given j with all $2j+1$ projections allowed, is[3]

$$K_j = \frac{(\nu_j + 2j)!}{\nu_j!\,(2j)!}\,. \tag{17}$$

The partial contributions ($s_j = \ln K_j$) to the black hole entropy ($S = \sum_j s_j$) that can potentially dominate the numerically large entropy, may correspond to the three cases: $j \ll \nu_j$, $j \gg \nu_j$, and $j \sim \nu_j \gg 1$. These contributions are:

$$j \ll \nu_j\,, \qquad s_j \approx 2j \ln \nu_j\,;$$

$$j \gg \nu_j\,, \qquad s_j \approx \nu_j \ln j;$$

$$j \sim \nu_j \gg 1, \qquad s_j \sim 4j \ln 2.$$

In all three cases the partial contributions to the entropy S are much smaller parametrically than the corresponding contributions $a_j \sim j\nu_j$ to the area $A = \sum_j a_j$. Thus, in all these cases $S \ll A$, so that with indistinguishable edges of the same j, we cannot make the entropy of a black hole proportional to its area [26].

Let us consider now the opposite assumption, that of completely distinguishable edges. In this case the total number of states is $K = \nu\,!$, with the microcanonical entropy $S = \nu \ln \nu$. In principle, this entropy can be made proportional to the black hole area A. The model (though not looking natural) could be as follows. Let us choose a large quantum number $J \gg 1$, and assume that the horizon area A is saturated by the edges with j in the interval $J < j < 2J$, and with "occupation numbers" $\nu_j \sim \ln J$. Then the estimates both for S and A are $\sim J \ln J$, and the proportionality between the entropy and the area can be attained.

However, though under the assumption of complete distinguishability the entropy can be proportional to the area, the *maximum* entropy for a given area is much larger than the area itself. Obviously, here the maximum entropy for fixed $A \sim \sum_j \sqrt{j(j+1)}\;\nu_j$ is attained with all j's being as small as possible, say, 1/2 or 1. Then, in the classical limit $\nu \gg 1$, the entropy of a black hole grows faster than its area: while $A \sim \nu$, $S = \nu \ln \nu \sim A \ln A$. Thus, the assumption of complete distinguishability is in conflict with the holographic bound, and therefore should be discarded[4].

Let us consider a third conceivable possibility, which is quite popular (see for instance [26]). Here the assumption is that the total number of states is

$$K = \prod_j (2j+1)^{\nu_j}, \tag{18}$$

with the entropy of the horizon surface given by

$$S = \sum_j \nu_j \ln(2j+1). \tag{19}$$

Obviously, in this case the maximum entropy $S_{max} \sim A$ is reached with the smallest possible value of j for each edge. Thus, we arrive here effectively to the "it from bit" picture which may look attractive.

However, it can be easily demonstrated that this scheme corresponds to the following assumption on the distinguishability of edges:

nonequal j, any m $\longrightarrow$ indistinguishable;
equal j, nonequal m $\longrightarrow$ distinguishable;
equal j, equal m $\longrightarrow$ indistinguishable.

Comparison of the first two lines in this table demonstrates that this assumption looks strange and unnatural.

Thus, the only reasonable assumption on the distinguishability of edges that may result in acceptable physical predictions (i.e., may comply both with the Bekenstein–Hawking relation and with the holographic bound) is:

nonequal j, any m $\longrightarrow$ distinguishable;
equal j, nonequal m $\longrightarrow$ distinguishable;
equal j, equal m $\longrightarrow$ indistinguishable.

4 Microcanonical Entropy of Black Holes

Under the last assumption, the number of states of the horizon surface for a given number ν_{jm} of edges with momenta j and their projections $j_z = m$, is obviously

$$K = \nu\,! \prod_{jm} \frac{1}{\nu_{jm}\,!}\,, \qquad \nu = \sum_j \nu_j = \sum_{jm} \nu_{jm}\,, \tag{20}$$

and the corresponding entropy is

$$S = \ln K = \ln(\nu\,!) - \sum_{jm} \ln(\nu_{jm}\,!)\,. \tag{21}$$

---

[3] Perhaps the simplest derivation of this formula is as follows. We are looking here effectively for the number of ways of distributing ν_j identical balls into $2j+1$ boxes. Then, the line of reasoning presented in [25], §54, results in equation (17). I am grateful to V.F. Dmitriev for bringing to my attention that this equation can be derived in this simpleminded way.

[4] There is no disagreement between our conclusion and those of [26, 27, 28]: what is there called complete distinguishability, corresponds to another assumption, the last one in this section.

The structures of the last expression and of formula (16) are so different that in a general case the entropy certainly cannot be proportional to the area. However, this is the case for the maximum entropy in the classical limit.

In the classical limit, with all effective "occupation numbers" being large, $\nu_{jm} \gg 1$, the entropy in the Stirling approximation is

$$S = \sum_{jm} \nu_{jm} \times \ln \left(\sum_{j'm'} \nu_{j'm'} \right) - \sum_{jm} \nu_{jm} \ln \nu_{jm} \,. \tag{22}$$

We calculate its maximum for a fixed area A, i.e., for a fixed sum

$$N = \sum_{jm}^{\infty} \sqrt{j(j+1)}\, \nu_{jm} = \text{const}\,. \tag{23}$$

The problem reduces to the solution of the system of equations

$$\ln \left(\sum_{j'm'} \nu_{j'm'} \right) - \ln \nu_{jm} = \mu \sqrt{j(j+1)}\,, \tag{24}$$

or

$$\nu_j = (2j+1)\, e^{-\mu\sqrt{j(j+1)}}\, \nu\,. \tag{25}$$

Here μ is the Lagrange multiplier for the constraining relation (23). Summing expressions (25) over j, and recalling that $\sum_j \nu_j = \nu$, we obtain:

$$\sum_{j=1/2}^{\infty} (2j+1)\, e^{-\mu\sqrt{j(j+1)}} = 1, \tag{26}$$

with the solution

$$\mu = 1.722. \tag{27}$$

On the other hand, when multiplying eq. (24) by ν_{jm} and summing over jm, with the constraint (23) we obtain the following result for the maximum entropy as a function of given values of N:

$$S_{\max} = 1.722\, N\,. \tag{28}$$

Now, with the Bekenstein–Hawking relation we find the value of the Barbero–Immirzi parameter:

$$\gamma = 0.274. \tag{29}$$

This derivation follows closely [24] (see also [7]).

It should be emphasized that this calculation is not special for LQG, but it may be applied (with obvious modifications) to a more general class of approaches to the quantization of surfaces. The following assumption is really necessary: the surface should consist of sites of different sorts, so that there are ν_i sites of each sort i, with a generalized effective quantum number r_i (here $\sqrt{j(j+1)}$), and a statistical weight g_i (here $2j+1$). Then in the classical limit, the maximum entropy of a surface can be found at least numerically, and it is certainly proportional to the area of the surface.

A few more comments are appropriate here. A nice feature of the obtained picture is that in the case considered, the occupation numbers ν_j have a sort of Boltzmann distribution (see eq. (25)), where the partial contributions $\sqrt{j(j+1)}$ to the horizon area correspond to energies, and μ is the analogue of the inverse temperature.

It is worth mentioning that, by substituting the occupation numbers given by eq. (25) into expression (23), we arrive at the conclusion that both the effective quantum number N, and with it the horizon area A, are proportional to an integer ν. This may create the impression of an equidistant area spectrum for a black hole. However, we should keep in mind that relation (25) for occupation numbers is approximate and it is valid only for $\nu_j \gg 1$ in the leading approximation to the Stirling formula. In fact, all ν_j are integers, and therefore the exact formula (23) does not correspond to an equidistant spectrum.

Finally, let us note that the correction to our simplest version (22) of the Stirling formula results in a correction $\sim \ln^2 A$ to the Bekenstein–Hawking relation (4) [24].

5 Again Loop Quantum Gravity and Holographic Bound

We come back now to the result of [9]. If we assume that the value (12) of the parameter γ is the universal one (i. e., it is not special to black holes, but refers to any quantized spherical surface), then the value (6) is not the maximum one in LQG for a surface of the area (7). This looks quite natural: with the transition from the unique choice made in [9], $j^{u(d)} = 1/2$, $j^{d(u)} = 0$, to a more extended and rich one, the number of quantum states should, generally speaking, increase. Hence its logarithm, which is the entropy of a quantized surface, increases as well.

To prove this statement we rewrite formula (6) as follows:

$$S_{1/2} = \ln 2 \sqrt{\frac{4}{3}}\, N = 0.800\, N\,, \quad N = \sqrt{\frac{3}{4}}\, \nu\,, \tag{30}$$

and consider the value of N, together with the horizon area A, as fixed ones. This is certainly smaller than the above result from eq. (28), $1.722\, N$.

As expected, in the general case with N given by eq. (23) with all values of j_i^u, j_i^d, j_i^{ud} allowed and $g_i = 2j_i^{ud} + 1$, the maximum entropy is even larger [24]

$$S_{\max} = 3.120\, N. \tag{31}$$

The corresponding value of the Barbero–Immirzi parameter in this case is

$$\gamma = 0.497. \tag{32}$$

Thus, there is an obvious conflict between the holographic bound and the result (30) advocated in [9].

We might try to avoid the conflict by assuming that the value (12) for the Immirzi parameter γ is specific for black holes only, while for other quantized surfaces γ is smaller. However, such a way out would be unattractive and unnatural.

Quite recently the results obtained in [9] were also criticized and revised in [29, 30]. It is instructive to compare the starting points of [29, 30] with ours. What both approaches have in common is the LQG expression (11) for the quantized horizon area A. But in other respects the difference between them is drastic.

Our approach is based on the natural and transparent physical requirement that the "occupation numbers" ν_{jm} (see formulae (19)–(22)) should guarantee the maximum black hole entropy S for a given horizon area A. Our additional assumption, formulated in section 3, is also natural and physically sound: only those edges that have the same jm are indistinguishable. Then the problem reduces to the straightforward and simple calculation presented in Section 4.

On the other hand, the approach of [29, 30] is based on the equidistance assumption (see [31, 32]) and formulated in [29] as follows: "the fixed classical area a is quantized in the following way:

$$a = 4\pi\gamma l_p^2\, k\,, \qquad k \in \mathbb{N}\,, \tag{33}$$

where k is arbitrary." However, as has been demonstrated already in Section 1, there are no physical arguments in favor of an equidistant spectrum. Thus, this assumption is in fact an *ad hoc* one.

Nevertheless, the equation derived in [30] for the Barbero–Immirzi parameter, which in our notations has the form

$$2\sum_{j=1/2}^{\infty} e^{-2\pi\gamma\sqrt{j(j+1)}} = 1 \tag{34}$$

is quite close to our equations (26) (μ in (26) is equal to $2\pi\gamma$). But I do not see any reason why the number of states for a given j should be here 2, instead of the usual $2j+1$ [5]. Still, since the sum is strongly dominated by the first term, with $j = 1/2$, the result 0.238 for γ, obtained in [30], is close to ours 0.274.

The conclusions can be summarized as follows.

1. The equidistant result of [9] is not appropriate.
2. The value of the Barbero–Immirzi parameter, $\gamma = 0.274$, and thus the spectrum of a quantized horizon in LQG (with expression (11) for area), are fixed uniquely by the holographic bound, i. e., by the requirement that among the spherical surfaces of a given area, it is the horizon surface that has the maximum entropy.

Acknowledgements

I am grateful to V.F. Dmitriev, V.M. Khatsymovsky, and G.Yu. Ruban for fruitfull discussions. This investigation was supported in part by the Russian Foundation for Basic Research through Grant No. 03-02-17612.

---

[5] The comment on formula (34) in [30] (also rewritten in our notation) is: "... if the number of states for a given spin j was $2j+1$ instead of 2 one would have in (34) $2j+1$ instead of 2 in front of the exponential and the appropriate γ would be equal to 0.273985635...".

References

[1] J.D. Bekenstein, *Lett. Nuovo Cimento*, **11**, 467 (1974).

[2] D. Christodoulu, *Phys. Rev. Lett.*, **25**, 1596 (1970).

[3] D. Christodoulu, R. Ruffini, *Phys. Rev.*, **D4**, 3552 (1971).

[4] J.D. Bekenstein, in T. Piran, R. Ruffini (eds) *Proceedings of the Eighth Marcel Grossmann Meeting on General Relativity*, World Scientific, Singapore, p.92 (1999).

[5] I.B. Khriplovich, *Phys. Lett.*, **B431**, 19 (1998).

[6] J.D. Bekenstein, V.F. Mukhanov, *Phys. Lett.*, **B360**, 7 (1995).

[7] I.B. Khriplovich, *Zh. Eksp. Teor. Fiz.*, **126**, 527 (2004) [*Sov. Phys. JETP*, **99**, 460 (2004)].

[8] J.A. Wheeler, in L. Keldysh, V. Fainberg (eds) *Sakharov Memorial Lectures in Physics*, Moscow, **2**, p.751 (1991).

[9] A. Ashtekar, J. Baez, A. Corichi, K. Krasnov, *Phys. Rev. Lett.*, **80**, 904 (1998).

[10] C. Rovelli, L. Smolin, *Nucl. Phys.*, **B442**, 593 (1995); erratum, ibid. B456, 753 (1995).

[11] A. Ashtekar, J. Lewandowski, *Class. Quantum Grav.*, **14**, 55 (1997).

[12] R. Loll, *Phys. Rev. Lett.*, **75**, 3048 (1995);
R. Loll, *Nucl. Phys.*, **B460**, 143 (1996).

[13] R. De Pietri, C. Rovelli, *Phys. Rev.*, **D54**, 2664 (1996).

[14] S. Frittelli, L. Lehner, C. Rovelli, *Class. Quantum Grav.*, **13**, 2921 (1996).

[15] J. D. Bekenstein, *Phys. Rev.*, **D23**, 287 (1981).

[16] L. Susskind, *J. Math. Phys.*, **36**, 6377 (1995).

[17] G. Immirzi, *Class. Quantum Grav.*, **14**, L177 (1997).

[18] C. Rovelli, T. Thiemann, *Phys. Rev.*, **D57**, 1009 (1998).

[19] C. Rovelli, *Phys. Rev. Lett.*, **77**, 3288 (1996).

[20] K.V. Krasnov, *Phys. Rev.*, **D55**, 3505 (1997).

[21] K.V. Krasnov, *Gen. Rel. Grav.*, **30**, 53 (1998).

[22] K.V. Krasnov, *Class. Quantum Grav.*, **16**, 563 (1999).

[23] C. Vaz, L. Witten, *Phys. Rev.*, **D64**, 084005 (2001).

[24] R.V. Korkin, I.B. Khriplovich, *Zh. Eksp. Teor. Fiz.*, **122**, 1 (2002) [*Sov. Phys. JETP*, **95**, 1 (2002)].

[25] L.D. Landau, E.M. Lifshitz, *Statistical Physics*, Pergamon Press, London, (1958).

[26] A. Alekseev, A.P. Polychronakos, M. Smedback, *Phys. Lett.*, **B574**, 296 (2003).

[27] A.P. Polychronakos, *Phys. Rev.*, **D69**, 044010 (2004).

[28] G. Gour, V. Suneeta, *Class. Quant. Grav.*, **21**, 3405 (2004).

[29] J. Lewandowski, M. Domagala, preprint gr-qc/0407051 (2004).

[30] K. Meissner, preprint gr-qc/0407052 (2004).

[31] A. Ashtekar, A. Corichi, K.V. Krasnov, *Adv.Theor.Math.Phys.*, **3**, 419 (2000).

[32] A. Ashtekar, J.C. Baez, K.V. Krasnov, *Adv.Theor.Math.Phys.*, 4, 1 (2000).

12

Gravitational Wave Experiments and the Baksan Project "OGRAN"

L. Bezrukov[1], S. Popov[2], V. Rudenko[2], A. Serdobolskii[2]
and M. Skvortsov[3]

[1]*Institute for Nuclear Research, Russian Academy of Sciences,
Moscow, Russia
e-mail: bezrukov@inr.ac.ru*
[2]*Sternberg Astronomical Institute, Moscow State University,
Moscow, Russia
e-mail: serg@sai.msu.ru, rvn@sai.msu.ru*
[3]*Institute of Laser Physics, Siberian Branch of the Russian
Academy of Sciences, Novosibirsk, Russia
e-mail: skv@laser.nsc.ru*

Abstract

A brief sketch of the present status of gravitational wave experiments is given. Attention is paid to recent observations with a gravitational detector network. We also present the project OGRAN for a combined optic-interferometrical and acoustical gravitation wave antenna planned for installation into the underground facilities of the Baksan Neutrino Observatory. We describe the general principles of the apparatus, expected sensitivity and current characteristics of the antenna prototype; planned ways for sensitivity improvement are also discussed.

Keywords: gravitational waves, gravitational detectors

1 Introduction

The current status of gravitational wave (GW) experiments is characterized by an international net of GW antennae of two types: optical interferometers on suspended mass-mirrors and cryogenic acoustical resonant bars with SQUID sensors. A "bar net" containing five independent detectors presented by the IGCE (see http://igec.lnl.infn.it) and ROG [1] collaborations is well developed and has a relatively long history of observation. The "interferometer family" composed by LIGO (http://www.ligo.org), VIRGO (http://www.virgo.infn.it), GEO [2] and TAMA [3] recently started the first individual and joint series of observations at intermediate sensitivity; work for improving it is in progress.

Modern cryogenic resonance bar detectors are able to register metric perturbations at the level of 10^{-21} in a narrow bandwidth of the order of a few hertz around the resonance frequency. Large base GW interferometers currently can detect GW-bursts with magnitude $\sim 10^{-19}$ and spectrum width of several hundred hertz. It is believed that in the near future free-mass interferometrical antennae will achieve wideband sensitivity of 10^{-22} which corresponds to the expected rate of events of at least one per day. Similar sensitivity might be reached by "big ball" detectors having a mass of 30–60 ton cooled down to 10 mK [4] with the advantage of increasing the expected rate of events due to the omni-directional antenna pattern. At present this type of detector exists as moderate scale prototypes [4, 5, 6] and funding of the full scale projects is not guaranteed. Thus we could say that there is some "scientific competition" between solid body resonance acoustical detectors on the one hand and wide band free-mass optical interferometers on the other. Despite a visible potential advantage of the interferometers (except for their cost which is ten times higher than that of bar detectors) both types in fact complement each other taking into account the different mechanisms of their interaction with gravitational radiation: generation of acoustical waves for the bar and excitation of mass-mirror's relative displacements for interferometers (in the last case the effect can be described also as variations of the optical wavelength). This difference and also the observation of some anomalous coincident events with bar detectors [7, 8] stimulated a discussion about the comparison of the cross sections of both types of gravitational antennae which started after supernova SN 1987A [9, 10] and was renewed recently [11] after results published by the ROG collaboration.

Joint observations with bars and interferometers at the current level of sensitivity were carried out in the past and continue to take place at present [12].

It has been mentioned (see, e.g., [13]) that the sensitivity threshold in GW searches is much higher than the potential sensitivity of any individual detector due to the "blind character of all sky searching" when coordinates and parameters of a hypothetical source are unknown and we have to run over all conceivable positions and signal templates in the data processing. This increases the "chance probability" of getting a false "noise event" or equivalently decreases the effective registering sensitivity. Although the declared aim of the present antenna net is an operation at the level of 10^{-21}, the typical effective threshold of "bursts sensitivity" from recent observations is still between 10^{-18} and 10^{-19} magnitude of metric perturbation. Such moderate sensitivity allows us to overlap sources located in the neighbourhood of the Galaxy with radius less than 1 Mpc, i.e. this type of GW experiment can be called a "search for rare events" hoping to have a "lucky chance" of nearby "relativistic star catastrophe".

In [14, 15] it was proposed that similar sensitivity might be achieved by a room temperature bar detector with an optical read-out using advanced optical technology (low noise optical frequency standards). The group of the cryogenic antenna AURIGA has developed a micro-gap optical transducer as a low noise alternative to the SQUID read-out [16]. A more sophisticated variant of the optical read-out, an optical FP-resonator extended along the bar body and having the same scale as the bar itself, is analysed in [14]. This construction can be considered as a combination of bar and interferometer unified in one detector and we can expect a more complex response under a GW excitation containing the acoustical and optical parts independently to some extent. This is the idea behind the Russian national project OGRAN (opto-acoustical gravitational antenna).

In what follows we review recent results collected by the existing net of GW antennae and present in detail the current status of the OGRAN project.

2 Results of GW searches

The best filtering procedure in searching for GW pulses (introduced by a pioneer of GW experiments J. Weber [17]) consists of a selection of coincident perturbations of widely spatially separated detectors. It is a straightforward method for effectively decreasing the local noise background. Even more, the coincidences guarantee the extraterrestrial nature of the registered signals. This procedure remains valid, improved by a preliminary reduction of the detection noise by optimal filtering assuming a certain type of the expected signal (see, for example, http://www.ligo-wa.caltech.edu).

The following "signal families" are often considered: a) relativistic binary coalescence, b) supernova explosions (or more general, collapses), c) pulsars, d) primordial stochastic GW background. However lacking "a prior information" (no data of source's position and parameters) the algorithm of matched filtering for such "blind search" becomes extremely cumbersome, requires a lot of computational time and power etc. It affects the ability of quick online detection and finally results in a decrease of the effective sensitivity. Nevertheless a number of such observations have been performed and the results are presented below.

i) Interferometers

The LIGO collaboration has published a series of articles after the first short observational period called "the first Science run, S1": 23/08–09/09/2002. Three interferometric detectors were involved: two at Hanford Observatory with 2 km (H1) and 4 km (H2) bases (http://www.ligo-wa.caltech.edu) and one 3000 km apart at Livingstone Observatory also with a 4 km long base (http://www.ligo-la.caltech.edu).

a) *Search for short GW bursts*

This program [18] is closer to the original Weber's algorithm of selection of very short coincident pulse-excitations at independent detectors. A new element introduced at present consists of the modelling of expected signals in two probable waveforms: a quasi δ-pulse with Gaussian envelope, and a similar pulse with a sine carrier – the "sine-Gaussian" pulse. The carrier frequency and pulse duration vary. So the "matched filtering" can be applied only as a multi-channel procedure. A sophisticated preliminary noise filtering and selection of "GW-burst candidates" is applied. It results in a step-by-step reduction of the useful observation time interval

[18]: 17 days (408 hours) (S1 time) – 96 hours (effective duty cycle) – 35.5 hours (triple-coincidence time).

This concluded in the determination of the "first upper limit from LIGO on GW bursts". Roughly it was stated that "with a confidence level 90% at frequencies (150–3000) Hz short GW bursts with duration (40–100) ms and spectrum amplitudes $h_f \sim (10^{-18}$–10^{-19} Hz$^{-1/2}$ can be expected at a rate of $R \leq 1.6$ day^{-1}". The observation was performed in several frequency windows $\Delta f \simeq 10^2$ Hz inside the wider bandwidth. Thus the threshold for admitted metric perturbations is $h \leq 10^{-17}$–10^{-18}. Below we will compare it with bar observations.

b) *Search for GW from binary neutron stars coalescence*

This program [19] is known as detection of "chirp signals", i.e. short GW pulses with sweeping carrier which is radiated by a relativistic binary during the last orbiting cycles of coalescence. A typical example of coalescence for a neutron star binary $2 \times 1.4\, M_{\odot}$ provides a chirp signal which would traverse the sensitive band of the interferometers (100–3000) Hz in 2 seconds. At its current sensitivity the Livingstone detector is able to register such signals from a distance $\sim$ 176 Kpc and the Hanford instrument from 46 Kpc. So only Milky Way events might be overlapped in this experiment with a small contribution $\sim$ 10% from the Magellanic Clouds.

The bank of templates used in this observation was modelled for variable parameters such as: i – the orbit inclination, α – the initial signal phase and masses $(1-3)M_{\odot}$. A "density" of templates was chosen so that a loss of SNR did not exceed 3%; it gave a total number of templates close to 2100. Noise characteristics were studied based on arbitrary selected lumps of noise playground; special sophisticated veto criteria were applied to avoid statistical bias.

During the S1 run no candidates for GW-chirps were found at the threshold SNR=8 corresponding to 90% confidence limit. Meanwhile a Monte-Carlo simulation with artificial chirp-signals confirmed an ability to detect coincident events. It was concluded that "a new upper limit of neutron star binaries coalescence rate for Milky Way like galaxies is $R \leq 1.7 \cdot 10^2$ year^{-1}" which, while better than previous results is still far from a realistic astrophysical prediction.

The designed sensitivity of LIGO set-ups would allow us to overlap binaries at a distance $\sim$ 21 Mpc. The optimistic theoretical prediction of galactic rate of coalescence $5 \cdot 10^{-4}$ year^{-1} [20] forecasts an integral rate of chirp-events of 0.25 year^{-1}.

c) *Search for continuous GW radiation (pulsars) [21]*

There are several theoretical scenarios forecasting GW emission by pulsars: NS-spin precession, oscillations of eigen modes, r-modes crust currents etc. It is believed that the most effective GW radiation would be produced by its non-zero ellipticity $\varepsilon < 1$, which is defined by the relative asymmetry of the inertia momentum in the plane normal to the NS rotational axis. Theoretical estimates in the literature of the ellipticity predict as a realistic value $\varepsilon \sim 10^{-7} - 10^{-8}$ although for NS with a very strong magnetic field it might be several orders higher [22]. In any case the expected GW radiation from galactic pulsars does not exceed $h \sim 10^{-24}$ metric perturbation.

During the S1 run the LIGO collaboration carried out a test of GW observations of the millisecond pulsar $J1939 + 2134$. This type of measurement was associated with long time accumulation of the narrow band $\Delta f \simeq 4$ Hz antenna's output noise centred at the double pulsar frequency $2f_{ps} \simeq 1284$ Hz. Before accumulation the raw output data were demodulated through a special time resampling and a very complex multichannel "matched filtering" procedure with an appropriate bank of templates varying through unknown parameters of the source (orbit inclination, polarisation angle and initial phase). There was no "coincidence procedure" in this measurement of the continuous signal from a well positioned source, the results of all the detectors were used independently and could be summed (also data of GEO set up were involved).

The lowest strain noise spectrum density of the L and H detectors in these measurements were at a level $\sim 10^{-22}$. The effective time duty cycle was $209\,h$ and $137\,h$ for the Hanford and Livingstone instruments respectively. The upper limit for the detected amplitude of the GW radiation from the pulsar $J1939 + 2134$ inferred from the S1 run data is $h \simeq (1.5 - 4.5)10^{-22}$ (taking into account the different interferometers involved). This limit does not represent a new qualitative astrophysical result but it was the first time when the measurement was performed for a well defined target – a particular pulsar with known coordinates and frequency. Estimation of its ellipticity resulted from this upper limit was large $\varepsilon < 3\,10^{-4}$ suggesting a huge magnetic field $\sim 10^{16}$ G.

d) *Detection of GW stochastic background*

This is a very ambitious program for the detection of a relic GW stochastic radiation produced in a much more early universe (at the "Planck time") than its electromagnetic analog CMBR [22, 27]. Such an objective looks more problematic with respect to the current interferometer sensitivity in comparison with the above mentioned programs. A conventional description of that background contains a special parameter $\Omega_g(f)$ which is the GW energy density per unit logarithmic frequency normalized to the critical energy of the "close universe" $\Omega_g(f) = (f/\rho_c)(d\rho_g/df)$. This parameter is related to the power spectrum of the GW strain $< h^2(f) >$ via the formula [23]: $< h^2(f) >= (3H^2/10\pi^2 f^3)\Omega_g(f)$, where $H \sim (1.5 \cdot 10^{10} y)^{-1}$ is the present day Hubble constant. Thus the largest $h(f)$ corresponds to the low frequency wing of the interferometer bandwidth, so for $f = 100$ Hz we estimate $h(f) \sim 5 \cdot 10^{-22}\sqrt{\Omega}_g\, Hz^{-1/2}$ with $\Omega_g \simeq$ const. In the Standard Cosmological Model the GW relic background is considered as an isotropic, stationary, Gaussian noise with $\Omega_g \leq 10^{-8}$ [27].

A detection method consists of the measurement of the cross-correlation variable combined from outputs of two independent detectors located in the same place. Then a correlated "GW noise signal" has to be accumulated in time at the background of independent intrinsic noises of both detectors. The sensitivity to the parameter Ω growths as $\Omega \sim |h|^2/\sqrt{\Delta f T}$ where T is the integration time of the measurement. This ideal theoretical scheme needs a correction associated with real positions of GW-interferometers separated by a known distance. The correction factor decreasing the sensitivity of the H and L interferometers is $\gamma \sim 0.1$ at a frequency $\sim$ 100 Hz.

During the S1 run only $\sim$ 100 hrs of coincident interferometric strain data were selected to establish a "LIGO upper limit" on the stochastic background of gravitational radiation. With 90% confidence limit at the frequency band 100–300 Hz the experimental estimation of $\Omega_g h_{100}^2 \leq 23$, where h_{100} is the Hubble constant in units of 100 km/sec/Mpc. This new limit is not too useful for testing astrophysical models but it is much better than the previous result from interferometers [23].

ii) Bars

Bar detector collaborations IGEC (an International net) and ROG (an Italian net) also published reports on the programs mentioned above. "Bar observations" have the advantage of much longer duty cycles compared to interferometers but they lack narrow bandwidth. Upper limits for different types of GW signal collected with bar detectors are briefly summarised below.

In a program of *short GW bursts search*, data from five cryogenic bars were analysed over a four years time interval 1997–2000 (there were three Italian set-ups, one of the US and one Australian, see, for example, http://igec.lnl.infn.it). The effective observation time was distributed as follows: during 4 years there were 1319 days when at least one detector was operating; 707 days with at least 2 detectors in simultaneous operation; 173 days with 3 detectors and 26 days with 4 detectors (5 detectors practically had no time of joint operation). In these experiments a strength of the expected GW bursts was quantified through its Fourier amplitude $h_f \simeq (1/4Lf^2)\sqrt{E/M}$ Hz^{-1} defined by E – the energy deposited by a GW burst in the detector with parameters: M – mass, L – length, f – mean resonance frequency. Four detectors had frequencies close to 900 Hz but the Australian detector NIOBE had $f \sim 700$ Hz. The matching filtering was performed referring to the δ-pulse modelling of GW bursts.

As a result of the blind search IGEC found "no coincidences below the threshold $h_f \simeq 10^{-20}$ Hz^{-1}; at the threshold itself the registered "noise event" rate was "at least one per year". It means a daily rate ot GW bursts above this threshold has to be less than $\sim 4 \times 10^{-4}$, that looks much stronger than the LIGO result. However to estimate the integral magnitude of metric perturbation (used in the LIGO approach) we have to multiply the Fourier amplitude by the burst's spectrum width. For "short bursts" the number is comparable to the bar's resonance frequency $\sim (10^2 - 10^3$ Hz. The event rate has also to be corrected by the ratio of bandwidths $\sim (2-3) \times 10^3$. Then the reduced IGEC upper limit is: "bursts with amplitude $h \leq 10^{-18} - 10^{-17}$ might have a rate $R \leq 1$ day^{-1}" comparable to the LIGO result mentioned above. At the same time the "robustness of the IGEC limit" is much higher than the LIGO one because the "multichannel (triple etc.) coincidences" exponentially decrease the "false alarm" error.

There is no special program for bar detectors to *search for GW signals from binary coalescences*. Due to the very narrow frequency bandwidth of these instruments such program naturally is unified with the "coincident bursts search".

In the program of *searching for periodic GW signals* the ROG collaboration has a stronger "experimental upper limit" than LIGO. In [24] a 2 day blind-sky search was performed with the EXPLORER's 1991 data. The search was centred at the frequency 922 Hz in the band ~ 1 Hz. It resulted in an upper limit of $\sim 3 \times 10^{-23}$ with 99% confidence level. The same data were used in [25] searching for periodic signals from the Galactic center in the bandwidth 0.06 Hz. The upper limit of the "incoherent searching" (i.e. without any template) was estimated as $\sim 3 \times 10^{-24}$ after 95 days of observation [25].

A similar searching targeting the Galactic center was performed by the ALLEGRO detector group [26] at frequencies 896.5, 920 Hz in the band ~ 1 Hz. An upper limit $h \leq 8 \times 10^{-24}$ was found.

These results are roughly one order of magnitude stronger than the LIGO one. But the LIGO observation was a well targeted coherent search with a particular source and family of templates covering a bank of unknown spin-down parameters. Such a search is physically more informative. It allows us in particular to estimate bounds for the pulsar's ellipticity (see above).

Finally the limits of the *GW stochastic background* were also derived from observations at the bar couple EXPLORER/NAUTILUS [27, 28]: it was found $\Omega_g \leq 6 \times 10$ at the very narrow frequency interval 907.15–907.25 Hz; where the factor "10" is deduced from observations but the factor "6" takes into account the distance between the bars ~ 600 km (i.e. it is proportional to the value of γ^{-1}).

Thus we can conclude that the upper limits established from recent experiments with GW antennae are still rough for checking the validity of the astrophysical models used to predict the power and rate of expected GW events. The effective threshold for the metric variations derived from these experiments is not below the level $h \sim 10^{-19}$ for "short burst" detection and $h \sim 10^{-22}$ for continuous signals. We can expect an improvement of these numbers from advanced interferometers in the near future. Meanwhile a special strategy of observation so called a "search for astro-gravity correlation" makes it reasonable to continue experiments with the available antennae even at their current sensitivity level.

3 Search for GW events associated with GRB

This is a particular version of the common program of "searching for GW events accompanied by other types of radiation" which can be registered by parallel observational channels such as neutrino telescopes, X-ray and cosmic particle detectors [29]. The strategy is based on the general principle of optimal filtering of weak signals: the more *a prior information* is accessible the better the quality of detection. A modern understanding of the nature of two astrophysical phenomena, GW events and Gamma-Ray Bursts (GRB), suggests that both may have common progenitors – superdense relativistic stars at the moment of some catastrophic process in their evolution as binary coalescence, stellar core collapse, fragmentation etc; (see for a GW review [30] and for GRB models [31] and references therein). Thus it is reasonable to look for anomalies in the output noise of gravitational detectors around time marks defined by the registered GRB. It strongly reduces the amount of stochastic data under processing compared with the "blind search" strategy and also provides a possibility to integrate weak GW signals through an optimal summing over many GRB events. Then we may hope to discover GW-bursts even with antennae of moderate sensitivity.

The problem of such parallel multichannel searching is considered in [32] for bar detectors and in [33] for interferometers. In [34, 35] a nontrivial problem of accumulation of GW–GRB signals was analysed. The effectiveness of the accumulation depends on a "prior information" of the GW burst's time position with respect to the GRB. Just in this point there is large uncertainty in the theoretical models.

Nevertheless some algorithms of the "cumulative detection" have been proposed that take into account the GRB time structure [36] and the signal processing typical of cryogenic bar detectors.

The first experimental tests for checking the hypothesis of the GW–GRB association are presented in [37, 38]. Still no positive results were reported. The current upper limit for a GW burst amplitude of this type of signal was established by the ROG group and is $h \leq 2.5 \times 10^{-19}$ [39].

4 Search for rare events

The lack of sensitivity of the available detectors forces experimentalists to follow programs to "search for rare events", i.e. to keep antennae in a long duty cycle relaying on a "lucky chance" of relativistic catastrophic processes in the neighbouring region $r \leq (50\text{–}100)$ Kpc of our Galaxy. This situation is similar to modern programs of neutrino astrophysics concerning a "searching for star collapse" by registering cosmic neutrino radiation. The sensitivity of all known neutrino detectors allows a ν-flux registration from sources located closer than 100 Kpc. Precedents of such events have taken place in the past years.

The first one occurred during SN1987A when some coincident signals registered by room temperature bars [7] were associated with neutrino bursts detected simultaneously by neutrino telescopes. Despite the very strong criticism [40, 41] this case had definitely demonstrated a reason for the "astro-gravity correlation" strategy and furthermore a reason to keep in continuous operation detectors even of moderate sensitivity.

The second case was reported recently by the ROG collaboration. Analysis *a posteriori* of the EXPLORER-NAUTILUS data collected in the year 2001 discovered an excess of statistically meaningful coincident excitations [42]. The excess observed in the sidereal time coordinate was concentrated around the four hour mark with two hours width. At this period the two bars were oriented perpendicular to the galactic plane and their sensitivity for galactic GW-sources was maximal. The energies of coincident events deposited in each bar were approximately equal and being recalculated to the radiated energy were estimated as $\sim 10^{-2} M_{\odot}c^2$ for sources in the Galaxy center. At the same time the rate of events was too high $\sim 200/y$ in comparison with conventional astrophysical forecasts. Afterwards the statistical significance of this observation was critically discussed [43], [44] but recent observations seem to confirm the first result. Detailed theoretical analysis [45] still did not lead to a definite physical explanation of the phenomenon: it looks as if there were unknown sources distributed in the galactic disc or a few (if not one) closely located invisible GW-burst repeaters. An exotic supposition of GW radiation produced by primordial black hole binaries contradicts the requirement of homogeneous distribution of such objects in the Halo. Thus observations have to be continued.

5 Opto-acoustical Gravitational Antenna

It follows from our brief review above that the current "effective upper limits" on cosmic GW pulses are $(10^{-19}\text{–}10^{-18})$ in terms of metric perturbation. It is interesting to note that interferometers achieved such sensitivity without deep cooling (only advanced versions are planned to have "cryogenic mirrors" [46]). This is the result of two technical ideas put into the construction of these set ups. The first one is to operate at frequencies apart from mechanical resonances of "high Q suspensions" where the brownian noise is strongly suppressed. The second one is to use a laser optical read out with a very small back action of the photon short noise: it allows us to amplify the opto-mechanical transformation of the mirror's displacements by increasing the laser power (note that up to now super cryogenic bars did not realize their potential sensitivity which is limited by the noise of the SQUID read out [47]).

This forms the basis of the Russian national project OGRAN having foreseen the construction of a room temperature bar combined with a FP-interferometer for the readout.

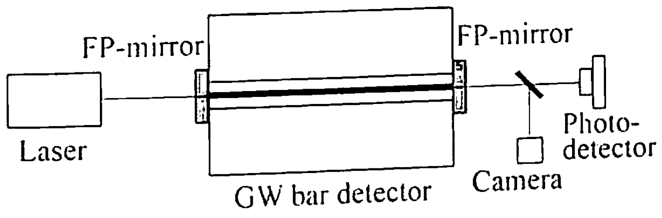


Figure 1: Schematic layout of the OGRAN detector.

The layout of OGRAN is shown in Figure 1. It consists of a bar with a tunnel along the central symmetry axis where a high finesse FP interferometer is formed by two mirrors attached to the bar ends. In contrast with the optical read out developed for AURIGA by the Legnaro group [16] this construction has an expanded optical cavity with bar's length instead of a microgap optical sensor attached at one end of the bar. This difference has a qualitative consequence: in general the response of such combined opto-acoustical antenna must contain not only acoustical excitation of the bar but also an "optical part" as a result of GW–EM interaction. Such antenna was considered in [14] where it is concluded that in a free-mass interferometer the "optical" and "acoustical" parts are undistinguishable, while for the "bar-interferometer" with mirrors following nongeodesic paths their difference has to be observable. Although the difference tends to zero in a "very long GW-approximation" this new important feature of the combined antenna has to be taken into account with an open mind.

The possibility of achieving a resolution $h \leq 10^{-18}$ at the bar with an optical sensor was proved in [15]. Briefly the argument is as follows: The realistic sensitivity of the bar antenna for GW bursts with duration τ is determined by

$$h_{min} \geq 2L^{-1}(kT_n/M\omega_{\mu}^2)^{1/2} \cdot (\omega_{\mu}\tau Q)^{-1/2}, \qquad (1)$$

where L, M, ω_{μ}, Q are the bar parameters length, mass, resonant frequency and quality factor; k is the Boltzmann constant. T_n is the effective noise temperature which depends on transducer and amplifier noise and couples with the physical temperature T through a noise factor F: $T_n = T \cdot F$. Substitution in (1) of typical parameter values: $L = 2$ m, $M = 10^3$ kg, $Q = 3 \times 10^5$, $\omega_\mu \simeq 10^4$ s^{-1}, $\tau = 10^{-3}$ s, $T = 300$ K results in

$$h_{min} \geq 1.5 \times 10^{-19} \cdot F^{1/2}. \tag{2}$$

The noise factor F is defined by the ratio of the real noise variance to the thermal noise variance in the antenna bandwidth $\Delta f \simeq \tau^{-1}$. For the optical interferometer read out

$$F = (2M/\tau)(G_e/G_b)^{1/2}, \tag{3}$$

where the spectral densities of brownian G_b and optical G_e noise are defined by $G_b = 2kTM\omega_\mu/Q$ and $G_e = B\omega_\mu^2(2\hbar\omega_e/\eta W)(\lambda_e/2\pi N)^2$; where ω_e, λ_e, W are the frequency, wavelength and power of the optical pump respectively; and η, N, B are the photodiode quantum efficiency, number of FP reflections and "excess noise factor" (the number of times real optical noise exceeds the short noise level). The optical parameters for OGRAN are $W = (1-3)\,Wt$, $B \simeq (1-10)$, $\lambda_e = 1.064\,\mu$, $\theta = 0.8$, $N = 10^3 - 10^4$ and one can estimate $F \simeq 1$, and forecast the sensitivity (2) to $h \sim 10^{-19}$.

It is important to emphasize that equation (1) assumes a whitening of the Brownian bar noise (a cut off of the bar resonance noise region) and that one operates at the "wings" of the thermal noise spectrum. A response to GW excitation out of resonance is very small but modern optical read out can pick it up.

An opto-acoustical detector can be associated with a high frequency stability laser as a source of the optical pump. A simple direct injection of the beam into the bar's FP resonator would require an unrealistic frequency stability as $\Delta\omega/\omega = h \sim 10^{-19}$ Hz$^{1/2}$. For this reason the scheme has to be designed as a "differential bridge" with automatic compensation for a large part of frequency drifts. In experimental optics two types of such "bridges" are known: the Michelson interferometer (as the interferometric GW antennae) and the so-called "comparator of optical standards" in which one narrow frequency EM source refers to a similar one and the slow drift of both might be corrected. This scheme was used by the Legnaro group [16] and it was chosen also for the OGRAN set up.

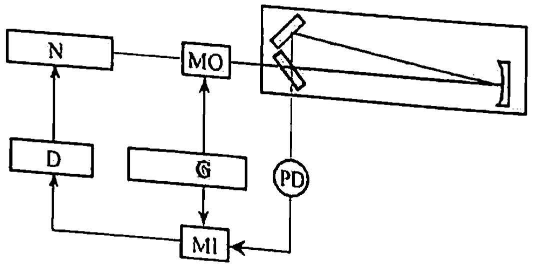


Figure 2: LPF-scheme of laser frequency stabilisation: MO – electro-optical phase modulator; MI – double balance mixer; D – laser frequency driver; G – Generator $f \sim 5$ MHz; N – ND: YAG-Laser; PD – photodetector.

The OGRAN collaboration is formed by three Russian Institutes: *i)* Sternberg Astronomical Institute of Moscow State University, *ii)* Institute for Nuclear Research RAS (Moscow), *iii)* Institute of Laser Physics Siberian Branch RAS (Novosibirsk). The goal is to install a large (~ 2.5 ton) opto-acoustical GW detector into an undeground camera at the Baksan Neutrino Observatory (INR) and to use it in a duty cycle in cooperation with the GW world net firstly at room temperature and then in its advanced cryogenic version. ILP, a center of high frequency stability optical standards, developed a specific variant of the laser frequency stabilisation by a high finesse optical cavity (Figure 2) which was selected for OGRAN.

The logical structure of the OGRAN measurement scheme contains a frequency coupled laser source and FP-cavity of the bar, so that the oscillation of the bar's length results in frequency variations of the output light beam which has to be measured by some discriminator based on a very stable external optical cavity. In principle this corresponds to the AURIGA optical sensor [16] but its realization becomes much more complex for the long (expanded) cavity than for the microgap one.

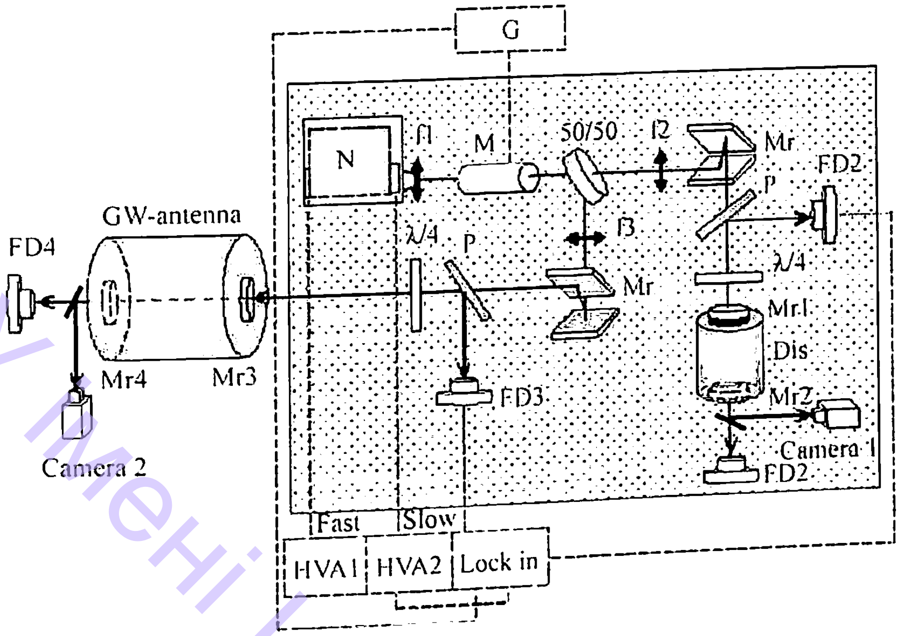


Figure 3: Layout of the OGRAN project: M – electro-optical modulator; P – polarizer; $\lambda/4$ – plate $\lambda/4$; FD – photodetector; G – Generator; N – ND: YAG-Laser; Dis – Discriminator; Mr – adjustable mirrors.

An OGRAN prototype was developed in SAI MSU. Its optical design is shown in Figure 3. Only small effective laser power ~ 2 mW was available, so two mirror FP cavities were utilized with Faraday isolators instead of the "three mirrors" resonator shown in Figure 2 (for large laser power the three mirror construction is preferable). In the mechanical part a small pilot model of the bar detector was manufactured with parameters: $M \simeq 50$ kg, $L = 50$ cm, $\omega_\mu = 2\pi \times 5$ kHz. The finesse of the FP-cavities of the bar and discriminator were equal $F \simeq 800$. The discriminator was thermoisolated and had a feedback loop at slow frequencies to tune the position of one mirror (attached with PZT to its end) keeping the optical resonance. Calibration experiments have shown the absolute sensitivity to the pilot bar oscillations at the level ~ $(1-2) \times 10^{-14}$ cm/Hz$^{1/2}$ [4]. The sensitivity was limited by the resonance thermal noise of the bar and approximately the same level of optical noise. This result is two orders of magnitude smaller than the designed level $\sim 10^{-16}$ cm/Hz$^{1/2}$.

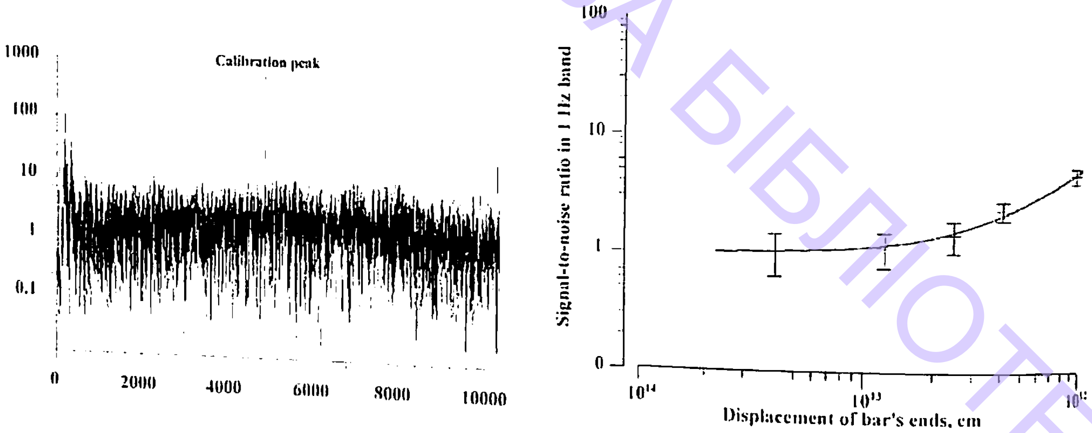


Figure 4: Spectral characteristics of the OGRAN pilot model: a) frequency noise distribution, b) displacement sensitivity.

At present a new 3 Wt single mode stabilized laser and mirrors with high reflectivity $R = 0.9997$ and small losses ~ 50 ppm are being prepared. This modification allows us to get 1.5 orders improvement in the registering amplitude for the pilot model out of the acoustical resonance. The big components of the real experiment are almost ready (vacuum chamber, bars) as well as the infrastructure for the project: the undeground laboratory in Baksan, factory hangar in SAI MSU for test measurements with big bars, etc.

The development plan foresees OGRAN operation at the level of $\sim 3 \times 10^{-19}$ with "room temperature bar" starting in 2006 and a parallel construction of the cryogenic OGRAN version with the final goal $h \sim 3 \times 10^{-22}$. For this the OGRAN collaboration hopes to use the experience and assistance of the Italian cryogenic bar group (ROG collaboration [1]).

Acknowledgements

The authors acknowledge the support and inspiration from the directors of their respective institutes, Academicians Sergei Bagaev, Viktor Matveev and Anatolij Cherepashchuk. We also thank our colleagues V. Braginskii, A. Manukin, V. Rubakov, V. Lipunov, K. Postnov and G. Bisnovatyi-Kogan for useful discussions. This research was supported by the Ministry of Sciences through the programs Neutrino Detectors and Fundamental Metrology and by the Grant RFBR N 04-02-17320.

Finally we would like to express a very deep sad feeling for the untimely loss in a tragic accident of our young, very talented colleague, co-author and friend Andrej Serdobolskii who made a valuable contribution to the OGRAN prototype.

References

[1] P. Astone et al., *Astropart. Phys.*, **10**, 83 (1999).

[2] B. Willke et al., *Class. Quant. Grav.*, **19**, N7, 1377 (2002).

[3] H. Tagoshi et al., *Phys. Rev. D*, **63**, 062001 (2001).

[4] G. Frossati, *J. Low Temp. Phys.*, **101**, 81 (1995).

[5] E. Coccia, J.A. Lobo, J.A. Ortega, *Phys. Rev.*, **52**, 3735 (1995).

[6] O.D. Aguiar, L.A. Andrade, J.L. Barroso et al., *Class. Quant. Grav.*, **21**, 457 (2004).

[7] E. Amaldi, P. Bonifazi, M.G. Castellano et al., *Europhys. Lett.*, **3**, 1325 (1987).

[8] P. Astone, D. Babusci, M. Bassan et al., *Class. Quant. Grav.*, **19**, 5449 (2002).

[9] J. Weber, *Found. Phys.*, **14**, 1185 (1984).

[10] G. Preparata, *Il Nuovo Cimento*, **B101**, 625 (1988).

[11] R. Sisto, A. Moleti, *Int. J. Mod. Phys.*, **13**, N4, 625 (2004).

[12] J.T. Whelan, E. Daw, I. S. Heng et al., preprint gr-qc/0308045 (2003).

[13] V.N. Rudenko, *Astronomy Reports*, **78**, 12, 1116 (2001).

[14] V. Kulagin, A. Polnarev, V. Rudenko, *JETP*, **64**, N6, 915 (1986).

[15] A. Gusev, V. Kulagin, V. Rudenko, *Gravitation and Cosmology*, **2**, N1, 68 (1996).

[16] L. Conti, G.A. Prodi, S. Vitale et al., in J. Tran Thanh Van, J. Dumarchez, S. Reynaud, C. Salomon, S. Thorsett, J.Y. Vinet (eds) *Gravitational waves and experimental gravity, Proc. 34th Ren. Moriond*, World Publish, Hanoi-Vietnam, p.85 (2000).

[17] J. Weber, *Phys. Rev.*, **117**, 306 (1960).

[18] B. Abbott, R. Abbott, R. Adhikari et al., preprint gr-qc/0312056 (2003).

[19] B. Abbott, R. Abbott, R. Adhikari et al., preprint gr-qc/0308069 (2003).

[20] V.M. Lipunov, preprint astro-ph/0406502 (2004).

[21] B. Abbott, R. Abbott, R. Adhikari et al., preprint gr-qc/0308050 (2003).

[22] B. Abbott, R. Abbott, R. Adhikari et al., preprint gr-qc/0312088 (2003).

[23] K. Compton, D. Nicolson, B. Schutz, *Proc. MG-7*, Stanford Univ. USA, W. Sci. Publ., 1078 (1994).

[24] P. Astone, K.M. Borkowskii, P. Jaranowski, A. Krolak, *Phys. Rev. D*, **65**, 022001 (2002).

[25] P. Astone et al., *Phys. Rev. D*, **65**, 022001 (2002).

[26] E. Mauceli, M.P. McHugh, W.O. Hamilton et al., preprint gr-qc/0007023.

[27] P. Astone et al., *Phys. Let. B*, **385**, 421 (1996).

[28] P. Astone et al., *Astron. Astrophys.*, **351**, 811 (1999).

[29] L.S. Finn, S.D. Mohanty, J.D. Romano, *Phys. Rev. D*, **60**, 121101 (1999).

[30] L. Grishchuk, V. Lipunov, K. Postnov et al., *Physics-Uspehi*, **44**, 1 (2001).

[31] T. Piran, *Phys. Rep.*, **314**, 575 (1999).

[32] A.V. Gusev, V.K. Milyukov, V.N. Rudenko et al., *Gravitational waves, Proc. 2th E. Amaldi Conf.*, World Sci., 512 (1998).

[33] M.T. Murphy, J.K. Webb, I.S. Heng, *MNRAS*, **316**, 657 (2000).

[34] V.N. Rudenko, *Gravitational waves and experimental gravity, Proc. 34th Ren. Moriond*, World Publish., 269 (2000).

[35] G. Modestino, A. Moleti, *Phys. Rev. D*, **65**, 022005 (2002).

[36] P. Bonifazi, G.V. Pallottino, A.V. Gusev et al., *Astron. Astroph. Trans.*, **22**, N4-5, 557 (2003).

[37] P. Astone et al., *Phys. Rev. D*, **66**, 102002 (2002).

[38] P. Tricarico et al., *Phys. Rev. D*, **63**, 082002 (2001).

[39] P. Astone et al., preprint astro-ph/0408544 (2004).

[40] C.A. Dickson, B.F. Schutz, *Phys. Rev. D*, **51**, N6, 2644 (1995).

[41] V. Rudenko, A. Gusev, V. Kravchuk, M. Vinogradov, *JETP*, **91**, N5, 845 (2000).

[42] P. Astone et al., *Class. Quant. Grav.*, **19**, 5449 (2002).

[43] S. Finn, *Class. Quant. Grav.*, **20**, L37 (2003).

[44] P. Astone et al., *Class. Quant. Grav.*, **20**, 785 (2003).

[45] E. Coccia, F. Dubath, M. Maggiore, preprint gr-qc/0405047 (2004).

[46] LCGT: K. Kuroda et al., *Int. J. Mod. Phys.*, **8**, N5, 557 (1999).

[47] P. Astone et al., *Astroparticle Phys.*, **7**, 231 (1997).

13

Gamma Ray Bursts: Theory Versus Observations

G.S. Bisnovatyi-Kogan[1,2]

[1] *Space Research Institute, Profsoyuznaya 84/32, Moscow*
117997, Russia
e-mail: gkogan@iki.rssi.ru
[2] *Joint Institute of Nuclear Research, Dubna, Russia*

Abstract

The present common view about the origin of gamma ray bursts (GRB) is that it is cosmological. This conclusion is based on statistical analysis, and on measurements of the redshifts of the GRB optical afterglows of long GRB. No correlation is found between redshift, GRB spectrum, and total GRB fluence. A comparison between KONUS and BATSE data revealed some differences in GRB statistics and hard X-ray line emission. A connection between GRBs and supernova (SN) explosions was found by comparing optical spectra of GRB afterglows and SNe. Different GRB models are reviewed, and a model based on the collapse of a massive star ending in the formation of a black hole surrounded by a massive accretion disk, seems the most attractive. Hard gamma-ray afterglows, prompt optical spectra, hard X-ray lines measurements could be important for further insight into GRB origin. The possible connection of short GRBs with soft gamma repeaters is discussed.

Keywords: gamma-ray astronomy, high energy astrophysics

1 Introduction

It is generally accepted now that cosmic gamma-ray bursts (GRB) have a cosmological origin. The first cosmological model, based on explosions in active galactic nuclei (AGN) was suggested in [37]. A mechanism for the formation of a GRB in the vicinity of a collapsing object based on neutrino-antineutrino annihilation was analyzed in [7]. GRB production in supernova explosions was suggested in [15]. Here we discuss the observational features of GRB, analyze difficulties and problems with their interpretation in the cosmological model and physical restrictions to their model. At the end of this paper we analyze a possible connection between short GRB and soft gamma repeaters (SGR).

2 GRB physical models

GRB models may be classified in two levels. The upper one is related directly to their observational appearance, and include three main types of model:

- fireball,
- cannon ball (or gun bullet),
- precessing jets.

The main restrictions are related to the next basic level of GRB models, which is related to the energy source, producing the huge energy output required by the cosmological GRB model. This level contains many different models:

1. (NS+NS) or (NS+BH) mergers. This mechanism was investigated numerically in [39, 40]. Gamma radiation is produced here by $(\nu, \tilde{\nu})$ annihilation, and the energy output is not enough to explain the most powerful GRBs even taking into account strong beaming. The energy emitted in the isotropic optical afterglow of GRB 990123 [1, 27] is about an order of magnitude larger than the total radiation energy output in this model.
2. Magnetorotational explosion. This mechanism, proposed in [34] for a cosmological GRB, had been suggested for a supernova explosion in [9]. Numerical calculations gave the efficiency of transforming rotational into kinetic energy at the level of a few percent [3, 4]. This is enough for the supernovae energy output but too low for cosmological GRBs.
3. Hypernova. This model, denoting very powerful supernovae, was suggested in [34], and became popular because traces of supernova explosions are believed to have been found in the optical afterglows of several GRBs [47, 18, 49]. GRBs generated by strongly magnetized rapidly rotating new born neutron stars were considered in [52]. Another hypernova model proposes the collapse of a massive core, the formation of a $\sim 20 M_\odot$ black hole surrounded by a massive disk with rapid accretion and the formation of a GRB [28]. This model seems to be the most promising. High mass W-R stars as GRB predecessors seem to be the most probable objects [22, 51].
4. Magnetized disks around rotating (Kerr) black holes (RBH). This is a model based on extracting rotating energy of a RBH when the magnetic field connects the RBH with the surrounding accretion disk or accretion torus [53].
5. In the model proposed in [41] a GRB is created by the pair-electromagnetic pulse from an electrically charged black hole surrounded by a baryonic remnant. A serious problem for this model is how to form such a strongly charged BH.

6. GRBs generated by shocks behind a neutron star after a SN explosion of its companion were proposed in [26].
7. GRBs from superconducting strings have been proposed in [8].
8. The transition from neutron to quark star as an energy source for GRBs was investigated in [6].

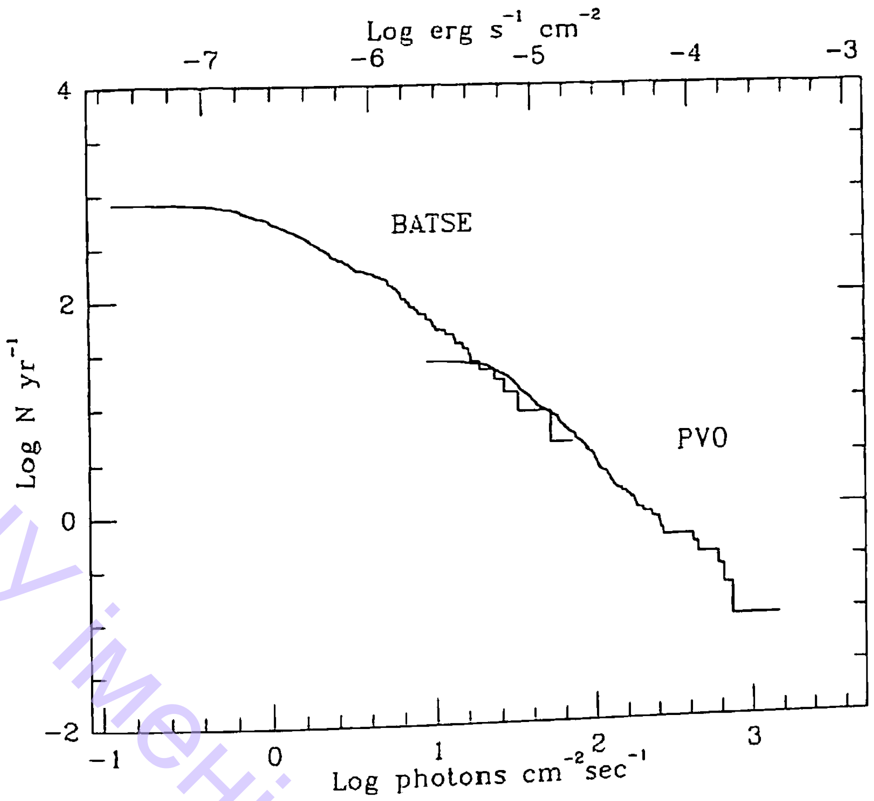


Figure 1: The log N–log P distribution, taken from [21], for the combined BATSE and PVO data. The strongest bursts give a 3/2 power law.

3 Basic Observational Data

3.1 Statistics

Statistical arguments in favor of the cosmological origin of GRBs are based on a visual isotropy of GRB distribution on the sky, combined with a strong deviation from the euclidean uniform distribution with slope 3/2 of the $\log N - \log S$ distribution obtained from BATSE observations [21]. Similar properties have been obtained from the KONUS experiment [29], where deviations from the 3/2 slope have been attributed to selection effects. Analysis of KONUS data taking into account selection effects [25] gives an average value $< V/V_{max} >= 0.45 \pm 0.03$; the value 0.5 corresponds to pure uniform distribution. KONUS data has been obtained

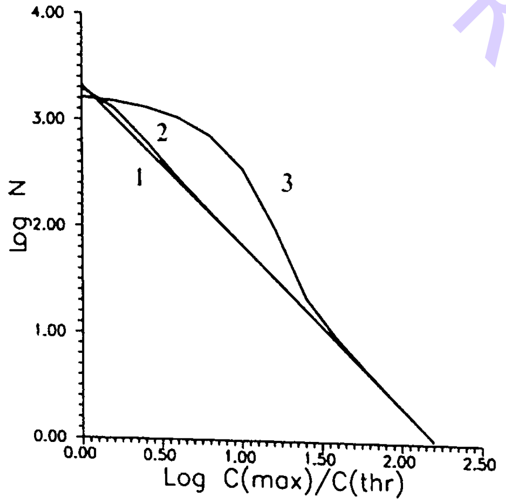


Figure 2: a) $[\log N - \log C(max)/C(thr)]$ in the presence of stochastic errors, distributed according to a normal distribution with an average error Δ_1 in units of a threshold [11]; 1: $\Delta_1 = 0$, line with slope 3/2; 2: $\Delta_1 = 1$; 3: $\Delta_1 = 10$. b) Same as a) for a lognormal distribution, Δ determines a logarithm of number of thresholds, as an average error; 1: $\Delta = 0$, line with a slope 3/2; 2: $\Delta = 1$; $C(max)$ is the peak intensity of the burst; $C(thr)$ is the corresponding threshold value.

in a constant background. Similar analysis [43] of BATSE data, obtained under a substantially variable background, gives $< V/V_{max} >= 0.334 \pm 0.008$. These two results seem to be in contradiction. KONUS sensitivity was only three times less than that of BATSE, where deviations from a uniform distribution are still large [21].

Detailed statistical analysis of BATSE data divided into four hardness classes and the calculation of their respective $< V/V_{max} >$ are given in [44]. The cosmological model predicts smaller values of $< V/V_{max} >$ for softer GRBs in the case of a uniform sample, because larger redshifts correspond to softer spectra. The result is quite the opposite, and soft GRBs have larger $< V/V_{max} >$ than hard ones, 0.47 and 0.27 respectively. The possibility of a decisive role of selection effects (incompleteness of data, statistical errors in the luminosity in the presence of a threshold) was analyzed in [24, 11]. A threshold for statistical errors modify the $\log N - \log S$ curve [11], so that it could be similar to the BATSE distribution (see Figure 1 and Figure 2, curve 3 (left)).

3.2 Optical afterglows and redshifted lines

The spectra of optical afterglows show redshifts up to 4.5, suggestive of the cosmological origin of GRBs and their enormous energy outputs. In most cases the redshifts have been measured in the faint host galaxies. A list of redshifts is given in Table 1, data from [20] plus total GRB fluences [13]). This table contains the trigger number and fluence from the 4B catalogue [33], and GRB fluence from other references.

It was shown in [34] that the properties of GRB afterglows are explained better if the GRB source is situated in a star forming region with high gas density. Inter-

Table 1: GRB Host Galaxies, Redshifts and Fluences (June 2001)

Trigger number	GRB	R mag	Redshift	Type^a	Fluence^e erg/cm^2
	970228	25.2	0.695	e	10^{-5}
6225	970508	25.7	0.835	a,e	$3.5 \cdot 10^{-6}(3{+}4)$
6350	970828	24.5	0.9579	e	$7 \cdot 10^{-5}$
6533	971214	25.6	3.418	e	$10^{-5}(3{+}4)$
6659	980326	29.2	$\sim$1?		$6.3 \cdot 10^{-7}(3{+}4)$
6665	980329	27.7	<3.9	(b)	$7.1 \cdot 10^{-5}(3{+}4)$
6707	980425^c	14	0.0085	a,e	$4.4 \cdot 10^{-6}$
6764	980519	26.2			$9.4 \cdot 10^{-6}$(all 4)
	980613	24.0	1.097	e	$1.7 \cdot 10^{-6}$
6891	980703	22.6	0.966	a,e	$5.4 \cdot 10^{-5}(3{+}4)$
7281	981226	24.8			$2.3 \cdot 10^{-6}(3{+}4)$
7343	990123	23.9	1.600	a,e	$5.1 \cdot 10^{-4}$
7457	990308^d	>28.5			$1.9 \cdot 10^{-5}(3 + 4)$
7549	990506	24.8	1.30	e	$2.2 \cdot 10^{-4}$
7560	990510	28.5	1.619	a	$2.6 \cdot 10^{-5}$
	990705	22.8	0.86	x	$\sim 3 \cdot 10^{-5}$
	990712	21.8	0.4331	a,e	
	991208	24.4	0.7055	e	$\sim 10^{-4}$
7906	991216	24.85	1.02	a,x	$2.1 \cdot 10^{-4}(3{+}4)$
7975	000131	>25.7	4.50	b	$\sim 10^{-5}$
	000214		0.37–0.47	x	$\sim 2 \cdot 10^{-5}$
	000301C	28.0	2.0335	a	$\sim 4 \cdot 10^{-6}$
	000418	23.9	1.1185	e	$1.3 \cdot 10^{-5}$
	000630	26.7			$2 \cdot 10^{-6}$
	000911	25.0	1.0585	e	$5 \cdot 10^{-6}$
	000926	23.9	2.0369	a	$2.2 \cdot 10^{-5}$
	010222	>24	1.477	a	brightest of BeppoSAX

[a] e = line emission, a = absorption, b = continuum break, x = X-ray;
[c] Association of this galaxy/SN/GRB is somewhat controversial;
[d] Association of the OT with this GRB may be uncertain;
[e] The number of BATSE peak channel is indicated in brackets, from [33], otherwise the estimation of bolometric fluence from other sources [13] is indicated.

action of a mighty GRB pulse with the surrounding dense medium ($n = 10^4$–10^5 cm^{-3}) creates a specific form of the optical afterglow. Calculations of the light curve and spectrum of such an afterglow have been done in [16, 5]. Some of the results are presented in Figure 3. The counterparts of cosmological GRBs due to the interaction of γ-radiation with the dense interstellar medium are "long-living" objects, existing for years. To distinguish GRB counterparts from a supernova event having similar energy output, it is necessary to take into account its unusual light curve and spectrum. In the optical region of the spectrum the strongest emission lines are H_α and H_β. In the case of a small cloud with an anisotropic GRB the duration of the optical emission decreases to months or weeks [5].

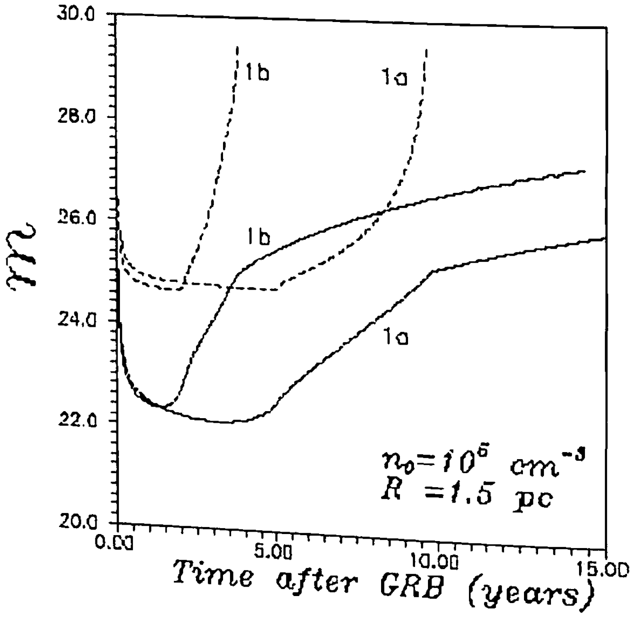


Figure 3: Magnitude of the counterparts (upper limit – solid line, lower limit – dashed line) as a function of time after burst for GRBs with total flux near the Earth $F_{\rm GRB} = 10^{-4}$ erg cm^{-2}: 1a) for E = 10^{52} erg; $n_0 = 10^5$ cm^{-3}; 1b) for E = 10^{51} erg; $n_0 = 10^5$ cm^{-3}, from [16].

3.3 Collimation

Strong collimation of the GRB radiation was suggested to avoid the otherwise huge energy production deduced from the observations. In the "cannon-ball" model [18] the bulk motion Lorentz factor is $\Gamma \approx 10^2$–10^3. This leads to a collimation factor $\Omega \approx 10^{-4}$–10^{-6}. The main restriction to the collimation angle follows from the analysis of the probability of the appearance of an orphan optical afterglow [38], which most probably has low or no collimation at all. If bursts are isotropic, it can be expected to detect ~ 0.2 afterglows. So the absence of orphan afterglows in a search suggests $\Omega_{opt}/\Omega_\gamma << 100$, which is enough to rule out the most extreme collimation scenarios. Radio variability and source counts have been used [35] to place a limit on the collimation angle $\theta_\gamma \geq 5°$. Radio afterglows appear in the non-relativistic phase of the GRB remnant evolution, so they are expected to radiate essentially

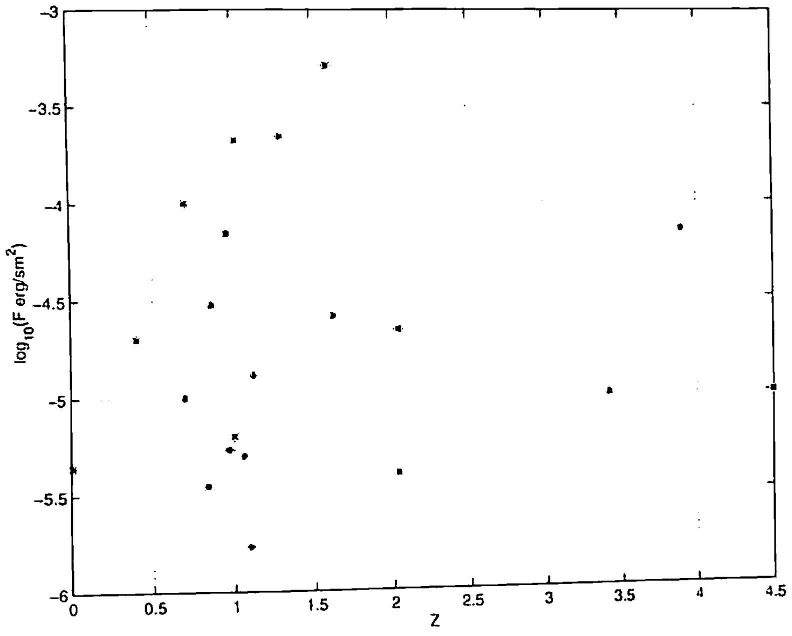


Figure 4: Fluence F versus redshift z for the GRBs from Table 1.

isotropically, and the orphan afterglow limits on ratio Ω_r/Ω_γ immediately imply a limit on Ω_γ itself.

A comparison of the redshifts and fluences from Table 1 does not reveal any correlation between distance and observed flux (see Figure 4). It is explained by collimation, and strong scattering in different beam sight angles. If the collimation is connected to the relativistic bulk motion [18], then a correlation is expected between the GRB duration and its power: stronger GRBs should be shorter. Absence of such correlation excludes collimation resulting from relativistic bulk motions.

3.4 Prompt optical afterglows

The afterglow of GRB 990123 was captured by optical observations 22 seconds after the onset of the burst [2, 1]. GRB 990123 was detected by BATSE on 1999 January 23.407594. The event was strong and consisted of a multi-peaked temporal structure lasting ≥100 s (see Figure 5), with significant spectral evolution. The T50 and T90 durations are 29.82 (± 0.10) s and 63.30 (± 0.26) s respectively. The maximum optical brightness of 8.95^m was reached 30 seconds after the GRB beginning, and after 95 seconds it was already of 14.5^m. So the gamma ray maximum almost coincides with the optical maximum. The observed optical luminosity, for a redshift $z = 1.61$, reaches $L_{opt} \approx 4 \cdot 10^{49}$ ergs/s, about five orders of magnitude brighter than the optical luminosity of any observed supernova. The energy of the prompt optical emission reaches 10^{51} ergs, and the isotropic gamma-ray flux is about $2.3 \cdot 10^{54}$ ergs, what exceeds the rest energy of the Sun [1, 27]. Other bright afterglows have been observed in GRB 021004 (15m, z=2.3), GRB 030329 (12.4m, z=0.168) and GRB 030418 (16.9m). (The brightest visual magnitude and redshift for each of the GRBs are given in brackets). The most remarkable afterglow ever observed was in GRB 030329 (see for example [42]), where probably the supernova was detected (judging by the features of the spectrum) [49]. Konus-Wind observations of GRB 030329

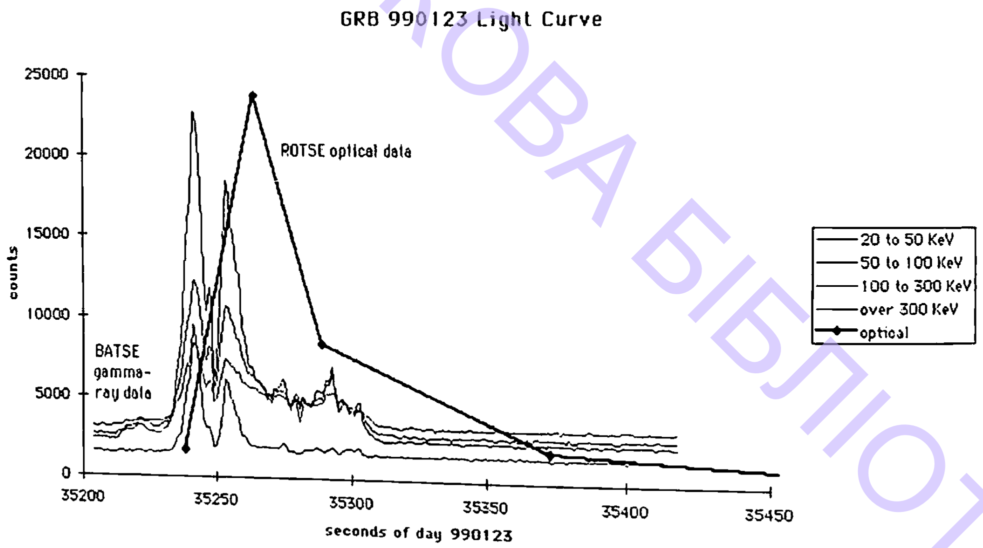


Figure 5: The thin lines represent BATSE γ-ray profile with 1024 ms resolution in different energy channels. The thick line represents the first few frames of ROTSE data, beginning 22 seconds after the initial trigger, from [36].

are presented in Figure 6. The light curve of the optical afterglow obtained at the Crimea observatory is shown in Figure 7 [36].

3.5 High-energy afterglow

EGRET observations on CGRO have shown that GRBs emit also very hard (up to 20 GeV) gamma photons [21]. The number of GRBs with detected hard gamma radiation is about ten; from them five bursts had registered photon energies over 100 MeV [45]. Hard gamma emission continues up to 1.5 hours in GRB 940217. Comparison of the angular aperture of EGRET and BATSE leads to the conclusion that hard gamma radiation could be observed in a large fraction (about one half) of all GRBs. The spectral slope in the hard gamma region lays between (−2) and (−3.7) and varies rapidly, becoming softer with time (as in GRB 920622, [46]). The non-pulsed Crab spectrum slope varies between (−1.78) and (−2.75) [32]. If GRBs are connected to SN explosions and neutron star formation, then residual oscillations of the neutron star may be responsible for the extended hard gamma ray afterglows [10, 50, 19].

3.6 Hard X-ray lines

Hard gamma-ray lines in GRB spectra have been discovered by KONUS [30] in 20–30% of the GRBs. They had been interpreted as cyclotron lines. These spectra had shown a distinct variability: the visible absorption decreases with time (Figure 8). In BATSE data such hard X-ray spectral features were found [17], with 13 statistically significant line candidates from 117 investigated GRBs (see spectrum of GRB 930916 in Figure 8). The conclusion from this paper is not definite. Note that many of the BATSE spectra used for the analysis were obtained well after the trigger and, according to [30], the lines are strongest at the beginning of the burst (see Figure 8).

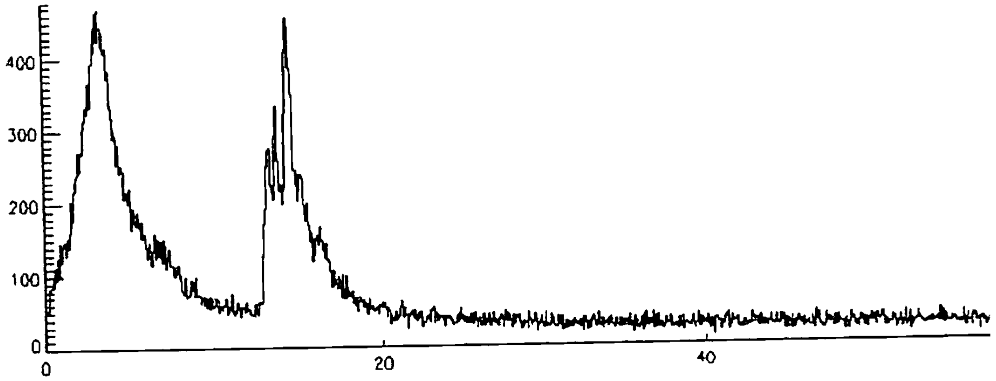


Figure 6: GRB 030329 in gamma-ray: Konus-Wind (UT)11:37:29. Fluence = 1.2 × 10^{-4} erg /cm^2, duration = 50 s, peak flux = 2.5 × 10^{-5} erg/cm^2 / s, from [36].

3.7 GRB and supernovae; short GRB and SGR

The following GRBs show evidence of their connection with SN: GRB 980425 ($z = 0.0085$, 40 Mpc), GRB 980326 ($z = 1$), GRB 011121 ($z = 0.365$), GRB 020405 ($z = 0.695$), GRB 030329 ($z = 0.169$). Some optical afterglows have a red bump during 15–75 days, which may be consistent with the underlying SN explosion (see, for example, [47]). The most intensive observations have been done for SN 2003dh/GRB 030329 [49, 48].

Statistical analysis reveals at least two separate samples of long ($>\sim 2$ s) and short bursts. Optical afterglows have been observed and redshift measurements have been obtained only for long bursts. Therefore, it is not excluded that short bursts have different (perhaps galactic) origin. Let us compare the properties of short GRBs with giant bursts from soft gamma-repeaters (SGR) inside the Galaxy. From larger distances only giant bursts would be registered, which could be attributed to short GRBs. The existence of giant bursts in SGRs (3 in 4 firmly known SGRs in the Galaxy and LMC) implies that SGRs in nearby galaxies may be observed as short GRBs. It is expected that there are more than 10 "short GRBs" of this type from M 31 and other close galaxies. The absence of any GRB projecting onto the local group galaxies may indicate that SGRs could be closer and less luminous objects than is currently accepted [12, 14].

4 Conclusions

There is no fully consistent GRB model: neither radiation, nor explosion. The short GRBs may have galactic origin and giant bursts in SGRs may be connected with short GRBs [31]. Critical observations are needed: spectra of prompt optical

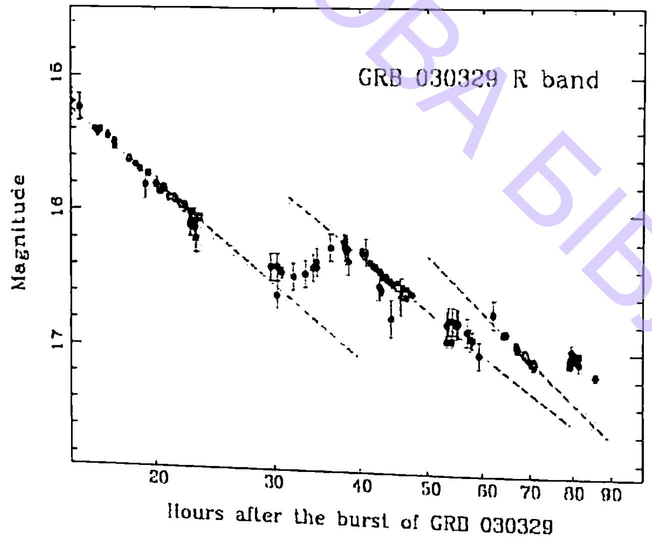


Figure 7: Variability in hours scale of the optical afterglow of GRB 030329, from [36].

afterglows; study of hard γ-ray afterglows; search for orphan optical afterglows in optical all sky monitoring. Cosmological GRB may originate from the collapse of a massive rotating star followed by the formation of a Kerr black hole with a massive magnetized disk and rapid accretion leading to a GRB. Some exotic models also cannot be excluded.

Acknowledgements

This work was supported by a RFBR grant 02-02-16900.

References

[1] C.W. Akerlof et al., *Nature*, **398**, 400 (1999).

[2] C.W. Akerlof, T.A. McKay, *IAU Circ.*, **7100**, 1 (1999).

[3] N.V. Ardeljan, G.S. Bisnovatyi-Kogan, S.G. Moiseenko, *Physics–Uspekhi*, 40, 1076 (1997).

[4] N.V. Ardeljan, G.S. Bisnovatyi-Kogan, S.G. Moiseenko, *Astron. Ap.*, **355**, 1181 (2000).

[5] M.V. Barkov, G.S. Bisnovatyi-Kogan, Preprint IKI Pr-2091 (2003).

[6] Z. Berezhiani, I. Bombaci, A. Drago, F. Frontera, A. Lavagno, *Astrophys.J.*, 586, 1250 (2003).

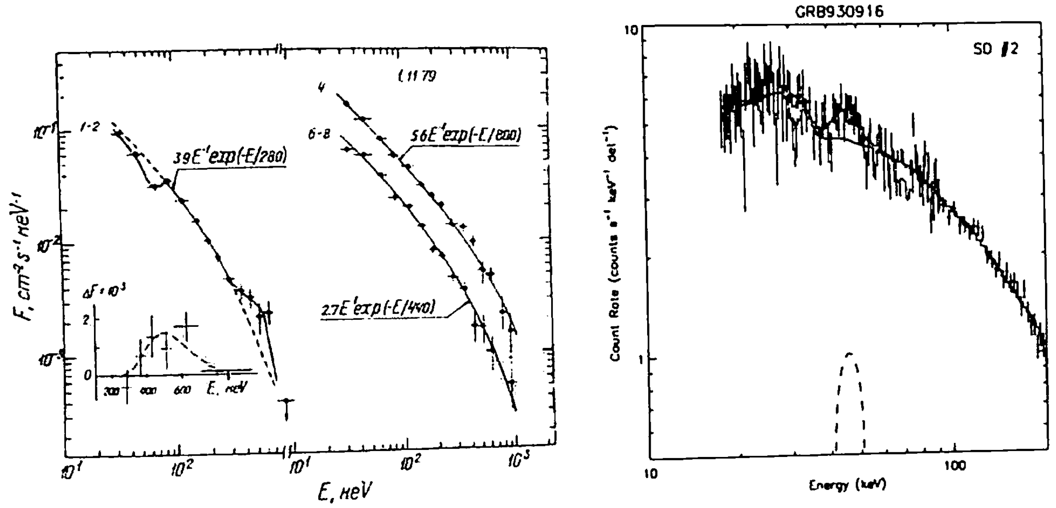


Figure 8: Left panel: Spectral evolution of GRB 791101. (1–2) – spectrum in the first 8 s with an absorption line at ≈ 65 keV and a broad emission feature at 350–650 keV; (4) – spectrum in the 4-th 4 s interval; (6–8) – spectrum summed over 6-th, 7-th and 8-th intervals, from [31]. Right panel: Data from the interval 22.144 to 83.200 s after the BATSE trigger of GRB 930916. The "bump" at 30 keV is expected from the K-edge of the iodine in NaI. A spectral feature is added to the model: an emission line at 45 keV improves χ^2 by 23.1. The solid histogram depicts the total count model; the dashed histograms show the continuum and line portions separately, from [17].

[7] V.S. Berezinsky, O.F. Prilutsky, *Astron. Ap.*, **175**, 309 (1987).

[8] V. Berezinsky, B. Hnatyk, A. Vilenkin, *Phys. Rev. D*, **64**, 043004 (2001).

[9] G.S. Bisnovatyi-Kogan, *Sov. Astron.*, **14**, 652 (1971).

[10] G.S. Bisnovatyi-Kogan, *Astrophys. J. Suppl.*, **97**, 185 (1995).

[11] G.S. Bisnovatyi-Kogan, *Astron. Ap.*, **324**, 573 (1997).

[12] G.S. Bisnovatyi-Kogan, *Mem. Soc. Astron. It.*, **73**, 318 (2002).

[13] G.S. Bisnovatyi-Kogan, preprint astro-ph/0310361 (2003).

[14] G.S. Bisnovatyi-Kogan, preprint astro-ph/0401369 (2004).

[15] G.S. Bisnovatyi-Kogan, V.S. Imshennik, D.K. Nadyozhin, V.M. Chechetkin, *Astrophys. Space Sci.*, **35**, 23 (1975).

[16] G.S. Bisnovatyi-Kogan, A.N. Timokhin, *Sov. Astron.*, **41**, 423 (1997).

[17] M.S. Briggs et al., *Astrophysical Letters and Communications*, **39**, 237 (1999).

[18] S. Dado, A. Dar, A.De Ru'jula, *Astron. Ap.*, **388**, 1079 (2002).

[19] K.Y. Ding, K.S. Cheng, *Month. Not. R.A.S.*, **287**, 671 (1997).

[20] S.G. Djorgovski et al., in E. Costa, F. Frontera, J. Hjorth (eds) *Gamma-Ray Bursts in the Afterglow Era, Proceedings of the International workshop held in Rome, CNR headquarters, 17–20 October, 2000*, Springer, Heidelberg Berlin, p.218 (2001).

[21] G.J. Fishman, C.A. Meegan, *Ann. Rev. Astron. Ap.*, **33**, 415 (1995).

[22] S.S. Gershtein, *Astronomy Letters*, **26**, 730 (2000).

[23] C.J. Hailey, F.A. Harrison, K. Mori, *Astrophys. J. Lett.*, **520**, L25 (1999).

[24] T.E. Harrison, W.R. Webber, B.J. McNamara, *Astron. J.*, **110**, 2216 (1995).

[25] J.,C. Higdon, M. Schmidt, *Astrophys. J.*, **355**, 13 (1990).

[26] Ya.N. Istomin, B.V. Komberg, *Astron. Zh.*, **79**, 1008 (2002).

[27] S. Kulkarni et al., *Nature*, **398**, 389 (1999).

[28] A.I. MacFadyen, S.E. Woosley, *Astrophys. J.*, **524**, 262 (1999).

[29] E.P. Mazets et al., *Sov. Astron. Lett.*, **6**, 318 (1980).

[30] E.P. Mazets et al., *Astrophys. Space Sci.*, **82**, 261 (1982a).

[31] E.P. Mazets S.V. Golenetskii, Iu.A. Gurian, V.N. Ilinskii, *Astrophys. Space Sci.*, **84**, 173 (1982b).

[32] R. Much et al., *Astrophys. Space Sci.*, **120**, 703 (1996).

[33] W.S. Paciesas et al., *Astrophys. J. Suppl.* **122**, 465, 497 (1999).

[34] B. Paczynski, *Astrophys. J. Lett.*, **494**, L45 (1998).

[35] R. Perna, A. Loeb, *Astrophys. J. Lett.*, **509**, L85 (1998).

[36] A.S. Pozanenko, *Private communication* (2003).

[37] O.F. Prilutsky, V.V. Usov, *Astrophys. Space Sci.*, **34**, 387 (1975)

[38] J.E. Rhoads, in V.G. Gurzadyan, R.T. Jantzen, R. Ruffini (eds) *The Ninth Marcel Grossmann Meeting. Proceedings of the MGIXMM. Meeting held at The University of Rome "La Sapienza", 2–8 July 2000*, **3, Part C**, 2419 (2002).

[39] M. Ruffert, H.-Th. Janka, *Astron. Ap.*, **338**, 535 (1998)

[40] M. Ruffert, H.-Th. Janka, *Astron. Ap.*, **344**, 573 (1999).

[41] R. Ruffini et al., *Astron. Ap.*, **359**, 855 (2000).

[42] V. Rumjantsev et al., *GRB Coordinates Network*, **2005**, 1 (2003).

[43] M. Schmidt, *Astrophys. J. Lett.*, **523**, L117 (1999).

[44] M. Schmidt, preprint astro-ph/0101163 (2001).

[45] E.J. Schneid et al., *Ann. NY Acad. Sci.*, **759**, 421 (1995).

[46] E.J. Schneid et al., *Astrophys. J.*, **453**, 95 (1995).

[47] V.V. Sokolov, in E. Costa, F. Frontera, J. Hjorth (eds) *Gamma-Ray Bursts in the Afterglow Era, Proceedings of the International workshop held in Rome, CNR headquarters, 17–20 October, 2000*, Springer, Heidelberg Berlin, p.136 (2001).

[48] V.V. Sokolov et al., *Bulletin of the Special Astrophysical Observatory RAS*, **56**, 5 (2004).

[49] K.Z. Stanek et al., *Astrophys. J. Lett.*, **591**, L17 (2003).

[50] A.N. Timokhin, G.S. Bisnovatyi-Kogan, H.C. Spruit, *Month Not. R.A.S.*, **316**, 734 (2000).

[51] A.V. Tutukov, A.M. Cherepashchuk, *Astronomy Reports*, **47**, 386 (2003).

[52] V.V. Usov, *Nature*, **357**, 472 (1992).

[53] M.H.P.M. van Putten, *Phys. Rep.* **345**, 1 (2001).

14

Cosmic Gamma Ray Bursts: The Brightest Stars of the Universe. In Memory of George Gamow: a Bright Physicist

M.G. Bernardini[1,2], C.L. Bianco[1,2], P. Chardonnet[1,3],
F. Fraschetti[1,4], R. Ruffini[1,2] and S.-S. Xue[1,2]

[1] *ICRA, International Center for Relativistic Astrophysics, Rome, Italy*
[2] *Università di Roma, La Sapienza, Italy*
[3] *Université de Savoie and LAPTH Annecy, France*
[4] *Università di Trento, Italy*

Abstract

Gamma ray bursts (GRBs) are the most powerful explosions of the universe. This field has focused the attention of the most powerful earth and space observatories: KECK, VLT, VLBA, CGRO-BATSE, BeppoSAX, Chandra, Rossi-XTE, XMM-Newton, HETE-2, HST, INTEGRAL and now SWIFT[1]. Thanks to this, theoreticians now have a more clear picture of this enigmatic phenomenon: it is presumably the end of a very massive star which is dying and, before forming a black hole, originates this spectacular event. Here, we will review some fundamental observational facts and present in detail our theory: the electromagnetic black hole (EMBH) model. This model, purely theoretical, uses the theory of General Relativity as well as the fundamental properties of Quantum Field theory. From this unification, we have developed a picture which allows us to interpret all the observational facts and, in addition, to make some remarkable predictions. The most interesting one (proposed three years ago) is the possibility of inducing a supernova during the formation of the GRB. Here we propose a new paradigm in order to interpret recent Chandra data: we called it the triptych paradigm. It consists of a natural association of three actors: the GRB, the supernova and the newborn

---

[1] See NASA High Energy Astrophysics Science Archive Research Center: http://heasarc.gsfc.gov

neutron star. We will show how the model's success in interpreting naturally some apparently complicated data. This prediction, to be tested observationally, will need new facilities in X-ray missions.

Keywords: gamma ray bursts: source, afterglow, ultra relativistic plasma, black hole physics, evolution of massive stars

1 From the Vela Satellites to the Swift Mission: 37 Years of Hunting GRBs

GRBs are sudden, intense flashes of gamma radiation characterized by very fast variability, with a time scale as short as milliseconds [1].

George Gamow would have certainly proposed new ideas about GRBs, the most powerful explosions of the universe. Although much progress has been made in the last few years, the origin of these explosions is still a great mystery in physics. This puzzle started during the Cold War. The Americans were afraid that the Soviets would launch a nuclear weapon on space and explode it on the Moon. In order to check that the treaty of non proliferation was respected, US launched the Vela satellites. The program was a nuclear detonation (NUDET) detection system setup and was designed to detect nuclear explosions in space and later in the air. The original Vela satellites were equipped only with sensors for space detection and were launched in three pairs between 1963 and 1965. They operated for at least five years, far beyond their nominal designed life of six months. A second generation called Advanced Vela were launched in 1967, 1969 and 1970. These satellites added "bangmeters" – optical sensors for detecting atmospheric tests – and had a nominal design life of 18 months, later rated with a seven year lifespan, and in operation for more than ten years, with the last one being turned off in 1984 – after 14 years of successful operation. Launched in couples, they were monitoring the gamma ray emission in space. Any nuclear weapon would have released an enormous quantity of gamma ray photons producing an intense flash.

During 1967, these satellites detected flashes of intense gamma radiation, and so the history of GRBs began.

Due to an extended orbit, and to the fact that they were working in groups it was possible to determine, by triangulation, using the measurement of the time delay, the direction of the sources. The answer opened a new enigma for physicists. This mysterious flash was not coming from the Earth or the Moon but from directions above the planet's orbits. These satellites detected more than 73 GRBs. This discovery was announced during a meeting in California in 1973 and a new field of research was born. In order to explain this sudden increase of gamma ray flashes, the astronomers proposed many possible explanations [2]. At that time, there were supposed to be galactic in origin. From the peak flux $l \sim 10^{-5}$ erg s^{-1} cm^{-2} it was possible to estimate the total luminosity L of these sources $L \sim 4\pi R^2 l \sim 10^{41}$ erg s^{-1}, for $R \sim 10$ kpc. The most popular explanation was related to neutron stars.

Nothing was very convincing because of the lack of data until the Compton Gamma Ray Observatory (CGRO) was launched in the beginning of the 90s. The BATSE experiment (Burst And Transient Source Experiment) launched on board of CGRO was dedicated to the study of GRBs and lasted more than 9 years.[2] The results are very impressive. More than 2000 bursts were detected, timescale recorded and energy spectrum measured [3]. From the map of the bursts in galactic coordinates it appears clearly that the distribution is highly isotropic. A great debate about the distance to GRBs opened:

- GRBs are at cosmological distance $R \sim Gpc \simeq 10^{27}$ cm and $L \sim 10^{51}$ erg s^{-1}
- GRBs are of galactic origin $R \sim 100\ kpc \simeq 10^{23}$ cm and $L \sim 10^{43}$ erg s^{-1}

BATSE data identified two classes: short and long GRBs. With a clearer picture new problems arise: bursts do not have all the same lightcurve. One important piece of the Rosetta Stone was discovered in February 1997 when the Dutch–Italian Satellite Beppo-SAX[3] observed the first afterglow of GRBs [4]. The discovery of long-lived afterglows in X-rays, optical and radio and the direct determination of the distance revolutionized the field [5–8]. The existence of a counterpart opened a new area in the GRBs field. The measurement of the first redshift opened a Pandora's box putting GRBs in the field of cosmology. They are at cosmological distance, and as a consequence their energy is colossal: the luminosity is a few 10^{54} erg s^{-1}. Using Einstein equation $E = mc^2$ it is equivalent to the sun converting into pure energy in a few seconds. Up to now more than 45 redshifts have been measured. In the first year of the discovery of GRBs, more theories than the number of objects detected up to now have been proposed. The enigmatic flashes discovered 40 years ago by the Vela satellites appear to have a key role in cosmology. They could be at the origin of the reionization of the universe at high redshift recently discovered by WMAP. The hope is to detect a burst at redshift greater than ~ 6 which corresponds to the furthermost quasar detected up to now [9].

2 Observational Properties of GRBs

Nowadays the GRB area is fast becoming one of the most important fields in astrophysics like the study of primordial fluctuations in cosmology. An unprecedented campaign of observations using state of the art telescopes on the Earth and in space has contributed to revolutionize the field. Nine satellites (CGRO, Beppo-SAX, Chandra, Rossi-XTE, XMM-Newton, HETE, HST, INTEGRAL and SWIFT) and all ground observatories are involved in this campaign. The possible link between GRBs and ultra high energy cosmic rays is analyzed by cosmic rays observatories like AUGER or neutrino telescopes like ANTARES and AMANDA. This unprecedented effort provides a harvest of data within an international network[4]. The rate of GRBs is approximately one per day. Since their intrinsic luminosity is very high, they are naturally the best probe of the cosmos at high z. Basically all the properties of GRBs come in three observational windows: the γ- and X-ray spectrum, the optical and the radio data. Observations of GRB afterglows span a broad range of

---

[2] See: http://www.batse.msfc.nasa.gov/batse/

[3] See: http://www.asdc.asi.it/bepposax/

[4] See IPN3 at: http://ssl.berkeley.edu/ipn3

Table 1: Measured GRB redshifts. See summary table of redshifts: http://www.mpe.mpg.de/~ jcg/grbrsh.html by Jochen Greiner.

GRB Name	z	GRB Name	z
soon	soon	050603	2.821
050525	0.606	050505	4.3
050502	3.793	050416	0.6535
050408	1.2357	050401	2.90
050319	3.24	050318	1.44
050315	1.949	050126	1.29
041006	0.716	040924	0.859
031203	0.105	040701X	0.2146
031203	0.105	030429X	2.65
030329	0.168	030328	1.52
030323	3.372	030226	1.98
021211	1.01	021004	2.3
020903X	0.25	020813	1.25
020405	0.69	020305	0.2?
020124	3.198	011211	2.14
011121	0.36	010921	0.45
010222	1.477	000926	2.066
000911	1.058	000418	1.118
000301C	2.03	000214	0.37-0.47
000210	0.846	000131	4.5
991216	1.02	991208	0.706
990712	0.434	990705	0.842
990510	1.619	990506	1.3
990123	1.60	980703	0.966
980613	1.096	980425	0.0085
980329	≤ 3.9	980326	1?
971214	3.42	970828	0.9578
970508	0.835	970228	0.695

frequencies from $\sim$ GHz to $\sim 10^9$ GHz in timescales from minutes to years after the burst.

2.1 γ and X-rays properties of GRBs

Since GRBs are very energetic phenomena, the X-ray window is essential to understand the underlying physics. It is also in this window that the most fundamental discoveries were made. The first observations and also two fundamental properties of GRBs are the light curve and the spectrum. The high energy spectrum is a broken power law suggesting a non-thermal spectrum like synchrotron radiation. The observed spectrum is well fitted by a power law (a "Band relation") which does not exclude a thermal spectrum. In fact a sum of black body spectra could also reproduce the observed spectrum with great accurary. This is one of the "apparent" paradoxes of this field, as we will see in detail in Section 4.3, and an example of how much care must be taken in interpreting the data. The substructures observed in the lightcurve and the origin of the spectral emission are still matters for discussion. There are two scenarios for the origin of the substructures: internal or external shocks. The natural collision of the ultrarelativistic shock waves with the interstellar matter favors the external shocks scenario. Considerations on timescales lead part of the community to reject this mechanism in favor of internal shocks. We will show in Section 4.2 numerical simulations of the EMBH model succeeding to interpret substructures with external shocks.

The high energy resolution spectrometers on board of satellites open a new window in GRB studies: X-ray spectroscopy. This new tool is essential in order to understand the environment and the nature of the burst. The presence of iron lines in emission and absorption indicates massive progenitor stars. X-ray spectroscopy also allows us to measure the distance of GRBs by the shift of the iron line [10], [11], as for GRB991216. The study of the absorption line tells us also the local hydrogen density and so the baryon loading. Spectroscopic studies of the afterglow emission could reveal the presence of elements, the outflow and a temperature estimate.

2.2 The optical properties of GRBs

Firm evidence for a cosmological origin of GRBs was established in 1997 from the absorption lines of the afterglow. The study of GRB afterglows is very important since their absorption spectra contain information on both the distance and the local environment where the explosion took place. The accurate determination of position in the X afterglow, allows a powerful optical telescope (HST, KECK, VLT) to scrutinize the very distant galaxy associated with GRBs. It is also a new method to determine the distance of the GRB jointly with X-ray spectroscopy. Redshifts of GRBs are determined mainly for the host galaxy. Djorgovski et al. have shown that GRBs are associated with faint galaxies ($R \sim 25$) at cosmological redshifts with $z \sim 1$ [12]. In addition, optical observations of host galaxies and their environments show possible association with supernovae and could help us in determining the nature of the progenitor star. In 1992, Dermer has shown the possibility of using GRBs as standard candles [13]. Gamma ray bursts would have advantages over Supernovae Ia. While the limit on SN searches using HST could be at redshift ~ 2, the GRBs have been measured at greater redshifts with more than 32 bursts over 55 having redshift greater than 1. GRBs could in principle be produced at very high redshifts and witness the first population of stars in the universe [14]. GRBs hence could help us to go through the unexplored window of the dark age of our universe. An open question is whether we can we use GRBs as standard candles. With the accumulation of redshifts measured, this question appears to be crucial even if the answer is not straightforward. Some relationships have been proposed in order to find this "golden rule" that could dethrone SN as the geometers of the universe.

2.3 The radio properties of GRBs

The characteristic size of a GRB image is ~1 micro-arsecond a day after the burst and cannot be resolved by existing optical telescopes. A recent campaign of observations in radio using the VLBI providing the first measurement of a burst in expansion allowed us to determine the size of GRB030329 [15]. The size of the afterglow is found to be 0.07 milliarsecond (mas) and 0.17mas 25 and 83 days after the burst respectively. Assuming a set of cosmological parameters $\Omega_\Lambda = 0.73$, $\Omega_M = 0.27$ and $H_0 = 71\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ the distance to this burst is 589 Mpc and one milliarcsecond corresponds to 2.85 pc. Taylor and colleagues [15] found that the polarization is less than 1 percent and concluded that no variant of the cannonball model is consistent with their proper-motion limits. A compact component was also observed at a distance of 0.28 mas 52 days after the burst. This additional component is not expected in the standard model. Hoping for a closer burst in the near future, this method could revolutionize GRB science as Beppo-SAX did with the redshift measurement. Seeing the GRBs in action inside the ISM with long base-line experiments will additionally test the theories.

The radio observations also showed that the supernova associated with GRB980425 was unusual with a very fast expanding shell [16]. Filliatre et al. have recently discovered the IR afterglow of the dark INTEGRAL burst GRB040422 with the VLT [17]. They concluded that it is possible to detect dark optical bursts by a more systematic use of 8 meter telescopes.

3 The EMBH Model of GRBs

We propose here a theory of GRBs that we call EMBH. Our approach is radically different because we do not intend just to fit the data. Our aim is to propose a global theory of this exceptional phenomenon using general relativity and quantum mechanics. We will see some unexpected consequences of the theory. More details can be found in [18], [19]. For the collapsar model and the supranova model see original articles [20], [21] and the review articles: [22] and [23].

- We proposed the existence of a flash before the GRB flash, called P-GRB for Proper GRB and identified it in some bursts.
- We gave a unified picture of all observed GRBs "short" and "long".
- We advanced the novel idea that the powerful explosion in GRBs could induce a supernova in their vicinity.
- Recently we explored the possibility that the X-ray flux observed in some past bursts is due to the cooling of a newborn neutron star.

For such a special phenomenon we need also an unusual theory. This theory of GRBs was developed on the concept of extracting energy from a black hole. Black holes are often seen as giant vacuum cosmos cleaner, but in the 70s it was also discovered that energy could be extracted from black holes. The basic concept of our theory is to use this theoretical possibility to power the engine of the GRBs [24]. There are different possibilities for the way in which energy can be extracted from black holes. We have used the mechanism of vacuum polarization around a Kerr black hole.

3.1 The foundations: energy extraction from a black hole

In this section we will explain in detail this concept and compute the energy we can extract this way. In 1971, in *"Introducing the Black Hole"* by R. Ruffini and J. Wheeler, the theorem was advanced that a black hole is characterized uniquely by three independent parameters: the mass-energy M, the angular momentum L and the charge Q making it an EMBH [25]. Such an ansatz, which came to be known as the "uniqueness theorem" turns out to be one of the most difficult physics and mathematics theorems to prove. The progress in the proof has been authoritatively summarized by B.Carter. The situation can be considered satisfactory from the point of view of the physical and astrophysical considerations. Nevertheless some fundamental mathematical and physical issues concerning the most general perturbation analysis of an EMBH are still the topic of active scientific discussion. Ruffini and Christodoulou (1971) have shown that the energy extractable from a EMBH is governed by the mass-energy formula

$$E_{BH}^2 = M^2c^4 = \left(M_{\rm ir}c^2 + \frac{Q^2}{2\rho_+}\right)^2 + \frac{L^2c^2}{\rho_+^2}. \tag{1}$$

$M_{\rm ir}$ is the irreducible mass, r_+ is the horizon radius and ρ_+ is the quasi-spheroidal cylindrical coordinate of the horizon evaluated at the equatorial plane [26]. Up to 50% of the mass-energy of an extreme EMBH can in principle be extracted using the vacuum polarization. From quantum mechanics, we have the possibility to create virtual pairs e^+e^-. In order to create real particles we need to give them energy. This will be achieved via the enormous electric field around a Kerr black hole. We can easily compute the strength of the electric field in order to create pairs. The condition is simply given by: $(\frac{\hbar}{m_e c}) \times (e\mathcal{E}_c) \sim m_e c^2$. Here m_e and e are respectively the mass and charge of the electron. It means that the vacuum could be polarized by an electric field larger than the critical value:

$$\mathcal{E}_c = \frac{m_e^2 c^3}{\hbar e}. \tag{2}$$

This can indeed occur in the field of a Kerr–Newmann EMBH. As a first simplifying assumption we have assumed absence of rotation with spherically symmetric distributions. The space-time is then described by the Reissner–Nordström geometry, whose spherically symmetric metric is given by

$$d^2s = -\left[1 - \frac{2GM}{c^2r} + \frac{Q^2G}{c^4r^2}\right] d^2t + \left(\left[1 - \frac{2GM}{c^2r} + \frac{Q^2G}{c^4r^2}\right]\right)^{-1} g_{rr}(r)d^2r + r^2d^2\theta + r^2\sin^2\theta d^2\phi. \tag{3}$$

The first new result we obtained is that the pair creation process does not occur at the horizon of the EMBH: it extends over the entire region outside the horizon in which the electric field exceeds the critical value. The electric field in the Reissner-Nordström geometry has only a radial component given by $\mathcal{E}(r) = \frac{Q}{r^2}$. This region extends from the horizon radius $r_+ = 1.47 \cdot 10^5 \mu(1 + \sqrt{1-\xi^2})$ in cm, out to the dyadosphere radius r_{ds} where the electric field $\mathcal{E}$ decreases to its critical value $\mathcal{E}_c$. We can show that: $r_{ds} = 1.12 \cdot 10^8 \sqrt{\mu\xi}$ cm, where we have introduced the dimensionless mass and charge parameters $\mu = \frac{M}{M_\odot}$, $\xi = \frac{Q}{(M\sqrt{G})} \leq 1$. The vacuum polarization process occurs as if the entire dyadosphere is subdivided into a concentric set of shells of capacitors each of thickness $\hbar/m_e c$ and each producing a number of e^+e^- pairs on the order of $\sim Q/e$. The second new result has been to realize that the local number density of e^+e^- pairs created in this region as a function of radius is given by

$$n_{e^+e^-}(r) = \frac{Q}{4\pi r^2 \left(\frac{\hbar}{mc}\right) e} \left[1 - \left(\frac{r}{r_{\rm ds}}\right)^2\right] \qquad (4)$$

and consequently the total number of e^+e^- pairs in this region is $N^\circ_{e^+e^-} \simeq \frac{Q-Q_c}{e} \times \left[1 + \frac{(r_{\rm ds}-r_+)}{\frac{\hbar}{mc}}\right]$ where $Q_c = \mathcal{E}_c r_+^2$. The total number of pairs is larger by an enormous factor $r_{\rm ds}/(\hbar/mc) > 10^{18}$ than the value Q/e which a naive estimate of the discharge of the EMBH would have predicted. Due to this enormous amplification factor in the number of pairs created, the region between the horizon and the dyadosphere radius $r_{\rm ds}$ is dominated by an essentially high density neutral plasma of e^+e^- pairs. We have defined this region as the dyadosphere of the EMBH from the Greek *duas*, *duadsos* for pairs. The energy density of the electron–positron pairs is given by $\epsilon(r) = \frac{Q^2}{8\pi r^4}\left(1 - \left(\frac{r}{r_{\rm ds}}\right)^4\right)$. The total energy of pairs converted from the static electric energy and deposited within the dyadosphere is then

$$E_{\rm dya} = \frac{1}{2}\frac{Q^2}{r_+}\left(1 - \frac{r_+}{r_{\rm ds}}\right)\left[1 - \left(\frac{r_+}{r_{\rm ds}}\right)^4\right]. \qquad (5)$$

As we will see in what follows this is one of the two fundamental parameters of the EMBH theory. In the limit $\frac{r_+}{r_{\rm ds}} \to 0$, leads to $E_{\rm dya} \to \frac{1}{2}\frac{Q^2}{r_+}$, which coincides with the energy extractable from EMBHs by reversible processes ($M_{\rm ir} =$ const.), namely $E_{BH} - M_{\rm ir} = \frac{1}{2}\frac{Q^2}{r_+}$. Due to the very large pair density and to the sizes of the cross-sections for the process $e^+e^- \leftrightarrow \gamma + \gamma$, the system is expected to thermalize to a plasma configuration for which $n_{e^+} = n_{e^-} \sim n_\gamma \sim n^\circ_{e^+e^-}$ where $n^\circ_{e^+e^-}$ is the total number density of e^+e^--pairs created in the dyadosphere.

3.2 The fundamental hydrodynamical equations of the EMBH Model

The next step is to develop a numerical code in order to follow the evolution of the e^+e^--pair plasma generated in the dyadosphere. For a complete description of the general relativistic treatment of the hydrodynamic expansion we recommend the following articles: [27] and [28]. The beauty of a theory is not only to interpret data by equations, it is also to make sense and to explain physically what happens. The EMBH model of GRB gives a full coherent picture of this exceptional phenomenon. In short this theory explains to us that what we call a GRB is the end of a massive star creating a black hole. In that sense the GRBs are the brightest stars of the universe. The key point is that different from normal stars where the physics is dominated by equilibrium equations, the characteristic of a GRB is a dynamical process. The timescales are also very different. The EMBH is really a tool which enlightens us to what happens near this dead star. This process which appears observationally very short, a few seconds in the γ ray window, is much longer in situ. We will see also that the observed spectrum which is observed as a breaking power law and thought by a large majority of physicists to be non thermal could be a back body spectrum! Precisely because GRBs are dynamical objects: the quantities we observe are very different from the quantities at the source. Here we want to stress that unlike other astrophysical objects, GRBs are very different, and in interpreting the data we must be very careful of what could be "optical illusions"! It is often written that "the timescales of the bursts are of the order of 0.1 to 100 seconds". In order to avoid confusion we will introduce two times: the arrival time t_a that we observe (i.e. 0.1–100 s), and the laboratory time t. For signals emitted by a pulse moving with velocity v in the laboratory frame, we have the following relation between the interval of arrival time Δt_a and the corresponding interval of laboratory time Δt:

$$\Delta t_a = \left(t_0 + \Delta t + \frac{R_0 - r}{c}\right) - \left(t_0 + \frac{R_0}{c}\right) = \Delta t - \frac{r}{c}. \qquad (6)$$

For simplicity in what follows we indicate by t_a the interval of arrival time measured from the reception of a light signal emitted at the onset of the gravitational collapse. Analogously, t indicates the laboratory time interval measured from the time of the gravitational collapse. In this case, Eq.(6) can be written simply as:

$$t_a = t - \frac{\int_0^t v(t')\,dt' + r_{ds}}{c} \times cos(\theta) + \frac{r_{ds}}{c}, \qquad (7)$$

where θ is the angle in the laboratory frame between the radial direction of each point of the ABM surface and the line of sight. We have chosen the dyadosphere radius r_{ds} as the value of r at $t = 0$. It is important to stress that, although there is a Lorentz gamma factor, Eq. (7) is not a Lorentz transformation, which by its own nature is linear and refers to a specific value of the Lorentz gamma factor at a given laboratory time. The transformation in Eq. (7) is nonlinear in the Lorentz gamma factor and depends on all the values of the gamma factor of the source from the time $t = 0$ to the laboratory time t [29].

We have assumed that, following the gravitational collapse process, a region of very low baryonic contamination exists in the dyadosphere all the way to the remnant of the progenitor star. The limit on such baryonic contamination is simply given by: $\rho_{B_c} \ll m_p n_{e^+e^-}(r)$ where ρ_{B_c} is the mass-energy density of baryons. Due to the high density of pairs $n_{e^+e^-}$, we can put a conservative limit near the dyadosphere radius $\rho_{B_c} \ll m_p n_{e^+e^-}(r) \simeq 3.2 \cdot 10^8$. Such conditions can be easily satisfied in the collapse to an EMBH, but not necessarily in a collapse to a neutron star. Consequently we have solved the equations governing a plasma composed solely of e^+e^--pairs and electromagnetic radiation, starting at time zero from the dyadosphere configurations corresponding to constant density. Numerical simulations from the Livermore code have shown very clearly the self organization of the expanding plasma in a very sharp pulse which we have defined as the pair-electromagnetic pulse (PEM pulse), in analogy with the EM pulse observed in nuclear explosions. The PEM pulse expands initially in a region of very low baryonic contamination created by the process of gravitational collapse. As it moves further out the baryonic remnant of the progenitor star is encountered. The baryonic matter remnant is assumed to be distributed well outside the dyadosphere in a shell of thickness Δ between an inner radius $r_{\rm in}$ and an outer radius $r_{\rm out} = r_{\rm in} + \Delta$ at a distance from the EMBH at which the original PEM pulse expanding in vacuum has not yet reached transparency. As an example we choose $r_{\rm in} = 100 r_{\rm ds}$ and $\Delta = 10 r_{\rm ds}$. The total baryonic mass $M_B = N_B m_p$ is assumed to be a fraction of the dyadosphere initial total energy ($E_{\rm dya}$). The total baryon-number N_B is then expressed as a function of the dimensionless parameter B given by

$$B = \frac{N_B m_p c^2}{E_{\rm dya}}, \tag{8}$$

where B value is in the range $10^{-8} - 10^{-2}$ and m_p is the proton mass. We shall see below the paramount importance of B in the determination of the features of the GRB.

As the PEM pulse reaches the region $r_{\rm in} < r < r_{\rm out}$, it interacts with the baryonic matter which is assumed to be at rest. In our simplified quasi-analytic model we make the following assumptions to describe this interaction:

- the PEM pulse does not change its geometry during the interaction;
- the collision between the PEM pulse and the baryonic matter is assumed to be inelastic;
- the baryonic matter reaches thermal equilibrium with the photons and pairs of the PEM pulse.

These assumptions are valid if: (i) the total energy of the PEM pulse is much larger than the total mass-energy of baryonic matter M_B, $10^{-8} < B < 10^{-2}$, (ii) the ratio of the comoving number density of pairs and baryons at the moment of collision $n_{e^+e^-}/n^\circ_B$ is very high (e.g., $10^6 < n_{e^+e^-}/n^\circ_B < 10^{12}$) and (iii) the PEM pulse has a large value of the gamma factor ($100 < \bar\gamma$). In the collision between the PEM pulse and the baryonic matter at $r_{\rm out} > r > r_{\rm in}$, we impose total conservation of energy and momentum. We consider the collision process between two radii r_2, r_1 satisfying $r_{\rm out} > r_2 > r_1 > r_{\rm in}$ and $r_2 - r_1 \ll \Delta$. The amount of baryonic mass acquired by the PEM pulse is $\Delta M = \frac{M_B}{V_B}\frac{4\pi}{3}(r_2^3 - r_1^3)$, where M_B/V_B is the mean-density of baryonic matter at rest. The conservation of total energy leads to the estimate of the corresponding quantities before (with "∘") and after such a collision

$$(\Gamma\bar\epsilon_\circ + \bar\rho^\circ_B)\bar\gamma^2_\circ \mathcal{V}_\circ + \Delta M = (\Gamma\bar\epsilon + \bar\rho_B + \frac{\Delta M}{V} + \Gamma\Delta\bar\epsilon)\bar\gamma^2\mathcal{V}, \tag{9}$$

where $\Delta\bar\epsilon$ is the corresponding increase of internal energy due to the collision and $\bar n^\circ_B = \frac{N_B}{V_B}$, $\bar\rho^\circ_B = m_p \bar n^\circ_B c^2$. Similarly momentum-conservation gives

$$(\Gamma\bar\epsilon_\circ + \bar\rho^\circ_B)\bar\gamma_\circ U^\circ_r \mathcal{V}_\circ = (\Gamma\bar\epsilon + \bar\rho_B + \frac{\Delta M}{V} + \Gamma\Delta\bar\epsilon)\bar\gamma U_r \mathcal{V}, \tag{10}$$

where the radial component of the four-velocity of the PEM pulse is $U^\circ_r = \sqrt{\bar\gamma^2_\circ - 1}$ and Γ is the thermal index. We then find

$$\Delta\bar\epsilon = \frac{1}{\Gamma}\left[(\Gamma\bar\epsilon_\circ + \bar\rho^\circ_B)\frac{\bar\gamma_\circ U^\circ_r \mathcal{V}_\circ}{\bar\gamma U_r \mathcal{V}} - (\Gamma\bar\epsilon + \bar\rho_B + \frac{\Delta M}{V})\right], \tag{11}$$

$$\bar\gamma = \frac{a}{\sqrt{a^2 - 1}}, \qquad a \equiv \frac{\bar\gamma_\circ}{U^\circ_r} + \frac{\Delta M}{(\Gamma\bar\epsilon_\circ + \bar\rho^\circ_B)\bar\gamma_\circ U^\circ_r \mathcal{V}_\circ}. \tag{12}$$

These equations determine the gamma factor $\bar\gamma$ and the internal energy density $\bar\epsilon = \bar\epsilon_\circ + \Delta\bar\epsilon$ in the capture process of baryonic matter by the PEM pulse.

After the engulfment of the baryonic matter of the remnant the plasma formed of e^+e^--pairs, electromagnetic radiation and baryonic matter expands again as a sharp pulse, namely the PEMB pulse. As the condition of transparency is reached the *injector phase* terminates. The electromagnetic energy of the PEMB pulse is released in the form of free-streaming photons - the proper GRB. The remaining energy of the PEMB pulse is released as an accelerated-baryonic-matter (ABM) pulse.

In analogy and by extension of the results obtained for the PEM and PEMB pulse cases, we also assume that the expansion of the ABM pulse through the ISM occurs keeping its width constant in the laboratory frame, although the results are quite insensitive to this assumption. We assume then that this interaction can be represented by a sequence of inelastic collisions of the expanding ABM pulse with a large number of thin and cold ISM spherical shells at rest with respect to the central EMBH. Each of these swept up shells of thickness Δr has a mass $\Delta M_{\rm ism}$ and is assumed to be located between two radial distances r_1 and r_2 (where $r_2 - r_1 = \Delta r \ll r_1$) in the laboratory frame. These collisions create an internal energy $\Delta E_{\rm int}$. We indicate by $\Delta\epsilon$ the increase in the proper internal energy density due to the collision with a single shell and by ρ_B the proper energy density of the swept up baryonic matter. This includes the baryonic matter composing the remnant around the central EMBH, already swept up in the PEMB pulse formation, and the baryonic matter from the ISM swept up by the ABM pulse: $\rho_B = \frac{(M_B + M_{\rm ism})c^2}{V}$. Here V is the ABM pulse volume in the comoving frame, M_B is the mass of the baryonic remnant and $M_{\rm ism}$ is the ISM mass swept up from the transparency point through the r in the laboratory frame: $M_{\rm ism} = m_p n_{\rm ism}\frac{4\pi}{3}\left(r^3 - r_\circ{}^3\right)$, where m_p is the proton mass and $n_{\rm ism}$ is the number density of the ISM in the laboratory frame. The energy conservation law in the laboratory frame at a generic step of the collision process is given by

$$\rho_{B_1}\gamma_1{}^2\mathcal{V}_1 + \Delta M_{\rm ism}c^2 = \left(\rho_{B_1}\frac{V_1}{V_2} + \frac{\Delta M_{\rm ism}c^2}{V_2} + \Delta\epsilon\right)\gamma_2{}^2\mathcal{V}_2, \tag{13}$$

where the quantities with the index "1" are calculated before the collision of the ABM pulse with an elementary shell of thickness Δr and the quantities with "2" after the collision, γ is the gamma factor and $\mathcal{V}$ the volume of the ABM pulse in the laboratory frame so that $V = \gamma\mathcal{V}$.

The momentum conservation law in the laboratory frame is given by

$$\rho_{B_1}\gamma_1 U_{r_1}\mathcal{V}_1 = \left(\rho_{B_1}\frac{V_1}{V_2} + \frac{\Delta M_{\text{ism}}c^2}{V_2} + \Delta\epsilon\right)\gamma_2 U_{r_2}\mathcal{V}_2, \tag{14}$$

where $U_r = \sqrt{\gamma^2 - 1}$ is the radial covariant component of the four-velocity vector.

We thus obtain

$$\Delta\epsilon = \rho_{B_1}\frac{\gamma_1 U_{r_1}\mathcal{V}_1}{\gamma_2 U_{r_2}\mathcal{V}_2} - \left(\rho_{B_1}\frac{V_1}{V_2} + \frac{\Delta M_{\text{ism}}c^2}{V_2}\right); \tag{15}$$

$$\gamma_2 = \frac{a}{\sqrt{a^2-1}}, \qquad a \equiv \frac{\gamma_1}{U_{r_1}} + \frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}\gamma_1 U_{r_1}\mathcal{V}_1}. \tag{16}$$

We can use for $\Delta\varepsilon$ the following expression

$$\Delta\varepsilon = \frac{E_{\text{int}_2}}{V_2} - \frac{E_{\text{int}_1}}{V_1} = \frac{E_{\text{int}_1} + \Delta E_{\text{int}}}{V_2} - \frac{E_{\text{int}_1}}{V_1} = \frac{\Delta E_{\text{int}}}{V_2} \tag{17}$$

because we have assumed a "fully radiative regime" and so $E_{\text{int}_1} = 0$. Substituting Eq.(16) in Eq.(15) and applying Eq.(17), we obtain:

$$\Delta E_{\text{int}} = \rho_{B_1}V_1\sqrt{1 + 2\gamma_1\frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}V_1} + \left(\frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}V_1}\right)^2} - \rho_{B_1}V_1\left(1 + \frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}V_1}\right); \tag{18}$$

$$\gamma_2 = \frac{\gamma_1 + \frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}V_1}}{\sqrt{1 + 2\gamma_1\frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}V_1} + \left(\frac{\Delta M_{\text{ism}}c^2}{\rho_{B_1}V_1}\right)^2}}. \tag{19}$$

These relativistic hydrodynamic (RH) equations have to be numerically integrated. These are the actual set of equations we have integrated in the EMBH theory.

3.3 Inside the EMBH: unveiling the secret of the GRB

The fundamental hydrodynamic equations described previously have been integrated numerically. The e^+e^- plasma moves outward from the EMBH reaching a very unique relativistic configuration: the plasma self-organizes in a sharp pulse which expands in the comoving frame exactly by the amount which compensates for the Lorentz contraction in the laboratory frame. The sharp pulse remains of constant thickness in the laboratory frame and self-propels outwards reaching ultrarelativistic regimes, with gamma factors larger than 10^2, in a few dyadosphere crossing times. We recall that, in analogy with the electromagnetic (EM) pulse observed in a thermonuclear explosion on Earth, we have defined this more energetic pulse formed by electron–positron pairs and electromagnetic radiation a pair-electromagnetic-pulse or PEM pulse. We have described the interaction of the PEM pulse with the baryonic remnant of mass M_B left over from the gravitational collapse of the progenitor star. We gave the details of the decrease of the gamma factor and the corresponding increase in the internal energy during the collision. The dimensionless parameter $B = M_Bc^2/E_{dya}$ which measures the baryonic mass of the remnant in units of the E_{dya} is the second fundamental free parameter of the EMBH theory. Then we have computed the further expansion of the e^+e^- plasma, after the engulfment of the baryonic remnant of the progenitor star. The evolution of the PEMB pulse is shown to depend only on E_{dya} and B: the PEMB pulse is degenerate in the mass and charge parameters of the EMBH and rather independent of the exact location of the baryonic matter of the remnant. The relevant thermodynamical quantities of the PEMB pulse, the temperature in the different frames and the e^+e^- pair densities are all computable at each time. The power of the EMBH model is to follow all parameters as a function of time.

As the condition of transparency is reached, the injector phase is concluded with the emission of a sharp burst of electromagnetic radiation and an accelerated beam of highly relativistic baryons. We recall that we have respectively defined the radiation burst (the P-GRB) and the accelerated-baryonic-matter (ABM) pulse [30]. By computing for a fixed value of the EMBH different PEMB pulses corresponding to selected values of B in the range $[10^{-8}$–$10^{-2}]$, it has been possible to obtain a crucial universal diagram. In the limit of $B \to 10^{-8}$ or smaller almost all E_{dya} is emitted in the P-GRB and a negligible fraction is emitted in the kinetic energy $E_{Baryons}$ of the baryonic matter and therefore in the afterglow. On the other hand in the limit $B \to 10^{-2}$ which is also the limit of validity of our theoretical framework, almost all E_{dya} is transferred to $E_{Baryons}$ and gives origin to the afterglow and the intensity of the P-GRB correspondingly decreases. We have identified the limiting case of negligible values of B with the process of emission of the so-called "short bursts". A complementary result reinforcing such an identification comes from the thermodynamical properties of the P-GRB: the hardness of the spectrum decreases for increasing values of B. The injector phase is concluded by the emission of the P-GRB and the ABM pulse, as the condition of transparency is reached. The beam-target phase, in which the accelerated baryonic matter (ABM) generated in the injector phase collides with the ISM, gives origin to the afterglow. Again for simplicity we have adopted a minimum set of assumptions:

- The ABM pulse is assumed to collide with a constant homogeneous interstellar medium of number density $n_{\text{ism}} \sim 1$ cm^{-3}. The energy emitted in the collision is assumed to be instantaneously radiated away (fully radiative condition). The collision and emission process is described using spherical symmetry, taking only the radial approximation and neglecting all the delayed emission due to off-axis scattered radiation.

- Special attention is given to computing numerically the power of the afterglow as a function of the arrival time using the correct governing equations for the space-time transformations in line with the RSTT paradigm. iii) Finally, approximate solutions are adopted in order to obtain the determination of the power law exponents of the afterglow flux and to compare and contrast them with the observational results as well as with alternative results in the literature.

4 The Results

We will now confront our theory with observational results and see how the EMBH model interprets the data. As we mentioned previously it is a kind of "dialogue" with the source since the apparent observation is sometimes different from what happens near the GRB.

4.1 The EMBH shapes of GRB

In this section we will give a very simple and consistent desciption of the bolometric light curve of the GRB by the EMBH model. We stress that GRB are dynamical processes and so the time-evolution is very important. In a first approximation we expand Eqs.(18, 19) to second order in the quantity $\frac{\Delta M_{\rm ism} c^2}{\rho_{B_1} V_1} \ll 1$. We obtain the following expressions:

$$\Delta E_{\rm int} = (\gamma_1 - 1)\,\Delta M_{\rm ism} c^2 - \frac{1}{2}\frac{\gamma_1^2 - 1}{M_B + M_{\rm ism}}\left(\Delta M_{\rm ism}\right)^2 c^2\,, \tag{20}$$

$$\Delta\gamma = -\frac{\gamma_1^2 - 1}{M_B + M_{\rm ism}}\Delta M_{\rm ism} + \frac{3}{2}\gamma_1\frac{\gamma_1^2 - 1}{\left(M_B + M_{\rm ism}\right)^2}\left(\Delta M_{\rm ism}\right)^2\,, \tag{21}$$

where we set $\Delta\gamma \equiv \gamma_2 - \gamma_1$ and used $\rho_{B_1} V_1 \equiv \left(M_B + M_{\rm ism}\right) c^2$. In the limit $\Delta E_{\rm int} \to dE_{\rm int}$, $\Delta\gamma \to d\gamma$, and $\Delta M_{\rm ism} \to dM_{\rm ism}$, neglecting also second order terms. Also

$$dM_{\rm ism} = 4\pi r^2 m_p n_{\rm ism} dr = 4\pi r^2 m_p n_{\rm ism} v dt, \quad v = \frac{dr}{dt}, \tag{22}$$

where the ISM number density is assumed for simplicity to be $n_{\rm ism} = 1\,{\rm cm}^{-3}$, we obtain:

$$dE_{\rm int} = (\gamma - 1)\, dM_{\rm ism} c^2\,, \tag{23}$$

$$d\gamma = -\frac{\gamma^2 - 1}{M_B + M_{\rm ism}} dM_{\rm ism}\,. \tag{24}$$

Equations (23, 24) are the limiting cases of Taub's hydrodynamical equations. They have been referred to in the GRB literature as the Blandford–McKee equations.

From Eqs. (23, 24), it follows that the emitted flux in the laboratory frame is given by:

$$\frac{dE}{dt} = 4\pi r^2 n_{\rm ism} m_p v\gamma\left(\gamma - 1\right) c^2, \tag{25}$$

and the corresponding flux in the detector at arrival time,

$$\frac{dE}{dt_a^d} = \left[\frac{dt}{dt_a^d}\frac{dE}{dt}\right]_{t=t(t_a^d)} = 4\pi n_{\rm ism} m_p c^2\left[v r^2 \gamma\left(\gamma - 1\right)\frac{dt}{dt_a^d}\right]_{t=t(t_a^d)}\,. \tag{26}$$

Just after the transparency condition is reached, the ISM matter involved is small and $\gamma \simeq \gamma_\circ$. We can then write the equation describing the ABM pulse motion in this phase as $r\left(t\right) = vt$ with $v \simeq c$. The following relation between

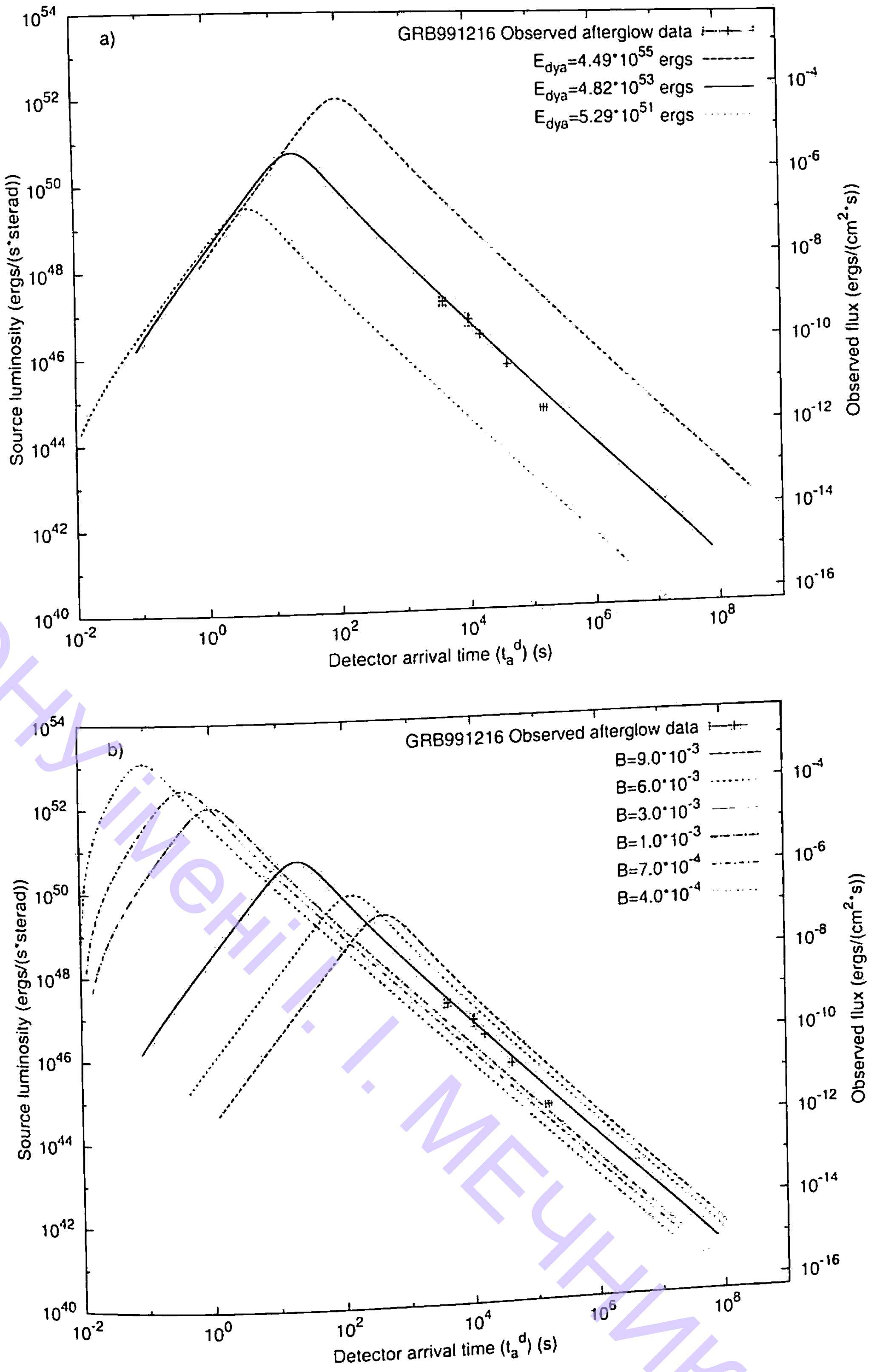


Figure 1: The light curve of GRBs as predicted in the EMBH theory is displayed as a function of the theory parameters: the energy of dyadosphere and the baryonic mass. We have used as an example 991216.

laboratory time and arrival time can be used: $t = 2\gamma_\circ{}^2 t_a = \frac{2\gamma_\circ{}^2}{1+z} t_a^d$. The luminosity in laboratory time is:

$$\frac{dE}{dt} \propto \gamma_\circ^2 n_{\rm ism} t^2 \tag{27}$$

and in arrival time:

$$\frac{dE}{dt_a^d} \propto \frac{\gamma_\circ^8 n_{\rm ism}}{(1+z)^3} \left(t_a^d\right)^2 . \tag{28}$$

The slope has a characteristic power of 2 in arrival time. Then $M_{\rm ism}$ is increasing and γ can no longer be considered constant and strongly decreases, but v is still almost constant, equal to c. As a consequence, we can still say that $r\,(t) = vt$ with $v = c$, and the relation between laboratory- and arrival- times is given by the following "effective" relation: $t \propto \left(t_a^d\right)^{0.20}$, as a result of the best fit of the numerical data in this region. We can show that the flux in the laboratory frame is given by

$$\frac{dE}{dt} \propto \gamma_P^2 t_P^6 n_{\rm ism} t^{-4} \tag{29}$$

and equal to

$$\frac{dE}{dt_a^d} \propto n_{\rm ism} \left(t_a^d\right)^{-1.6} \tag{30}$$

in arrival time in agreement with the observations. It is believed that there are two classes of GRB: those shorter and those longer than 2 seconds, called "short" and "long" respectively. The EMBH model explains naturally these observational facts as simply related to the parameters B of the model. It also reproduces the fact that the "short" spectrum is harder and the "long" one is soft. In our model, short bursts are P-GRB: for a value of B typically less than $\sim 10^{-5}$, the intensity of the P-GRB is dominant over the afterglow and then the bursts appear shorter. In the opposite case, for B larger than $\sim 10^{-4}$, the P-GRB is negligible and all the intensity is in the afterglow.

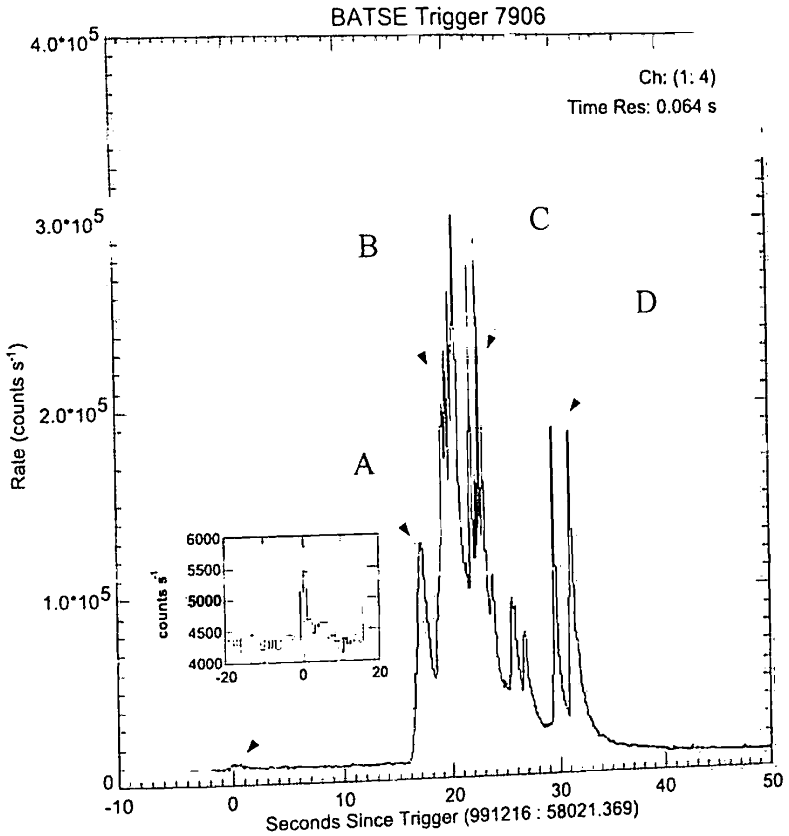
4.2 Interpreting the substructures

At the third GRB meeting in Rome, G. Ghisellini pointed out some open problems in GRB. In particular one associated with the origin of substructures: "are internal shocks efficient enough?" [31].

It was often claimed that the solution was obtained by using the internal shock model. Here we will compute the lightcurve in the EMBH theory assuming external shocks. The afterglow is emitted as the ABM pulse ploughs through the interstellar matter engulfing new baryonic material. Previously, we have adopted the simplified assumption that the interstellar medium is at constant density ($n_{ism} = 1$ cm^{-3}). Consequently, the afterglow emission is very smooth in time. We want to see if in this framework we can also explain most of the time variability observed by BATSE, all of which except for the P-GRB should correspond to the beam-target phase in the IBS paradigm.

In the comoving frame the energy per unit volume and per solid angle is

$$\left(\frac{dE}{dV d\Omega}\right)_\circ = \frac{\Delta\varepsilon}{4\pi} \tag{31}$$



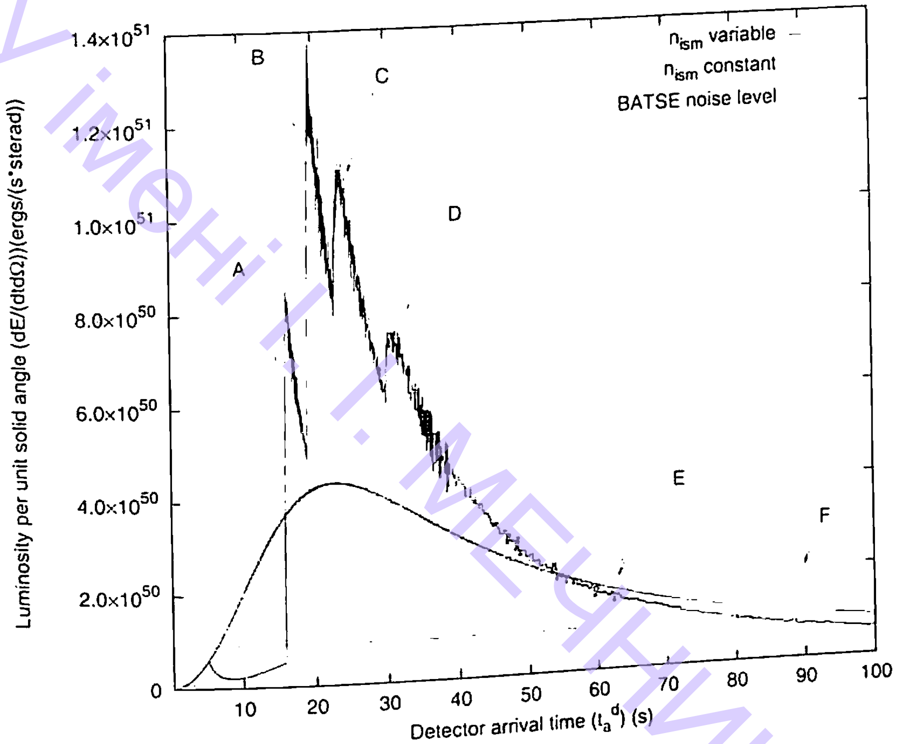


Figure 2: The numerical computation of the structures originating from the external shocks is shown for the burst 991216 where we have reproduced also the BATSE lightcurve. It is clear that the EMBH is able to explain the high variation due to collision of the ABM pulse with the interstellar matter.

with $\Delta\varepsilon$ the internal energy density developed in the collision, as the emission is isotropic in this frame. The total number of photons emitted is an invariant quantity independent of the frame used. Thus we can compute this quantity as seen by an observer in the comoving frame (which we denote with the subscript "o") and by an observer in the laboratory frame (which we denote with no subscripts). We find

$$\frac{dN_\gamma}{dtd\Omega d\Sigma} = \int_{shell} \left(\frac{dN_\gamma}{dtd\Omega d\Sigma}\right)_\circ \Lambda^{-3} \cos\vartheta\,, \tag{32}$$

where $\cos\vartheta$ comes from the projection of the elementary surface of the shell on the direction of propagation and $\Lambda = \gamma(1-\beta\cos\vartheta)$ is the Doppler factor introduced in the two following differential transformations

$$d\Omega_\circ = d\Omega \times \Lambda^{-2} \tag{33}$$

for the solid angle transformation and

$$dt_\circ = dt \times \Lambda^{-1} \tag{34}$$

for the time transformation. The integration in $d\Sigma$ is performed over the visible area of the ABM pulse at laboratory time t, namely with $0 \leq \vartheta \leq \vartheta_{max}$ and ϑ_{max} defined by $cos(\vartheta_{max}) = \frac{v}{c}$. An extra Λ factor comes from the energy transformation:

$$E_\circ = E \times \Lambda\,. \tag{35}$$

Thus finally we obtain:

$$\frac{dE}{dtd\Omega d\Sigma} = \int_{shell} \left(\frac{dE}{dtd\Omega d\Sigma}\right)_\circ \Lambda^{-4} \cos\vartheta\,, \tag{36}$$

where we clearly identify $\left(\frac{dE}{dtd\Omega d\Sigma}\right)_\circ$ as the energy density in comoving frame up to a factor $\frac{v}{4\pi}$ (see Eq. (31)). Then we have:

$$\frac{dE}{dtd\Omega} = \int_{shell} \frac{\Delta\varepsilon}{4\pi} v \cos\vartheta\, \Lambda^{-4}\, d\Sigma\,, \tag{37}$$

where the integration in $d\Sigma$ is performed over the ABM pulse visible area at laboratory time t, namely with $0 \leq \vartheta \leq \vartheta_{max}$.

Equstion (37) gives us the energy emitted toward the observer per unit solid angle and per unit laboratory time t in the laboratory frame. What we really need is the energy emitted per unit solid angle and per unit detector arrival time t_a^d, so we must use the complete relation between t_a^d and t given in Eq.(7). First we multiply the integrand in Eq.(37) by the factor (dt/dt_a^d) to transform the energy density generated per unit of laboratory time t into the energy density generated per unit arrival time t_a^d. Then we integrate with respect to $d\Sigma$ over the *equitemporal surface* of constant arrival time t_a^d instead of the ABM pulse visible area at laboratory time t. The analog of Eq. (37) for the source luminosity in detector arrival time is:

$$\frac{dE_\gamma}{dt_a^d d\Omega} = \int_{EQTS} \frac{\Delta\varepsilon}{4\pi} v \cos\vartheta\, \Lambda^{-4} \frac{dt}{dt_a^d} d\Sigma\,. \tag{38}$$

The results are given in Figure 2 for the burst 991216. See also: [32] and [33]. We obtain, in agreement with observations:

1. the intensity of the A, B, and C peaks as a function of the ISM inhomogneities;
2. the fast rise and exponential decay shape for each peak;
3. a continuous and smooth emission between the peaks.

Interestingly, the signal from shells E and F, which have a density inhomogeneity comparable to A, are undetectable. This is due to a variety of relativistic effects and partly to the spreading in the arrival time which for A, corresponding to $\gamma = 303.8$ is 0.4 s, while for E (F) corresponding to $\gamma = 57.23$ (56.24), is 10.2 s (10.6 s). In the case of D, agreement with the arrival time is reached, but not the double-peaked structure. The ABM pulse visible area diameter at the moment of interaction with the D shell is $\simeq 10^{15}$ cm, equal to the extension of the ISM shell. Under these conditions, the concentric shell approximation does not hold anymore: the disagreement with the observations simply manifests the need for a more detailed description of the three-dimensional nature of the ISM cloud.

4.3 The spectrum

We have shown that the observed spectrum is a broken power law known as Band relation [34]. The fireball model assumes that electrons are accelerated by shock into a power-law distribution $N_e \simeq \gamma_e^{-p}$ where p is the spectral index. It is also assumed that a considerable magnetic field builds up behind the shock, characterized by a fraction ϵ_B of equipartition. The spectrum is explained naturally by synchrotron radiation of these electrons using the two assumptions above.

Here we adopt three basic assumptions: a) the resulting radiation as viewed in the comoving frame during the afterglow phase has a thermal spectrum; b) the ISM swept up by the front of the shock wave, with a Lorentz gamma factor between 300 and 2, is responsible for this thermal emission; and c) the expansion occurs with spherical symmetry. These three assumptions are different from the ones adopted in the literature, in which the afterglow emission is believed to originate from synchrotron emission in the production of the shock or reverse shock generated when the assumed jet-like ejecta encounters the external medium. The structure of the shock is determined by mass, momentum and energy con- servation, i.e., the constancy of the specific enthalpy, which are standard conditions in shock rest frames and have already been used in our derivation. The only additional free parameter of our model is the size of the effective emitting area in the shock wave front (A_{eff}). The temperature T of the black body in the co-moving frame is:

$$T = \left(\frac{\Delta E_{int}}{4\pi r^2 \Delta\tau\sigma R}\right), \tag{39}$$

where $R = \frac{A_{eff}}{A_{ABM}}$ is the ratio between the effective emitting area and the ABM pulse surface A_{ABM}, ΔE_{int} is the internal energy developed in the collision with the ISM in a time interval $\Delta\tau$ in the co-moving frame and σ is the Stefan–Boltzmann constant. The ratio R, which is *a priori* a function that varies as the system evolves, is evaluated at every given laboratory time t. All the subsequent steps are now uniquely determined by the equations of motion of the system. The basic tool in this calculation involves the definition of the EQuiTemporal Surfaces (EQTS) for the relativistic expanding ABM pulse as seen by an asymptotic observer.

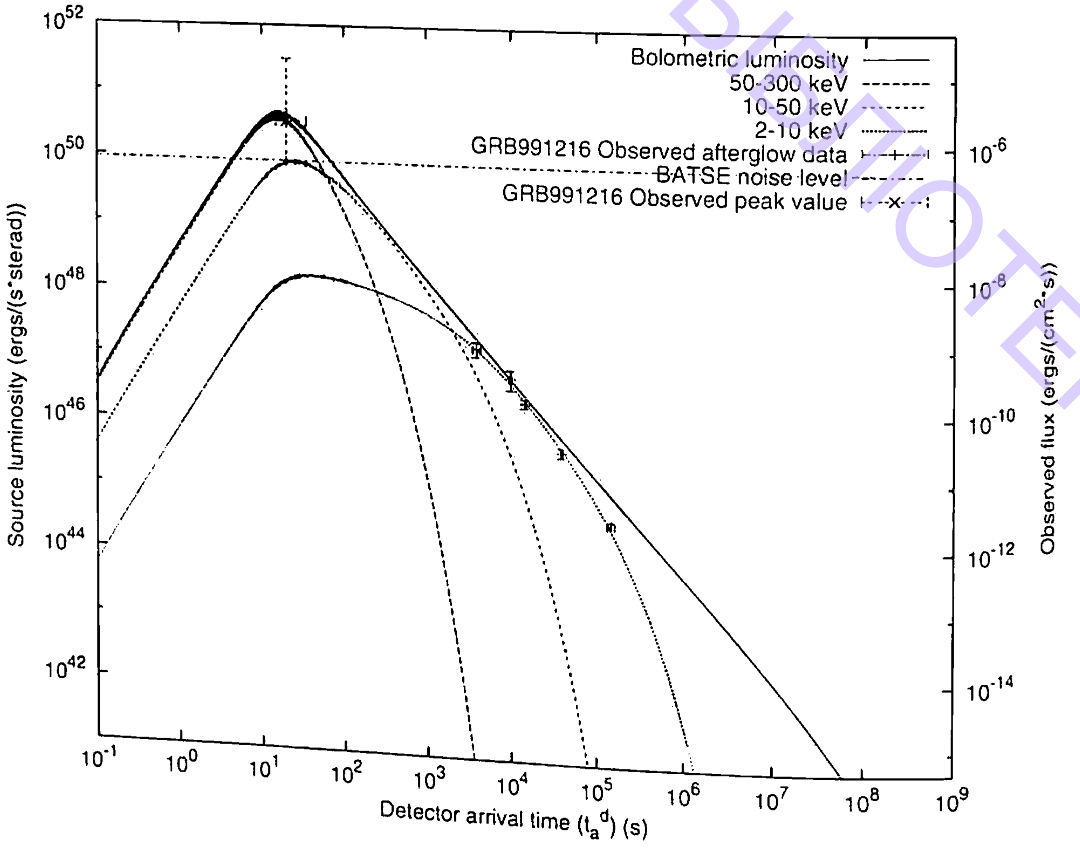


Figure 3: Best fit of the afterglow data of GRB 991216. The solid curve is the bolometric luminosity. The three dotted curves correspond to the luminosities in the bands 50300 keV, 1050 keV and 210 keV respectively. Near the E-APE, where the BATSE data are present, almost all the luminosity is in the 50300 keV band.

We are now ready to evaluate the source luminosity in a given energy band. The source luminosity at a detector arrival time a, per unit solid angle $d\Omega$ and in the energy band [1, 2] is given by

$$\frac{dE_\gamma}{dt_a d\Omega} = \int_{EQTS} \frac{\Delta\epsilon}{4\pi} vcos(\theta)\Lambda^{-4} \frac{dt}{dt_a} W(\nu_1, \nu_2, T_{arr}) d\Sigma, \tag{40}$$

where $\Delta\epsilon = \Delta E_{int}/V$ is the energy density released in the interaction of the ABM pulse with the ISM inhomogeneities measured in the comoving frame, $\Lambda = \gamma(1 -$ $vccos(\theta))$ is the Doppler factor, $W(\nu_1, \nu_2, T_{arr})$ is an effective weight required to evaluate only the contributions in the energy band $[\nu_1, \nu_2]$, $d\Sigma$ is the surface element of the EQTS at detector arrival time t_a on which the integration is performed (see

also [19]) and T_{arr} is the observed temperature of the radiation emitted from $d\Sigma$:

$$T_{arr} = \frac{T}{\gamma(1 - vc\cos(\theta))} \frac{1}{1+z}. \tag{41}$$

The effective weight $W(\nu_1, \nu_2, T_{arr}$ is given by the ratio of the integral over the given energy band of a Planckian distribution at a temperature T_{arr} to the total integral aT^4_{arr}:

$$W(\nu_1, \nu_2, T_{arr} = \frac{1}{aT^4_{arr}} \int_{\nu_1}^{\nu_2} \rho(T_{arr}, \nu) d(\frac{h\nu}{c})^3, \tag{42}$$

where $\rho(T_{arr}, \nu)$ is the Planckian distribution at temperature T_{arr}:

$$\rho(T_{arr}, \nu) = \frac{2}{h^3} \frac{h\nu}{\exp^{h\nu kT_{arr}} - 1}. \tag{43}$$

We can now proceed to the best fit of the observed data using GRB 991216 as the prototype. Such an estimate is perfectly well defined from a theoretical point of view, although from a numerical point of view the integration on all the EQTS and the associated relativistic transformations have raised time consuming difficulties. Almost 10^8 paths with different temperatures and different Lorentz boosts had to be considered in the integration over the EQTS. We give in Figure 3 the results for the three energy bands 50300 keV (BATSE), 210 keV (R-XTE, Chandra) and 1050 keV. It is most remarkable that the best fit is obtained simply by a factor R, which is monotonically varying in the range: $5.01 \times 10^{-12} \leq R \leq 3.01 \times 10^{-8}$ in correspondence with the beginning of the afterglow emission and the last observation by Chandra at ~ 37 hr after the GRB. We point out the perfect agreement with the data obtained by BATSE Rapid Burst Response7 in the energy range 50300 keV. The fit of the data obtained by the R-XTE and Chandra satellites in the energy range 210 keV, is very impressive.

4.4 The tryptich paradigm

The relation between GRB and supernovae is one of the most complex aspects to be addressed by our model, which needs the understanding of new fields of general relativistic physics in relation to yet unexplored many-body solutions in a substantially new astrophysical scenario. As we will show in two systems GRB980425/SN1998bw and GRB030329/SN2003dh there is in each one, the possibility of an astrophysical triptych formed by:

1) the formation of the black hole and the emission of the GRB,
2) the gravitational collapse of an evolved companion star, leading to a supernova [35],
3) a clearly identified URCA source whose nature appears to be of great interest.

This new astrophysical scenario presents new challenges:

a) The identification of the physical reasons of the instability leading to the gravitational collapse of a $10M_\odot$ star, giving origin to a black hole. Such an implosion must occur radially with negligible mass of the remnant ($B \leq 10^{-2}$).

b) The identification of the physical reasons for the instability leading to the gravitational collapse of an evolved companion star, giving origin to a supernova.

c) The theoretical issues related to the URCA sources, from the physics of black holes, to the physical processes occurring in the expanding supernova remnants, to the very exciting possibility that we are observing for the first time a newly born neutron star. The main effort next is to show that the detailed understanding we have reached for the GRB phenomenon and its afterglow allows us to state convincingly, that the URCA source, contrary to what appears currently in the literature, is not part of the GRB or of its afterglow.

Currently it is attempted to explain both the supernova and the GRB (GRB 980425 / SN 1998bw) as two aspects of a single phenomenon assuming that the GRB takes its origin from a specially strong and yet unobserved supernova process: a hypernova [36–38]. We have taken a very different approach, following Ciceros classic aphorism "divide et impera", which was adopted as the motto of the Roman empire. In this case, which is a very complex system, we plan to divide and identify the truly independent physical constituents and conquer the understanding of the underlying astrophysical process, leading to an unexpected and much richer scenario.

In addition to the source GRB 980425 and the supernova SN1998bw, two X-ray sources have been found by BeppoSAX in the error box for the location of GRB 980425: S1 and S2 ([39]), which have been traditionally interpreted either as a background source or as a part of the GRB afterglow. Our approach has been to first comprehend the entire afterglow of GRB 980425 within our theory. This allows the computation of the luminosity in given energy bands, the spectra, the Lorentz gamma factors, and more generally the dynamical aspects of the source. Having characterized the features of GRB 980425, we can gradually approach the remaining scenario, disentangling the GRB observations from those of the supernova and then both the GRB and the supernova observations from those of the sources S1 and S2. This leads to a natural identification of distinct events and to their autonomous astrophysical characterization. Our best fit for GRB 980425 corresponds to $E_{dya} = 1.1 \times 10^{48}$ erg, $B = 7 \times 10^{-3}$ and the ISM average density is found to be ~ 0.02 particle cm^{-3}. The plasma temperature and the total number of pairs in the dyadosphere are respectively T = 1.028MeV and $N_{e\pm} = 5.3 \times 10^{53}$. The P-GRB is below the threshold and in the case of this source is not observable. The characteristic parameter $\mathcal{R}$, defining the filamentary structure of the ISM, monotonically decreases from 4.81×10^{-10} to 2.65×10^{-12}. The results are given in Figure 4 where the bolometric luminosity is represented together with the optical data of SN1998bw, the source S1 and the source S2. It is clear that GRB 980425 is separated both from the supernova data and from the sources S1 and S2. While the occurrence of the supernova in relation to the GRB has already been discussed within the GRB-Supernova Time Sequence (GSTS) paradigm, we would like to address here the nature of the source S1 which we have named, in celebration of the work of Gamow and Shoenberg, URCA-1. It is clear, from the theoretical predictions of the afterglow luminosity, that URCA-1 cannot be part of the afterglow. There are three different possibilities for such source: 1) Its possible relation to the black hole formed during the process of gravitational collapse leading to the GRB emission. 2) Its possible relation to emission originating in the early phases of the expansion of the supernova remnant. 3) The very exciting possibility that for the first time we are observing a newly born neutron star out of the supernova phenomenon. While some general considerations will be discussed in the conclusions, we would like to stress here the paramount importance of following the further time history of URCA-1 and of the source S2. If, as we propose, S2 is a background source, its flux should be practically constant in time, unrelated to the GRB 980425 / SN1998bw system.

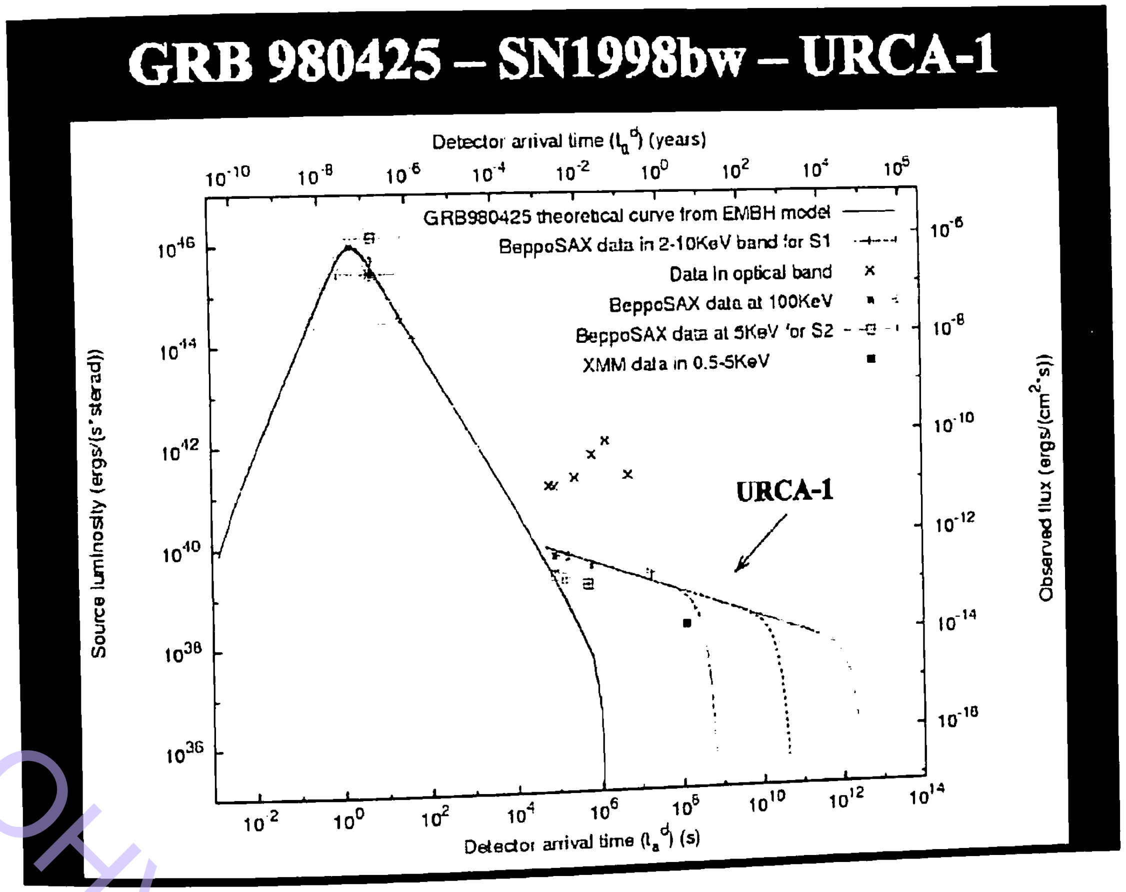


Figure 4: The tryptich paradigm in case of GRB980425 and SN1998bw. We see clearly that the new X-ray data are not related to the afterglow light curve as predicted by the EMBH model.

5 Conclusions and Outlook

The EMBH model of GRB has provided a remarkably successful framework in which to interpret observations of gamma-ray bursts, among them, the discovery of the P-GRB in the GRB991216, predicted by the EMBH model. Globally this theory is robust in explaining different bursts where the fireball model requires us to differentiate between bursts. A clear case is the GRB980425 where people used to classify it as a "different class". Another example is the "short" and the "long" as two separate classes of GRB. Instead we saw how the EMBH offers a natural framework in which all bursts find an interpretation. We impose two parameters at the begining for t=0 and the numerical code computes the light-curve, the temperature from which we will deduce the spectrum. Therefore it is not possible to "adjust" the curve after computation to fit better the data.

We are optimistic about the future of this theory and the success in explaining the data shows us that we are able to observe energy extraction from black holes as predicted by the EMBH model.

Acknowledgements

P. Chardonnet would like to thank the organizing committee for their kind hospitality and for this nice celebration in the centennial of the birth of George Gamow. P. Chardonnet is also grateful to the ICRA organization for financial support.

References

[1] R. Klebesadel, I.B. Strong, *ApJ.*, **182**, L85 (1973).

[2] A.K. Harding, *Phys. Reports.*, **206**, 328 (1991).

[3] C.A. Meegan et al., *Nature*, **355**, 143 (1992).

[4] F. Frontera, in M. Feroci, F. Frontera, N. Masetti, L. Piro (eds) *Third Rome Workshop on GRB in the afterglow era*, ASP Conf. Ser., **312**, 3 (2003).

[5] E. Costa et al., *Nature*, **387**, 783 (1997).

[6] J. van Paradijs et al., *Nature*, **386**, 868 (1997).

[7] S.G. Djorgovski et al., *Nature*, **387**, 876 (1997).

[8] M.R. Metzger et al., *Nature*, **387**, 878 (1997).

[9] D.Q. Lamb, and Z. Haiman in M. Feroci, F. Frontera, N. Masetti, L. Piro (eds) *Third Rome Workshop on GRB in the afterglow era*, ASP Conf. Ser., **312**, 287 (2003).

[10] L. Piro, in M. Feroci, F. Frontera, N. Masetti, L. Piro (eds) *Third Rome Workshop on GRB in the afterglow era*, ASP Conf. Ser., **312**, 149 (2003).

[11] J. Reeves, D. Watson, J. Osborne, P. O'Brien, and K. Pounds, in M. Feroci, F. Frontera, N. Masetti, L. Piro (eds) *Third Rome Workshop on GRB in the afterglow era*, ASP Conf. Ser., **312**, 181 (2003).

[12] S.G. Djorgovski et al., *Proceeding MGIX*, Word Scientific, p. 315 (2001).

[13] C.D. Dermer, *Phys. Rev. Lett.*, **68**, 1799 (1992).

[14] S.G. Djorgovski et al., in M. Feroci, F. Frontera, N. Masetti, L. Piro (eds) *Third Rome Workshop on GRB in the afterglow era*, ASP Conf. Ser., **312**, 249 (2003).

[15] G.B. Taylor, D.A. Frail, E. Berger, S.R. Kulkarni, *ApJ*, **609**, L1 (2004).

[16] S.R. Kulkarni et al., *Nature*, **395**, 663 (1998).

[17] F. Filliatre et al., *A& A*, **438**, 793 (2005).

[18] R. Ruffini, C.L. Bianco, P. Chardonnet, F. Fraschetti and S.S. Xue, *Int. Journ. Mod. Phys. D*, **12**, 17 (2003).

[19] R. Ruffini et al., *Proceeding X Marcel Grossmann*, World Scientific, in press, astro-ph/ 0503475 (2005).

[20] S.E. Woosley, *ApJ*, **405**, 273 (1993).

[21] M. Vietri, and L. Stella *ApJ*, **507**, L45 (1998).

[22] G.J. Fishman, and G.J. Meegan, *ARAA*, **33**, 415 (1995).

[23] P. Mészáros, *ARAA*, **40**, 137 (2002).

[24] T. Damour and R. Ruffini *Phys. Rev. Lett.*, **35**, 463 (1975).

[25] R. Ruffini and J.A. Wheeler *Physics Today* , **24**, 30 (1971).

[26] D. Christodoulou, and R. Ruffini, *Phys. Rev. D*, **4**, 3552 (1971).

[27] R. Ruffini, J.D. Salmonson, J.R. Wilson and S.S. Xue, *A&A*, **350**, 334 (1999).

[28] R. Ruffini, J.D. Salmonson, J.R. Wilson and S.S. Xue, *A&A*, **359**, 855 (2000).

[29] R. Ruffini, C.L. Bianco, P. Chardonnet, F. Fraschetti and S.S. Xue, *ApJ*, **555**, L107 (2001).

[30] R. Ruffini, C.L. Bianco, P. Chardonnet, F. Fraschetti and S.S. Xue, *ApJ*, **555**, L113 (2001).

[31] G. Ghisellini, in M. Feroci, F. Frontera, N. Masetti, L. Piro (eds) *Third Rome Workshop on GRB in the afterglow era*, ASP Conf. Ser., **312**, 319 (2003).

[32] R. Ruffini, C.L. Bianco, P. Chardonnet, F. Fraschetti and S.S. Xue, *Nuovo Cimento*, **116**, 99 (2001).

[33] R. Ruffini, C.L. Bianco, P. Chardonnet, F. Fraschetti and S.S. Xue, *ApJ*, **581**, L19 (2002).

[34] G. Ghirlanda, A. Celottiet, and G. Ghisellini, *A& A*, **393**, 409 (2002).

[35] R. Ruffini, C.L. Bianco, P. Chardonnet, F. Fraschetti and S.S. Xue, *ApJ*, **555**, L117 (2001).

[36] B. Paczynski, *ApJ*, **494**, L45 (1998).

[37] S.R. Kulkarni et al., *Nature*, **395**, 663 (1998).

[38] K. Iwamoto et al., *Nature*, **395**, 672 (1998).

[39] E. Pian et al., *ApJ*, **536**, 778 (2000).

15

Mechanisms of Supernova Explosions

V.M. Chechetkin, M.V. Popov and S.D. Ustyugov

Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

Abstract

Three-dimensional hydrodynamical simulations of the development of a large-scale instability accompanying deflagration in the degenerate cores of rotating stars – progenitors of supernovae type II and Ia – are presented. Large-scale convective processes are important as the deflagration front propagates. It is assumed that the supernova explosion is strongly nonspherical; a large-scale front structure emerges and propagates most rapidly along the rotational axis. The role of beta processes and neutrino radiation in SN II is discussed.

Keywords: large-scale, convective, instability, supernovae, explosions

1 Introduction

It is usual to classify supernovae according to their optical spectra. They are classified as type I (SN I) if hydrogen lines are absent from their spectra and type II (SN II) if hydrogen lines are present. In turn, SNe I are subdivided into types Ia, Ib, and Ic. SNe Ia are distinguished by the presence of a silicon absorption line at about 6150Å in the period following the explosion and by strong iron emission lines at later times. Conversely, silicon lines are absent from the initial spectra of SNe Ib and SNe Ic. Relatively strong helium lines (especially near 5876Å) are typical of SNe Ib, while they are not observed (or are very faint) in SNe Ic. Type II, Ib, and Ic supernovae are believed to be the products of the explosions of massive single stars (SN II) or binary systems (SN Ib and SN Ic). The progenitors of SNe Ia are white dwarfs with masses close to the Chandrasekhar limit, $M_{ch} \sim 1.44 M_{\odot}$, consisting of a mixture of carbon and oxygen nuclei and a highly degenerate electron-positron gas.

Here we study type II and type Ia supernovae. For both cases we carried out 3D numerical simulations of large-scale convective instability in the star cores taking into account their rotation. For SN II we used an equation of state for proto-neutron matter. In this type of supernova beta processes and neutrino radiation play an important role. For type Ia we used an equation of state for fully ionized matter that describes the electron–positron component in terms of Fermi–Dirac statistics using various asymptotics and the ion component in an ideal-gas approximation.

2 Initial Conditions and Numerical Technique for Type II Supernovae

When calculating the distributions of density and temperature inside the protoneutron star, we took the central density and temperature to be $\rho_c = 2 \times 10^{14}$ g cm^{-3} and $T_c = 10^{11}$ K. We apply the equation of state from [1] as a tabulated dependence of the pressure on density and entropy. We specify the relative number density of electrons to be constant and equal to 0.35. We adopt a Gaussian excess for entropy distribution at the initial time near the center ($\rho_0 = 0$), $S = S_0 + (S' - S_0)exp\{-(\mathrm{r} - \mathrm{r}_0)^2/b^2\}$, and associate it to some nonequilibrium process. We determine the initial entropy (S_0) from the central temperature and density to be $S_0 = 1.6327\ k_B$/nucleon. The maximum entropy in the center is $S' = 2.8$, and the parameter $b = 0.02$ was chosen such that the size of the region of the entropy perturbation is one-fifth of the total size of the computational region.

The hydrodynamical equations that we used to model processes inside the protoneutron star in the three-dimensional case have the form

$$\begin{cases} \rho \dfrac{d\mathbf{v}}{dt} = -\mathrm{grad}\, p - \dfrac{\rho G M}{r^3}\mathbf{r}, \\ \dfrac{d\rho}{dt} + \rho\, \mathrm{div}\, \mathbf{v} = 0, \\ \dfrac{dE}{dt} = T\dfrac{dS}{dt} + \dfrac{p}{\rho^2}\dfrac{d\rho}{dt}, \\ \dfrac{dS}{dt} = 0. \end{cases} \tag{1}$$

Here ρ is density, $\mathbf{v}$ is the velocity of matter, p is the pressure, E is the energy and S, the entropy. All quantities are functions of the three spatial coordinates and time.

We use dimensionless physical parameters as follows: density in units of $\rho_0 = 2 \times 10^{14}$ g cm^{-3}, length in units of $L_0 = 2 \times 10^7$ cm, time in units of $t_0 = 7.74 \times 10^{-3}$ s, mass in solar masses, $M = 2 \times 10^{33}$ g and temperature in units of 10^{11} K.

In the computations we used an explicit, conservative, Godunov-type TVD scheme. This means that the system is written in a so-called divergent form, where the symbolic vector ($\mathbf{U}$) is present in the partial time derivative, and the flux vector ($\mathbf{F}$) is present in the spatial derivative [2]:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0. \tag{2}$$

The three-dimensional space is separated into cubic cells with a constant grid size. The values of all density-vector variables ($\rho, \rho u, \rho S$) correspond to the cell center, while the fluxes of these variables ($\rho u, \rho u^2 + p, \rho S u$) are calculated for the cell boundaries. We calculate the variables at time t' using the solutions of the one-dimensional problem for each spatial direction. To preserve the second-order approximation, we

apply a cyclic rearrangement of these directions. Afterwards, we add a source function to the equation, associated with the action of the gravitational field of the star, which we take to be constant in time.

3 Hydrodynamic Simulations for Type II Supernovae

The initial equilibrium configuration was obtained by using an iterative method [3], which was modified for an arbitrary equation of state. We performed calculations for two cases. The first one has no rotation. We used it to test the resulting equilibrium state of the star. The second one has a slow rigid-body rotation, with a chosen ratio of rotational kinetic energy T to gravitational energy $|W|$, $T/|W| = 0.01$. We calculated the kinetic and potential energies for integrated quantities derived from the computed density profile. The rotational period of the star, which corresponds to a ratio of the kinetic to potential energies of 1%, is 14 ms. We chose the coordinate system such that the plane of rotation coincides with the Oxy axis. In this way the angular-rotation vector of the star has only one component, $\Omega_z = \Omega = const$.

The results of the calculations are presented in Figure 1. Contrary to the models considered in [4] [5], two bubbles form in our simulations after ~ 3 ms and become elongated in opposite directions along the rotation axis. Four additional bubbles arise slightly later, after ~ 5 ms, in the plane of rotation of the protoneutron star. This entropy distribution disrupts the symmetry of the picture obtained in [5]. The bubbles on the rotation axis are the first to tear away from the hot core and float up toward the surface. This occurs because the density changes more rapidly along the rotation axis. Cooler material sinks toward the center of the protoneutron star in the space between these two bubbles. The bubbles in the plane of rotation break away form the hot core somewhat later, and also rise toward the surface. Our computations indicate that, after the first one, additional smaller bubbles form and also begin to float towards the surface.

4 Neutrino Transport in Type II Supernovae

Let us consider a bubble at the central part of a star, having a size of 6×10^5 cm and a density of about 2×10^{14} g cm^{-3}, similar to the density of the surrounding medium. The bubble temperature is substantially higher than the temperature of the surrounding medium. To simplify the model we assume that initially the bubble is composed of iron nuclei ($A = 56$, $Z = 26$) and free ultrarelativistic electrons. At densities of 2×10^{14} g cm^{-3} such a medium is characterized by intense beta processes.

In a uniform, isotropic approximation, the kinetic equation describing the evolution of the neutrino distribution in a bound region whose characteristic size d, density, and partial concentrations of components vary with time, can be written in the form

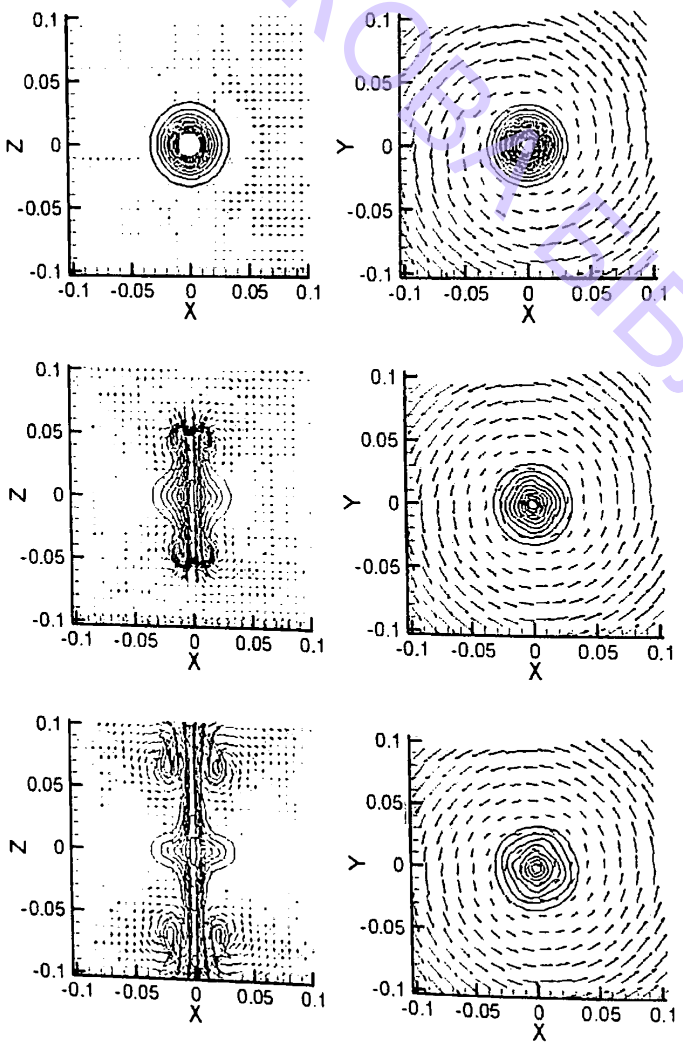


Figure 1: Levels of entropy and velocity fields in the meridional and equatorial planes.

$$\frac{\partial f(p,t)}{\partial t} = 4\pi\Big[(1-f(p,t))\int\limits_0^\infty dp'(p')^2 f(p',t)K^{in}(p,p',t) - f(p,t)\int\limits_0^\infty dp'(p')^2(1-f(p',t))K^{out}(p,p',t)\Big] + \frac{d\rho(t)}{\rho(t)dt}f(p,t) + (1-f(p,t))S(p,t) - f(p,t)Y(p,t) - \frac{c}{d(t)[1+\gamma(p,t)]}(f(p,t)-f_g(p,p_g)). \quad (3)$$

Here, the functions $K^{in}(p,p',t)$ and $K^{out}(p,p',t)$ depend on details of the process of neutrino scattering by electrons, and $S(p,t)$ and $Y(p,t)$ are sources and sinks of

neutrinos, determined by the direct and inverse beta processes. The term containing the logarithmic derivative of the density is responsible for variations in the neutrino distribution due to the changing dimensions of the region where the neutrinos are concentrated. The last term describes neutrino escape through the boundary of a bubble. Here, c is the sound velocity, $d(t)$ is the representative size of a bubble, $\gamma(p,t)$ is the opacity of matter. The value $\gamma = 0$ corresponds to free propagation. We normalized the distribution function as follows:

$$n(t) = 4\pi(2\pi\hbar)^{-3}\int\limits_0^\infty dp'(p')^2 f(p',t). \quad (4)$$

This equation for the neutrino distribution function must be supplemented by an equation describing the evolution of the electron number density

$$\frac{dn_e(t)}{dt} = 4\pi\int\limits_0^\infty dp\,p^2[-(1-f(p,t))S(p,t) + f(p,t)Y(p,t)] + \frac{d\rho}{\rho\,dt}n_e(t), \quad (5)$$

and also by the relation between the particle number densities on the one hand, and the density of the medium, on the other:

$$m_n\left[n_n(t) + \frac{A}{Z}n_e(t)\right] = \rho(t). \quad (6)$$

We used the following expressions for γ in the simulations:

$$\gamma(x,t) = 4\pi\int\limits_0^\infty dy y^2(1-f(y,t))[K^{out}(x,y,t) + K^{in}(x,y,t)], \quad (7)$$

and

$$\gamma(x,t) = 4\pi\int\limits_0^\infty dy y^2[(1-f(y,t))K^{out}(x,y,t) + K^{in}(x,y,t)], \quad (8)$$

where we introduced the notations $x = p/p_F(0)$, $y = p'/p_F(0)$. The function $K^n(x,y,t)$ describes the process of neutrino scattering by atomic nuclei. The factor $(1-f(y,t))$ takes into account the Pauli principle. The difference between (7) and (8) is whether to include neutrino scattering by atomic nuclei or not. For an isotropic neutrino distribution in a bubble, Pauli principle is not important for scattering by atomic nuclei. In this process only the direction of neutrino changes, but it does not matter if their distribution is isotropic. Calculations with these two variants of $\gamma(x,t)$ show the influence of the isotropy of neutrino distribution on the results. More detailed discussion on neutrino transfer can be found in [5].

5 Results of the Neutrino Transport Simulation

The results of our numerical simulations are presented in Figures 2–3, where we used equations (7) and (8) for the function $\gamma(x,t)$.

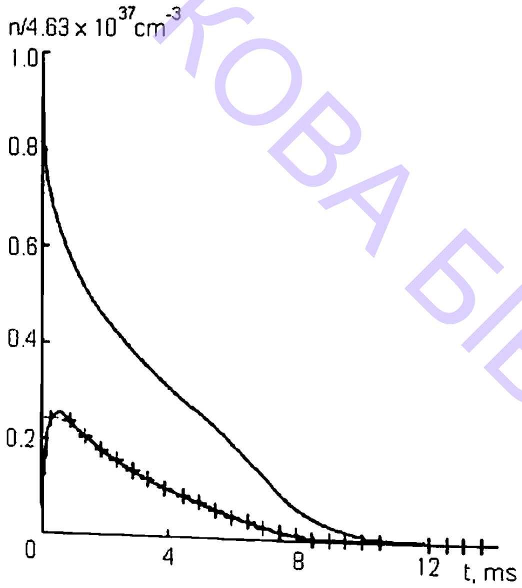


Figure 2: Time dependence of the number density of neutrinos (marked curve) and electrons (unmarked curve).

Let us estimate the spectrum of the neutrinos emitted by a supernova in the adopted scenario. In the case when equation (8) applies, the spectra from the neutrino emission from the bubble and from the entire supernova should be identical as the surrounding medium is optically thin (Figure 4 c). It is impossible to estimate accurately the neutrino spectrum emitted by the supernova in the first case (equation (7) because of high opacity, and we must simulate the neutrino transport over the entire stellar zone in order to obtain a correct description. We can only present here the hardest (Figure 4 a) and softest (Figure 4 b) possible spectra.

6 Estimates of Neutrino Radiation

At the initial time, the mass of matter with excess entropy ($s_{max} = 2.5$) is $0.07M_\odot$. After 3.5 ms, $0.02M_\odot$ of this material approaches the boundary of the neutrinosphere, where the density decays to $\rho = 10^{11}$ g cm^{-3}, and becomes transparent to the neutrinos. The density of these neutrinos is comparable to the density of electrons with mean energy ~ 60 MeV. In this case, the intensity of the neutrino emission can be estimated as

$$L = (0.04M_\odot \times 60\ \mathrm{MeV})/(\mu m_n \times 3.5 \times 10^{-3}) \sim 4 \times 10^{54}\ \mathrm{erg\,s^{-1}},$$

where μ is the mean molecular mass per electron in the absence of electron–positron pairs. The fraction of neutrino energy absorbed per gram of material in the shock front is

$$\frac{d\epsilon}{dt} = \epsilon_\nu n_\nu n_e \langle\sigma v\rangle/\rho \sim L\sigma n_1/\rho R^2\ \mathrm{erg\ g^{-1}\,s^{-1}},$$

where R is the radius of the shock, $\sigma = g^2\epsilon_\nu^2/\hbar^4 c^4 \sim 10^{-44}(\epsilon_\nu/m_e c^2)^2$ cm^2 is the cross-section of weak interactions, $g = 1.4102 \pm 0.0012 \times 10^{-49}$ erg cm^3 is the

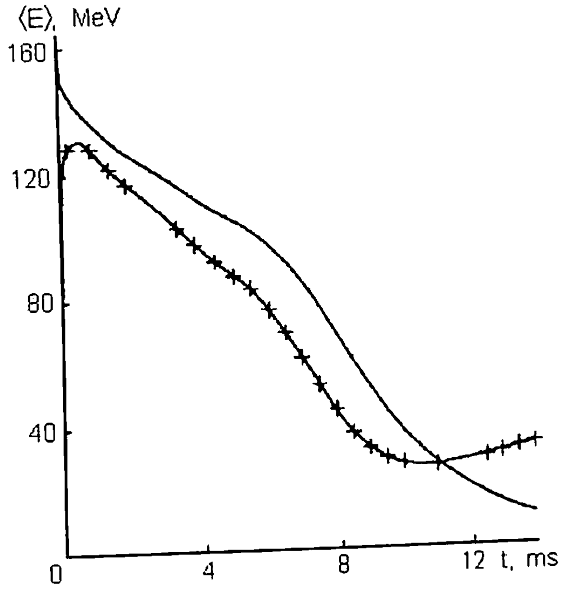


Figure 3: Time dependence of the average energies of neutrinos (marked curve) and electrons (unmarked curve).

Fermi constant of weak interaction, $n_1 = 1.688 \times 10^{28} T_9^3$ cm^{-3} is the density of electron–positron gas in the shock wave when the temperature corresponds to an ultrarelativistic gas with $\rho \leq 10^5 T_9^3$ g cm^{-3} (see [6]), $T_9 = T/10^9$ K, ρ is the density of the material in the shock, ϵ_ν is the mean neutrino energy, and n_ν is the density of neutrinos in the shock. For $T_9 = 100$, $\rho = 10^8$ g cm^{-3}, and $R = 10^7$ cm, we obtain $d\epsilon/dt = 0.97 \times 10^{27}$ erg g^{-1} s^{-1}. This is much more than the neutrino losses from the shock front: $d\epsilon_\nu/dt \sim 6 \times 10^{10} T_9^6 \sim 6 \times 10^{22}$ erg g^{-1} s^{-1} (see [6]), i.e., the large-scale convection could support a diverging shock wave, leading to the ejection of the supernova envelope.

7 Initial State of the Problem and Numerical Technique for Type Ia Supernovae

When a star less massive than $8M_\odot$ leaves the main sequence and becomes a red giant, a degenerate carbon–oxygen core begins to form. This core represents a developing white dwarf. Energy is released by the star energy sources, with the innermost shell being located at the core–envelope interface. The mass of the core increases due to the inflow of deflagration products from this interface. The temperature and density of the core gradually increase if nuclear reactions do not occur. Ultimately, a temperature of $\sim 3 \times 10^8$ K is reached at the center of the star, and deflagration of the CO mixture sets in. The development of a thermal carbon–oxygen flash begins, which eventually results in the thermonuclear explosion of the star. During the deflagration, an iron core forms. It consists of the products of thermonuclear burning of carbon and oxygen. The thermonuclear energy is released, increasing the entropy of the core material [7, 8]. Thermal and mechanical equilibrium are disrupted and pulsations begin to develop giving rise to large-scale hydrodynamic instabilities. It

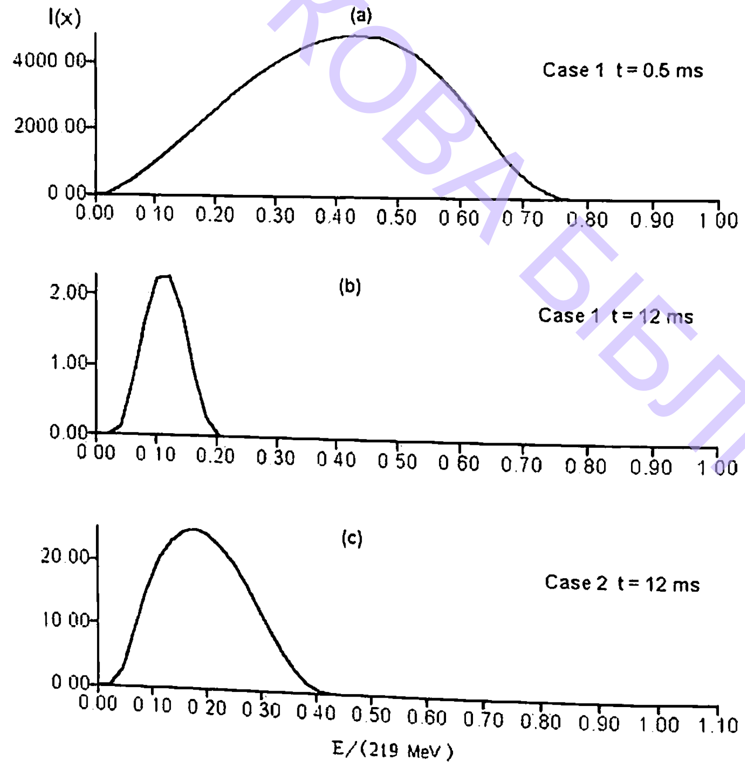


Figure 4: Spectra of the emitted neutrinos $I(x)$.

was shown by Imshennik et al. [9] that a deflagration regime is characteristic of the thermonuclear burning of a degenerate carbon–oxygen core.

For numerical modeling of type Ia supernovae, instead of (1) we used the set of equations

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \operatorname{div} \mathrm{m} = 0, \\ \dfrac{\partial (m_i)}{\partial t} + \dfrac{\partial \Pi_{ik}}{\partial x_k} = \rho g_i, \\ \dfrac{\partial (\rho S)}{\partial t} + \operatorname{div}(\rho S \mathrm{v}) = 0, \end{cases} \tag{9}$$

where $\Pi_{ik} = P\delta_{ik} + \rho v_i v_k$, $\mathrm{m} = \rho \mathrm{v}$ is the momentum, and g is the free-fall acceleration. We chose a spherical coordinate system (r,θ,ϕ) with Lame coefficients $(h_1, h_2, h_3) = (1, r, r\sin\theta)$. To create a numerical scheme we used a divergence form of equation (9) analogous to (2).

All the quantities are dimensionless using the following characteristic parameters: radius $R_0 = 1 \times 10^8$ cm, density $\rho_0 = 2 \times 10^9$ g cm^{-3}, pressure $p_0 = 1.2132 \times 10^{27}$ dyn cm^{-2}, entropy $S_0 = k_b/\mu = 8.3141 \times 10^7$ cm^2 s^{-2}K, where k_b is the Boltzmann constant and μ is the atomic mass unit, free-fall acceleration $g_{r0} = p_0/(R_0 \cdot \rho_0) = 6.0663 \times 10^9$ cm s^{-2}, velocity $v_0 = \sqrt{p_0/R_0} = 3.4832 \times 10^9$ cm s^{-1}, and time $t_0 = R_0/v_0 = 0.0287$ s.

The dimensionless radius of the computational domain is 1.5. However, the initial equilibrium configuration was calculated for a larger radius of 1.8. In this way, we take into account the gravitational pull of the outer layers. The radius of the iron core formed by the time of the onset of the convective instability was assumed to be 0.25. The background entropy (the entropy of the rotating CO sphere) is $S_0 = 0.3564$, which corresponds to a temperature at the center of the star of $T_0 = 1 \times 10^8$ K and a density of $\rho_0 = 2 \times 10^9$ g cm^{-3}. The angular rotational speed is $\Omega_0 = 2.0732$ s^{-1}. At $t = 0$, we pass to the equation of state for iron in the central region, $r \leq 0.25$, and increase the entropy by a factor of ten. The values $r_{core} = 0.25$ and $S_{core} = 10 \cdot S_0$ are free parameters of the problem. They are determined by the deflagration process, which controls the speed of propagation of the front, and by the dynamics of the core pulsations. We do not consider the relationship between these processes here.

8 Hydrodynamics Simulations for Type Ia Supernovae

We performed our computations on a $N_\phi \times N_\theta \times N_r$: $40 \times 80 \times 40$ grid with a Courant number $C_{cour} = 0.8$ using classical boundary conditions. In the course of the evolution, iron is mixed with unburned carbon and oxygen. Accordingly, we calculated the pressure using the expression $p = (1 - \alpha)p_{CO} + \alpha p_{Fe}$. That is, we assume a superposition of two equations of state. Since the entropy depends on the amount of matter that has burned (the iron content of the mixture), the coefficient $\alpha = \alpha(S)$ is a function of the entropy. We use the linear approximation $\alpha = (S - S_0)/(S_{core} - S_0)$.

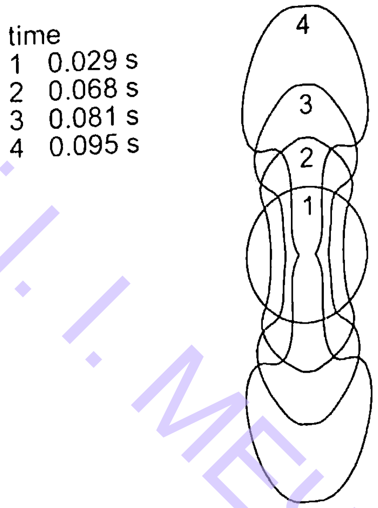


Figure 5: Iron core evolution under the effect of large-scale convection.

Figure 5 shows entropy contours at the boundary of the iron core. Since the calculations implement a locally adiabatic approximation – i.e., they assume conservation of entropy at any point of the Lagrangian coordinate space – these entropy levels characterize the transfer of material. The entropy contours illustrate the departure of the core shape from the initial, nearly spherically symmetric configuration up until the time when the core breaks up and the regions of outflow acquire a jet-like structure.

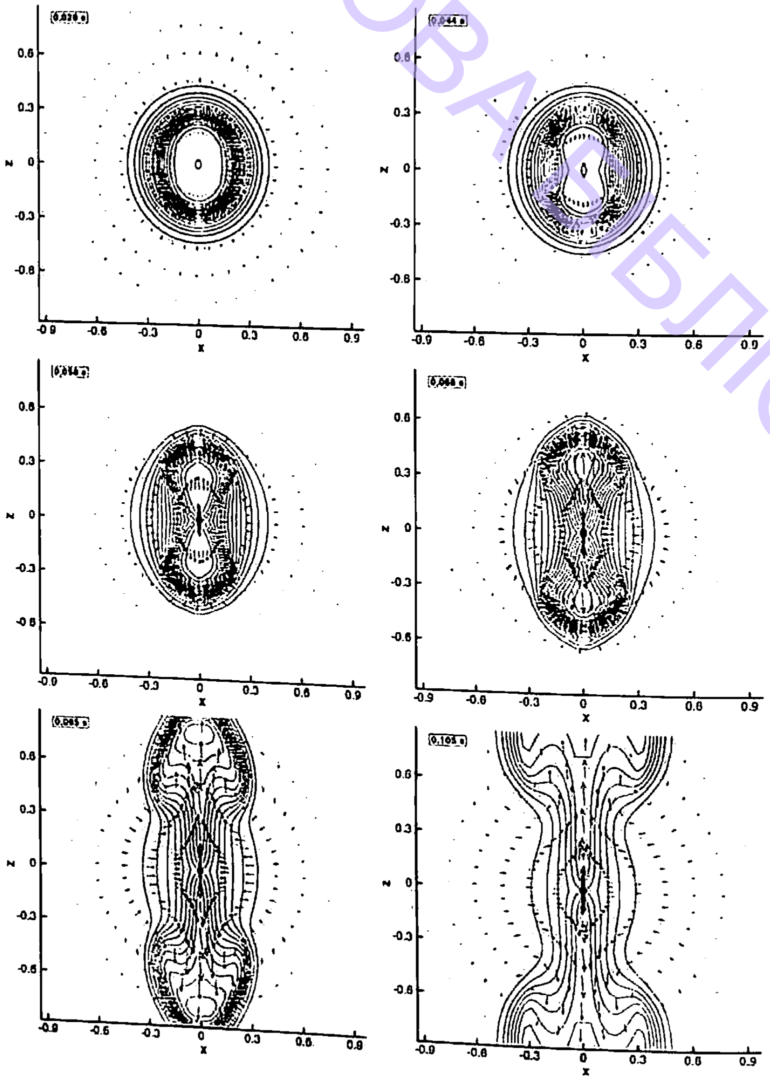


Figure 6: Contours of entropy and the momentum field.

Figure 6 illustrates the development of the explosion shown by the entropy contours and the momentum field in the meridional plane. Two bubbles consisting of a mixture of the deflagration products with the initial carbon and oxygen begin to rise along the rotational axis. We can see from the figure that, at 0.05–0.07 s, the core loses its ellipsoidal shape, and its inner part begins to resemble two oppositely directed jets. Note the formation of two toroidal eddies encircling the jets. This flow structure is due to the mushroom instability of a jet penetrating into a medium. Another interesting feature is the formation of streams from the unburned layers of the core into the central region. These streams have an equatorial configuration. The last two graphs in Figure 6 ($t = 0.085$ and 0.105 s) exhibit only the jet pattern. During the explosion, virtually all the burned material emerges at the surface, with a carbon–oxygen mixture again appearing at the center.

The bubbles expansion velocity becomes supersonic with Mach number of about 1.5. This should result in the formation of a shock wave with an entropy jump behind the shock front. However, we cannot trace the formation of this shock since we use hydrodynamical equations in a form that does not admit discontinuous solutions for the entropy (the local adiabaticity approximation).

The density in the central region again grows during convection due to the transport of material from the outer layers to the center. This is due to the fact that the unburned primary material has a lower temperature. Therefore, the conditions required for deflagration are restored again in the centre, and a cyclic formation of emerging large-scale structures is possible.

9 Discussion and Conclusions

We have carried out numerical simulations of the development of a large-scale instability of the thermonuclear-deflagration front during the explosion of type II and type Ia supernovae. Large-scale structures form in the rotating presupernova and rise from the center to the outer layers of the star. This process is of paramount importance for understanding the explosion mechanism. The propagation of the deflagration front in a supernovae is a strongly non-spherically symmetric process, and a largescale front structure emerges, travelling most rapidly along the rotational axis. Fresh thermonuclear fuel continuously arrives at the center of the core, which should result in new flashes of burning. The fraction of burned fuel remains an open question. This quantity affects the peak height of the light curve of the supernova. Another question concerns the production of chemical elements and the interpretation of light curves for an explosion lacking spherical symmetry. Khokhlov [10] carried out numerical simulations of the development of large-scale structures in a non-rotating supernova core. These resemble the structures that, according to [11], emerge during the development of large-scale convection.

A number of mechanisms for feeding a type II supernova shock wave via neutrino emission from the core of the star have been proposed. One of these is based on the convective transport of neutrinos inside or outside the forming neutron star region [12, 13]; others are based on the decrease of the opacity to neutrinos at nuclear densities [14, 15] or on the effects of nuclear-neutrino interactions [16, 17], which increase the mean free path of neutrinos in the opaque core. However, in spite of the increase in the neutrino luminosity resulting from these mechanisms, their development requires a long time interval (~ 1 s) and a number of specific conditions (see above) must be satisfied. The mechanism we have considered here is based on the development of large-scale hydro-dynamical instability inside a rotating protoneutron star, and has several advantages over previous models. It can provide a rapid burst (over a time $\sim 10^{-2}$ s) of high-energy neutrinos emission, which can give the required boost to the energy of the shock.

Our results are indirectly supported by observations of SN 1987A, which is located in the Large Magellanic Cloud. SN 1987A is a type II supernova and, to all appearances, a similar explosion mechanism was realized there. Optical images of SN 1987A were obtained in 1995–2002 using the Hubble Space Telescope with violet, yellow, and red filters (see http://cfawww.harvard.edu/cfa/oir/Research/sins.html). It is likely that there is no compact object at the center, so that the presupernova star exploded completely. During the destruction of this star, an elongated structure aligned with the rotational axis and normal to the plane of the ring was observed. This corresponds to the development of thermal instability in the degenerate CO core of the presupernova during its complete destruction. The results of our three-dimensional hydrodynamic calculations are consistent with the observational data just described.

Acknowledgements

This work was supported by INTAS (grant 01-491), the Russian Foundation for Basic Research (project code 03-02-16548), and the Program "Leading Scientific Schools of Russia" (grant NSh-1029.2003.2).

References

[1] J.M. Lattimer, F.D. Swest, *Nucl. Phys. A*, **535**, 331 (1991).

[2] A.A. Samarskii, Yu.P. Popov, *Chislennye Melody (Numerical Methods)*, Nauka, Moscow (1980).

[3] K. Hachisu, *Astrophys. J., Suppl. Ser.*, **62**, 461, (1986).

[4] V.M. Chechetkin, S.D. Ustyugov, A.A. Gorbunov, V.I. Polezhaev, *Pis'ma Astron. Zh.*, **23**, 34 (1997).

[5] M.V. Sazhin, S.D. Ustyugov, V.M. Chechetkin, *JETP Lett.*, **64**, 871 (1996).

[6] W. Fowler, F. Hoyle, *Nucleosynthesis in Massive Stars and Supernovae*, University of Chicago Press, Chicago (1965).

[7] L.N. Ivanova, V.S. Imshennik, V.M. Chechetkin, *Astron. Zh.*, **54**, 661 (1977).

[8] L.N. Ivanova, V.S. Imshennik, V.M. Chechetkin, *Astron. Zh.*, **54**, 1009 (1977).

[9] V.S. Imshennik et al., *Pis'ma Astron. Zh.*, **25**, 250 (1999).

[10] A.M. Khokhlov, preprint astro-ph/0008463 (2000).

[11] V.M. Chechetkin, S.D. Ustyugov, A.A. Gorbunov, V.I. Polezhaev, *Astron. Lett.*, **23**, 30 (1997).

[12] A. Burrows, *Astrophys. J.*, **318**, 57 (1987).

[13] R.W. Mayle, J.R. Wilson, *Astrophys. J.*, **334**, 909 (1988).

[14] A. Burrows and R.F. Sawyer, *Phys. Rev. C: Nucl. Phys.*, **58**, 554 (1998).

[15] G.G. Raffelt, D. Seckel, G. Sigl, *Phys. Rev. D: Part. Fields*, **54**, 2784 (1996).

[16] M. Prakash et al., *Phys. Rep.*, **280**, 1 (1997).

[17] S. Reddy, M. Prakash, J.M. Lattimer, *Astrophys. J.*, **478**, 689 (1997).

16

The Supercritical X-ray Binary SS433: Results of INTEGRAL Observations and a Coordinated Campaign

A.M. Cherepashchuk

Sternberg Astronomical Institute, Moscow, Russia
e-mail: cher@sai.msu.ru

Abstract

Results of coordinated X-ray and optical observations of SS433 support the model of SS433 as a supercritically accreting massive X-ray binary system containing a black hole.

Keywords: supercritical accretion, X-rays, massive binary systems

1 Introduction

SS433 is a massive eclipsing X-ray binary system at an advanced evolutionary stage containing a relativistic object which is surrounded by a supercritical precessing accretion disk. From the central parts of the accretion disk emanate two collimated precessing relativistic ($v \simeq 0.26c$) jets (e.g. [13, 2, 8]).

In this system the A5-7I star fills its Roche Lobe, mass transfer occurs on thermal time scale ($\dot{M} \simeq 10^{-4} M_{\odot} yr^{-1}$) and accretion onto the relativistic object is supercritical [16]. Both the precession period ($P_{prec} = 162^{d}.5$) and the disk inclination angle to the orbital plane ($\sim 20°$) keep on average stable over tens of years. The orbital period ($P_{orb} = 13^{d}.082$) is constant and the shape of the optical light curve varies significantly with precessional phase [10]. The precession period $P_{prec} = 162^{d}.5$ is observed in the Doppler shifts of optical and X-ray emission lines and in the optical light variations of the binary system [12, 14, 3].

The main puzzles of SS433 that still remain to be solved include: (1) the nature of the relativistic object, (2) the mechanism of collimation and acceleration of matter in the jets, (3) the nature of the stable precessional phenomenon in this eclipsing X-ray binary system.

INTEGRAL observations of SS433 in 2003 discovered a hard (up to ~ 100 KeV) X-ray spectrum in this supercritically accreting X-ray binary [4, 5] suggesting the presence of an extended hot region at the centre of the supercritical accretion disk. These new data can be compared with those obtained at different energies: soft X-rays (2–10 KeV, the ASCA data), soft and medium X-rays (1–35 KeV, the GINGA data), hard X-rays (20–70 KeV, the INTEGRAL data), and optical. Optical observations of SS433 were carried out simultaneously with X-ray observations at the 6-meter telescope of SAO RAS (high signal-to noise optical spectroscopy), at the 0.6-meter telescope of the Crimean Station of Sternberg Astronomical Institute (*BVR*-photometry) and at the 1.5-meter Russian–Turkish telescope at TUBITAK observatory, Turkey (*BVR* photometry). Results of a coordinated multiwavelength observational campaign of SS433 are written in [5, 6]. Here we summarize the basic results of these investigations.

2 X-ray Observations

The reduction method of the X-ray data and corresponding software are described in [15]. A map of the sky around SS433 obtained by IBIS/ISGRI on board INTEGRAL is presented in Figure 1. SS433 is clearly detected on both X-ray ranges: 25–50 KeV and 50–100 KeV. The IBIS/ISGRI spectrum of SS433 collected over 67–69 INTEGRAL orbits in May 2003 is presented in Figure 2.

The best fit of this spectrum in the range 25–100 KeV is for an expotential cut-off with $E_c = 14 \pm 2$ KeV. A power-law approximation $\sim E^{\alpha}$ with photon index $\alpha \approx 2.7 \pm 0.13$ can not be excluded. The X-ray light curve for SS433 in the range 25–50 KeV is presented in Figure 3.

The integrated hard X-ray luminosity of SS433 in May 2003 was $L_x(18-60)$ KeV$\approx 4 \cdot 10^{35}$ erg s^{-1}, $L_x(60-120)$ KeV $\simeq 2 \cdot 10^{35}$erg s^{-1} (assuming a distance of 5 kpc to SS433). It is about 10% of the soft X-ray jet luminosity. An X-ray eclipse at hard energies is observed to be a little narrower than the optical eclipse. In contrast to the classical eclipsing X-ray binaries (like Cen X-3, Vela X-1 etc.) the width of the X-ray eclipse for SS433 increases with increasing energy (see Figure 4). The eclipse depth for SS433 in hard X-rays is at least 80% which is about twice the depth of the eclipse in soft X-rays 4.6–27 KeV (see Figure 4). The amplitude of precessional variability of SS433 in 25–50 KeV is also about two times that in soft X-rays and reaches ~ 80%, therefore the ratio $A_{pr} = \frac{F_{max}}{F_{min}}$ for the 25–50 KeV range is ~ 4.

In the case of thin relativistic jets the expected amplitude of precessional variability due to relativistic beaming only is $A_{pr} < 1.5$. So the large (~ 4 times) precession variability found in the hard X-ray range 25–50 KeV can not be explained by jet relativistic beaming alone and should be basically caused by geometric screening of the inner disk structure producing the hard X-rays by the outer edge of the precessing geometrically thick accretion disk.

The primary eclipse in hard X-rays was observed on May 9–11, 2003. Also we have used the data obtained on May 10, 2003 kindly provided by Dr. Diana Hannikainen. The observed hard X-ray eclipse falls to the precession phase $\psi_{pr} =$ 0.09 (according to the ephemeris by [10]), fairly close to the T3 phase (moment of

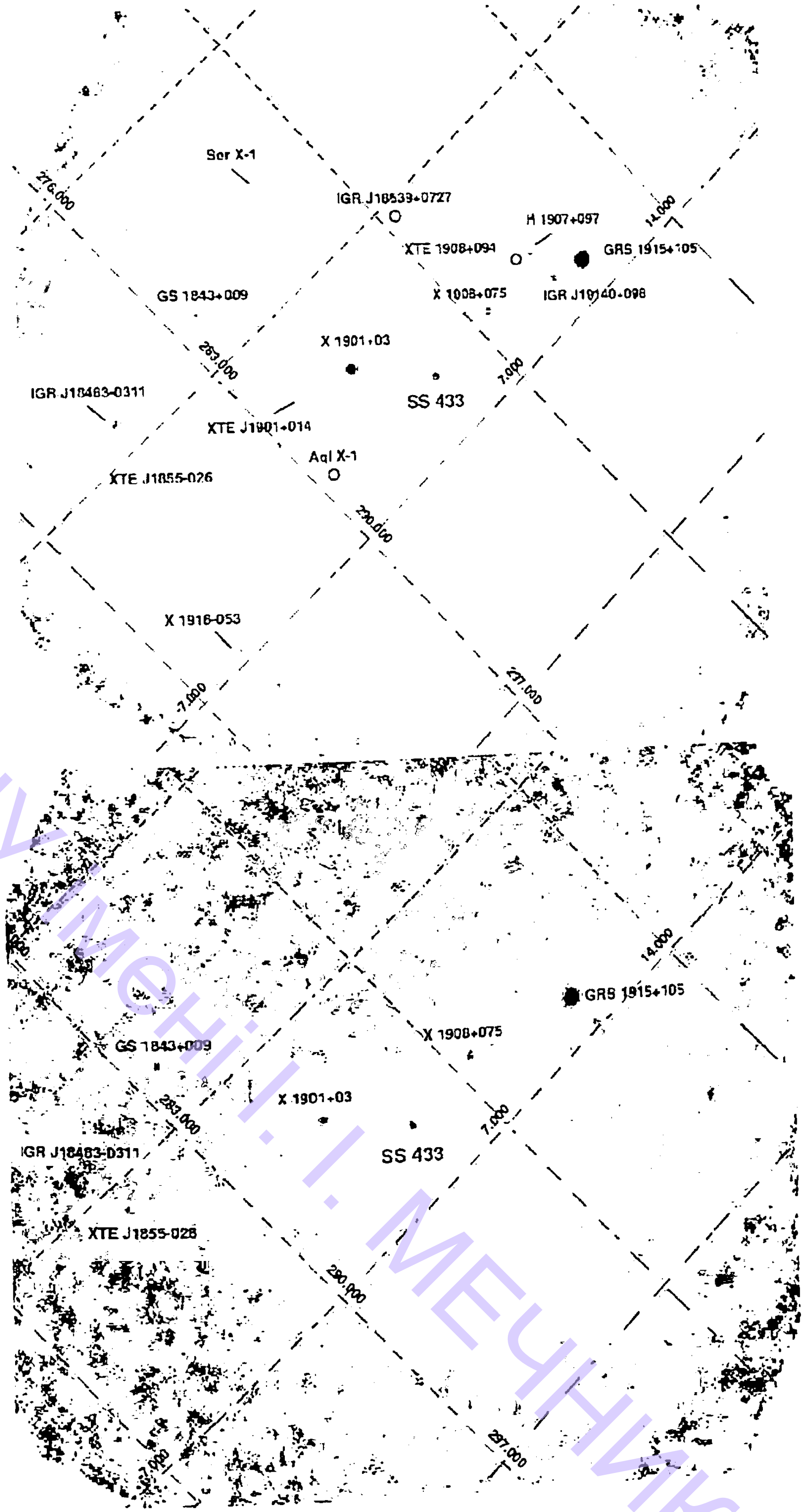


Figure 1: Map of the sky around SS433 obtained by the IBIS/ISGRI telescopes of the INTEGRAL space X-ray observatory on May 2003. 25–50 KeV energy band (top panel) and 50–100 KeV energy band (bottom panel); (from [4]).

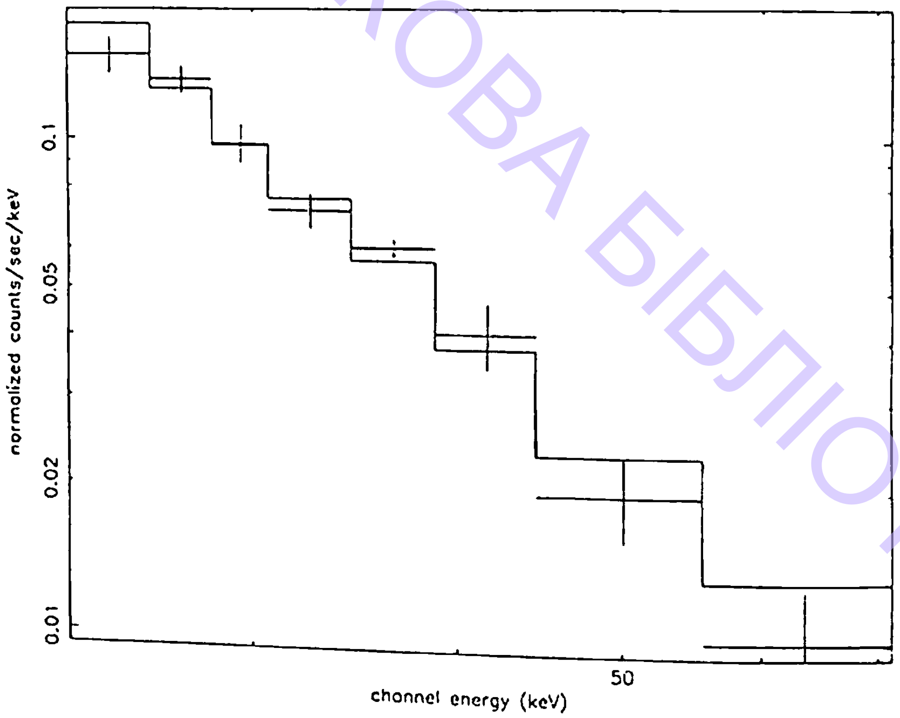


Figure 2: IBIS/ISGRI spectrum of SS433 collected over INTEGRAL orbits 67–69. The best fit power law photon index (solid line) is $\alpha = 2.69 \pm 0.13$; reduced $\chi^2 = 0.63$ for 7 degrees of freedom, CL=95%, see [5].

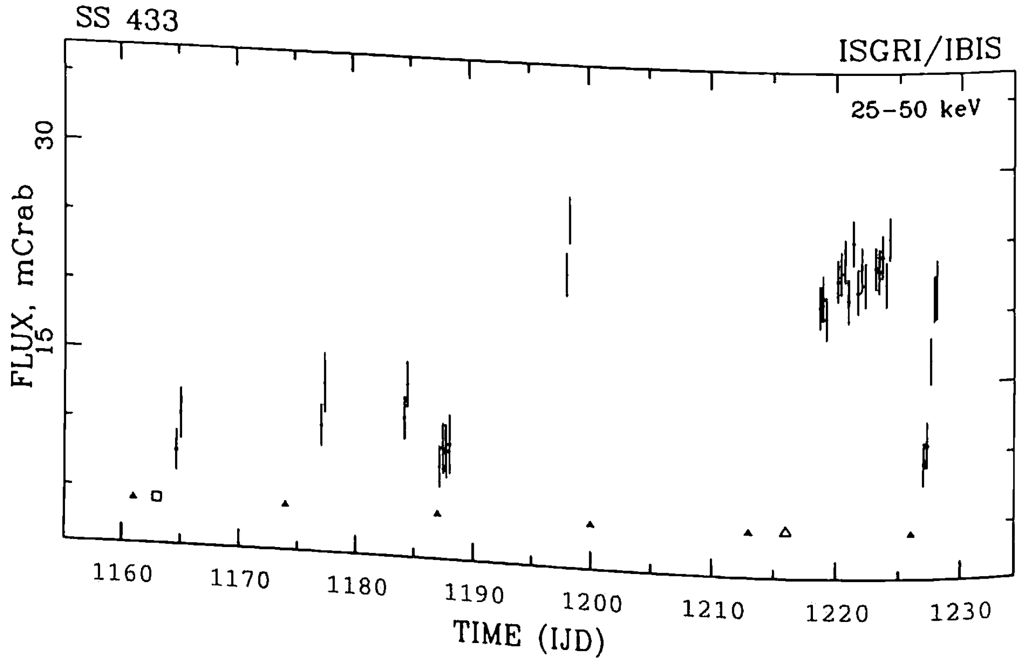


Figure 3: Light curve ofSS433 (IBIS/ISGRI telescope, 25–50 KeV) in March-May 2003 (from [4]). Filled triangles indicate optical eclipse minima according to [10]. The open triangle and square indicate precessional phases of face-on and cross-over, respectively. IJD corresponds to INTEGRAL Julian Date: MJD-51544.

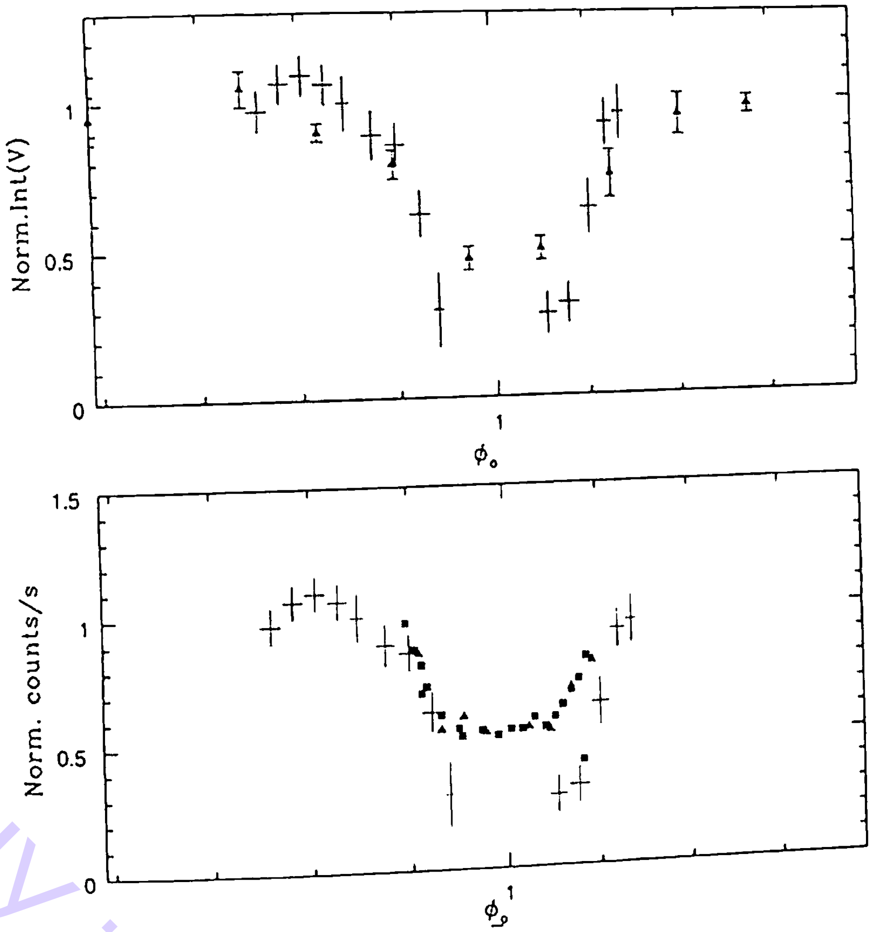


Figure 4: Upper panel: The IBIS/ISGRI 25–50 keV eclipse light curve averaged over 20 ks superimposed on the mean V-light curve of SS433 (filled triangles) simultaneously obtained at the SAI Crimean Laboratory 0.6-meter telescope, normalized to the maximum flux. Bottom panel: The same light curve normalized to the maximum flux of GINGA (filled squares, 5–27 keV) and ASCA (6–10 keV, triangles) light curves [11, 17]. For more details see [5].

maximum separation of the Doppler shifting emission lines, disk maximum opening). The X-ray eclipse in hard X-rays was interpreted in the framework of the model of a precessing thick accretion disk containing an extended hot corona in its central part [1]. Only the corona is assumed to emit in X-rays, while the star and the disk eclipse it in due course during the orbital and precession motion. Precessing variability observations provide a "vertical" scan of the emitting structure, restricting the height of the corona and the opening angle of the thick accretion disk.

The orbital (eclipse) variability observations scan the emitting structure "horizontally", restricting the extension of the hot corona in the plane of the disk and the mass ratio of the components $q = m_x/m_v$ (m_x and m_v are the mass of the relativistic component and optical star respectively). Therefore, the analysis of both precessional and eclipse variability allows us to restore the structure of the

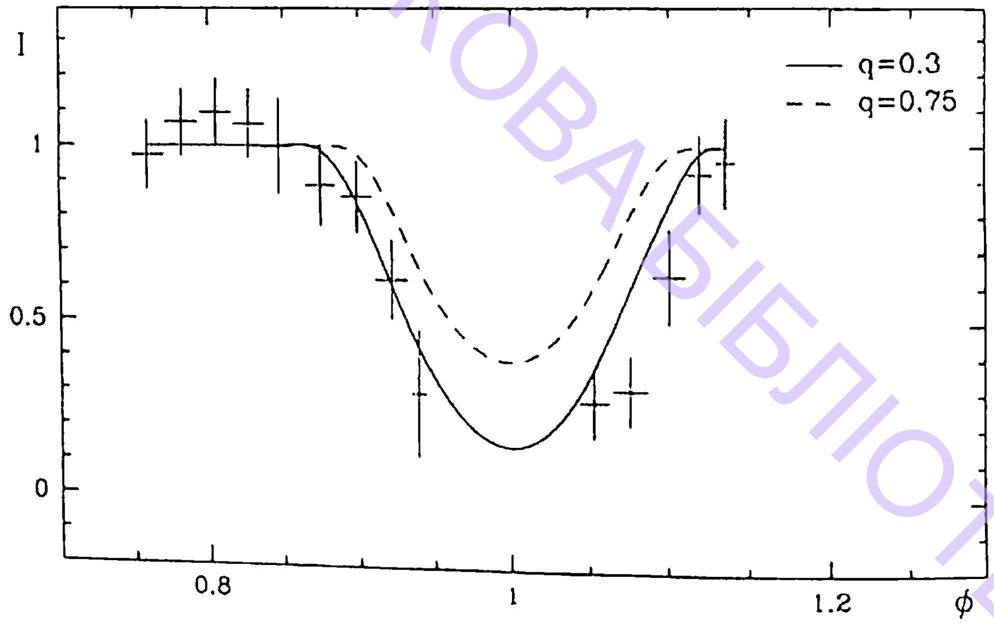


Figure 5: Best fit of the IBIS/ISGRI X-ray eclipse light curve by the binary system model with mass ratio $q = m_x/m_v = 0.3$ (solid line) and $q = 0.75$ (dashed line).

region in the accretion disk center where the hard X-rays are produced, as well as the mass ratio q. The best solution was obtained for $q = 0.2$–0.3 (see Figure 5). The height of the corona is $\sim 0.1a$ (where $a \sim 80R_{\odot}$ is the separation between the components of the system), the "horizontal" extension of the corona is $\sim 0.3a$. The origin and stability of such a corona in the supercritical accretion disk needs further investigation.

3 Optical Observations

The primary minimum of the optical V-light curve observed simultaneously with the hard X-ray observations, coincides with the middle of the hard X-ray flux eclipse. Optical spectroscopy of SS433 was obtained at the 6-meter telescope of the Special Astrophysical Observatory during 6 nights in April 28th and May 9–13th 2003, simultaneously with the INTEGRAL observations at the precessional phase corresponding to the maximum disk opening angle. The spectral resolution is ~ 3 Å (1.2 Å per pixel), and the signal-to-noise ratio at $\lambda 4250$ Å is $\sim$60–80. The determined spectral class of the optical star A5-A7I agrees with the results by [7].

A heating effect of the companion atmosphere is discovered in the absorption lines. During the disk eclipse egress, the low-excitation lines strongly weaken. The stellar hemisphere illuminated by the bright accretion disk probably has a temperature of about 20000 K. The spectral class A5-A7I is for the non-heated hemisphere of the normal star because the most reliable photospheric absorption spectrum is detected in the middle of the primary eclipse when the bright accretion disk is eclipsed by the body of the normal star. The mean radial velocity curve of the optical star of SS433 measured from 22 individual absorption lines is presented in

---

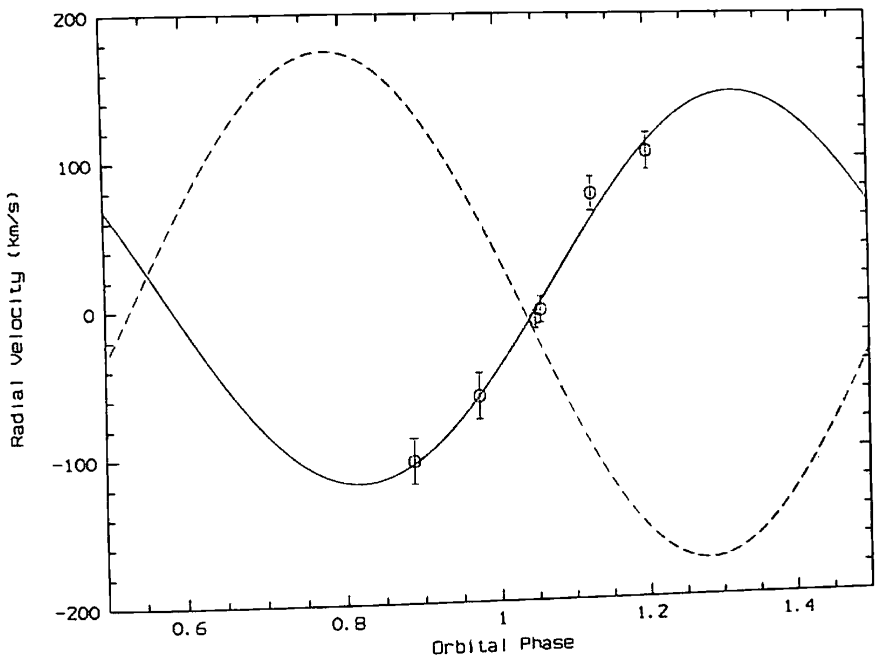


Figure 6: Mean radial velocity curve of the optical A5-A7I companion of SS433 measured from 22 individual weak absorption lines. The accretion disk radial velocity curve as measured by the HeIIλ4686 Å emission [9] is shown by the dashed line. From paper [5].

Figure 6. The full semiamplitude of the radial velocity curve of the optical star is $K_v = 132 \pm 9$ km s^{-1}. This result confirms an earlier determination of K_v in paper [7], also obtained at the maximum opening phases.

4 Basic Parameters of the Binary System

Comparison of radial velocities of the accretion disk ($K_x = 175$ km s^{-1}, [9]) and of the optical star $K_v = 132 \pm 9$ km s^{-1} yields a mass ratio for the SS433 binary system $q = m_x/m_v = K_v/K_x = 0.75$. This value of q is an upper limit because the observed value of K_v should be corrected for the strong heating effect. The value of q derived from the analysis of the hard X-ray eclipse is $q = 0.2$–0.3. Taking the value of $K_v < 132$ km s^{-1} yields a mass function of the optical star $f_v(m) = \frac{m_x^3 sin^3 x}{(m_x+m_v)^2} < 3.1$ solar mass.

5 Conclusions

The basic conclusions reached in [5, 6] are the following. It is shown that the width of the X-ray eclipse in SS433 increases with increasing X-ray energy. This is opposite to the case of classical X-ray binary systems and presumably reflects the complicated structure of supercritical accretion disk. A hot extended corona may be suggested in the supercritical accretion disk. The origin and nature of this corona is a special problem which needs further investigation. The analysis of the X-ray eclipse and the precession variability in the hard X-ray range 25–50 KeV, together with the results of optical spectroscopy yields a value for the mass of the relativistic object $m_x = 9\,M_\odot$ which supports the model of SS433 as a massive X-ray binary with a supercritically accreting black hole.

6 Acknowledgements

The author thanks R.A. Sunyaev, S.N. Fabrika, K.A. Postnov, S.V. Molkov, I.E. Panchenko, N.I. Shakura, A.N. Timokhin, E.V. Seifina, E.A. Antokhina and I.F. Bikmaev for discussions and collaboration. The work was supported through grants by the Leading Scientific Schools of Russia NSh-388.2003.2 and RFBR 02-02-17524.

References

[1] E.A. Antokhina, E.V. Seifina, A.M. Cherepashchuk, *Soviet Astron.*, **36**, 143 (1992).

[2] A.M. Cherepashchuk, in R.A. Sunyaev (ed) *Sov. Sci. Rev. Ap. Space Phys.*, **70**, 1 (1988).

[3] A.M. Cherepashchuk, *Space Sci. Rev.*, **102**, 23 (2002).

[4] A.M. Cherepashchuk et al., *Astron. Astroph.*, **411**, L441 (2003).

[5] A.M. Cherepashchuk, R.A. Sunyaev, S.N. Fabrika et al., preprint (astro-ph/0405338) (2004).

[6] A.M. Cherepashchuk, R.A. Sunyaev, S.N. Fabrika et al., *Astron. Astroph.*, accepted (2005).

[7] D.R. Gies, W. Huang, M.V. McSwain, *Astrophys.J.*, **578**, 67 (2002).

[8] S.N. Fabrika, *Ap. Space Phys. Rev.*, **12**, 1 (2004).

[9] S.N. Fabrika, L.V. Bychkova, *Astron.Astroph.*, **240**, L5 (1990).

[10] V.P. Goranskii, V.F. Esipov, A.M. Cherepashchuk, *Astron.Rep.*, **42**, 209 (1998).

[11] N. Kawai, M. Matsuoka, H.-C. Pan, G.C. Stewart, *Publ. Astron. Soc. Japan*, **41**, 491 (1989).

[12] T. Kotani, N. Kawai, M. Matsuoka, W. Brinkmann, *Publ. Astron. Soc. Japan*, **48**, 619 (1996).

[13] B. Margon, *ARAA*, **22**, 507 (1984).

[14] Y.L. Marshall, C.R. Canizares, N.S. Schulz, *Astrophys. J.*, **564**, 941 (2002).

[15] M.G. Revnivtsev, R.A. Sunyaev, D.A. Varshalovich et al., *Pis'ma Astron. Zh.*, **30**, 430 (2004).

[16] N.I. Shakura, R.A. Sunyaev, *Astron. Aph.*, **24**, 337 (1973).

[17] W. Yuan, N. Kawai, W. Brinkmann, M. Matsuoka, *Astron. Aph.*, **279**, 451 (1995).

17

Axisymmetric Stationary Flows in Compact Objects

V.S. Beskin

P.N. Lebedev Physical Institute, Moscow, Russia
e-mail: beskin@lpi.ru

Abstract

The Grad–Shafranov approach allows us to analyze the two-dimensional structure of a wide class of axisymmetric stationary flows in the vicinity of compact astrophysical objects including the most general case corresponding to the axisymmetric stationary MHD flow in the Kerr metric. Several examples are presented where it is shown how this approach can be used to obtain the quantitative description of the real transonic flows in the vicinity of rotating and moving black holes.

Keywords: Grad–Shafranov equation, accretion, black holes

1 Introduction

Many astrophysical sources are axisymmetric and stationary to a good accuracy. These include both accreting neutron stars and black holes, axisymmetric stellar (solar) winds, jets from young stellar objects, and ejection of particles from magnetospheres of rotating neutron stars. It cannot be ruled out that such magnetohydrodynamic flows also play an important role in other galactic sources, e.g., microquasars. The latter are regarded as candidates for black holes not to mention active galactic nuclei where the electrodynamical processes in the vicinity of the rotating supermassive black holes are considered as the most reasonable model of their central engine [1]. So, it is not surprising that ideal magnetohydrodynamics, which allows sufficiently simple formalization of the problem, is actively applied when describing these flows.

The point is that due to axial symmetry and stationarity (as well as the ideal freezing-in condition), in the most general case it is possible to introduce five integrals of motion which are constant at axisymmetric magnetic surfaces. This remarkable fact allows us to separate the problem of finding the poloidal field structure (the poloidal flow structure in the hydrodynamics) from the problem of particle acceleration and the structure of electric currents. The solution of the latter task for a given poloidal field can be obtained in terms of quite simple algebraic relations. It is important that such an approach can be straightforwardly generalized to flows in the vicinity of the rotating black holes, as the Kerr metric is also axially symmetric and stationary.

On the other hand, it is much more difficult to find the two-dimensional poloidal magnetic field structure (the hydrodynamical flow structure). First of all, this is due to the complex structure of the equation describing axisymmetric stationary flows. In the general case, it is a nonlinear equation of the mixed type, which changes from elliptical to hyperbolical at singular surfaces and in addition contains integrals of motion in the form of free functions. Later on, the equations describing axially symmetric stationary flows were called Grad–Shafranov equations after the authors who in the late 1950s formulated an equation of such a type in connection with controlled thermonuclear fusion [2, 3]. This equation, however, was originally related to equilibrium static configurations only and required strong revision when it was generalized to the transonic case. The full version of such an equation was formulated by L.S. Soloviev in 1963 in the third volume of *Problems of Plasma Theory* [4]. The full relativistic version in the Kerr metric for nonzero temperature was formulated in [5].

2 Grad–Shafranov Equation

Let us consider the axisymmetric stationary plasma flow in the vicinity of a rotating black hole, i.e., in the Kerr metric [1]:

$$\mathrm{d}s^2 = -\alpha^2\mathrm{d}t^2 + g_{ik}(\mathrm{d}x^i + \beta^i\mathrm{d}t)(\mathrm{d}x^k + \beta^k\mathrm{d}t), \qquad (1)$$

where

$$\alpha = \frac{\rho}{\Sigma}\sqrt{\Delta}, \qquad \beta^r = \beta^\theta = 0, \qquad \beta^\varphi = -\omega = -\frac{2aMr}{\Sigma^2},$$
$$g_{rr} = \frac{\rho^2}{\Delta}, \qquad g_{\theta\theta} = \rho^2, \qquad g_{\varphi\varphi} = \varpi^2. \qquad (2)$$

Here α is the lapse function (gravitational redshift) vanishing on the horizon

$$r_\mathrm{g} = M + \sqrt{M^2 - a^2}, \qquad (3)$$

ω is the angular velocity of local nonrotating observers (the so-called Lense–Thirring angular velocity), and

$$\Delta = r^2 + a^2 - 2Mr, \qquad \rho^2 = r^2 + a^2\cos^2\theta,$$
$$\Sigma^2 = (r^2 + a^2)^2 - a^2\Delta\sin^2\theta, \qquad \varpi = \frac{\Sigma}{\rho}\sin\theta. \qquad (4)$$

As usual, M and a are the black hole mass and angular momentum per unit mass ($a = J/M$) respectively. Here indices without hats denote components of vectors with respect to the coordinate basis $\partial/\partial r$, $\partial/\partial\theta$, and $\partial/\partial\varphi$, and indices with hats correspond to their physical components. Finally, below we shall use the system of units where $c = G = 1$.

In what follows we shall also use the 3 + 1 split language [1]. Within this approach, the physical quantities are expressed in terms of three-dimensional vectors which would be measured by observers moving around the rotating black hole with angular velocity ω (so-called ZAMOs – zero angular momentum observers). The convenience of the 3 + 1 split language is connected with the fact that the representation of many expressions has the same form as in the flat space. On the other hand, all thermodynamic quantities are determined in the comoving reference frame.

Now, we shall demonstrate how the five 'integrals of motion', which are constant at the magnetic surfaces, can be derived in the general case of axisymmetric stationary flows. It is convenient to introduce the scalar function $\Psi(r, \theta)$ which has the meaning of a magnetic flux. As a consequence, the magnetic field is defined in the following way:

$$\mathbf{B} = \frac{\nabla\Psi \times \mathbf{e}_{\hat\varphi}}{2\pi\varpi} - \frac{2I}{\alpha\varpi}\mathbf{e}_{\hat\varphi}, \qquad (5)$$

where $I(r, \theta)$ is the total electric current inside the region $\Psi < \Psi(r, \theta)$.

As usual, we assume that the magnetosphere contains sufficient amount of plasma to satisfy the freezing-in condition which, using the 3 + 1 split language, preserves the form $\mathbf{E} + \mathbf{v} \times \mathbf{B} = 0$. On the other hand, the stationarity (as well as the condition for zero longitudinal electric field) implies that the field $\mathbf{E}$ can be written as

$$\mathbf{E} = -\frac{\Omega_\mathrm{F} - \omega}{2\pi\alpha}\nabla\Psi. \qquad (6)$$

By substituting relation (6) into the Maxwell equations, it is easy to verify that the condition $\mathbf{B}\cdot\nabla\Omega_\mathrm{F} = 0$ is satisfied, i.e., that Ω_F must be constant at the magnetic surfaces (Ferraro's isorotation law):

$$\Omega_\mathrm{F} = \Omega_\mathrm{F}(\Psi). \qquad (7)$$

Next, the Maxwell equation $\nabla\cdot\mathbf{B} = 0$, the continuity equation, and the freezing-in condition allow us to write the four-velocity of matter $\mathbf{u}$ in the form

$$\mathbf{u} = \frac{\eta}{\alpha n}\mathbf{B} + \gamma(\Omega_\mathrm{F} - \omega)\frac{\varpi}{\alpha}\mathbf{e}_{\hat\varphi}, \qquad (8)$$

where $\gamma = 1/\sqrt{1 - v^2}$ is the Lorentz factor of matter (measured by ZAMOs), and the quantity η is the particle flux to magnetic flux ratio. Due to the relationship $\Psi(r, \theta) = \mathrm{const}$ as well, $\nabla\cdot(\eta\mathbf{B}_p) = 0$, η must be constant at the magnetic surfaces i.e.

$$\eta = \eta(\Psi). \qquad (9)$$

The next two integrals of motion result from our assumption that the flow is axisymmetric and stationary. This yields the conservation law of the energy E and the z-component of angular momentum L_z:

$$E = E(\Psi) = \frac{\Omega_\mathrm{F} I}{2\pi} + \mu\eta(\alpha\gamma + \omega u_\varphi); \qquad (10)$$

$$L = L(\Psi) = \frac{I}{2\pi} + \mu\eta\varpi u_{\hat{\varphi}}, \tag{11}$$

where $\mu = (\rho_m + P)/n$ is the relativistic enthalpy (ρ_m is the internal energy density, and P is the pressure). Finally, in the axially symmetric case the isentropy condition yields

$$s = s(\Psi), \tag{12}$$

so that the entropy per particle, $s(\Psi)$, is the fifth integral of motion.

The five integrals of motion $\Omega_{\rm F}(\Psi)$, $\eta(\Psi)$, $s(\Psi)$, $E(\Psi)$, and $L(\Psi)$, as well as the poloidal magnetic field $\mathbf{B}_{\rm p}$, allow us to find the toroidal magnetic field $B_{\hat{\varphi}}$ and all other plasma parameters:

$$\frac{I}{2\pi} = \frac{\alpha^2 L - (\Omega_{\rm F} - \omega)\varpi^2(E - \omega L)}{\alpha^2 - (\Omega_{\rm F} - \omega)^2\varpi^2 - \mathrm{M}^2}; \tag{13}$$

$$\gamma = \frac{1}{\alpha\mu\eta}\frac{\alpha^2(E - \Omega_{\rm F}L) - \mathrm{M}^2(E - \omega L)}{\alpha^2 - (\Omega_{\rm F} - \omega)^2\varpi^2 - \mathrm{M}^2}; \tag{14}$$

$$u_{\hat{\varphi}} = \frac{1}{\varpi\mu\eta}\frac{(E - \Omega_{\rm F}L)(\Omega_{\rm F} - \omega)\varpi^2 - L\mathrm{M}^2}{\alpha^2 - (\Omega_{\rm F} - \omega)^2\varpi^2 - \mathrm{M}^2}, \tag{15}$$

where

$$\mathrm{M}^2 = \frac{4\pi\eta^2\mu}{n}. \tag{16}$$

It is easy to see that M^2 is proportional (with the factor α^2) to the Mach number squared of the poloidal velocity $u_{\rm p}$ with respect to the Alfvén velocity $u_{\rm A} = B_{\rm p}/\sqrt{4\pi n\mu}$, i.e., $\mathrm{M}^2 = \alpha^2 u_{\rm p}^2/u_{\rm A}^2$.

Since $\mu = \mu(n, s)$, definition (16) allows us to express the concentration n (and hence the specific enthalpy μ) as a function of η, s, and M^2. This means that along with the five integrals of motion, the expressions for I, γ, and $u_{\hat{\varphi}}$ depend only on one additional quantity, namely the Mach number M. To determine the Mach number M, it is necessary to use the obvious relation $\gamma^2 - \mathrm{u}^2 = 1$, which, owing to eqs. (14) and (15), can be rewritten in the form

$$\frac{K}{\varpi^2 A^2} = \frac{1}{64\pi^4}\frac{\mathrm{M}^4(\nabla\Psi)^2}{\varpi^2} + \alpha^2\eta^2\mu^2, \tag{17}$$

where

$$A = \alpha^2 - (\Omega_{\rm F} - \omega)^2\varpi^2 - \mathrm{M}^2 \tag{18}$$

and

$$\begin{aligned} K = \alpha^2\varpi^2(E - \Omega_{\rm F}L)^2 &\left[\alpha^2 - (\Omega_{\rm F} - \omega)^2\varpi^2 - 2\mathrm{M}^2\right] \\ &+ \mathrm{M}^4\left[\varpi^2(E - \omega L)^2 - \alpha^2 L^2\right]. \end{aligned} \tag{19}$$

As for the Grad–Shafranov equation itself, i.e., the equilibrium equation for magnetic field lines, it can be written in the form

$$\begin{aligned} &\frac{1}{\alpha}\nabla_k\left\{\frac{1}{\alpha\varpi^2}[\alpha^2 - (\Omega_{\rm F} - \omega)^2\varpi^2 - \mathrm{M}^2]\nabla^k\Psi\right\} + \frac{\Omega_{\rm F} - \omega}{\alpha^2}(\nabla\Psi)^2\frac{\mathrm{d}\Omega_{\rm F}}{\mathrm{d}\Psi} \\ &+ \frac{64\pi^4}{\alpha^2\varpi^2}\frac{1}{2\mathrm{M}^2}\frac{\partial}{\partial\Psi}\left(\frac{G}{A}\right) - 16\pi^3\mu n\frac{1}{\eta}\frac{\mathrm{d}\eta}{\mathrm{d}\Psi} - 16\pi^3 nT\frac{\mathrm{d}s}{\mathrm{d}\Psi} = 0, \end{aligned} \tag{20}$$

where

$$G = \alpha^2\varpi^2(E - \Omega_{\rm F}L)^2 + \alpha^2\mathrm{M}^2L^2 - \mathrm{M}^2\varpi^2(E - \omega L)^2, \tag{21}$$

and the derivative $\partial/\partial\Psi$ acts on the integrals of motion only. Finally, expressing in eq. (20) the terms $\nabla_k\mathrm{M}^2$ according to eq. (19) we obtain [5]

$$\begin{aligned} &A\left[\frac{1}{\alpha}\nabla_k\left(\frac{1}{\alpha\varpi^2}\nabla^k\Psi\right) + \frac{1}{\alpha^2\varpi^2(\nabla\Psi)^2}\frac{\nabla^i\Psi\cdot\nabla^k\Psi\cdot\nabla_i\nabla_k\Psi}{D}\right] \\ &+ \frac{\nabla'_k A\nabla^k\Psi}{\alpha^2\varpi^2} - \frac{A}{\alpha^2\varpi^2(\nabla\Psi)^2}\frac{1}{2D}\nabla'_k F\nabla^k\Psi + \frac{\Omega_{\rm F} - \omega}{\alpha^2}(\nabla\Psi)^2\frac{\mathrm{d}\Omega_{\rm F}}{\mathrm{d}\Psi} \\ &+ \frac{64\pi^4}{\alpha^2\varpi^2}\frac{1}{2\mathrm{M}^2}\frac{\partial}{\partial\Psi}\left(\frac{G}{A}\right) - 16\pi^3\mu n\frac{1}{\eta}\frac{\mathrm{d}\eta}{\mathrm{d}\Psi} - 16\pi^3 nT\frac{\mathrm{d}s}{\mathrm{d}\Psi} = 0. \end{aligned} \tag{22}$$

Here

$$D = \frac{A}{\mathrm{M}^2} + \frac{\alpha^2}{\mathrm{M}^2}\frac{B_{\hat{\varphi}}^2}{B_{\rm p}^2} - \frac{1}{u_{\rm p}^2}\frac{A}{\mathrm{M}^2}\frac{c_{\rm s}^2}{1 - c_{\rm s}^2}, \tag{23}$$

$$F = \frac{64\pi^4}{\mathrm{M}^4}\frac{K}{A^2} - \frac{64\pi^4}{\mathrm{M}^4}\alpha^2\varpi^2\eta^2\mu^2, \tag{24}$$

and the gradient ∇'_k denotes the action of ∇_k under the condition that M is fixed. Let us stress that in eq. (22) the pressure P, the temperature T, the sound velocity $c_{\rm s}$, and the specific enthalpy μ are to be expressed via an equation of state in terms of the entropy $s(\Psi)$ and the square of the Mach number M^2. In turn, the quantity M^2 is to be considered as a function of $(\nabla\Psi)^2$ and the integrals of motion,

$$\mathrm{M}^2 = \mathrm{M}^2\left[(\nabla\Psi)^2, E(\Psi), L(\Psi), \eta(\Psi), \Omega_{\rm F}(\Psi), s(\Psi)\right]. \tag{25}$$

The latter relation is the implicit form of eq. (17). The stream equation (22) coupled with definitions (5) – (11) is the desired equation for the poloidal field which contains only the magnetic flux Ψ and the five integrals of motion $\Omega_{\rm F}(\Psi)$, $\eta(\Psi)$, $s(\Psi)$, $E(\Psi)$, and $L(\Psi)$ depending on it.

Equation (22) is a second-order equation linear with respect to the highest derivatives. It changes its type from elliptical to hyperbolical at singular surfaces where the poloidal velocity of matter becomes equal to either fast or slow magnetosonic velocity ($D = 0$), or to the cusp velocity ($D = -1$). Although at the Alfvénic surface, $A = 0$, the type of equation does not change, the Alfvénic surface does represent a singular surface of the Grad–Shafranov equation because a regularity condition must be satisfied there.

3 Examples

Bondi–Hoyle accretion. As a first example, we consider the hydrodynamic accretion onto a moving back hole (the Bondi–Hoyle accretion), which is one of the classical problems of modern astrophysics [1]. First of all, let us formulate the hydrodynamical limit of the Grad–Shafranov equation, where we can neglect the electromagnetic field contribution. In this case, it is convenient to introduce a new potential $\Phi(\Psi)$ satisfying the condition $\eta(\Psi) = \mathrm{d}\Phi/\mathrm{d}\Psi$. Using definition (8) we obtain

$$\alpha n \mathbf{u}_{\mathrm{p}} = \frac{1}{2\pi\varpi}(\nabla\Phi \times \mathbf{e}_{\hat{\varphi}}). \tag{26}$$

Surfaces $\Phi(r,\theta) = \mathrm{const}$ define the streamlines of matter.

In the hydrodynamic limit, there are only three integrals of motion. These are the energy flux and the z-component of the angular momentum:

$$E(\Phi) = \mu(\alpha\gamma + \varpi\omega u_{\hat{\varphi}}); \tag{27}$$

$$L(\Phi) = \mu\varpi u_{\hat{\varphi}}, \tag{28}$$

as well as the entropy $s = s(\Phi)$. Now the algebraic Bernoulli equation (17) takes the form

$$(E - \omega L)^2 = \alpha^2\mu^2 + \frac{\alpha^2}{\varpi^2}L^2 + \frac{\hat{\mathrm{M}}^4}{64\pi^4\varpi^2}(\nabla\Phi)^2, \tag{29}$$

where the 'Mach number' squared $\hat{\mathrm{M}}^2$ is defined as $\hat{\mathrm{M}}^2 = 4\pi\mu/n$. Then the Grad–Shafranov equation (20) can be rewritten in the form [5]

$$-\frac{1}{\alpha}\nabla_k\left(\frac{\hat{\mathrm{M}}^2}{\alpha\varpi^2}\nabla^k\Phi\right) - 16\pi^3 nT\frac{\mathrm{d}s}{\mathrm{d}\Phi}$$
$$+\frac{64\pi^4}{\alpha^2\varpi^2\hat{\mathrm{M}}^2}\left[\varpi^2(E-\omega L)\left(\frac{\mathrm{d}E}{\mathrm{d}\Phi} - \omega\frac{\mathrm{d}L}{\mathrm{d}\Phi}\right) - \alpha^2 L\frac{\mathrm{d}L}{\mathrm{d}\Phi}\right] = 0, \tag{30}$$

where now

$$D = -1 + \frac{1}{u_{\mathrm{p}}^2}\frac{c_s^2}{1-c_s^2}. \tag{31}$$

As we see, eq. (30) contains only one singular surface, i.e., the sonic surface, determined from the condition $D = 0$.

To construct the solution corresponding to the Bondi–Hoyle accretion, it is possible to seek the solution of the Grad–Shafranov equation for the flux function $\Phi(r,\theta)$ in the form of a small perturbation of the spherically symmetric flow in the reference frame moving with the black hole

$$\Phi(r,\theta) = \Phi_0[1 - \cos\theta + \varepsilon_1 f(r,\theta)]. \tag{32}$$

Here we introduce a small parameter

$$\varepsilon_1 = \frac{v_\infty}{c_\infty} \tag{33}$$

which defines the ratio of the black hole velocity to the velocity of sound at infinity. For a nonmoving gravity center we return to the spherically symmetric flow. Thus, the Grad–Shafranov equation can be linearized:

$$-\varepsilon_1\alpha^2 D\frac{\partial^2 f}{\partial r^2} - \frac{\varepsilon_1}{r^2}(D+1)\sin\theta\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\frac{\partial f}{\partial\theta}\right) + \varepsilon_1\alpha^2 N_r\frac{\partial f}{\partial r} = 0, \tag{34}$$

where

$$N_r = \frac{2}{r} - \frac{\mu^2}{E^2 - \alpha^2\mu^2}\frac{M}{r^2}. \tag{35}$$

It is extremely important that according to (26) and (31),

$$D + 1 = \frac{\alpha^2\mu^2}{E^2 - \alpha^2\mu^2}\cdot\frac{c_{\mathrm{s}}^2}{1 - c_{\mathrm{s}}^2}, \tag{36}$$

so the factor α^2 enters every term of eq. (34). Hence, eq. (34) has no singularity at the horizon. In particular, it means that it is not necessary to specify any boundary conditions for $r = r_{\mathrm{g}}$. It is not surprising because the horizon corresponds to the supersonic region which cannot affect the subsonic flow. As a result, the complete solution can be presented in the form

$$\Phi(r,\theta) = \Phi_0[1 - \cos\theta + \varepsilon_1 g_1(r)\sin^2\theta)], \tag{37}$$

where the radial function $g_1(r)$ is the solution of the ordinary differential equation.

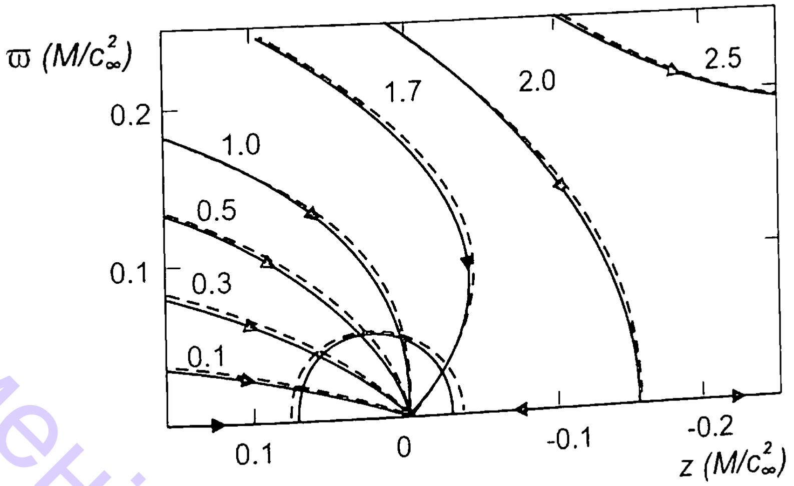


Figure 1: Bondi–Hoyle accretion onto moving black hole.

This means that we have constructed the analytical solution of the problem, i.e., obtained the full description of the flow structure. For example, the sonic surface has now the nonspherical form

$$r_*(\theta) = r_*\left[1 + \varepsilon_1\left(\frac{\Gamma+1}{5-3\Gamma}\right)k_2\cos\theta\right], \tag{38}$$

where the numerical coefficient $k_2 = r_* g_1'(r_*)$ is expressed through the derivative of the radial function $g_1(r)$ at the sonic point. As shown in Figure 1, the analytical solution fully agrees with the numerical calculations [8] in spite of the parameter $\varepsilon_1 = 0.6$ here being quite large. Here $\Gamma = 4/3$, and the numbers alongside the curves denote values Φ/Φ_0; the dashed lines show the streamlines and the form of the sonic surface obtained numerically.

As can be easily seen, outside the capture radius

$$R_{\mathrm{c}} \approx \varepsilon_1^{-1/2} r_* \tag{39}$$

our main assumption, i.e., the smallness of the deviation from the spherically symmetric flow, is not valid. Nevertheless, the solution found remains correct. This remarkable property is due to the Grad–Shafranov equation becoming linear for constant concentration n. But as we learn from the spherically symmetric Bondi accretion, at large distances $r \gg r_*$ from the sonic surface the density of the accreting matter is virtually constant. Accordingly, the concentration is constant for a homogeneous flow as well. As a result, under the condition that $R_c \gg r_*$, which holds true for $\varepsilon_1 \ll 1$, near and beyond the capture radius (where the perturbation $\sim \varepsilon_1 g_1(r)$ becomes comparable to unity) eq. (30) becomes linear. So that the sum of the two solutions, homogeneous and spherically symmetric ones, is also a solution.

Thin transonic disk. As a next example, we consider the internal two-dimensional structure of a thin accretion disk. Here, for simplicity we consider the case of a nonrotating (Schwarzschild) black hole [9]. We recall that according to the standard model the accreting matter forms an equilibrium disk rotating around the gravitational center with the Keplerian velocity $v_{\rm K}(r) = (GM/r)^{1/2}$. The disk will be thin provided that its temperature is sufficiently small ($c_s \ll v_{\rm K}$) since the vertical balance of the gravity force and the pressure gradient implies that

$$H \approx r \frac{c_{\rm s}}{v_{\rm K}}. \tag{40}$$

The General Relativity effects result in two important properties: the absence of stable circular orbits for $r < r_0 = 3r_{\rm g}$ and the transonic regime of accretion. The first point means that the accreting matter passing the marginally stable orbit approaches the black hole horizon sufficiently fast, namely, in the dynamical time $\tau_{\rm d} \sim [v_r(r_0)/c]^{-1/3} r_{\rm g}/c$. It is important that such a flow is realized in the absence of viscosity. The second statement results from the fact that up to the marginally stable orbit the flow is subsonic while at the horizon the flow is to be supersonic.

Up to now in the majority of studies devoted to thin accretion disks the procedure of vertical averaging was used, where the vertical four-velocity $u_{\hat\theta}$ was assumed to be zero [10]. As a result, the vertical component of the dynamic force $nu^b\nabla_b(\mu u_a)$ in the Euler equation was postulated to be unimportant up to the horizon. For this reason the disk thickness was determined by the pressure gradient even in the supersonic region near the black hole [11]. Here I am going to demonstrate that the assumption $u_{\hat\theta} = 0$ is not correct. As for the Bondi accretion, the dynamic force is to be important in the vicinity of the sonic surface.

Figure 2 shows the structure of a thin accretion disk after passing the marginally stable orbit $r = 3r_{\rm g}$ obtained by solving eq. (30) numerically for $c_0 = 10^{-2}$, $u_0 = 10^{-5}$, i.e., in the presence of a small parameter $\varepsilon_2 = u_0/c_0 \ll 1$. The solid lines correspond to the range of parameters $u_{\rm p}^2/c_0^2 < 0.2$. The dashed lines indicate the extrapolation of the solution to the sonic-surface region. In the vicinity of the sonic surface the flow has the form of the standard nozzle.

As we see, the flow structure near the sonic surface is far from being radial. The appearance of the narrow waist has a simple physical meaning. Indeed, the density remains almost constant for subsonic flow while the radial velocity increases from u_0 to $c_* \sim c_0$, i.e., for $\varepsilon_2 \ll 1$ it changes over several orders of magnitude. As a result, the disk thickness H should change in the same proportion owing to the

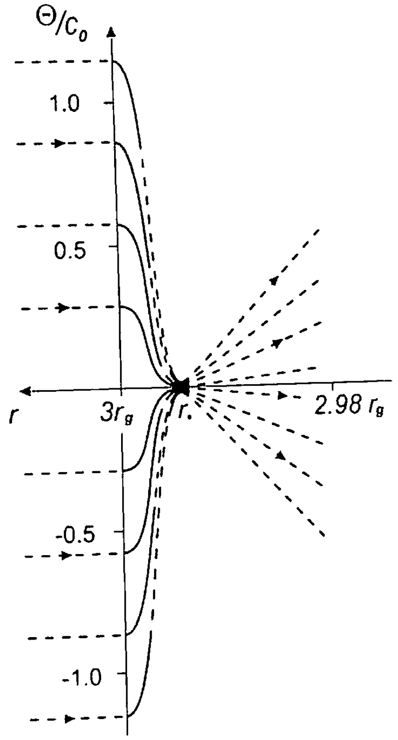


Figure 2: Nozzle shape of a thin transonic disk near a marginally stable orbit.

continuity equation as well,

$$H(r_*) \approx \frac{u_0}{c_0} H(3r_{\rm g}). \tag{41}$$

Here it is extremely important that both components of the dynamical force become comparable with the pressure gradient near the sonic surface:

$$\frac{u_{\hat\theta}}{r}\frac{\partial u_{\hat\theta}}{\partial\theta} \approx u_{\hat r}\frac{\partial u_{\hat\theta}}{\partial r} \approx \frac{\nabla_{\hat\theta} P}{\mu} \approx \frac{c_0^2}{u_0^2}\frac{\theta}{r}, \tag{42}$$

where the angle θ is counted off from the equatorial plane.

In other words, if there appears a nonzero vertical velocity component, the dynamical term $(\mathbf{v}\nabla)\mathbf{v}$ cannot be neglected in the vertical force balance near the sonic surface [9]. It is clear that this property remains valid for arbitrary radial velocity of the flow, i.e., even when the transverse contraction of the disk is not so pronounced. Taking dynamical forces into account causes two additional degrees of freedom to appear, which relate to the higher derivatives in the Grad–Shafranov equation. This also leads to extra conclusions independent of the value of ε_2. In thin accretion disks, the critical condition at the sonic surface does not fix the accretion rate any more; it determines the bending of the streamlines near the sonic surface.

Finally, the inclusion of the vertical velocity inevitably leads to the appearance of a small longitudinal scale $\delta r_\parallel \approx H_*$ in the vicinity of the sonic surface, which for a thin disk proves to be much smaller than the distance to the black hole for any value of the parameter u_0/c_0. In the standard one-dimensional approach, this scale does not emerge. As for the supersonic region (and, in particular, the region near the horizon), the disk thickness here will be determined not by the pressure gradient, but by the form of ballistic trajectories.

The Blandford–Znajek process. Finally, we discuss the energy loss of a rotating black hole embedded in an external magnetic field – the so-called Blandford-Znajek process [12]. This process is considered to be the preferential mechanism of energy release in active galactic nuclei, microquasars, and even cosmological gamma-ray bursters [13]. The main idea is based on the analogy with the energy transfer in the internal regions of radio pulsar magnetospheres. Indeed, let us suppose that there is a regular external magnetic field in the vicinity of a rotating black hole, and that the electric current I flows along magnetic field lines. Then, the electric field $\mathbf{E}$, which is induced by plasma rotating with angular velocity $\Omega_{\rm F}$, and the toroidal magnetic field $\mathbf{B}_{\varphi}$, which is due to the longitudinal current I, generate the electromagnetic energy flux (the Poynting vector flux) carrying the energy away along the magnetic field lines.

Of course, by definition, general relativity effects are important near the black hole. Consequently, it is not obvious that the pulsar analogy can be useful in all cases. For example, in radio pulsars, the braking of neutron stars results from the Ampére force acting on the star surface. This force results from the surface currents shorting the electric currents flowing in the pulsar magnetosphere [14]. In the case of black holes such currents cannot lead to deceleration as the event horizon is not a physically preferred surface, though surface currents themselves can be formally introduced in the framework of the so-called membrane paradigm [1].

Indeed, let us consider the well-known condition at the horizon (the absence of infinite electromagnetic fields in the reference frame comoving with a free falling observer) $E'_0 \to (E_0 + B_{\varphi})/\alpha < \infty$. It means that $E_0 + B_{\varphi} \to 0$ which can be rewritten in the form

$$4\pi I(\Psi) = [\Omega_{\rm H} - \Omega_{\rm F}(\Psi)] \, \frac{r_{\rm g}^2 + a^2}{r_{\rm g}^2 + a^2 \cos^2\theta} \sin\theta \left(\frac{{\rm d}\Psi}{{\rm d}\theta}\right)_{r_{\rm g}} . \tag{43}$$

Here $a = J/M$ is the rotation parameter. This relation was used [12] as the "boundary condition on the horizon". But it is clear that the horizon is not causally connected with the outer space. For this reason the conclusion was made that there is no energy flux along magnetic field lines passing through the black hole horizon [15], and hence a black hole cannot work as a unipolar inductor extracting the rotation energy by electromagnetic stresses.

However, a recent more accurate analysis [16] (in which, in fact, the first solution of the Grad-Shafranov equation for nonzero particle mass in the Kerr metric was obtained) indicates that the braking torque acts in the plasma generation region above the black hole horizon. Such a torque appears due to the action of long-range gravitomagnetic forces which penetrate into the regions causally connected with the outer space. As a result, it was shown that for a finite mass of particles in the very vicinity of the horizon, there is a hyperbolic region of the Grad–Shafranov equation which is altogether absent in the force-free approximation. Hence, the full version of the Grad–Shafranov equation needs no boundary condition on the horizon. Actually, this property was already demonstrated above by the example of the Bondi–Hoyle accretion. Thus, in this case it is impossible to consider relation (43) as a boundary condition. Relation (43) is automatically true for any solution of the Grad–Shafranov equation which can be extended up to the horizon.

On the other hand, in the force-free approximation, when the Grad–Shafranov equation remains elliptical up to the black hole horizon, extra condition (43) is to be included into consideration [17]. But as was demonstrated [7], this condition is actually the manifestation of the critical condition on the fast magnetosonic surface. In reality this condition is specified on the surface which does not coincide with the event horizon, and, hence, is in the causal connection with the outer magnetosphere.

Thus, the Blandford–Znajek mechanism of electromagnetic energy extraction from rotating black holes faces no causality problem. As in the pulsar magnetosphere, if there is enough secondary plasma to screen the longitudinal electric field, its charge density and electric currents produce the flux of electromagnetic energy propagating from the central star to infinity. For the same reason, a rotating black hole embedded into an external magnetic field works as a unipolar inductor extracting its energy of rotation by the flux of the electromagnetic energy.

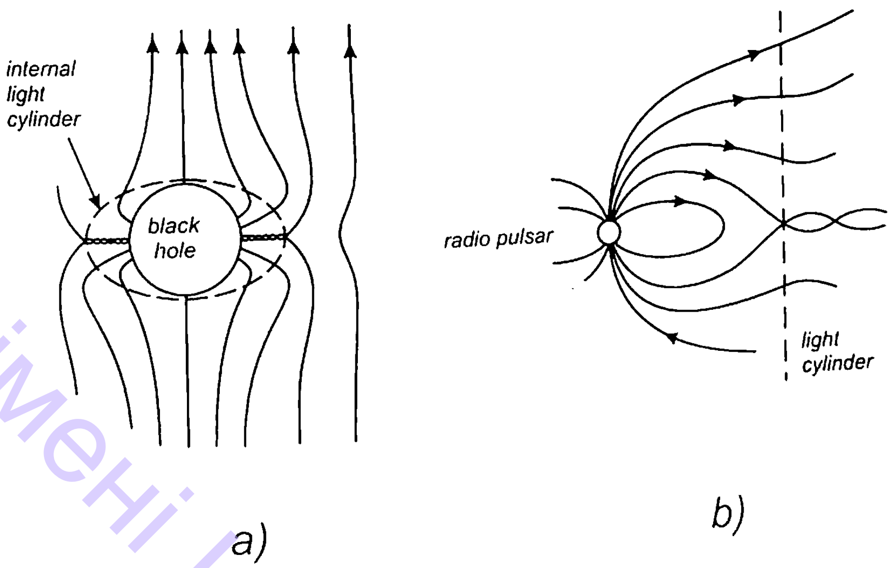


Figure 3: Magnetospheres of a rotating black hole (a) and a radio pulsar (b).

It should be emphasized, however, that the rate of the energy release needed to explain the luminosity of active galactic nuclei can be achieved only for the extreme black hole mass $\sim 10^9 M_{\odot}$, the extreme magnetic field $B \sim B_{\rm Edd}$, and the extreme angular velocity $a \sim M$. Therefore, some papers have recently appeared in which the efficiency of the Blandford–Znajek process in real astrophysical conditions was questioned [18]. In particular, it was pointed out that for rapid rotation the Wald solution for the vacuum magnetosphere leads to the magnetic field being pushed out from the horizon into the ergosphere [1], which causes the appearance of the additional factor $(1 - a^2/M^2) \to 0$ in the expression of the energy loss. But as shown in Figure 3a, in the black hole magnetosphere filled with plasma, all magnetic field lines crossing the surface of the internal light cylinder $\alpha^2 = (\Omega_F - \omega)^2 \varpi^2 + {\rm M}^2$ are to cross the black hole horizon as well, that is why to an order of magnitude the energy loss for extremely rotating black holes coincides with the maximum loss. Indeed, here the situation is to be fully analogous to the pulsar magnetosphere where the field lines passing the external light cylinder do not intersect the equatorial plane (Figure 3b). The magnetic field structure shown in Figure 3a has been recently obtained numerically [19].

Acknowledgments

The author is greatly indebted to the Organizing Committee of the Gamov Conference for the hospitality. This work was supported by Grant No 02-02-16762 from the Russian Foundation for Basic Research.

References

[1] K.S. Thorne, R.H. Price, D.A. Macdonald, *Black Holes. The Membrane Paradigm*, Yale University Press, New Haven (1986).

[2] V.D. Shafranov, *Sov. Phys. JETP*, **6**, 545 (1958).

[3] H. Grad, *Rev. Mod. Phys.*, **32**, 830 (1960).

[4] L.S. Soloviev, in M.A. Leontovich (ed) *Review of Plasma Physics*, **3**, Consultants Bureau, p.1 1967.

[5] V.S. Beskin, V.I. Pariev, *Physics Uspekhi*, **36**, 529 (1993).

[6] V.S. Beskin, Yu.N. Pidoprygora, *Physics JETP*, **80**, 575 (1995).

[7] V.S. Beskin, *Physics Uspekhi*, **40**, 659 (1997).

[8] R. Hunt, *Mon. Not. Roy. Astron. Soc.*, **198**, 83 (1979).

[9] V.S. Beskin, R.Yu. Kompaneetz, A.D. Tchekhovskoy, *Astronomy Letters*, **28**, 543 (2002).

[10] B. Paczyński, G.S. Bisnovatyi-Kogan, *Acta Astron.*, **31**, 283 (1981).

[11] M.A. Abramowicz, A. Lanza, M.J. Percival, *Astrophys. J.*, **479**, 179 (1997).

[12] R. Blandford, R.L. Znajek, *Mon. Not. R. Astron. Soc.*, **179**, 433 (1977).

[13] H.K. Lee, R.A.M.J. Wijers, G.E. Brown, *Phys. Rep.*, **325**, 83 (2000).

[14] V.S. Beskin, *Physics Uspekhi*, **42**, 1071 (1999).

[15] B. Punsly, F.V. Coroniti, *Astrophys. J.*, **350**, 518 (1990).

[16] V.S. Beskin, I.V. Kuznetsova, *Nuovo Cimento B*, **115**, 795 (2000).

[17] D.A. Uzdensky, *Astrophys. J.*, **603**, 652 (2004).

[18] M. Livio, G.I. Ogilvie, J.E. Pringle, *Astrophys.J.*, **512**, 100 (1999).

[19] S. Koide, *Phys. Rev. D*, **67**, 104010 (2003).

18

Super Star Clusters and Super Star Cluster Winds

S. Silich

Instituto Nacional de Astrofísica Optica y Electrónica,
AP 51, 72000 Puebla, México
e-mail: silich@inaoep.mx

Abstract

This paper discusses the tremendous progress made in the last decade in our observational and theoretical understanding of young, massive, compact star clusters. It also summarizes the observational properties of such clusters in the IR, radio, visible and X-ray regimes and show that in many cases they lead to extreme star cluster parameters: the energy release rate and stellar mass concentration. Then the existing theoretical models are critically discussed. In particular the limitations given by radiative gas cooling on the standard adiabatic Chevalier & Clegg model are considered and a new, radiative super star cluster wind model is formulated that seems to be more adequate for very compact and massive star clusters. The paper demonstrates that because of cooling the stationary wind solution vanishes when star cluster energy deposition rates exceed the threshold limit. This implies that, contrary to common believe, most massive and compact star clusters are not able to generate superwinds. Finally the expected appearance of super star cluster winds in the X-ray regime are discussed.

Keywords: star clusters, star cluster winds, hydrodynamics, X-rays

1 Introduction

Despite the phenomenon of massive and extremely compact star clusters being known for a long time from the ground-based observations of galaxies with violent star formation (e.g. R136 in 30 Doradus, [2, 3]), it took the power of Hubble Space Telescope (HST) to detect hundreds of such clusters in merging and starburst galaxies and to prove that Super Star Clusters (SSC) are not a scarce, unique phenomenon, but rather a regular mode of star formation in many galaxies. They were also discovered in the nuclei of spiral galaxies, circumnuclear rings, and are now observed in many wavelengths, from the infrared (IR) and radio to the visible and X-ray regimes.

Early results came from Holtzman et al. [4] HST study of the Perseus A central, merging elliptical galaxy NGC 1275. This revealed a population of bright, compact ($\leq$ 15 pc) point-like sources with masses exceeding 10^5 M$_\odot$ and ages less than a few hundred million years. These seem to be stable against tidal destruction and were proposed to be present day proto-globular clusters. Further on tremendous progress has been made in the observational and theoretical understanding of similar objects which were found in many distant merging and interacting systems, starburst galaxies and circumnuclear star-forming rings (see [5, 16] and references therein). It seems that the formation of the SSCs does not depend on the host galaxy metallicity [31], but are usually related to present day or past episodes of star formation.

The ages of SSCs spread from a few million years to several hundred million years. They have very high luminosities which in many cases surpass the luminosity of the R136 cluster. Nevertheless little is known about their internal structure. The best defined observational parameters are SSC half-light (or effective) radii, which are usually found within the range (3–10) pc [5] and SSC masses. The first mass-estimates from population synthesis models show that typical values fall into the range 10^5–10^6 M$_\odot$. Several direct measurements of the line-of-sight stellar velocity dispersion for SSCs in nearby dwarf galaxies NGC 1569 and NGC 1705 (see, for example, [6, 7]) gave similar values. However recent day dynamical mass estimates shifted the upper mass limit to a larger value. Walcher et al. [32] VLT determination of the stellar velocity dispersion for 9 SSCs in the nuclei of bulge-less spiral galaxies in combination with light profiles derived from the HST photometry shows that masses of the nuclear star clusters may exceed the 10^7 M$_\odot$ limit and reach 6×10^7 M$_\odot$.

SSCs are observed at different wavelengths. NGC 5253 supernebula is an impressive example of the multi-waveband study of a young SSC that seems still to be covered by its natal cloud. NGC 5253 is a special starburst galaxy whose radio continuum is almost entirely thermal. This suggests a very young starburst which still may be at the pre-supernovae stage. The supernebula was first observed at the VLA by Turner at al. [29] at 1.3 cm and 2 cm wavelength. To distinguish between the three possible supernebula models (SNR, AGN and HII region) Gorjian et al. [11] obtained 11.7 and 18.7 μm images of NGC 5253 with Keck telescope and concluded that IR-radio flux ratio favors the HII region hypothesis. The optically thick ionization-bounded model requires around 4×10^{52} s^{-1} Lyc photons to ionize a very compact (less than 2 pc in diameter) and dense ($n_e \sim 4 \times 10^4$ cm $^{-3}$) HII region. This provides more than half of the radio emission from the galaxy and implies that a (5–7) $\times 10^5$ M$_\odot$ star cluster is embedded into less than 1 pc radius volume. The Brackett line full width half maximum, FWHM = 76 km s^{-1}, suggests that, despite the supersonic internal motion, the super-nebula remains gravitationally bound (see [30]).

As the evolutionary time grows, the stellar feedback leads to the disruption of the parent molecular cloud by the intense ionization radiation, stellar winds and supernova explosions. The random supernova and individual stellar wind collisions lead to the thermalization of the released kinetic energy and to the super-nebula gas

heating. This in turn provides a high thermal overpressure that results in the ejected gas outflow with supersonic velocity, the super star cluster wind. This should be observed in the X-ray regime.

Indeed, X-ray emission from nearby starburst galaxies includes both, thermal diffuse component and unresolved discrete point sources. However the origin of the unresolved emission is not clear at the moment.

For example, Chandra and XMM-Newton background subtracted data revealed seventy discrete sources within the D_{25} ellipse of NGC 5253 [26]. 6 of them are suspected to be due to the local bubbles driven by individual SSCs. Most of them are very soft and have thermal spectra.

On the other hand, Kaaret et al. [12] provided a careful examination of the Chandra X-ray Observatory data archive together with HST NICMOS infrared images for three nearby starburst galaxies, M82, NGC 1569 and NGC 5253. The X-ray luminosities were calculated for the 0.3–8.0 keV energy band with the X-ray luminosity detection limit in between 5×10^{35} erg s^{-1} and 3×10^{36} erg s^{-1}. The positions of the 66 discrete X-ray sources then have been compared with locations of the 121 most luminous clusters found in these galaxies. The X-ray sources are preferentially located close to the star clusters. Nevertheless a significant displacement from the nearest clusters was regularly found that led to the conclusion (see [12]) that most of the observed discrete sources are X-ray binaries ejected from the parent dense stellar cluster.

On the other hand, X-ray emission from the massive star clusters was definitely observed in the Milky Way central region. Yusef-Zadeh et al. [33, 15] reported the first detections of X-ray emission from the Arches and Quintuplet clusters located within the central 30 pc of the Milky Way galaxy. The observed X-ray emission represents a mixture of two components: the diffuse emission and discrete point-like sources. The overall absorption-corrected luminosity within the (0.5–8) keV energy band from the more massive Arches cluster is $L_x \approx (0.5\text{–}3) \times 10^{35}$ erg s^{-1} with approximately 40% of the flux coming from the diffuse component and 60% from the discrete sources. The best spectral fit requires high, (4–5) times solar, metallicity of the X-ray plasma. The diffuse component is suggested to be from the Arches wind outflow. However a contribution from the scattered fluorescent emission from the nearby molecular cloud cannot be ruled out. The comparison of the astrometric positions for IR and X-ray sources indicates a systematic offset in the Chandra pointing that also should be considered for distant galaxies.

2 The Models

2.1 Adiabatic wind model

A prototype model for super star cluster winds was proposed in [10]. It assumes that all energy deposited by massive stars and supernova explosions is thermalized inside the star cluster volume by random collisions between individual stellar winds and supernova ejecta. This produces a high thermal overpressure which accelerates the ejected material and finally blows it out of the star cluster volume with high velocity. Three parameters completely define the Chevalier & Clegg adiabatic solution (the distribution of the wind density, temperature, thermal pressure and expansion velocity). They are: the energy and mass deposition rates, L_{SC} and $\dot{M}_{SC}$, and the radius, R_{SC}, of the star cluster. In the adiabatic case the central gas temperature T_c is fixed by the ratio of the energy deposition rate L_{SC} to the mass deposition rate $\dot{M}_{SC}$, $T_c \sim L_{SC}/\dot{M}_{SC}$, and may reach 10^7–10^8K. This implies that the thermalized outflow should be observed in the X-ray regime. Cantó et al. [8] confirmed Chevalier & Clegg analytic results by means of 3D adiabatic numerical calculations.

X-ray emission from the central Galactic clusters was predicted in [10]. Raga et al. [18] derived the X-ray luminosity of Arches cluster from their 3D numerical model. They took as an input model a star cluster with 60 identical massive stars, assumed a mean individual stellar wind terminal speed V_∞ = 1500 km s^{-1} and found reasonable agreement with data reported in [33]. Later on Stevens & Hartwell [24] compared Chevalier & Clegg model with Chandra observations of five nearby young massive clusters: R136 in 30 Doradus, NGC 3603, NGC 346, The Rosette Nebula and the Arches cluster. For the calculations they adopted different Arches cluster parameters: the total mass deposition rate $\dot{M}_{sc} = 7.3 \times 10^{-4}$ M$_\odot$ yr^{-1} and the average individual stellar wind terminal speed V_∞ = 2810 km s^{-1}. The last value was derived from observations of the individual massive stars in the field of the Arches cluster (see [13, 14]). Stevens & Hartwell [24] concluded that the temperature of the hot X-ray plasma is below the Chevalier & Clegg value and that observed X-ray luminosities are systematically above the standard model predictions.

2.2 Radiative wind model

The Chevalier & Clegg model has recently been revised by Silich et al. (see [21, 22]) who demonstrated that in the case of compact and massive star clusters the adiabatic conditions may not apply because of strong radiative cooling. In the radiative case the proper wind central values (T_c and ρ_c) must be selected by iterations from the condition that the outflow crosses the star cluster surface with the local sound speed. In the radiative case the wind central temperature, T_c, and the wind central density, n_c, are not independent as in the adiabatic solution [9]. They are related by (see [22]):

$$n_c = \sqrt{\frac{q_e - \frac{q_m}{\gamma-1} c_c^2}{\Lambda(T_c)}} = q_m^{1/2} \sqrt{\frac{\frac{V_{A\infty}^2}{2} - \frac{c_c^2}{\gamma-1}}{\Lambda(T_c)}}, \qquad (1)$$

where q_e and q_m are the energy and mass deposition rates per unit volume ($q_e = 3\dot{E}_{sc}/4\pi R_{sc}^3$; $q_m = 3\dot{M}_{sc}/4\pi R_{sc}^3$), γ is the ratio of specific heats, and $\Lambda(T)$ is the cooling function. This implies that in the radiative case T_c is always below the adiabatic value ($T_{c,ad} = \frac{\gamma-1}{\gamma} \frac{\mu}{k} \frac{q_e}{q_m}$), where μ is the mean mass per particle and k is the Boltzmann constant. Equation (1) also implies that the central wind pressure cannot be arbitrarily large, as it is in the adiabatic case. Therefore the sonic point radius, R_{sonic}, which in the adiabatic case is in direct proportion to the central wind pressure ($R_{sonic} \sim P_c/(2q_eq_m)^{1/2}$) cannot be arbitrarily large. The last statement may be reformulated (see [22]) in terms of the critical, threshold energy rate. If the mechanical energy deposition rate exceeds the critical value, the sonic point moves

under the star cluster radius and the stationary wind solution vanishes. We can find the threshold energy rate from the condition that the maximum possible sonic point radius approaches the star cluster radius, $R_{sonic} = R_{SC}$. Figure 1 displays the threshold energy rate for two different values of the adiabatic wind terminal speed, that is for two different ratios $L_{SC}/\dot{M}_{SC} = V_{A\infty}^2/2$ (see [22, 28]).

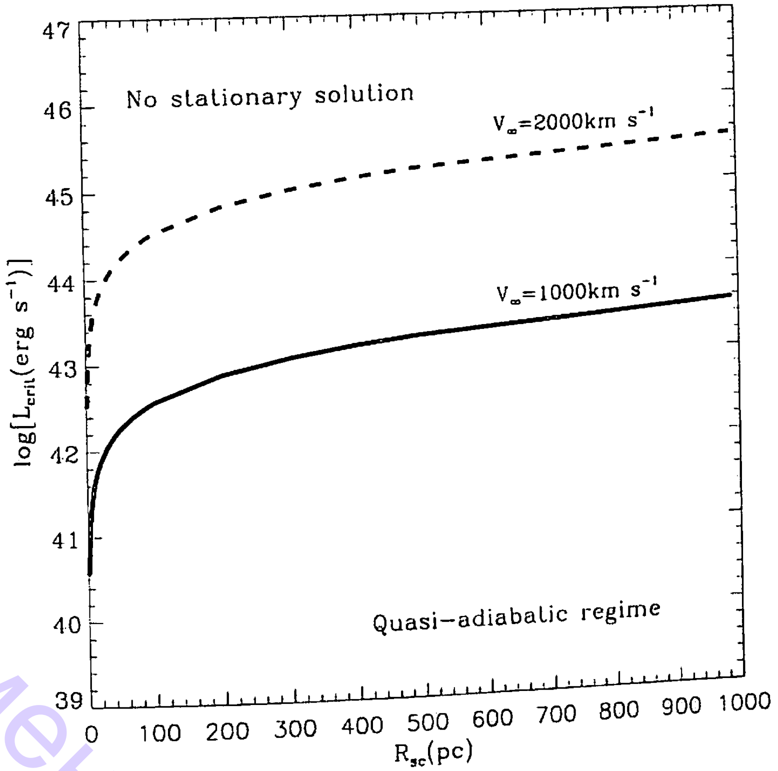


Figure 1: The threshold energy input rate for two different ratios of $L_{SC}/\dot{M}_{SC} \sim V_{A\infty}^2$. The solid line represents the threshold energy rate for $V_{A\infty}$ = 1000 km s^{-1}. The dashed line is the same for $V_{A\infty}$ = 2000 km s^{-1}. The thermal stationary wind solution does not exist if L_{SC} exceeds the threshold value.

Winds driven by SSCs whose power is close (but still below) the critical value, evolve in the strongly radiative regime. In this case cooling modifies drastically the adiabatic wind temperature distribution bringing the boundary of the X-ray zone and the photo-ionized envelope closer to the star cluster surface (Figure 2). This leads to the concentration of the diffuse X-ray emission inside a small volume whose radius, R_{Xcut}, is only a few times larger than the star cluster radius R_{SC}. Because of the sharp temperature drop (see Figure 2) and r^{-2} density distribution, the emission measure of the high velocity photo-ionized gaseous envelope may be much larger than it is expected from the adiabatic model. It was suggested (see [22]) that this may be observed as a broad ($\sim$ 1000 km s^{-1}) emission line component of low intensity.

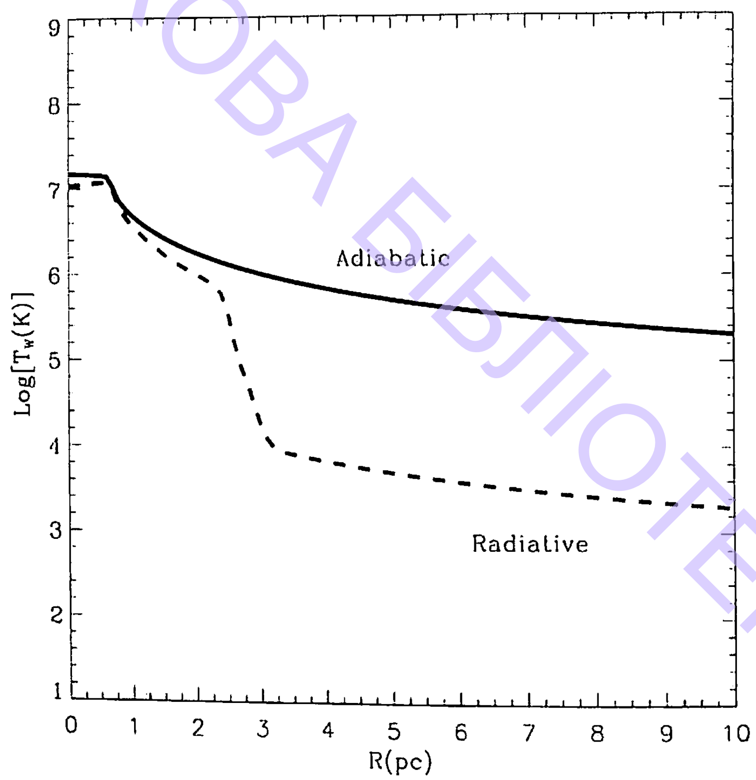


Figure 2: The stationary wind temperature distributions for adiabatic (solid line) and radiative (dashed line) solutions. The star cluster parameters are the same for both cases.

3 What is Going on in the Supercritical Regime?

When the energy deposition rate exceeds the critical value and the stationary wind solution vanishes, $\sim$ 30% of the deposited energy is radiated away from the star cluster volume [22]. This implies that in the supercritical regime cooling reduces the gradient of the thermal pressure which is not able now to support the stationary wind condition: $\dot{M}_{SC} = 4\pi R^2_{SC}\rho(R_{SC})u(R_{SC})$. Nevertheless this does not imply that catastrophic cooling of the ejected material occurs immediately. The lack of balance between the amount of mass ejected by supernovae and the amount of mass which flows away from the cluster, results in the accumulation of the ejected material inside the star cluster and a gradual growth of the gas density. This phase has been discussed qualitatively in [28]. It was suggested that cooling leads first to the formation of the photo-ionized super-nebula whose typical temperature is $\leq 10^4$ K. The super-nebula expands slowly (with expansion velocity a few tens of km s^{-1}) into the surrounding medium. This stage continues until the massive stars producing the UV photons leave the main sequence and evolve into supernovae. It ends up with the further decrease of the ejected gas temperature and mass accumulation. Simple estimations show that the characteristic time-scale for mass accumulation, $\tau \approx 2\pi(\rho_2 - \rho_1)R^3_{SC}V^2_{A\infty}/3L_{SC} \sim 10^5$ yr., is rather short (see [28]). However this scenario is waiting for confirmation by full hydrodynamical calculations.

4 The Ejected Gas Metallicity

The impact of cooling is highly dependent on the gas composition. The metallicity of the ISM is dependent on the amount of metals produced by massive stars and their ability to be mixed with the rest of the matter by atomic or turbulent diffusion (see [27, 17]). In the case of SSC winds all metals ejected by supernova explosions remain bound within the wind and are separated from the interstellar gas until the ejected material meets the reverse shock, generated by the collision of the high velocity gas with the surrounding medium, or being ejected into the intergalactic medium and then dispersed throughout intergalactic space. This implies that wind metallicity is completely defined by the massive stars metal yields and by the amount of hydrogen ejected by supernovae and individual stellar winds, e.g. by the super star cluster parameters: the initial mass function (IMF), stellar cluster lower and upper mass cut-offs and by the evolutionary tracks of massive stars. If these parameters are known, we can calculate the differential star cluster metal yields and thus the time evolution of the ejected gas metallicity (see, for example, [20]):

$$Z_w = \frac{M_{met,ej}}{Z_{\odot,met}M_{ej}}, \tag{2}$$

where Z_w is measured in Solar units, $M_{met,ej}$ is the mass of a particular element which is deposited per unit time interval by supernova explosions and stellar winds, and M_{ej} is the total amount of mass deposited by the star cluster per unit time. Figure 3 displays the ejected gas metallicity if oxygen is used as a metallicity tracer (see [28]). The high metallicity of the ejected material suggests an effective wind cooling at least within the first 10 Myrs of its evolution. The enhanced abundances also affect an X-ray appearance of run-away material (see discussion in [25, 20]).

5 X-ray Luminosity

The X-ray luminosity of the star cluster wind is:

$$L_x = 4\pi \int_0^{R_{xcut}} n_e n_i \Lambda_x(T, Z) r^2 \mathrm{d}r, \tag{3}$$

where R_{xcut} is the X-ray zone cutoff radius, n_e and n_i are electron and ion number densities and $\Lambda_x(T, Z)$ is the emissivity of the hot plasma in the X-ray energy band. We can easily obtain L_x if the temperature and density distributions and the outflow chemical composition are known. All X-ray models discussed so far (see [24, 8, 18, 22]) were focused on the analysis of nearby clusters of moderate mass and required more or less complicated numerical calculations that failed to pick out the main parameters which are responsible for the X-ray appearance of the star cluster winds. Meanwhile we can integrate (3) analytically and obtain a simple analytical formula for the X-ray luminosity from the star cluster winds using several very reasonable simplifications. Indeed, we know from the adiabatic model [9] and our radiative semi-analytical calculations that both the density and the temperature of the outflow remain almost constant (within a factor of three) inside the star cluster radius. Outside of the star cluster they rapidly approach their asymptotic r^{-2} and

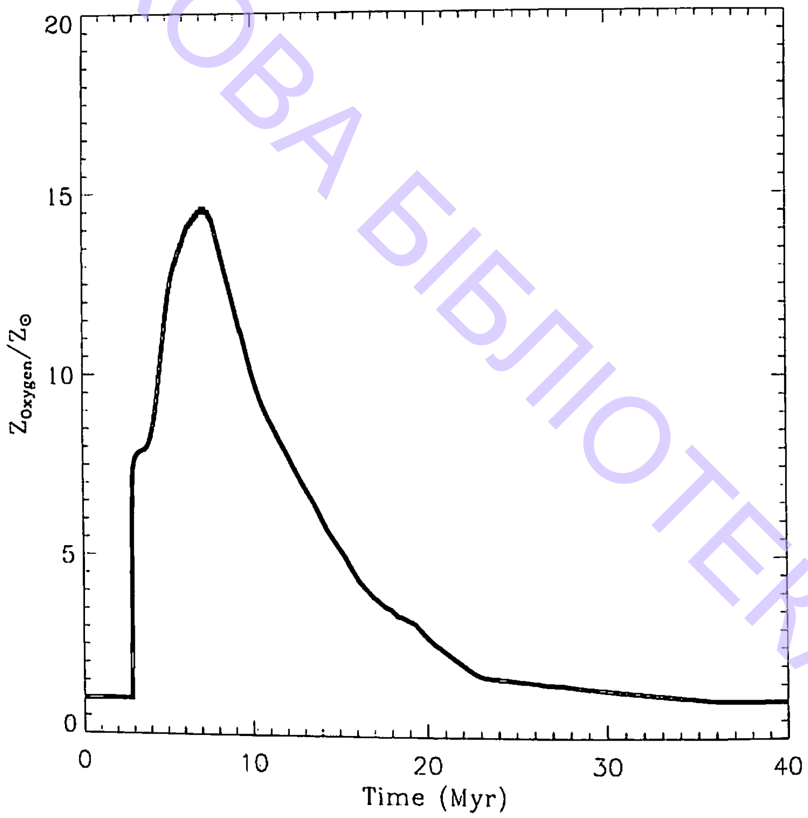


Figure 3: The metallicity of the coeval cluster ejecta (if oxygen is used as a tracer) as function of the evolutionary time, t. The IMF is assumed to be Salpeter with $1\,M_\odot$ and $100\,M_\odot$ mass cutoffs.

$r^{-4/3}$ profiles. We also know that even in the strong radiative regime the central wind temperature remains close to the adiabatic value and that the contribution from the outside cluster gas to the total X-ray luminosity is rather small due to the rapid decline in density. All of these facts allow us to adopt, for a crude estimate, the constant temperature, $T_w = T(R_{SC})$, and the homogeneous, $\rho_w = \alpha_\rho \rho(R_{SC})$, density distribution inside the star cluster, where the fiducial coefficient, $\alpha_\rho \approx 2.0$, takes into consideration the deviation of density from the homogeneous distribution. Under these simplifications we can obtain (see [23]) from equation (3)

$$L_x = 3.8 \times 10^{34} \Phi(T_w, Z_w) \frac{L^2_{SC,38}}{R_{SC,1} V^6_{1000}}, \tag{4}$$

where $\Phi(T, Z)$ is the normalization function, Z_w is the ejected gas metallicity, $L_{SC,38}$ is the star cluster energy deposition rate in units of 10^{38} erg s^{-1}, $R_{SC,1}$ is the star cluster radius in units of 1 pc and V_{1000} is the wind adiabatic terminal speed ($V_{A,\infty} = (2L_{SC}/\dot{M}_{SC})^{1/2}$) in units of 1000 km s^{-1}. The normalization coefficient in equation (4) suggests that $\Phi(T, Z) = 1$ when the X-ray emissivity, Λ_x, is $\Lambda_x = 3 \times 10^{-23}$ erg cm^3 s^{-1}. Figure 4 shows that the predictions of this simple analytic formula are in excellent agreement with our semi-analytical calculations and may be used for crude estimates of the X-ray luminosity from the super star cluster winds.

Equation (4) shows that the X-ray luminosity from the star cluster winds scales as the square of the star cluster power and is highly dependent on the wind terminal speed, i.e. on the ratio of the energy to the mass deposition rate. That is, the terminal wind velocity (or star cluster mass deposition rate) becomes the key parameter we have to know to calculate the star cluster wind X-ray properties.

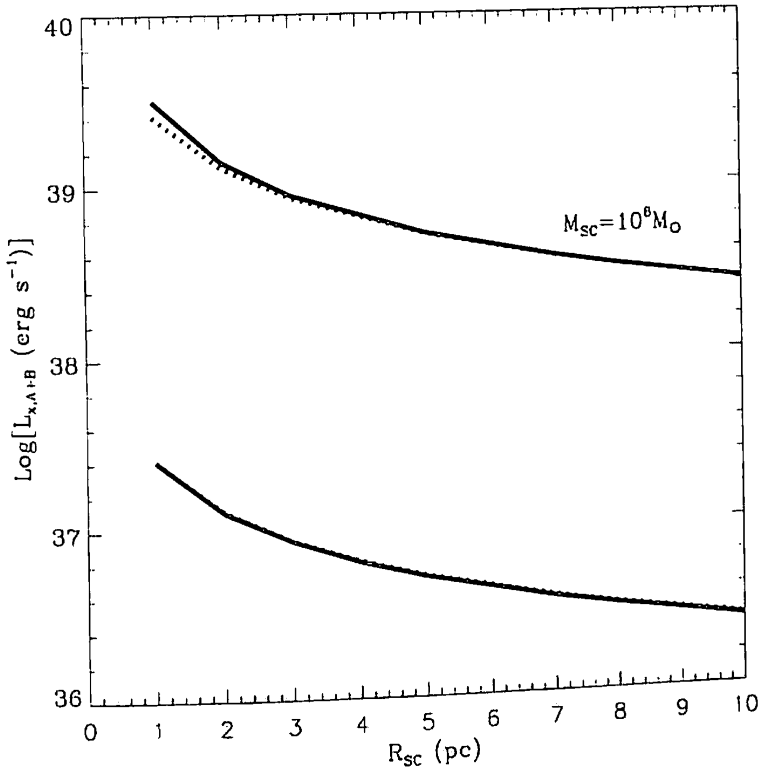


Figure 4: X-ray luminosity of the star cluster winds. Solid lines present results from numerical calculations which are compared to the analytical formula (4) (dotted lines) for star clusters with masses $10^6\,M_\odot$ and $10^5\,M_\odot$. The adiabatic wind terminal speed is $V_{A,\infty} = 1000$ km s^{-1} in both cases and the ejected gas metallicity is solar.

6 Concluding Remarks and Discussion

Super star clusters, massive and compact units of star formation, represent a regular, if not a dominant, mode of star formation in many nearby starburst and merging galaxies. Despite impressive progress over the last decade in observational and theoretical studies of these objects, many fundamental questions related to the formation and evolution of these objects and their impact on the host galaxy remain open.

This paper briefly reviews the basic observational characteristics of super star clusters defined so far at visible, IR and X-ray wavelengths and focuses on the discussion of the fast gaseous outflows generated by supernova explosions and random collisions of individual stellar winds inside the volume occupied by a super star cluster.

It is thoroughly discussed how properties of the super star cluster winds are dependent on the star cluster parameters. It is demonstrated that for very compact and massive star clusters radiative cooling becomes a dominant ingredient of the outflow physics.

The basic results from a new radiative super star cluster wind model (see [22, 28]) and clear limitations of the classical adiabatic [9] star cluster wind model in the star cluster energy deposition rate – star cluster radius parameter space are formulated. It is demonstrated that for low mass and less concentrated star clusters the impact of cooling is negligible. However radiative cooling modifies drastically the internal structure of the super star cluster winds if energy deposition rates approach the threshold limit. In this case cooling brings the inner boundaries of the hot X-ray zone and warm photo-ionized envelope closer to the star cluster edge. This leads to the specific predictions for the observational appearance of the star cluster winds in the X-ray and emission line regimes. In particular, to a high concentration of the X-ray emission and a flat-top low luminosity emission line component (see [22, 19]).

One of the most intriguing results from the radiative wind model is the existence of the threshold energy input rate. Above this limit, which is likely to be surpassed by the most massive and compact SSCs, the stationary wind solution vanishes. It is not clear at the moment how super-critical outflows evolve. Tenorio-Tagle et al. [28] suggested that the violation of the stationary wind conditions by radiative cooling leads to the mass accumulation inside the volume occupied by the super star cluster. This transient phase ends up as a slowly expanding, high density, photo-ionized super-nebula instead of the fast super-galactic wind which is usually advocated to be the major result of violent star formation.

The threshold mechanical energy deposition rate and the expected appearance of the SSC winds are highly dependent on the $L_{SC}/\dot{M}_{SC}$ ratio (or on the adiabatic terminal speed of the wind $V_{A\infty} \sim (L_{SC}/\dot{M}_{SC})^{1/2}$) and on the metallicity of the ejected material. Indeed, radiative losses of energy and the X-ray luminosity of the outflow are in direct proportion to the square of the wind density ρ_w^2. In turn, $\rho_w \sim \dot{M}_{SC}/R_{SC}^2 V_\infty \sim L_{SC}/V_\infty^3$. This implies that the energy loss rate $Q \sim n_w^2 \Lambda(Z,T) R_{SC}^3$ and the wind X-ray luminosity, $L_x \sim n_w^2 \Lambda_x(Z,T) R_{SC}^3$, are in inverse proportion to V_∞^6. That is, V_∞ becomes a key parameter we have to know to model observational properties of the star cluster winds. In this respect it is worth noting that the standard model [9] is not consistent with the X-ray emission from several nearby SSCs and with the X-ray properties of the super-galactic wind from the M82 galaxy (see [24, 25]). It was suggested that an additional mass loading from interstellar clouds being destroyed by the fast outflows may improve the model.

Another important issue is the impact of the ejected gas high metallicity. Despite the fact that the enhanced metallicity may significantly contribute to the wind luminosity and modify the wind properties, it is still ignored in most numerical calculations.

Despite impressive progress in studying super star clusters and their gaseous outflows in the past ten years, there is still much that remains uncertain. For example, what is going on inside a star cluster volume at the local scale (at places where the individual stellar winds and supernova ejecta collide)? What is the fraction of the mechanical energy supplied by supernovae and stellar winds which is converted into the ejected gas thermal energy? How do outflows evolve in the super-critical regime when the stationary solution is inhibited by strong radiative cooling? How does the interaction with the proto-cluster material modify the appearance of the super star cluster winds in the X-ray regime? How large is the observed super star cluster wind terminal speed? I believe all of these issues remain stimulating for new observational programs to be developed and will be debated during the forthcoming five years that separate us from the next Gamow meeting.

Acknowledgments

I am pleased to thank my collaborators, Guillermo Tenorio-Tagle, Casiana Muñoz-Tuñón and Ary Rodríguez Gonzalez for a number of stimulating discussions which led us step by step to a better understanding of the driving mechanisms, inherent physics and observational properties of super star clusters and their gaseous outflows. This study has been supported by CONACYT – México research grant 47534-F.

References

[1] Y.-H. Chu, M.-M. Mac Low, *Astrophys. J.*, **365**, 510 (1990).

[2] J. Melnick, *Astron. Astrophys.*, **153**, 235 (1985).

[3] R.W. O'Connell, J.J. Mangano, *Astrophys. J.*, **221**, 62 (1978).

[4] J.S. Holtzman et al., *Astron. J.*, **103**, 691 (1992).

[5] L.C. Ho, *Rev.MexAA, Conf. Ser.*, **6**, 5 (1997).

[6] L.C. Ho, A.V. Filippenko, *Astrophys. J.*, **466**, L83 (1996a).

[7] L.C. Ho, A.V. Filippenko, *Astrophys. J.*, **472**, 600 (1996b).

[8] J. Cantó, A.C. Raga, L.F. Rodriguez, *Astrophys. J.*, **536**, 896 (2000).

[9] R.A. Chevalier, A.W. Clegg, *Nature*, **317**, 44 (1985).

[10] R.A. Chevalier, R. *Astrophys. J.*, **397**, L39 (1992).

[11] V. Gorjian, J.L. Turner, S.C. Beck, *Astrophys. J.*, **554**, L29 (2001).

[12] P. Kaaret et al., *MNRAS*, **348**, L28 (2004).

[13] C.C. Lang, D.F. Figer, W.M. Goss, M. Morris, *Astron. J.*, **118**, 2327 (1999).

[14] C.C. Lang, W.M. Goss, L.F. Rodriguez, *Astrophys. J.*, **551**, L143 (2001).

[15] C. Law, F. Yusef-Zadeh, *Astrophys. J.*, **611**, 858 (2004).

[16] V.P. Melo, C. Muñoz-Tuñón, J. Maíz-Apellániz, G. Tenorio-Tagle, *Astrophys. J.*, **619**, 270 (2005).

[17] M.S. Oye, *MNRAS*, **339**, 849 (2003).

[18] A.C. Raga, P.F. Velázquez, J. Cantó, E. Masciadri, L.F. Rodriguez, *Astrophys. J.*, **559**, L33 (2001).

[19] A. Rodríguez-González, S. Silich, G. Tenorio-Tagle, in R. Mujica, R. Maiolino (eds) *Multiwavelenght AGN Surveys, Proceedings of the Guillermo Haro Conf.*, World Scientific, Singapore, p.259 (2004).

[20] S. Silich, G. Tenorio-Tagle, C. Muñoz-Tuñón, L.M. Cairos, *Astron. J.*, **123**, 2438 (2002).

[21] S. Silich, G. Tenorio-Tagle, C. Muñoz-Tuñón, *Astrophys. J.*, **590**, 791 (2003).

[22] S. Silich, G. Tenorio-Tagle, A. Rodríguez Gonzalez, C. Muñoz-Tuñón, *Astrophys. J.*, **610**, 226 (2004).

[23] S. Silich, G. Tenorio-Tagle, G.A. Añorve-Zeferino, *Astrophys. J.*, **635**, 1116 (2005).

[24] I.R. Stevens, J.M. Hartwell, *MNRAS*, **339**, 280 (2003).

[25] D. Strickland, I.R. Stevens, *MNRAS*, **314**, 511 (2000).

[26] L.K. Summers et al., *MNRAS*, **351**, 1 (2004).

[27] G. Tenorio-Tagle, *Astron. J.*, **111**, 1641 (1996).

[28] G. Tenorio-Tagle, S. Silich, A. Rodríguez Gonzalez, C. Muñoz-Tuñón, *Astrophys. J.*, **620**, 217 (2005).

[29] J.L. Turner, S.C. Beck, P.T.P. Ho, *Astrophys. J.*, **532**, L109 (2000).

[30] J.L. Turner, S.C. Beck, L.P. Crosthwaite, J.E. Larkin, I.S. McLean, D.S. Meier, *Nature*, **423**, 621 (2003).

[31] L. Vanzi, *Astron. Astrophys.*, **408**, 523 (2003).

[32] C. J. Walcher et al., *Astrophys. J.*, **618**, 237 (2005).

[33] F. Yusef-Zadeh et al., *Astrophys. J.*, **570**, 665 (2002).

19

The Stellar Population of Active Galatic Nuclei

R. Terlevich[1], E. Terlevich[1], J.P. Torres-Papaqui[1], R. Cid Fernandes[2], J. Melnick[3], D. Kunth[4], Q. Gu[5] and A. Bressan[6]

[1] *INAOE, Tonantzintla, Puebla, Mexico*
[2] *Departamento de Física, UFSC, Florianópolis, Brazil*
[3] *ESO, Alonso de Cordova 3107, Santiago, Chile*
[4] *IAP, Boulevard Arago 98bis, Paris, France*
[5] *Department of Astronomy, Nanjing University, Nanjing, China*
[6] *Astronomical Observatory, Padova, Italy*

Abstract

The understanding of how galaxies and inparticular their nuclei form and evolve has changed in a rather unexpected way in recent years. Massive black holes appear to populate the nucleus of virtually every massive enough galaxy, while the nuclei of active galaxies are now known to be the hosts of massive starforming regions. Black holes and starburst clusters coexist and are ubiquitous in the nuclear regions of galaxies. Observational evidence is abundant.

We will review it and describe our efforts to quantify the stellar fabric of the nuclei of active galaxies, and the description of their star formation history using a large sample of near-UV/optical spectra and state-of-the-art stellar population synthesis models and machine learning techniques.

Keywords: stellar populations, starbursts, active galactic nuclei, Seyfert galaxies

1 Starbursts and Active Galactic Nuclei

In recent years we have witnessed a dramatic increase in the amount of evidence that Massive Black Holes populate the central regions of most galaxies and also that *nuclear* star formation (SB) may play an important role in Active Galactic Nuclei (AGN).

If an SB-AGN connection exists, and if the nuclear SB has different characteristics for different AGN types, this will be an important realization both for galaxy (and AGN) formation and evolution models.

We will summarize the existing evidence for massive star formation in AGN and report our own efforts to characterise it using population synthesis methods.

1.1 Spectral evolution of a young cluster

A serious observational difficulty to detect the characteristic optical stellar absorptions from a young stellar population embedded in ionized gas, becomes immediately obvious when we look at Figures 1 and 2.

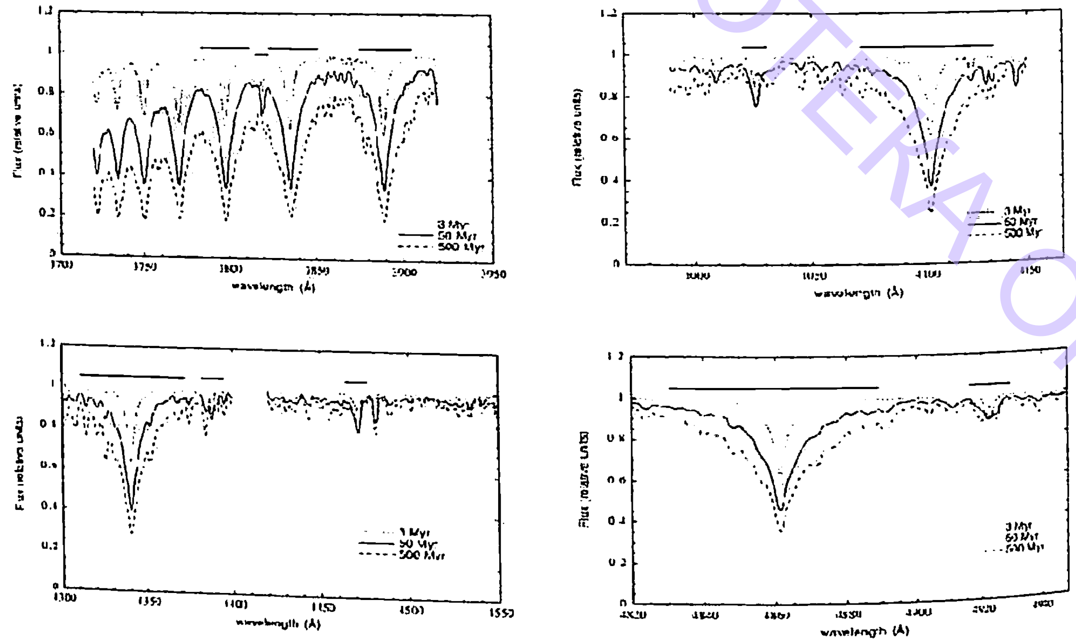


Figure 1: Predicted synthetic spectrum showing the Balmer and HeI lines from 3700 to 5000Å for an instantaneous burst. The cluster masses follow a Salpeter distribution from 1 to 80 $M_\odot$ for three different ages: 3, 50 and 500 Myr. From [4].

Hydrogen Balmer absorptions in young star clusters are very narrow when the stars are young and broaden substantially with age. Figure 1 (from [4]) shows the evolution of the Balmer and HeI optical absorption lines for cluster synthetic spectra. Their models follow an instantaneously formed massive star cluster as it evolves from 3 to 500 Myr, when the stars in the cluster have a Salpeter mass distribution with lower mass 1 and upper mass 80 $M_\odot$.

The integrated optical spectrum of a young massive starforming region (HII Region) is shown in Figure 2. It is clear how such spectrum is dominated by a wealth of unresolved emission lines, mostly permitted lines of HI (the Balmer series) and HeI and forbidden lines of common elements [OII],[OIII],[NII],[SII],[SIII], etc. To better distinguish the faint emission lines, we show the magnified blue region of the spectrum. It is clear that the strong emission Balmer and HeI lines from the ionized gas will fill the stellar absorptions difficulting the use of "classical" methods for the characterisation of the stellar population. The blue continuum will also tend to dilute other absorption features. As the star cluster ages though, the emission lines become weaker and the stellar absorptions begin to emerge.

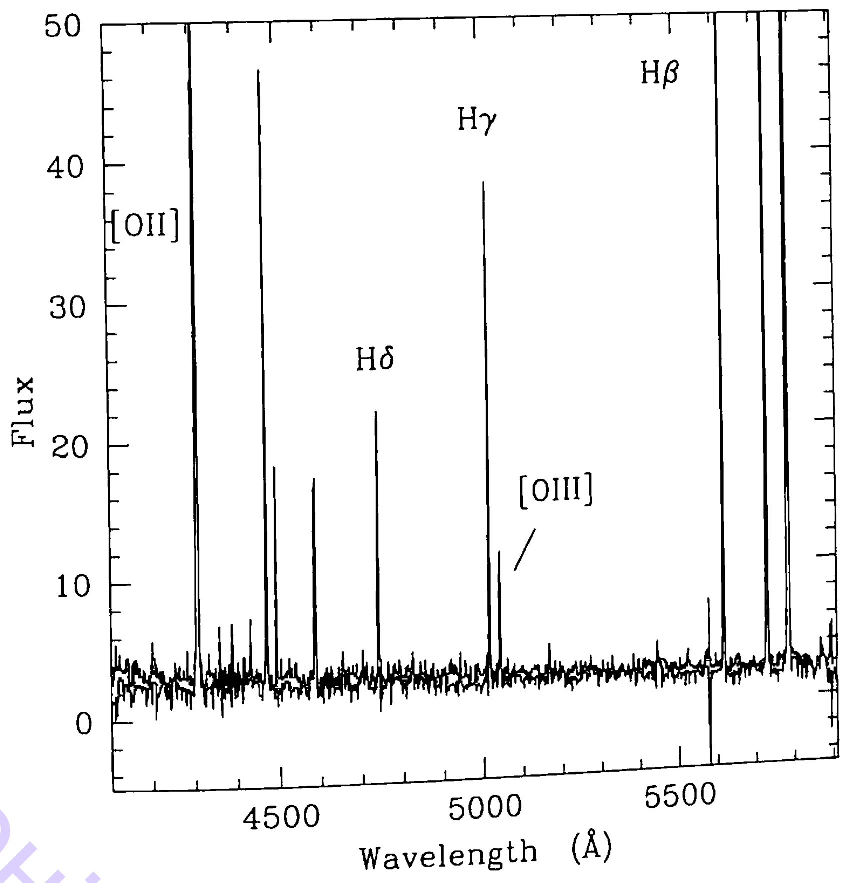


Figure 2: The UV-blue part of the spectrum of a very young SB galaxy where the Balmer series and HeI lines in emission are clearly seen, as well as the strong forbidden lines of [OII] and [OIII] (spectrum from SDSS DR3). Even in this magnified plot no absorption lines are detected.

Fifteen years ago, the detection of strong CaII triplet 8494,8542,8662 Å absorptions in a large sample of Seyfert type 2 and some type 1 nuclei provided the first *direct* evidence for a population of red supergiant stars in their nuclear regions[14]. These results are reproduced in Figure 3.

Of course, the resolution of those observations was not good enough to discriminate the nucleus from the bulge population. We could only say then that the red spectrum was dominated by the light coming from the red supergiants.

Today, with the advent of UV sensitive CCDs, precision studies have become possible in the near UV where the emission is less strong and the higher order Balmer absorptions from young massive stars are prominent.

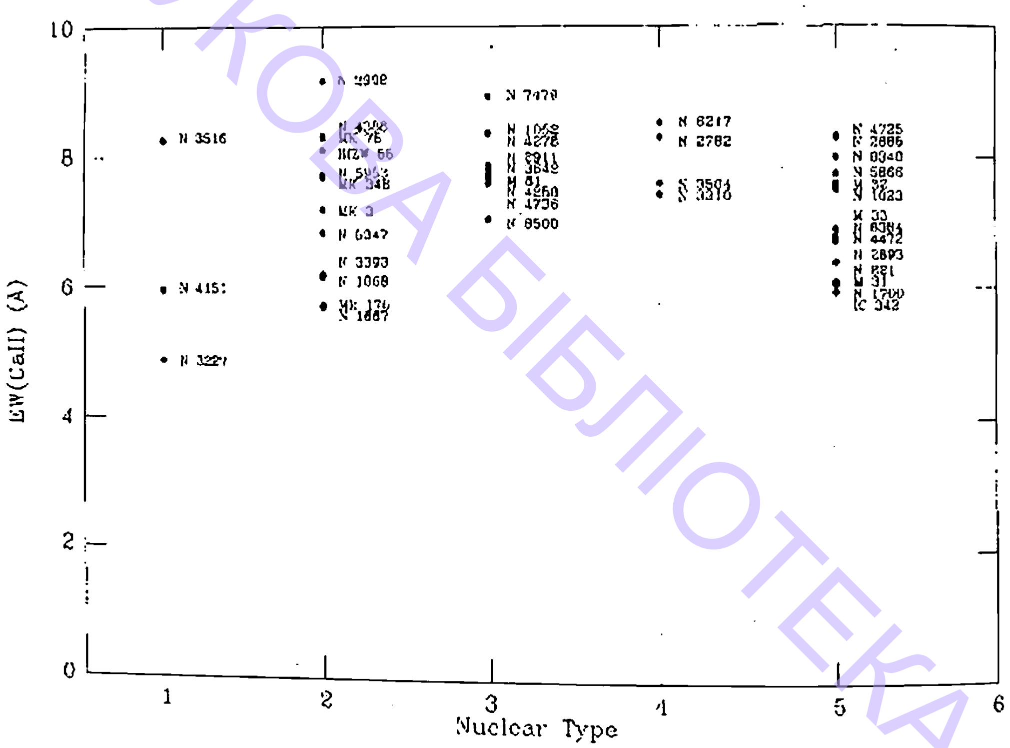


Figure 3: Near-IR CaII triplet strength as a function of nuclear type; 1: Sey 1, 2: Sey 2, 3: LINER, 4: Starburst, 5: Normal. From [14].

1.2 Recent results with HST

New results have been coming to light in recent yeas, based on HST observations which thanks to their high spatial resolution guarantee that we are observing the light that dominates the nuclear spectrum. The conclusion from the group formed by Heckman, Leitherer, González-Delgado and collaborators, is that all the UV to near-IR continuum light in at least some Seyfert 2 galaxies, all the ones observed with HST so far, is due to a nuclear starburst (SB), which is resolved in the UV.

In the case of the very luminous Seyfert 2 Mk477, seen as a Seyfert 1 in polarized light, the nuclear SB is at least as luminous as the Seyfert 1 component. Other Seyfert 2 galaxies also have comparable SB luminosities to those of their ionizing engines. Thus, according to the unified model, an observer situated along the axis of the torus will detect comparable contributions to the optical continuum coming from the SB and the Seyfert 1 or broad line region (BLR) component (Heckman et al. 1995; 1997).

1.3 More recent results

As already mentioned, the improvement in UV sensitivity of optical detectors has led to the detection of higher order Balmer absorptions in the nuclei of many AGNs [1–3, 5–9, 13]. These are the optical lines least contaminated by emission. Near UV spectra of type 2 Seyferts showing the range of strengths encountered in the upper Balmer series and in CaII H and K absorption lines can be seen in Figure 4 (from[7, 8]).

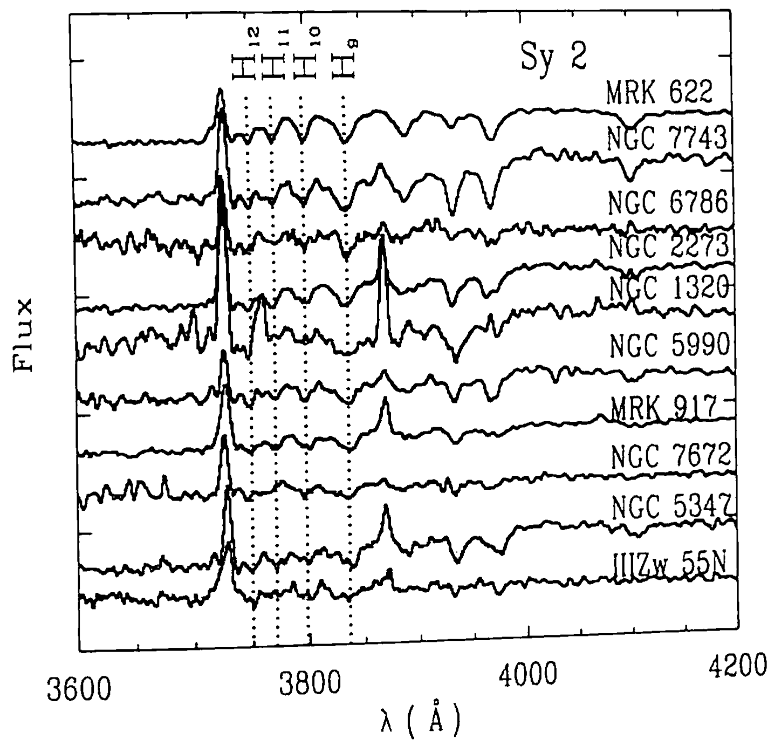


Figure 4: Optical UV spectra of Seyfert 2 galaxies showing with various strengths the upper Balmer series in absorption.These are ESO 1.5m data from Joguet's PhD Thesis (1990).

Joguet's Thesis data comprised ground based, near-UV spectroscopy of the nucleus of about 100 Southern Seyfert 2 galaxies. He found that

- about 50% have high order Balmer lines in absorption;
- about 25 % have strong emission lines and only two weak absorption lines, CaII K and the G band; and
- about 25 % show many metal absorption lines, and a reddish continuum.

Similar results have been obtained by Cid Fernandes and collaborators [2, 3, 12]. Thus the study of a representative sample of Seyfert 2 nuclei indicates that about half show the direct spectroscopic signature of hot massive young stars at optical/near-UV wavelengths.

2 Extreme Objects

2.1 Nuclear starbursts in type 1 AGN

Recently we have started to detect high order Balmer lines in Seyfert type 1 galaxies. This finding suggests a very compact and luminous SB as the nuclear light in these type 1 Seyferts is dominated by a basically unresolved nuclear component.

We have observed a sample of 50 nearby Seyfert 1 galaxies with the ESO-NTT and the Mexican 2.12 m telescope at Cananea, Sonora (OAGH). Representative spectra showing the higher order Balmer series absorption lines are shown in Figures 5 and 6.

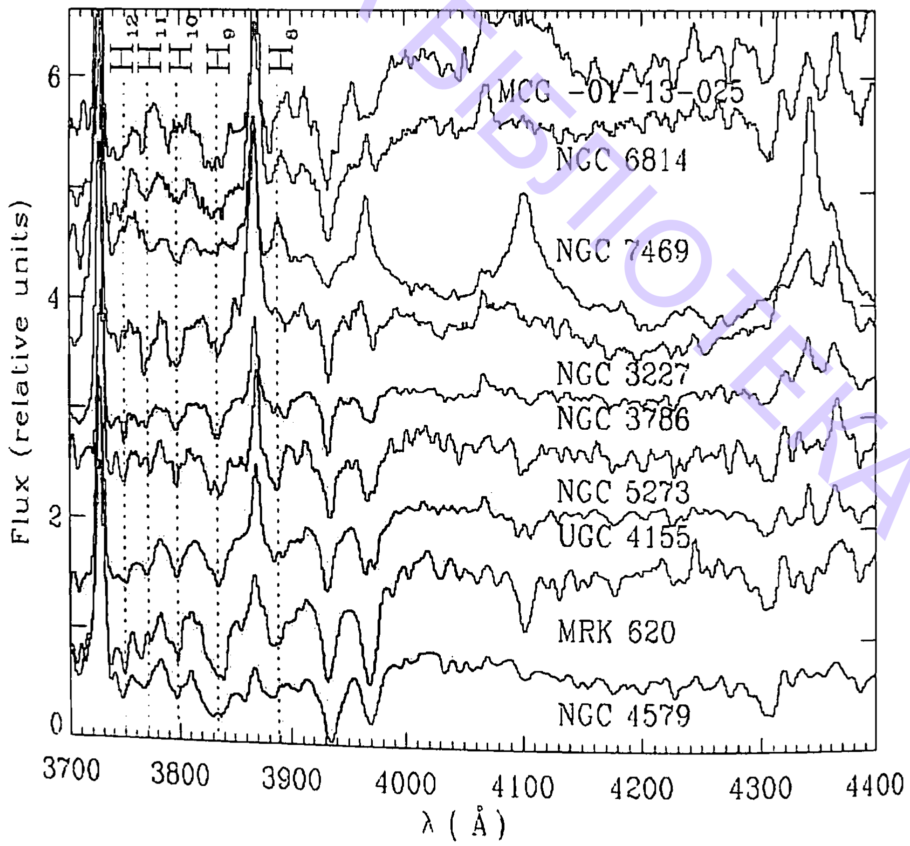


Figure 5: Optical UV spectra of Seyfert 1 galaxies showing with various strengths the upper Balmer series in absorption. These are ESO-NTT and OAGH data from Torres-Papaqui's PhD Thesis (2005).

The fraction of Seyfert 1 nuclei with Balmer absorption seems lower than in type 2 Seyferts, but contamination by the light from the BLR makes the detection of stellar absorptions in type 1 Seyferts much more difficult than in type 2.

2.2 Nuclear starbursts in powerful radio galaxies

The analysis of the nearest powerful FR II radiogalaxies has discovered spatially extended Balmer absorption [9] with strengths characteristic of B-A giant or super-giant stars at the center of at least two of them Hydra-A and 3C285 [1].

The ages derived for the young population are between 7 and 40 Myr. Four more nearby FRIIs (although with poorer data) also show Balmer line strengths which are systematically larger than those found in elliptical galaxies. The Ca II triplet in absorption was detected in all six radio galaxies.

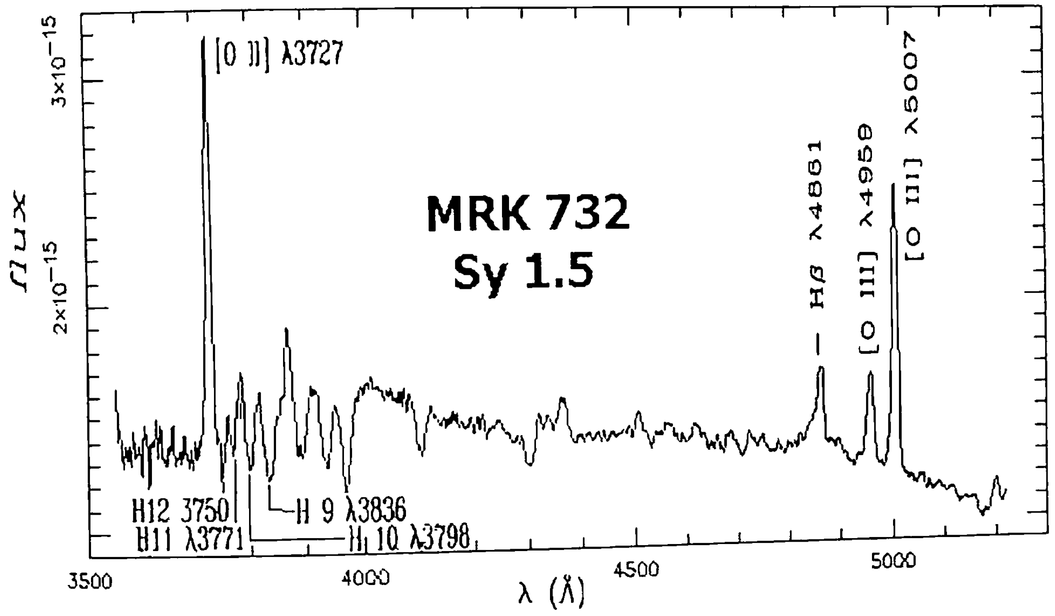


Figure 6: More detailed optical UV spectrum of a Seyfert 1 galaxy showing the upper Balmer series in absorption. OAGH data from Torres-Papaqui's PhD Thesis (2005).

Similar results have been found for other powerful radiogalaxies by Tadhunter and collaborators [13].

Figure 7 shows the optical spectrum of the radio galaxy Hydra A from [1].

3 How to Study the Nuclear SB?

The main conclusion at this stage is that there is compelling evidence that at least in the near Universe a large fraction of AGN, both radio-quiet and radio-loud have a nuclear young cluster or Starburst with luminosities comparable or even larger than that of the Broad Line Region.

What are the properties of these nuclear starforming regions, and, can we estimate the age of these young stellar clusters?

As well as the problem discussed above of the filling of stellar absorptions with emission lines from the ionized gas, stellar absorptions in Seyfert galaxies suffer dilution by the nuclear non-stellar component. This dilution makes the use of traditional methods based on absorption line strength analysis extremely difficult.

We have developed an alternative method that uses the profile of those absorption lines whose main broadening mechanism is related to the surface gravity of the stars (Terlevich et al. in preparation).

The hydrogen Balmer series and the CaII H and K lines are excellent examples of this effect, both sets of lines becoming increasingly broad as the stellar population ages. The broadening of the Balmer series with age can be easily seen in Figure 1.

Thus, in the near UV spectra of Seyferts – available to optical detectors thanks

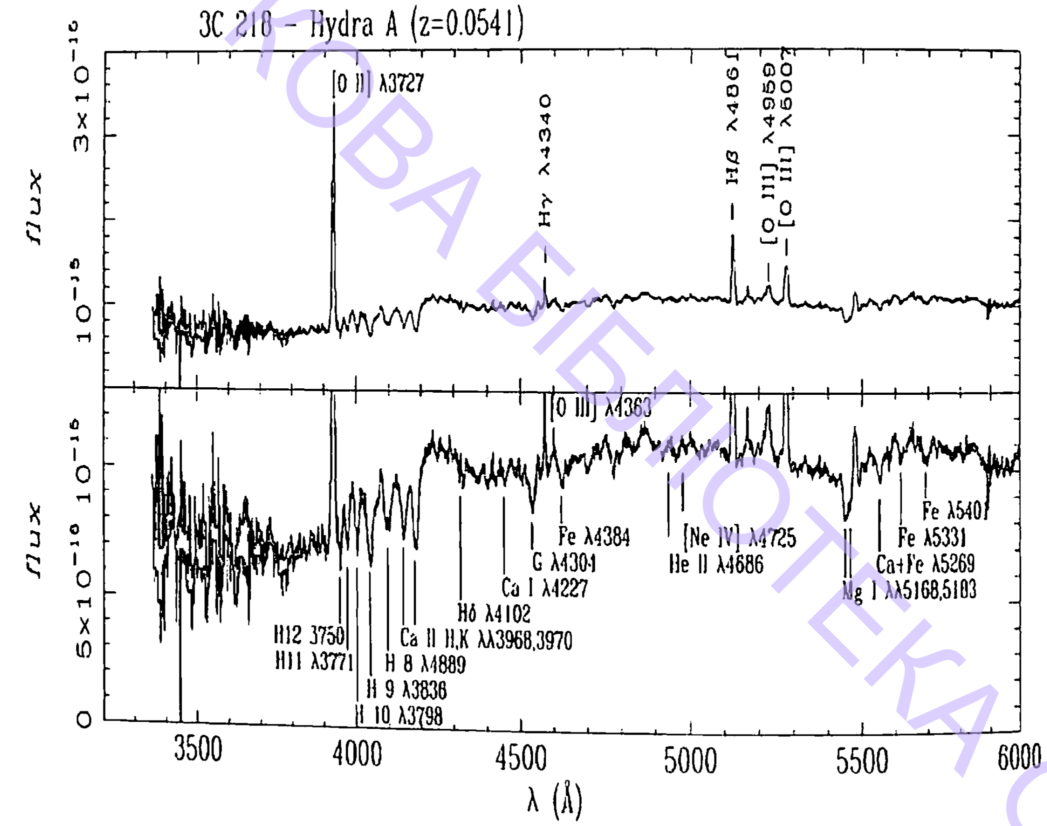


Figure 7: Optical spectrum of the powerful radio galaxy Hydra A showing the upper Balmer series in absorption. (Data from[1]).

to their recently improved UV sensitivity – the main characteristics of recent star formation are strong emission lines plus the higher order Balmer absorptions while strong CaII absorption is typical of the older bulge population. These are the lines least contaminated by emission.

3.1 Properties of nuclear star-forming regions

In the following sections, we will detail the methods used to obtain properties of these nuclear starforming regions, including the estimated age of these young stellar clusters.

As already mentioned, stellar absorptions in Seyfert galaxies suffer dilution by the nuclear non-stellar component. This dilution makes the use of traditional methods based on absorption line strength analysis extremely difficult.

We have developed an alternative method that uses the profile of those absorption lines whose main broadening mechanism is related to the surface gravity of the stars. The hydrogen Balmer and the CaII K lines are excellent examples of this effect, both sets of lines becoming increasingly broad as the stellar population ages. Recent star formation is characterized by strong emission lines plus Balmer absorptions while strong Ca II absorption is typical of the older bulge population.

3.2 High resolution cluster synthetic spectra

A UV-blue (850–4700 Å) high resolution library of theoretical stellar spectral energy distribution has been published by Rodríguez-Merino et al. 2005. The grid is called UVBLUE, and is based on LTE Kurucz models. It covers seven metallicity ranges, T_{eff} from 3,000 to 50,000 K and $\log(g)$ from 0.0 to 5.0 in steps $\Delta\log(g) = 0.5$. It spans the H-R diagram in a comprehensive way at high resolution, making it a most useful tool for the analysis of integrated stellar populations in galaxies.

We compared the FWHM and EW of the H10 and CaII K lines measured from our observed Seyfert 1 galaxies with the values obtained using the high resolution (0.2 Å/px) models by Rodríguez-Merino et al. 2005 mentioned above. The resolution of the models had to be degraded to reproduce the velocity dispersion of the observed galactic nuclei.

We used these age estimators applied to the nuclear stellar components, to determine the age of the starforming region and of the bulge component in our Seyfert 1 spectra (Torres-Papaqui PhD Thesis, INAOE 2005).

Preliminary results indicate that the nuclear Starbursts in AGN are very young and that the bulge component is also relatively young.

The advantage of this method is that it is relatively insensitive to extinction and to dilution by the non-stellar component.

A caveat, though, is that for this analysis we can include only those AGN with detected absorption lines. About 25% of type 2 AGN are dominated by strong emission lines. There are indications that in this subset an even younger Starburst is present but the evidence is not so compelling due to the emission line contamination of the stellar absorptions as in any young Starburst.

3.3 Simple stellar population synthesis

A different approach to characterize the nuclear stellar population consists of using simple stellar population (SSP) synthesis to fit the continuum spectrum of the emission line galaxies (e.g. [3]).

The star formation history of a galaxy is then represented by several SSP modules in a combination of strengths, metallicities and ages, left to evolve. They are also affected by extinction. The SSP represent instantaneous bursts of star formation.

We applied this method [3] to Benoit Joguet's Thesis data of 79 galaxies (65 Seyfert 2) observed at the ESO 1.5m telescope in La Silla with a longslit spectrograph. The CCD detector used had a spectacular UV response (Joguet et al. 2001).

The extracted 1-D spectra represent about 200 central parsecs in each galaxy. The base for the SSP synthesis is formed by standard models plus an added power law featureless continuum (FC) in different proportions. Emission lines are masked out before the fitting. An example of such a fitting is shown in Figure 8 for Mrk 883. This galaxy needs an almost 50% contribution to the light by a FC, which could be either a non-stellar continuum or an obscured SB.

The obtained results for the central (median: 174 pc) region of 65 Sey2 can be summarized as follows:

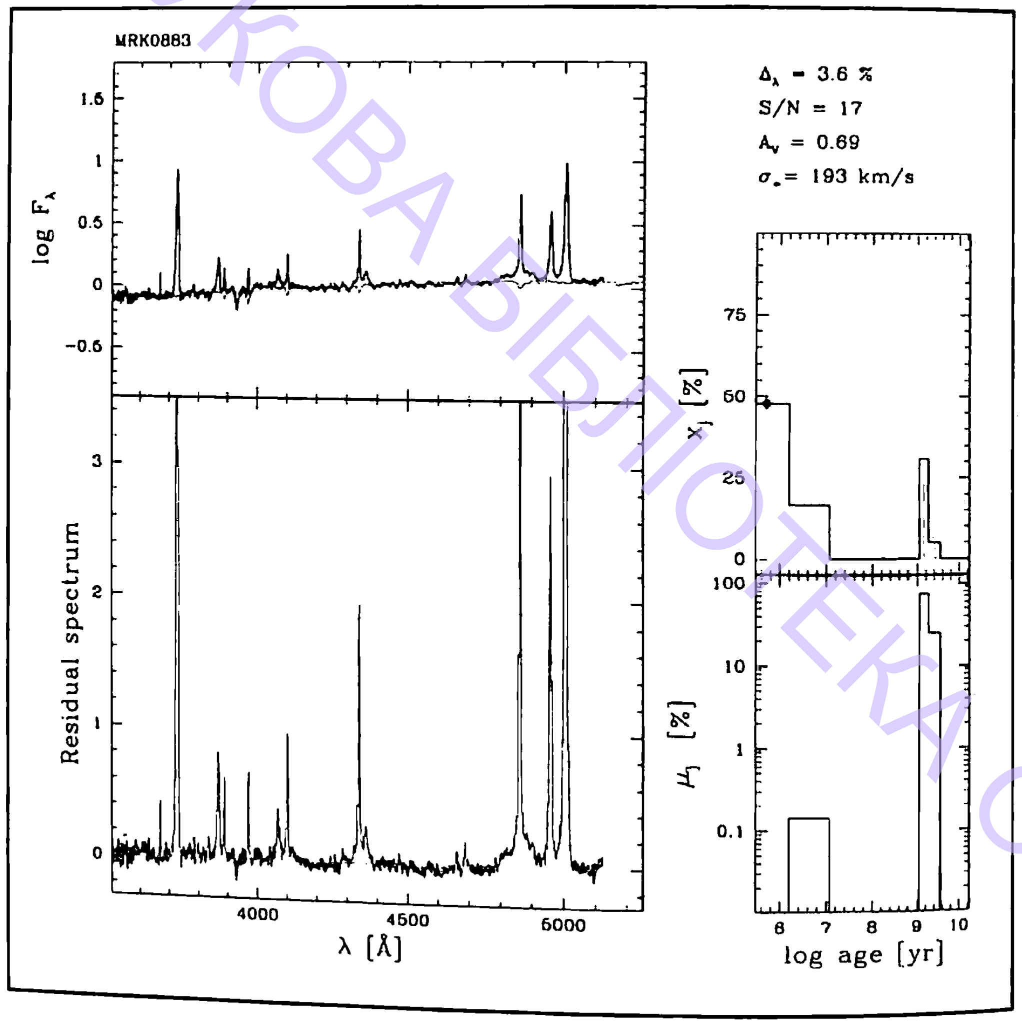


Figure 8: Star formation history from a SSP combination fitting to Mrk 883 (Joguet's data). The emission lines have been masked out. At the top left, the original spectrum and the fit. The subtraction (data - fit) shown at the bottom left represents the emission line residual spectrum. The column at the right gives the contribution in light (top) and in mass (bottom) of the different components. This particular galaxy needs a 50% contribution to the light by a Featureless Continuum (FC). From [3].

The models have been constructed using 10 SSP BC03 models, having solar and 2.5 solar metallicities, plus a power law FC of varied intensity. The galaxies present heterogeneous star formation histories.

Some spectra are dominated by old stars (age>2.5Gyr) contributing up to 80% of the optical light.

Some show a strong component of intermediate age stars (100 Myr < age < 1.4 Gyr).

Young clusters (age <25 Myr) are ubiquitous contributing in some cases to more than 50% of the light at 4020Å and in several cases, up to 20%.

A strong FC component is also present that could be a genuine monster or a dusty young burst.

At least 3 of the 4 components are present in most of the galaxies with significant strength (more than 10%) in any one galaxy.

We conclude that a simple picture of an elliptical galaxy plus a power law (used many times before) does not apply to the bulk of Seyfert 2s.

4 Conclusions and Future Work

There is substantial direct evidence of massive young stellar clusters present in the central regions of nearby AGN, both low luminosity radio quiet and radio loud.

- ∼50% of nearby Seyfert 2s, some 30% of observed powerful FRII radiogalaxies and about 20% of Seyfert 1 show the high Balmer series in absorption, denoting the presence of a young star cluster.
- The analysis of the profile of the higher Balmer and CaII absorption lines (insensitive to both dilution and extinction) provides a new way to estimate the ages of stellar populations, and hence a rough picture of the star formation histories.
- A sum of state-of-the-art evolutionary synthesis models plus a FC can reproduce the stellar population in the central 50-500 pc of a large, well defined, homogeneous sample of Seyfert 2 galaxies.
- Young starbursts, intermediate age and old populations appear in significant and mixed proportions.
- WR features were discovered in the pure emission (absorption subtracted) spectra of 5 out of 65 galaxies (making dating much more precise for the young component). The pure emission spectrum allows the study of the physical properties of the ionized gas.
- 17 galaxies known to have BLR (in direct or polarized light) show also strong FC in our fitting.
- Dusty young starbursts are identified for having a strong FC but no evidence for a BLR.
- One third to a half of nearby Seyfert 2 have experienced significant star formation in the recent past. As a class they live in a much younger stellar environment than LINERs and normal galaxies.
- No correlation was found between host morphology (as per HST optical to near-IR and ground based observations) and nuclear stellar population.
- Most interestingly, central stellar velocity dispersions are obtained from the fitting procedure as a bonus.

We are now applying Cid Fernandes synthesis population method to the sample of Seyfert 1 galaxies from Torres-Papaqui thesis, and also a new fitting method that uses machine learning algorithms (Solorio et al. 2005) that will allow fast and precise analysis of stellar populations for large number of objects.

Acknowledgments

We are grateful to the organizers who managed to make this conference a great success.

References

[1] I. Aretxaga, E. Terlevich, R. Terlevich, G. Cotter & A.I. Díaz, *MNRAS*, 325, 636 (2001).

[2] R. Cid Fernandes, T. Heckman, H.R. Schmitt & T. Storchi-Bergmann, *ApJ*, **558**, 81 (2001).

[3] R. Cid Fernandes, Q. Gu, J. Melnick, E. Terlevich, R. Terlevich, D. Kunth, R. Rodrigues Lacerda & B. Joguet, *MNRAS*, **355**, 273 (2004).

[4] R.M. González-Delgado & C. Leitherer, *ApJS*, **125**, 479 (1999).

[5] T.M. Heckman, R.M. González-Delgado, C. Leitherer, G.R. Meurer, J. Krolik, A.S. Wilson, A. Koratkar & A. Kinney, *ApJ*, **482**, 114 (1997).

[6] T. Heckman, J. Krolik, G. Meurer, D. Calzetti, A. Kinney, A. Koratkar, C. Leitherer, C. Robert & A. Wilson, *ApJ*, **452**, 549 (1995).

[7] B. Joguet, PhD Thesis, Institute d'Astrophysique de Paris (2000).

[8] B. Joguet, D. Kunth, J. Melnick, R. Terlevich & E. Terlevich, *Astron. Astrophys.*, **380**, 19 (2001).

[9] J. Melnick, Gopal-Krishna, & R. Terlevich, *Astron. Astrophys.*, **318**, 337 (1997).

[10] L.H. Rodríguez-Merino, M. Chávez, E. Bertone & A. Buzzoni, *ApJ*, **626**, 411 (2005).

[11] T. Solorio, O. Fuentes, R. Terlevich & E. Terlevich, preprint astro-ph/0507527 (2005).

[12] T. Storchi-Bergmann, R.M. González-Delgado, H.R. Schmitt, R. Cid Fernandes & T. Heckman, *ApJ*, **559**, 147 (2001).

[13] C. Tadhunter, T.G. Robinson, R.M. González-Delgado, K. Wills, & R. Morganti, *MNRAS*, **356**, 480 (2005).

[14] E. Terlevich, A.I. Díaz & R. Terlevich, *MNRAS*, **242**, 271 (1990).

20

Cosmic Rays, Structure and Dynamics of the Heliosphere

L.I. Miroshnichenko

IZMIRAN, Troitsk, Moscow Region, 142190, Russia
e-mail: leonty@izmiran.rssi.ru

Abstract

A number of the most important aspects of the physics of Cosmic Rays (Galactic Cosmic Rays, Anomalous Cosmic Rays, and Solar Cosmic Rays) are briefly reviewed. Especially addressed are the local interstellar spectrum of Galactic Cosmic Rays (GCR), the termination shock and the outer boundary of the modulation volume, the structure of the heliospheric magnetic field and its implication for particle propagation within the heliosphere and the dynamics of the heliospheric boundary over the solar cycle. In this context, the Anomalous Cosmic Rays (ACR) are of special interest. These energetic particles are, most probably, accelerated at the so-called heliospheric shock, which terminates the supersonic solar wind expansion.

Keywords: Sun, heliosphere, cosmic rays, termination shock, particle acceleration

1 Introduction

The three-dimensional region around the Sun, controlled by the solar wind and its embedded magnetic field, is called the heliosphere (Figure 1). This plasma bubble shields the planets from direct irradiation with both energetic interstellar particles (galactic cosmic rays, GCRs) and particles energized in the boundary region of the heliosphere (anomalous cosmic rays, ACRs). Also the flux of interstellar neutral atoms at the orbits of the terrestrial planets (Mercury, Venus, Earth and Mars) is significantly reduced due to filtering in the heliospheric interface and photo- and charge-exchange ionization upon approaching the inner heliosphere.

The interaction of the supersonic solar wind with the local interstellar medium (LISM) leads to a transition of the solar wind from supersonic to subsonic speeds at the heliospheric termination shock. Such a transition may also occur for the interstellar wind at the heliospheric bow shock; the heliopause is the boundary layer between the interstellar and solar wind.

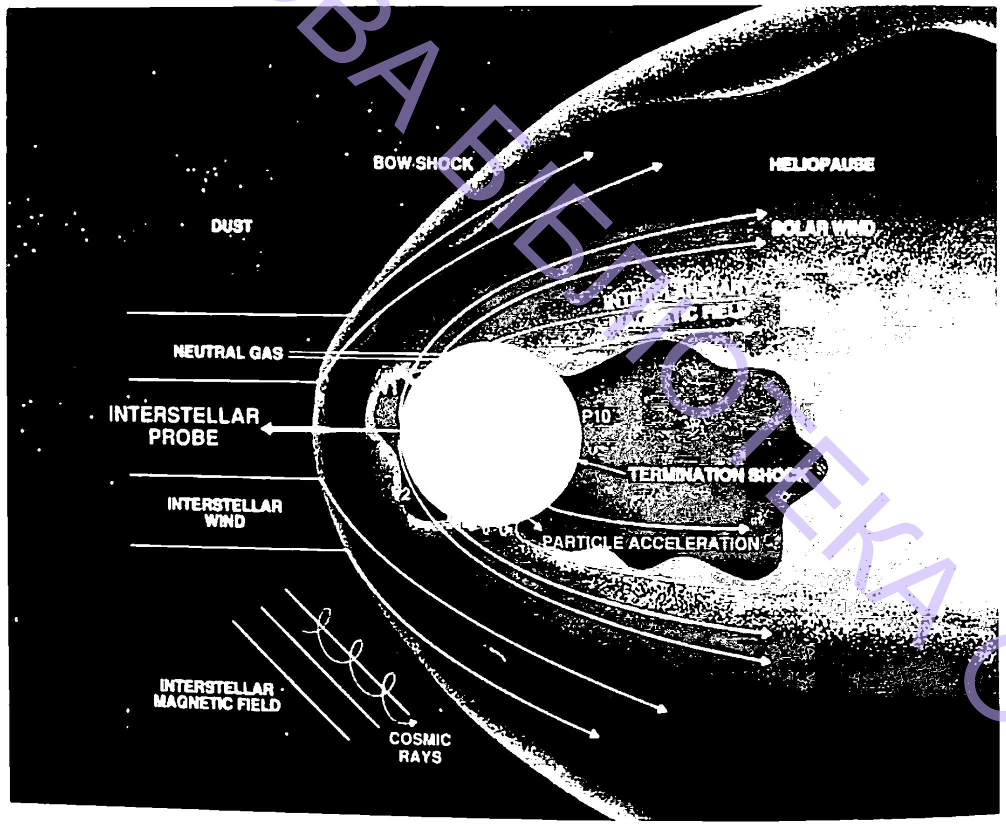


Figure 1: An artist sketch of the heliosphere (http://interstellar.jpl.nasa.gov/), with the current positions of *Pioneer 10* (P10), *Voyager* 1 (V1), and *Voyager 2* (V2).

The exact geometry and dimension of the heliosphere is uncertain, and the first estimates of those parameters came many years ago (early 1960s) from the studies of 11-year modulation of GCRs (e.g. [1]). Two important facts have been established: 1) the heliosphere (or more exactly, the region of cosmic ray modulation) is far from spherical; 2) its characteristic (effective) radius is within 90–110 AU. Several other attempts have been made to estimate the heliospheric scale size lately (e.g. [2]). Five different methods of inferring the distance to the termination shock lead to estimates that cluster in the range of 90±10 AU [3].

Galactic cosmic rays entering our heliosphere encounter an outward-flowing solar wind carrying a turbulent magnetic field that can be approximated by an Archimedean spiral. Due to this interaction the energy spectrum of GCR measured at the Earth's orbit is different from the local interstellar spectrum (LIS). Moreover, in the early 1970s it was found that the helium spectra at solar minimum is much flatter than the hydrogen spectra. These anomalous cosmic rays (ACRs) turned out to be a very interesting phenomenon from the point of view of heliospheric studies (see, e.g., [4]).

In the inner heliosphere, valuable information about the physics and dynamics of the heliospheric plasma can be derived from the studies of so-called Forbush decreases of GCRs, coronal mass ejections (CMEs) and SCRs (e.g., [5]). Some important data are obtained from the studies of SCR propagation in the interplanetary magnetic field (IMF) and possible particle acceleration at interplanetary plasma turbulence and shock waves, especially in the case of so-called "super-events" (e.g. [6]) and "rogue events" (e.g. [7, 8]), in an analogy to rogue ocean waves having unusually large amplitudes. To illustrate existing problems and recent achievements in this field, we give below several of the most spectacular examples of the heliospheric phenomena mentioned above.

2 Galactic Cosmic Rays

The Earth's atmosphere provides a shielding mass, at sea level, of about 1033 g cm^{-2}. Cosmic rays impacting on the atmosphere generate a nuclear cascade as illustrated in Figure 2 [9]. This figure gives a concept of primary cosmic radiation in space and of secondary cosmic radiation propagating through the atmosphere. In the left of this figure, main methods of cosmic ray measurements are mentioned. The center of the figure lists the substantial air mass that provides a shield against radiation in space, and the right part is a conceptual illustration of the nuclear interactions that occur in the atmosphere. The minimum energy a proton must have to initiate a nuclear interaction sequence that may be detectable at sea level is approximately 450 MeV. Higher energy protons (in the GeV energy range) will generate an atmospheric nuclear cascade, and these high-energy cascading particles have enough energy to continue the process to sea level. Hard components (muons) of secondary cosmic rays may be detected deeply under the terrestrial surface.

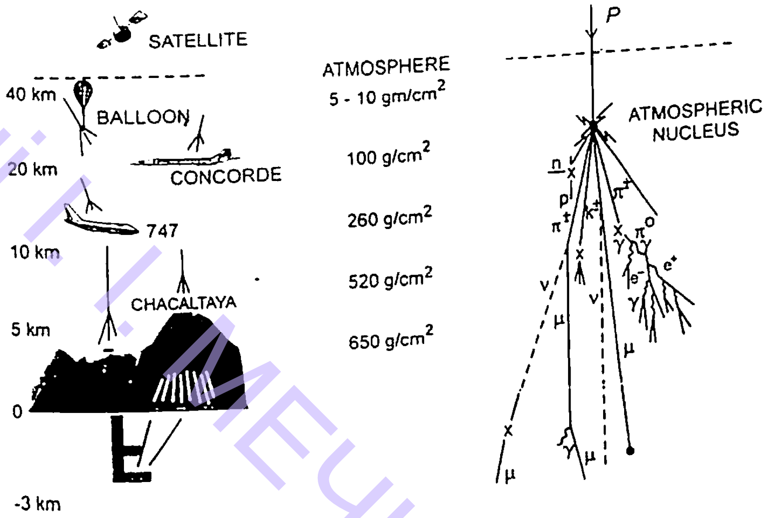


Figure 2: Primary cosmic rays in space, secondary cosmic radiation cascade in the terrestrial atmosphere, and hard muons at different underground depths [2].

There are four different ways to describe the spectrum of the components of the cosmic radiation [10]: (1) by particles per unit rigidity; (2) by particles per energy-per-nucleon; (3) by nucleons per energy-per-nucleon; (4) by particles per energy-per-nucleus. Propagation (and probably also acceleration) through cosmic magnetic fields depends on gyroradius or magnetic rigidity (R), ρ, multiplied by the magnetic field strength, B.

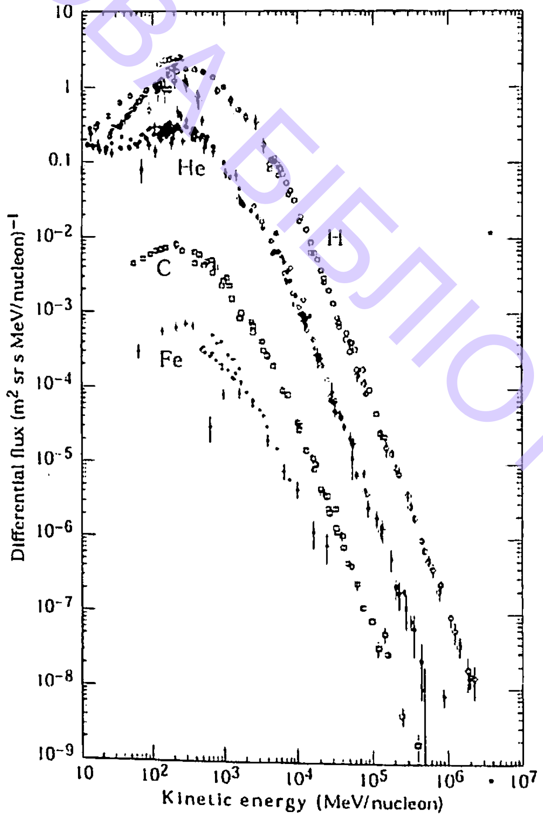


Figure 3: Differential energy spectrum of the major components of the galactic cosmic rays near the Earth [11]: H – hydrogen, He – helium, C – carbon, Fe – iron.

On the other hand, fragmentation of nuclei propagating through the interstellar gas depends on energy per nucleon, since that quantity is approximately conserved when a nucleus breaks up on interaction with the gas. Further, production of secondary cosmic rays in the atmosphere depends on the intensity of nucleons per energy-per-nucleon, approximately independently of whether the incident nucleons are free protons or bound in nuclei. At last, air shower experiments that use the atmosphere as a calorimeter generally measure a quantity that is related to the total energy per particle. The units of differential intensity, D, are [particles cm^{-2} s^{-1} sr^{-1} ϵ^{-1}], where ϵ represents the units of one of the four variables listed above.

Figure 3 shows the major components as a function of energy at a particular epoch of the solar cycle. The turn-up of the helium flux below ~60 MeV/nucleon is due to additional flux of the anomalous ^{4}He component [11] (see also [10]). In Section 4 we describe the behaviour of the anomalous cosmic rays (ACRs) in more detail.

New evidence appeared during the last years suggesting that the Sun has a global magnetic field (GMF) with dipole characteristics. Its magnitude at the Sun's poles does not exceed 1–2 G. This field is compressed in the meridional plane, strongly extended in radial direction and separated on two hemispheres (northern and southern ones) at the magnetic equator. As a physical boundary between them, there is a neutral current sheet similar to that in the tail of the terrestrial magnetosphere. This picture corresponds to the modern concept of the heliomagnetosphere confirmed by numerous observations of magnetic fields and particles in the interplanetary space.

The magnetic fields of the sunspots and of the solar GMF change their polarities with a periodicity of about 11 and 22 years, respectively, as it was found in [12] for 1953–1977. A more comprehensive picture of the GMF reversal is shown in Figure 4, compiled in [5] from data on the GMF polarity, temporal variations of the mean monthly sunspot numbers, W, and intensity of GCRs (count rate at the Climax neutron monitor) since 1951 up to now. The periods of the Sun's GMF reversals are also shown: 1957–1958, 1969–1972, 1980–1981, and 1991–1992. The most recent reversal seemed to have occured in 2000–2001.

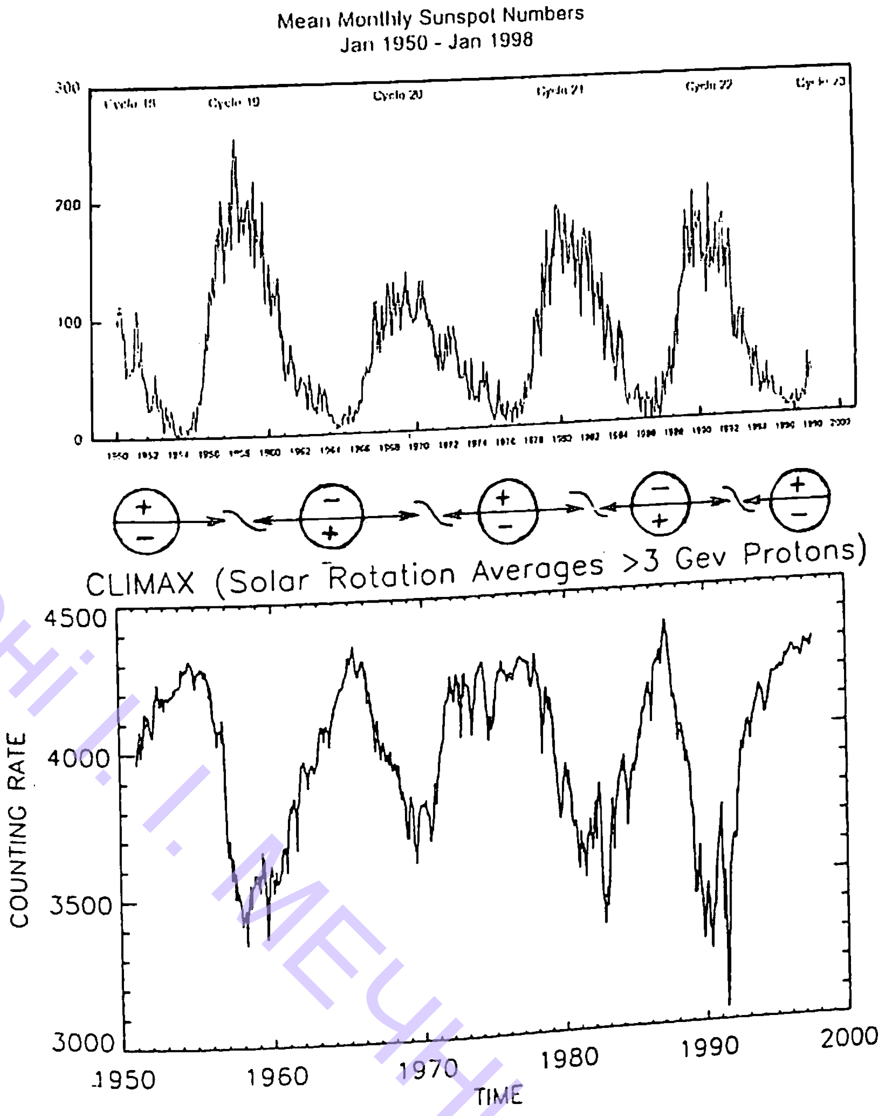


Figure 4: Mean monthly sunspot numbers (*top panel*, SGD, 1998, No.641), solar magnetic polarity (*center panel*) and solar rotation averages of counting rate at the Climax neutron monitor [13] (*bottom panel*) since 1951.

3 Coronal Mass Ejections and Solar Cosmic Rays

As a simple illustration of complexity of the solar activity phenomena under discussion, we describe briefly the first observation of the seismic response [14] to a solar flare by the Michelson Doppler Imager (MDI) on board Solar and Heliospheric Observatory (*SOHO*) on 9 July 1996. This flare was the only significant X-ray flare observed in 1996 (at the minimum of solar activity). Though this was a fairly moderated flare classified as X2.6/1B (10°S, 30°W), it was followed by observations of very spectacular CME (Figure 5) [15] and energetic electrons (44 keV – 10 MeV) recorded at the same spacecraft [16]. The electron spectrum of the event was a single power law between 0.1–2.0 MeV. A relatively sharp break occurs at about 2 MeV, leading to a harder spectrum at higher energies. Such a spectral shape is typical for impulsive flares with associated soft X-ray emission.

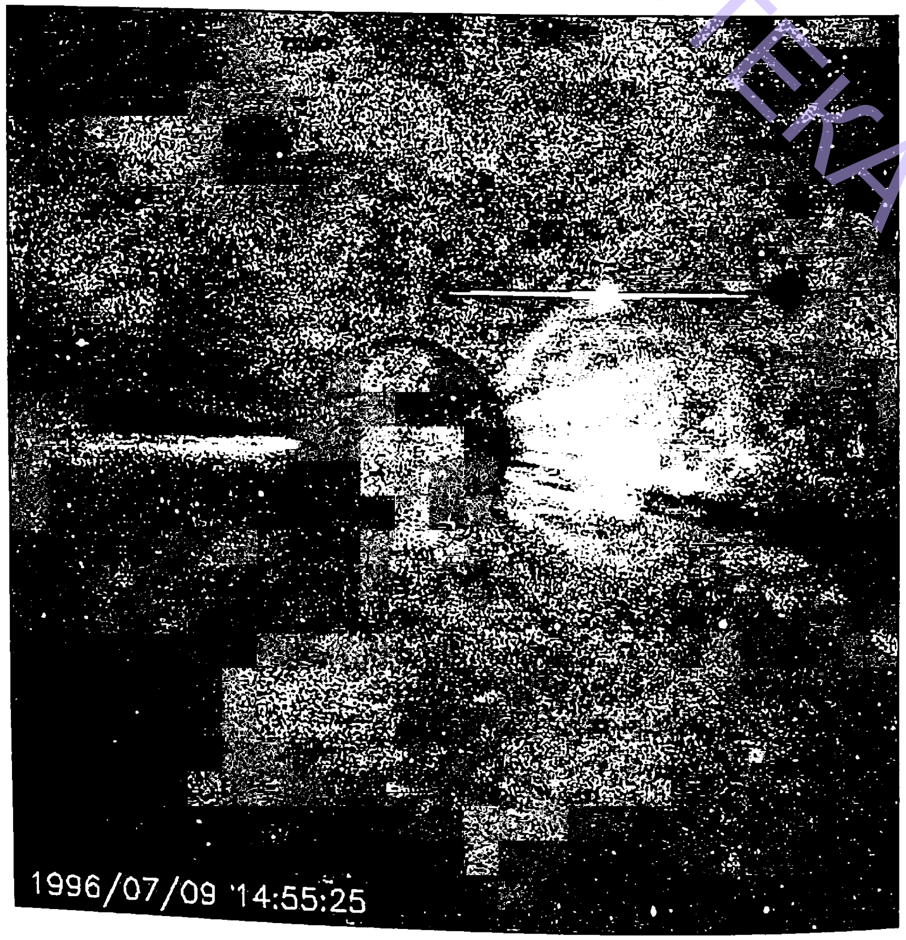


Figure 5: A coronal mass ejection observed by LASCO on board of *SOHO* spacecraft at 15:38:41 UT on 9 July 1996 [15]. The large dark circle in the center is the occulting disk of the coronograph (it extends to 3 solar radii). The imbedded white circle (with grid) indicates the position of the Sun. The background white (and black) dots represent stars and cosmic ray hits.

Additional analysis of radio observations shows [16] that it is quite possible that in fact two different populations of energetic electrons were detected after this flare.

As reported in [14], a circular wave packet propagated after the flare to a distance of 120,000 km from the flare site with an average speed of about 50 km s^{-1}. The amplitude of the flare ripples was approximately 50 m s^{-1}, which is higher than the amplitude predicted by the thick target model of solar flares. This particular event provides an excellent illustration of a very complicated picture of solar energetic phenomena followed by the production of energetic solar particles.

Rogue events are associated with multiple shocks and CMEs. In particular, from observation of two converging shock waves accompanied by an energetic particle event with unusually high and long-lasting intensities in August 1972, it was proposed (see references in [7, 8]) the first order Fermi acceleration between converging interplanetary shocks as a fast and highly efficient acceleration mechanism. Subsequently, time periods with unusually high intensities related to multiple CME shocks as well as other particle events with high intensities between pairs of shocks have been identified [7]. Well-known examples of rogue SEP events at the Earth occurred on 14 July 1959, 4 August 1972, 19 October 1989, and 14 July 2000. Rogue events also have been observed in the inner heliosphere: with *Helios 1* on 4 November 1980 at 0.5 AU and with *Ulysses* in March 1991 at 2.5 AU.

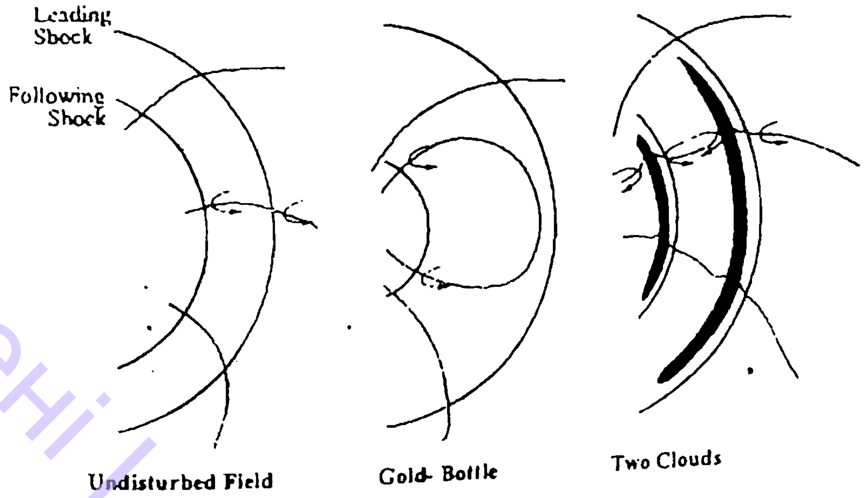


Figure 6: Possible geometries for the acceleration of energetic particles between two converging shocks produced by the Sun [8].

Rogue events cannot result in a simple compression of the medium between two converging shocks; although the distance between converging shocks would decrease with r as the shocks propagate outwards, the cross-section of the flux tube increases with r^2, leading to a net increase in volume between shocks. Figure 6 shows different geometries for a pair of shocks and the IMF. The left hand side shows the simplest case: two shocks and a background Archimedean spiral magnetic field. Particles traveling from the upstream region towards the shock partly are reflected at the shock because the magnetic compression across the shock front creates a magnetic mirror. Thus particles can be swept by the shock. However, reflection of particles approaching the shock from its downstream medium cannot be understood because the particle then experiences a diverging field. Therefore, although the shocks would sweep-up the particles, there would be no particle storage in the volume between them and consequently no first order Fermi acceleration.

To avoid this problem, the mechanism of a Gold bottle (middle section of Figure 6) was proposed: a large loop extends ahead of the shock and once in the upstream medium particles have a chance (depending on their pitch angle) of being reflected back and forth along the field line. As the shock expands, the length of the field line in the upstream medium is reduced and the particles are accelerated by a Fermi I process. Although large loops extending beyond 1 AU have been observed, it is suggested in [8] that this configuration might not be the only one to explain rogue events, in particular, since a typical rogue event has been observed by *Ulysses* at a distance of 2.5 AU [17]. Considering the calculations regarding the capability of a magnetic cloud to separate the particle populations upstream and downstream of a magnetic cloud, [8] suggests the scenario depicted on the right hand side of Figure 6: particles then are reflected repeatedly between the following shock (particles with small pitch angles passing the shock at the cloud) and the magnetic cloud behind the leading shock. The authors proposed a numerical model that allows simulating the effect of pairs of CME/shock on particle populations. They concluded that (a) the magnetic cloud following the leading shock is of outmost importance for the creation of high particle intensities, (b) the shocks need not converge to create an intensity enhancement, and (c) the trailing cloud is required to reduce intensities after the passage of the shock pair.

4 Anomalous Cosmic Rays

The anomalous component of cosmic rays (ACR) in the heliosphere was discovered more than 20 years ago. It applies to H, He, and the heavy ions of C, N, O, Ne, Ar and, possibly, S, Si and Fe with energies up to ~50 MeV/nucleon that are detected in the interplanetary space and Earth's magnetosphere. The chemical composition of this component differs from the solar and galactic abundances of the elements, displaying some excess of the C/O ratio. Figure 7 demonstrates typical energy spectra of energetic oxygen nuclei measured by various *ACE* instruments [18]. Other nuclear species tend to have similar spectra. Solid curves correspond to steady-state components of the interplanetary particle population; light dashes pertain to transient phenomena; heavy dashed curve denotes postulated quiet-time background of supra-thermal solar particles.

Due to the orbit of the Sun around the galactic center the interstellar environment at the boundary of the heliosphere is changing, implying changes in the structure and extension of the heliosphere as well as in the intensity of the cosmic rays and neutral gas fluxes. The neutral gas component of the local interstellar medium (LISM) flows almost unhindered into the heliosphere and, in the case of sufficiently high fluxes, can change the chemistry of planetary atmospheres.

On the other hand, significant fractions of GCR from LISM and the ACR produced in the outer heliosphere reach the environment and surface of the terrestrial planets, moons, asteroids and meteoroids. Besides an influence on planetary magnetospheres (e.g. through current systems) and atmospheres (e.g. through cloud formation), there is a production of radioactive elements due to direct imprints in the surfaces of celestial bodies. These so-called cosmogenic elements serve as clocks providing information about the past of the solar system.

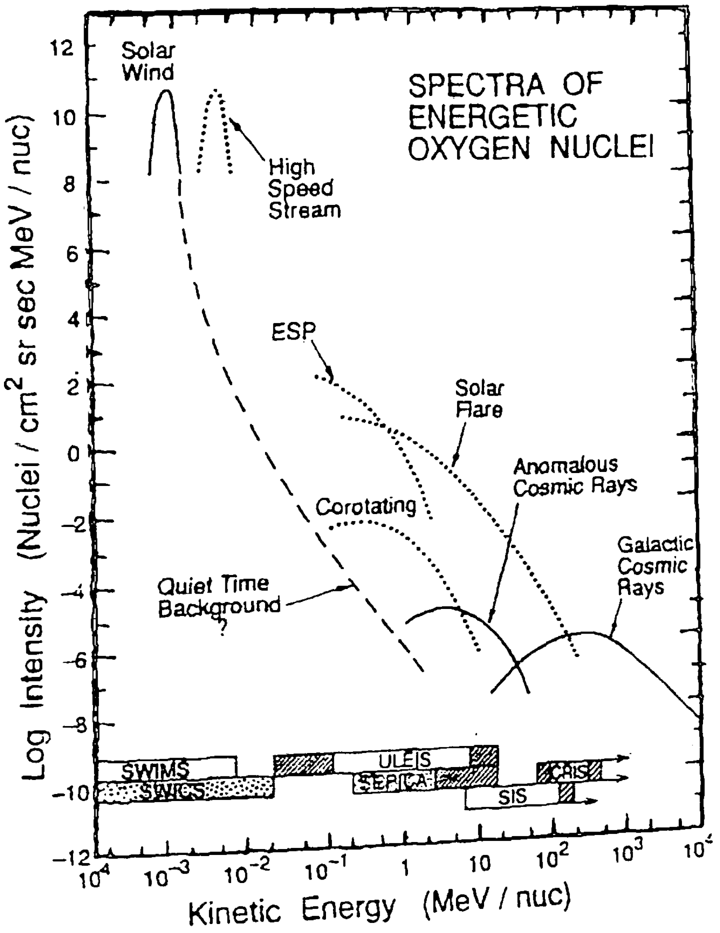


Figure 7: Typical energy spectra of various sources of energetic oxygen nuclei are shown, along with energy coverage for various *ACE* instruments [18]. Solid curves = steady-state components; light dashes = transient phenomena; heavy dashed = postulated quiet-time supra-thermal solar particles.

Thus, the surface of such Earth-like celestial bodies (terrestrial bodies) can be considered as "cosmochronic" archives. The legibility and reliability of these cosmochronic archives depend on the knowledge of the production rates of cosmogenic elements, i.e., on the corpuscular irradiation history of the surfaces and, thus, depend on the intensity of the cosmic rays flux in the past. This intensity is determined by the structure of the heliosphere because both the interstellar spectra of GCR and the source spectra of ACR are modulated by heliospheric transport processes.

There are many processes involved in the production of anomalous cosmic rays [4]. Figure 8 shows only those connected to the physics of ACRs in the outer heliosphere. Dotted lines with open arrows correspond to neutral atoms of the LISM that enter the heliosphere due to the relative motion of the Sun with respect to the LISM. Some neutral species (mainly hydrogen and oxygen) experience a filtering (*1*) in the (inner) heliosheath due to charge exchange with solar wind protons. An element-specific fraction of those neutrals (open circles) entering the heliosphere gets ionized by charge-exchange, photo-ionization or electron impact. The resulting pick-up ions (PUIs) are pre-accelerated while they are convected with the solar wind (*2*) towards the heliospheric shock. There, a certain fraction of PUIs gets injected (*3*) into the process of diffusive shock acceleration and is transformed into ACRs.

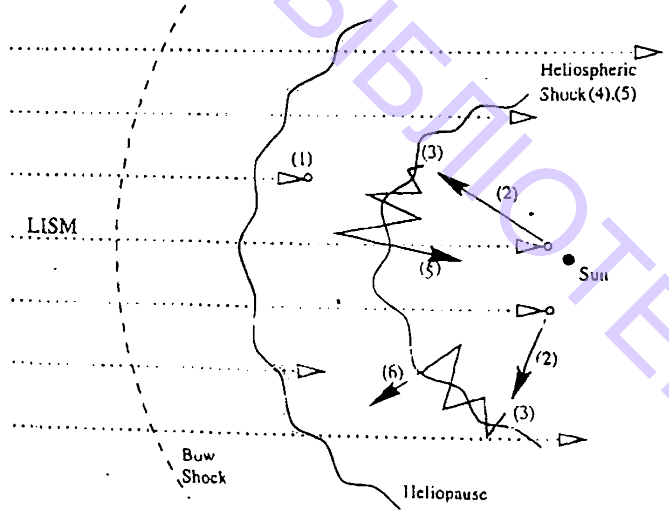


Figure 8: Processes resulting in the production of anomalous cosmic rays [4].

During their acceleration (solid lines repeatedly crossing the shock) the ACRs modify the structure of heliospheric shock (*4*). Eventually, after sufficient energization, the ACRs diffuse and drift either back into the region inside the asymmetric three-dimensional heliospheric shock surface (*5*) or into the heliosheath (*6*) where they, like those PUIs not being accelerated at the shock, serve as a source distribution of so called energetic neutral atoms (ENAs).

Extrapolating the gradients observed during the 1980s inside 49 AU outward of the shock source led to a best fit shock location of 84±5 AU [2]. The location of the termination shock will vary as the solar wind dynamic pressure varies over the solar cycle. The authors [19] have incorporated the temporal variations observed by *Voyager 2* into an MHD model that includes interstellar PUIs. It was found that the location of the termination shock varies by about 20 AU over the solar cycle. The minimum distance is predicted to occur near the solar maximum.

As shown in Figure 9, the shock is presently moving inward. As *Voyager 1* approaches the shock the low energy ACR spectra should rapidly increase within a few AU of the shock. To illustrate the time period over which shock encounters may occur, Figure 9 shows the predictions from the previous solar cycles delayed by 10, 20, and 30 years. Comparison with the *Voyager 1* trajectory suggests the possibility of one or more encounters with the termination shock by 2005. If there has been no encounter by then, the shock will likely again be moving outward with the arrival of increased pressure and it may be two to five years more before it moves back within *Voyager's* range.

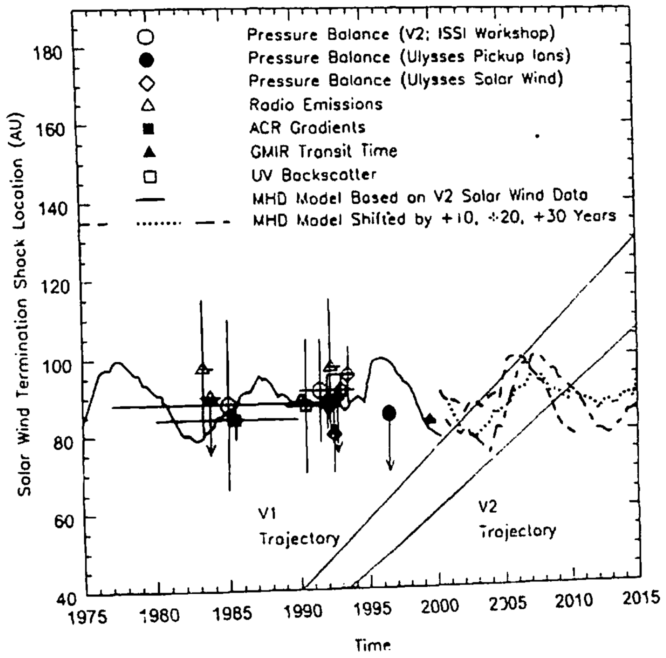


Figure 9: Estimates of the location of the solar wind termination shock by different methods [2]. Also shown are the trajectories of *V1* and *V2*.

5 Concluding Remarks

Much of the current knowledge about the physics of ACR has been reviewed in [2,4]. These energetic particles are, most probably, accelerated at the heliospheric shock that terminates the supersonic solar wind expansion. The relevance of heliospheric physics was pointed out in [2] as a vital link between basic (plasma) physics and (extra-heliospheric) astrophysics as well as a "test bed" for some astrophysical concepts.

In general, the interstellar fluxes of cosmic rays are reduced within the heliosphere, mainly inside the heliospheric shock that, according to current understanding, represents the main modulation boundary for GCR. The intensity of the anomalous component is also modulated by solar activity. However, modulation models based on this concept are just beginning to include realistic heliospheric configuration (see, e.g., [2,4]). It is, therefore, difficult to predict on the basis of such models the cosmic ray flux when the heliospheric shock is closer to the Sun and, thus, the main modulation is reduced. Because the solar wind flow and the heliospheric magnetic field structure inside the shock surface might change significantly, the heliospheric structure as a whole is modified and the cosmic ray modulation might be very different from what is observed nowadays.

Acknowledgements

This work was supported in part by the Russian Foundation of Basic Research (projects 02-02-39032, 03-02-96026), by the Federal Purpose Scientific and Technical Program (Section I, Project 4), and by a President's Grant of the Russian Federation (project 1445.2003.2). I am also very grateful to the Organizing Committee of GMIC-100 for inviting me as a plenary speaker as well as to my colleagues from Odessa National University and Odessa Astronomical Observatory for their hospitality.

References

[1] G.F. Krymsky, *Modulation of Cosmic Rays in the Interplanetary Space*, Nauka, Moscow (1969).

[2] E.C. Stone, A.C. Cummings, in W. Droege, H. Kunow and M. Scholer (eds) *Proc. 27th Int. Cosmic Ray Conf.*,, Copernicus Gesellschaft, Katlenburg-Lindau, Hamburg, **10**, p.4263 (2001).

[3] B. Heber, in R. Schlikeiser (ed) *Proc. 27th Int. Cosmic Ray Conf., Invited, Rapporteur, and Highlight Papers* (Special Volume), Copernicus Gesellschaft, Katlenburg-Lindau, Hamburg, p.118 (2001).

[4] H. Fichtner, *Space Sci. Rev.*, **95**, 639 (2001).

[5] L.I. Miroshnichenko, *Solar Cosmic Rays*, Kluwer Academic Publishers, Dordrecht (2001).

[6] E.W. Cliver, W. Droge, R. Muller-Mellin, *J. Geophys. Res.*, **98**, 15231 (1993).

[7] M.-B. Kallenrode, E.W. Cliver, in W. Droege, H. Kunow and M. Scholer (eds) *Proc. 27th Int. Cosmic Ray Conf.*, Copernicus Gesellschaft, Katlenburg-Lindau, Hamburg, **8**, p.3314 (2001).

[8] M.-B. Kallenrode, E.W. Cliver, in W. Droege, H. Kunow and M. Scholer (eds) *Proc. 27th Int. Cosmic Ray Conf.*, Copernicus Gesellschaft, Katlenburg-Lindau, Hamburg, **8**, p.3318 (2001).

[9] M.A. Shea, D.F. Smart, *Space Sci. Rev.*, **93**, 187 (2000).

[10] T.K. Gaisser, T. Stanev, *European Physical Journal C*, **3**,132 (1998).

[11] J.A. Simpson, *Annals of the New York Academy of Sciences*, **655**, 95 (1992).

[12] Yu.I. Stozhkov, in G.E. Kocharov (ed) *IX Leningrad Internat. Seminar on Cosmophysics*, Phys.–Techn. Inst., Leningrad, p.39 (1978).

[13] J.R. Jokipii, J. Kota, in M.S. Potgieter, B.C. Raubenheimer, D.J. van der Walt (eds) *Proc. 25th Int. Cosmic Ray Conf., Invited, Rapporteur, and Highlight Papers*, World Scientific, Singapore - New Jersy - London - Hong Kong, **8**, 151 (1997).

[14] V.V. Zharkova, A.G. Kosovichev, in S.G. Korzennik, A. Wilson (eds), *Proc. SOHO 6/GONG 98 Workshop, Structure and Dynamics of the Interior of the Sun and Sun-like Stars. Special Publication ESA SP-418*, European Space Agency, Noordwijk, p.661 (1998).

[15] E.W. Cliver, D.F. Webb, *Eos Trans. AGU*, **78**, No.10, 107 (1997).

[16] H. Sierks, I. Elendt, W. Droge et al., in M.S. Potgieter, B.C. Raubenheimer, D.J. van der Walt (eds) *Proc. 25th Int. Cosmic Ray Conf.*, World Scientific, Singapore – New Jersey – London – Hong Kong, **1**, p. 297 (1997).

[17] T.R. Sanderson, R.G. Marsden, A.M. Heras et al., *Geophys. Res. Lett.*, **19**, 1263 (1992).

[18] T.L. Garrard, E.R. Christian, R.A. Mewaldt et al., in M.S. Potgieter, B.C. Raubenheimer, D.J. van der Walt (eds) *Proc. 25th Int. Cosmic Ray Conf.*, World Scientific, Singapore – New Jersey – London – Hong Kong, **1**, 105 (1997).

[19] Y.C. Wang, L.F. Burlaga, *Geophys. Res. Lett.*, **27**, 1607 (2000).

II Cosmology and Gravitation

1

On the Higher Codimension Braneworld

S. Kanno and J. Soda

Department of Physics, Kyoto University, Kyoto 606-8501, Japan
e-mail: sugumi@tap.scphys.kyoto-u.ac.jp,
jiro@tap.scphys.kyoto-u.ac.jp

Abstract

We study a codimension 2 braneworld in the Einstein–Gauss–Bonnet gravity. In the linear regime, we show that the conventional Einstein gravity can be recovered on the brane. While, in the nonlinear regime, we find corrections due to the thickness and the bulk geometry.

Keywords: braneworld, codimension 2, thickness

1 Introduction

The old idea that our universe may be a braneworld embedded in a higher dimensional spacetime is renewed by recent developments in string theory which can be consistently formulated only in 10 dimensions. Instead of 10 dimensions, however, most studies of braneworld cosmology have been devoted to 5-dimensional models. Needless to say, it is important to investigate the higher codimension braneworld. In this paper, we investigate a codimension 2 braneworld in the Einstein–Gauss–Bonnet gravity taking into account the thickness. By examining the structure of the singularity in the equations of motion, we find a possibility to treat the thickness within the context of the distributional source. We name it a quasi-thick braneworld.

2 Quasi-thick Braneworld

We consider a codimension 2 braneworld with a positive tension in the 6-dimensional bulk spacetime described by the action

$$S = \frac{1}{2\kappa^2}\int d^6x\sqrt{-g_{(6)}}\left[R+\alpha R^2_{\rm GB}\right] - \int d^4x\sqrt{-g}\,\sigma + \int d^4x\sqrt{-g}\,\mathcal{L}_{\rm matter}\,, \qquad (1)$$

where κ^2 is the 6-dimensional gravitational constant, $g_{(6)\mu\nu}$ and $g_{\mu\nu}$ are the 6-dimensional bulk and the 4-dimensional brane metrics, respectively. Here, $\mathcal{L}_{\rm matter}$ is the Lagrangian density of matter on the brane, and σ is the brane tension. The Gauss–Bonnet (GB) term is given by $R^2_{\rm GB} = R^{abcd}R_{abcd} - 4R^{ab}R_{ab} + R^2$. The Latin indices $\{a, b, \cdots\}$ and the Greek indices $\{\mu, \nu, \cdots\}$ are used for tensors defined in the bulk and on the brane, respectively.

We will assume that a 6-dimensional metric has axial symmetry, which reads

$$ds^2 = dr^2 + g_{\mu\nu}(r, x^\mu)dx^\mu dx^\nu + L^2(r, x^\mu)d\theta^2 \ . \tag{2}$$

This assumption corresponds to the Z_2 symmetry in the Randall–Sundrum braneworld model. Here we have introduced polar coordinates (r, θ) for the two extra spatial dimensions, where $0 \le r < \infty$ and $0 \le \theta < 2\pi$. As we locate a four-dimensional brane at $r = 0$, which is a string like defect, we must take the boundary condition $\lim_{r\to 0} L(r, x^\mu) = 0$, $\lim_{r\to 0} L'(r, x^\mu) = \text{const.}$, where the prime denotes derivatives with respect to r. The first condition realizes the 4-dimensional brane at $r = 0$. And the second condition allows the existence of the conical singularity. The structure of conical singularity is a 2-dimensional delta function $\delta(r)/L$.

Possible components of the energy-momentum tensor T_{ab}, which could be balanced with the singular part of Einstein tensor, are given by

$$T_{ab}(\text{singular}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & T_{\mu\nu}\frac{\delta(r)}{2\pi L} + S_{\mu\nu}\delta(r) & 0 \\ 0 & 0 & S_{\theta\theta}\delta(r) \end{pmatrix} , \tag{3}$$

where $T_{\mu\nu}$ represents the conventional brane matter, while $S_{\mu\nu}$ and $S_{\theta\theta}$ describe the extra matter which mimics the thickness of the braneworld. In this way, the thickness can be treated within the context of a distributional source, which we dub quasi-thickness.

3 Effective Theory

3.1 Linear Regime

We now consider a perturbation around the vacuum solution

$$ds^2 = dr^2 + \eta_{\mu\nu}dx^\mu dx^\nu + c^2 r^2 d\theta^2 \ , \tag{4}$$

where c is a constant of integration. For $c \neq 1$, we have a conical singularity at $r = 0$. Then the matching condition determines the deficit angle in terms of the brane tension as $1 - c = \kappa^2\sigma/2\pi$. Note that $c < 1$ because the tension σ is positive. Here we will discuss only the scalar perturbation (see our paper [1] for the case of vector and tensor perturbations). The final result is written by

$$ds^2 = dr^2 + (1 + h_{\mu\nu})\,\eta_{\mu\nu}dx^\mu dx^\nu + L^2 d\theta^2 \ , \tag{5}$$

where

$$\begin{aligned} h_{\mu\nu} &= \frac{C_0}{2} + \frac{3}{r}Z_1(r, x^\mu) - \frac{3}{2}Z_2(r, x^\mu) + \int_0^r dr \int_0^r dr\delta g_{rr,\mu\nu} - 2\chi_{,\mu\nu}\, r \ , \\ \delta g_{rr} &= \frac{C_0}{2} - \frac{3}{r}Z_1(r, x^\mu) - \frac{3}{2}Z_2(r, x^\mu) \ , \end{aligned} \tag{6}$$

$$\begin{aligned} L^2 &= c^2 r^2 \left[1 - \frac{3}{2}C_0 - \frac{3}{r}Z_1(r, x^\mu) + \frac{9}{2}Z_2(r, x^\mu)\right] - c^2 r \int_0^r dr\delta g_{rr} \\ &+ 2c^2 r\chi \ . \end{aligned} \tag{7}$$

Here, we have defined

$$Z_1(r, x^\mu) = \int d^4p\; e^{ip\cdot x}A(q)J_1(qr) \ , \ Z_2(r, x^\mu) = \int d^4p\; e^{ip\cdot x}A(q)qJ_2(qr) \ . \tag{8}$$

Now, the brane is located at $r = 0$ in this coordinate system. Note that the distortion function $\chi(x^\mu)$ takes into account the fact that the brane is not necessarily located at constant r in the coordinate system which we used to solve equations in the bulk.

Since we have solved the bulk geometry, we next discuss the matching conditions. Using the energy-momentum tensor (3), matching conditions reduce to the effective theory on the brane. Using eq. (7), the deficit angle can be obtained from $\lim_{\epsilon\to 0} L'(r = \epsilon) = c(1 - C_0)$. This tells us that the deficit angle is perturbed on the brane at the linear level, but it is constant. From eq. (6), the extrinsic curvature in this limit gives $\lim_{\epsilon\to 0} K_{\mu\nu}(r = \epsilon, x^\mu) = \chi_{,\mu\nu}$. From this result, we see that the extrinsic curvature is nonzero near the brane, if the distortion field $\chi(x^\mu)$ exists. In the coordinate system where we have solved the equations of motion in the bulk, the location of the brane is specified by $\chi(x^\mu)$ and the resulting shape of the braneworld becomes the deformed cylinder in which the distortion is represented by $\chi(x^\mu)$ (see Figure 1). At first sight, this deformed cylinder looks like a 5-dimensional hypersurface. However, at the location of the brane ($r = \epsilon$ in the Gaussian normal gauge), the circumference radius of the cylinder vanishes because of $\lim\limits_{\epsilon\to 0} L(r = \epsilon, x^\mu) = 0$ there. Hence, the deformed cylinder is the 4-dimensional braneworld.

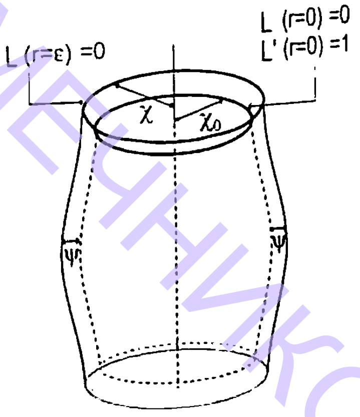


Figure 1: Schematic picture for the quasi-thick braneworld.

Now, we must specify $K_{\mu\nu}(r = 0, x^\mu)$ which is not known *a priori*. Without losing generality, we can write $K_{\mu\nu}(r = 0, x^\mu) = \chi_{0,\mu\nu}(x^\mu) + \rho_{\mu\nu}(x^\mu)$, $\rho^\mu{}_\mu = 0$, where we decomposed the extrinsic curvature at $r = 0$ into the traceless part $\rho_{\mu\nu}(x^\mu)$ and the trace part $\Box\chi_0(x^\mu)$. As the field $\chi(x^\mu)$ represents the location of the brane $r = \epsilon$, it is natural to regard $\chi_0(x^\mu)$ as the location of $r = 0$. Hence, $\psi(x^\mu) \equiv \chi(x^\mu) - \chi_0(x^\mu)$ can be interpreted as the thickness of the braneworld.

The matching condition of (μ, ν) component leads to

$$G^\mu{}_\nu = \frac{\kappa^2}{8\pi\alpha(1-c)} T^\mu{}_\nu + \delta^\mu_\nu \frac{cC_0}{4\alpha(1-c)} , \tag{9}$$

where the second term is the cosmological constant. Equation (9) proves that Einstein gravity is recovered for the codimension 2 braneworld in the case of the linearized gravity. This fact is consistent with the result in [2]. Considering the other components of the matching condition and conservation law of energy-momentum tensor, we get the relation $\chi_0 = c\chi$. Since $c < 1$, $\psi = \chi - \chi_0 = (1-c)\chi$ is positive. Hence, it is legitimate to interpret ψ as the effective thickness. It should be noted that the effective thickness ψ satisfies $\Box\psi = -\frac{\kappa^2}{3} S^\mu{}_\mu$. This is reminiscent of the radion in the codimension 1 braneworld.

3.2 Nonlinear Regime

The importance of the quasi-thickness can be manifest in the next order calculations. Up to the second order, we expect

$$\begin{aligned} G^\mu{}_\nu \quad = \quad & \frac{\kappa^2}{8\pi\alpha(1-c+cC_0)} T^\mu{}_\nu + \frac{cC_0 - \overset{(2)}{L}{}'}{4\alpha(1-c+cC_0)} \delta^\mu_\nu \\ & + c\Big[\chi^{;\mu}{}_{,\alpha}\chi^{;\alpha}{}_{,\nu} - \Box\chi\chi^{;\mu}{}_{,\nu} + \frac{1}{2}\delta^\mu_\nu \left\{(\Box\chi)^2 - \chi^{;\alpha}{}_{,\beta}\chi^{;\beta}{}_{,\alpha}\right\}\Big] \\ & - \frac{1}{1-c}\left[\rho^\mu{}_\alpha\rho^\alpha{}_\nu - \delta^\mu_\nu \rho^\alpha{}_\beta\rho^\beta{}_\alpha\right] , \end{aligned} \tag{10}$$

where the corrections in the last two lines carry the information of thickness of the brane. Another correction comes from $\overset{(2)}{L}{}'$ carries the effects of bulk, that is, the deformed deficit angle. This needs the next order analysis. Thus, it turns out that some corrections due to the quasi-thickness can be expected at the second order.

4 Conclusion

We have shown that the Gauss-Bonnet gravity gives a consistent framework discussing the quasi-thick codimension 2 braneworld.

Acknowledgements

This work was supported in part by Grant-in-Aid for Scientific Research Fund of the Ministry of Education, Science and Culture of Japan No. 155476 (SK) and No.14540258 (JS) and also by a Grant-in-Aid for the 21st Century COE "Center for Diversity and Universality in Physics".

References

[1] S. Kanno, J. Soda, *Journal of Cosmology and Astroparticle Physics*, 07, 002 (2004).

[2] P. Bostock, R. Gregory, I. Navarro, J. Santiago, *Phys. Rev. Lett.*, 92, 221601 (2004).

2

6D Brane with Gravitational Trapping Potential

M. Gogberashvili[1] and D. Singleton[2,3]

[1]*Andronikashvili Institute of Physics, 6 Tamarashvili Str.*
Tbilisi 0177, Georgia
e-mail: gogber@hotmail.com
[2]*Physics Dept., CSU Fresno, Fresno, CA 93740-8031, USA*
[3]*Center for Gravitation and Fundamental Metrology,*
VNIIMS, 3-1 M. Ulyanovoy Str., Moscow, 119313, Russia
e-mail: dougs@csufresno.edu

Abstract

We present a new (1+3)-brane solution to Einstein equations in (1+5)-space. As distinct from previous models this solution is free of singularities. The gravitational potential transverse to the brane is an increasing (but not exponentially) function and asymptotically approaches a finite value. The solution localizes the zero modes of all kinds of matter fields and Newtonian gravity on the brane.

Keywords: 6D brane, gravitational trapping, non-singular solution

The scenario where our world is associated with a brane embedded in a higher dimensional space-time [1–4] has attracted a lot of interest. In the models of [3, 4] gravitons, which are allowed to propagate in the bulk, are confined on the brane because of a warped geometry. However, there are difficulties with the choice of a natural trapping mechanism for some matter fields. For example, in the existing (1 + 4)-dimensional models spin 0 and spin 2 fields are localized on the brane with an exponentially decreasing gravitational warp factor, spin 1/2 field are localized with an increasing factor [5], and spin 1 fields are not localized at all [6]. For (1+5)-dimensions spin 0, 1 and 2 fields are localized on the brane with a decreasing warp factor and spin 1/2 fields again are localized with an increasing factor [7].

For reasons of economy and to avoid charge universality obstruction [8] we would like to have a universal gravitational trapping mechanism for all fields. In [9–11] we found such a solution of the 6-dimensional Einstein equations in (2+4)- and (1+5)-spaces, which localized all kind of bulk fields on the brane. These solutions contain non-exponential scale factors, which increase from the brane, and asymptotically approach a finite value at infinity. In [12] the solution of [10] was generalized to the case of n dimensions.

In this paper we present a new solution to the Einstein equations in $(1+5)$-space, which, similar to the models [9, 10], localizes all kind of fields, but is free of singularities. Because of this feature of the solution we are able to fix the free parameters of the model by setting realistic boundary conditions.

The 6-dimensional Einstein equations with stress-energy tensor T_{AB} are

$$ {}^6R_{AB} - \frac{1}{2} g_{AB}\, {}^6R = \frac{1}{M^4}\left(\Lambda g_{AB} + T_{AB}\right) . \tag{1}$$

Capital Latin indices run over $A, B, ... = 0, 1, 2, 3, 5, 6$. M is the fundamental scale, ${}^6R$ is the scalar curvature, Λ is the cosmological constant. All of these quantities are six dimensional.

As in [9, 10] we choose the ansatz for the 6-dimensional metric as

$$ds^2 = \phi^2(r)\eta_{\alpha\beta}(x^\nu)dx^\alpha dx^\beta - \lambda(r)(dr^2 + r^2 d\theta^2) , \tag{2}$$

where the Greek indices $\alpha, \beta, ... = 0, 1, 2, 3$ refer to 4-dimensional coordinates. The metric of ordinary 4-space, $\eta_{\alpha\beta}(x^\nu)$, has the signature $(+,-,-,-)$. The functions $\phi(r)$ and $\lambda(r)$ depend only on r, and thus are cylindrically symmetric in the transverse polar coordinates $(0 \le r < \infty, 0 \le \theta < 2\pi)$.

The stress-energy tensor T_{AB} is assumed to have the form

$$T_{\mu\nu} = -g_{\mu\nu}F(r), \qquad T_{ij} = -g_{ij}K(r), \qquad T_{i\mu} = 0 . \tag{3}$$

To solve eqs. (1) we take the 4-dimensional Einstein equations to have the ordinary form without a cosmological term *i.e.* $R_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}R = 0$. With ansätze (2) and (3) the $\alpha\alpha$, rr, and $\theta\theta$ components of Einstein field equations (1) respectively become (here prime $= \partial/\partial r$)

$$\begin{aligned} 3\frac{\phi''}{\phi} + 3\frac{\phi'}{r\phi} + 3\frac{(\phi')^2}{\phi^2} + \frac{1}{2}\frac{\lambda''}{\lambda} - \frac{1}{2}\frac{(\lambda')^2}{\lambda^2} + \frac{1}{2}\frac{\lambda'}{r\lambda} &= \frac{\lambda}{M^4}[F(r) - \Lambda] , \\ \frac{\phi'\lambda'}{\phi\lambda} + 2\frac{\phi'}{r\phi} + 3\frac{(\phi')^2}{\phi^2} &= \frac{\lambda}{2M^4}[K(r) - \Lambda] , \\ 2\frac{\phi''}{\phi} - \frac{\phi'\lambda'}{\phi\lambda} + 3\frac{(\phi')^2}{\phi^2} &= \frac{\lambda}{2M^4}[K(r) - \Lambda] , \end{aligned} \tag{4}$$

In [10] the source functions $F(r)$ and $K(r)$ outside the core $r > \epsilon$ were taken to have the form (f is some constant)

$$F(r > \epsilon) = K(r > \epsilon) = \frac{f}{\phi^2} , \tag{5}$$

and the following $\phi(r)$ and $\lambda(r)$ were found to be solutions of the Einstein equations

$$\phi(r) = a \tanh\left[\frac{b}{2}\ln\left(\frac{r}{c}\right)\right] = a\frac{r^b - c^b}{r^b + c^b} , \qquad \lambda(r) = \frac{2bc^b\rho^2 r^{b-2}}{(r^b + c^b)^2} , \tag{6}$$

where

$$f = \frac{3\Lambda a^2}{5} , \qquad b = \frac{a\Lambda\rho^2}{5M^4} , \tag{7}$$

ρ and c are integration constants with units of length, and $\phi(r \to \infty) \to a = \sqrt{5f/(3\Lambda)}$.

In [10] it was shown that the solution (6) provides a universal, gravitational trapping for all kinds of matter fields. However, in this model we did not specify source functions on the brane $0 \le r \le \epsilon$ and there were a large number of free parameters. Now we want to choose the source functions $F(r)$ and $K(r)$ everywhere, so that the solution ϕ will localize all kind of physical fields and be a regular function in the full 6-dimensional space-time.

To construct the non-singular solution we take the source functions $F(r)$ and $K(r)$, as

$$F(r) = \frac{f_1}{2\phi^2} + \frac{3f_2}{4\phi} , \qquad K(r) = \frac{f_1}{\phi^2} + \frac{f_2}{\phi} , \tag{8}$$

where f_1, f_2 are constants. For these source functions it was found [11] that

$$\phi = \frac{c^b + ar^b}{c^b + r^b} , \qquad \lambda = \frac{(a-1)bc^b\rho^2 r^{b-2}}{(c^b + r^b)^2} , \tag{9}$$

where

$$b = \frac{\rho^2\Lambda}{10M^4}(a - 1) , \qquad f_1 = -\frac{3\Lambda}{5}a , \qquad f_2 = \frac{4\Lambda}{5}(a + 1) , \tag{10}$$

and again ρ and c are integration constants with units of length. Note that for $b \ge 2$ we have $\phi(r \to \infty) \to a$ and $\phi'(r \to \infty) \to 0$. Also $\lambda(r \to \infty) \to 0$ so the effective geometry is 4-dimensional at $r = \infty$. Near the origin $r = 0$ we require that ϕ has the following boundary conditions

$$\phi(r \to 0) \approx 1 + dr^2 , \qquad \phi'(r \to 0) \approx 2dr , \tag{11}$$

where d is some constant. The above conditions fix $b = 2$.

The width of the brane ϵ corresponds to the inflection point of the function ϕ. Thus the condition $\phi''(r = \epsilon) = 0$ fixes c in (9) as $c^2 = 3\epsilon^2$. Finally if we require that we have a 6-dimensional Minkowski metric on the brane, $\lambda(r = 0) = 1$, we can fix also the integration constant ρ as $\rho^2 = \frac{3\epsilon^2}{2(a-1)}$. Putting all this together the solution in (9) becomes

$$\phi = \frac{3\epsilon^2 + ar^2}{3\epsilon^2 + r^2} , \qquad \lambda = \frac{9\epsilon^4}{(3\epsilon^2 + r^2)^2} , \tag{12}$$

Note that λ does not depend on a.

In [11] it was shown that the brane width can be expressed in terms of the bulk cosmological constant and fundamental scale

$$\epsilon^2 = \frac{40M^4}{3\Lambda} . \tag{13}$$

It can be shown [11] that the solution given by (12) satisfies the condition of classical stability that the total momentum of the brane matter in the direction of the extra dimensions be zero [14]. Finally in [11] it was shown that the solution in (12) traps fields of spin $0, 1/2, 1$ to the brane and reproduces 4D Newtonian gravity on the brane.

Even though the coordinate r runs from 0 to ∞ the proper distance along r is finite

$$s = \int ds = \int_0^\infty \sqrt{\lambda(r)}dr = \int_0^\infty \frac{3\epsilon^2}{3\epsilon^2 + r^2}dr = \frac{\pi\sqrt{3}\epsilon}{2}. \tag{14}$$

Thus this solution is more closely related to the two brane model with a finite distance between the branes (first reference in [4]) rather than the one brane model with an infinite extra dimension ([3] and second reference in [4]).

To summarize, we have shown that for a realistic form of the brane stress-energy, there exists a static, non-singular solution of the 6-dimensional Einstein equations, which provides a gravitational trapping of 4-dimensional gravity and matter fields on the brane.

Acknowledgements

This work is supported by a 2003 COBASE grant.

References

[1] N. Arkani-Hamed, S. Dimopoulos, G. Dvali, *Phys. Lett.*, **B429**, 263 (1998);

[2] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, G. Dvali, *Phys. Lett.*, B436, 257 (1998).

[3] M. Gogberashvili, *Int. J. Mod. Phys.*, **D11**, 1635 (2002).

[4] L. Randall, R. Sundrum, *Phys. Rev. Lett.*, **83**, 3370 (1999).

[5] B. Bajc, G. Gabadadze, *Phys. Lett.*, **B474**, 282 (2000).

[6] A. Pomarol, *Phys. Lett.*, **B 486**, 153 (2000).

[7] I. Oda, *Phys. Rev.*, **D 62**, 126009 (2000).

[8] S.L. Dubovsky, V.A. Rubakov, P.G. Tinyakov, *Journal of High Energy Physics*, **0008**, 041 (2000).

[9] M. Gogberashvili, P. Midodashvili, *Phys. Lett.*, **B 515**, 447 (2001).

[10] M. Gogberashvili, D. Singleton, *Phys. Rev.*, **D69** 026004 (2004).

[11] M. Gogberashvili, D. Singleton, *Phys. Lett.*, **B582**, 95 (2004).

[12] I. Oda, *Phys. Lett.*, **B571**, 235 (2003).

[13] P. Midodashvili, preprint hep-th/0308051 (2003).

[14] M. Gogberashvili, *Mod. Phys. Lett.*, **A14**, 2025 (1999).

3

Accelerating Multidimensional Cosmologies with Scalar Fields

V. Baukh

Department of Theoretical Physics, Odessa National University
2 Dvoryanskaya Street, 65026 Odessa, Ukraine
e-mail: bauch_vgr@ukr.net

Abstract

We study multidimensional cosmological models with a higher-dimensional product manifold $R \times M_0 \times M_1$ (where M_0 and M_1 are spherical and flat spaces, correspondingly) in the presence of a minimal free scalar field. The dimensions of spaces $M_i\,(i = 0, 1)$ are not fixed and the choice of an external (our) space in the product manifold is arbitrary. Dynamical behaviour of the model is analyzed both in Einstein and Brans–Dicke conformal frames. For a number of particular cases, it is shown that external space–time undergoes an accelerated expansion.

Keywords: accelerated cosmologies, Einstein frame, Brans–Dicke frame, scalar field

1 Introduction

Distance measurements of type Ia supernovae (SNe,Ia) [1] as well as cosmic microwave background (CMB) anisotropy measurements [2] performed in recent years give strong evidence for the existence of dark energy – a smooth energy density with negative pressure which causes an accelerated expansion of the Universe at the present time. The challenge to theoretical cosmology consists in finding a natural explanation of this acceleration (dark energy).

In this paper we consider a multidimensional cosmological model with a higher-dimensional product manifold which consists of spherical and flat spaces with dimensions d_0 and d_1 correspondingly in the presence of a minimally coupled free scalar field. We investigate this model to find under which circumstances the external spacetime undergoes an accelerated expansion with simultaneous compactification of the internal space.

To start with, let us consider a cosmological model with factorizable geometry

$$g = -\exp[2\gamma(\tau)]d\tau \otimes d\tau + a_0^2(\tau)g_0 + a_1^2(\tau)g_1\,, \tag{1}$$

which is defined on the manifold

$$M = R \times M_0 \times M_1 \tag{2}$$

and with minimal free scalar field φ. We assume that the factors M_i are Einstein spaces: $R[g_0] = R_0 > 0$, $R[g_1] = 0$. The action for considered model reads

$$S = \frac{1}{2\kappa^2}\int\limits_M d^Dx\sqrt{|g|}R + \int\limits_M d^Dx\sqrt{|g|}\left[-\frac{1}{2}e^{-2\gamma(\tau)}\partial_M\varphi\partial_N\varphi\right] + S_{YGH}\,, \tag{3}$$

where S_{YGH} is the York–Gibbons–Hawking boundary term.

It can be easily seen [3] that in the harmonic time gauge $\gamma = \gamma_0$ (see [4]) the equations of motion have the following solutions:

$$\begin{aligned}
a_0(\tau) &= a_{(c)0}\frac{\exp\left(-\frac{1}{d_0-1}\sqrt{\frac{d_1(d_0-1)}{D-2}}p^1\tau\right)}{\left[\cosh\left(\sqrt{\frac{(d_0-1)2\varepsilon}{d_0}}\tau\right)\right]^{\frac{1}{d_0-1}}}\,,\\
a_1(\tau) &= a_{(c)1}\exp\left(\frac{1}{d_1}\sqrt{\frac{d_1(d_0-1)}{D-2}}p^1\tau\right)\,,\\
\varphi(\tau) &= p^2\tau + q^2\,,
\end{aligned} \tag{4}$$

where $a_{(c)0,1}$, $p^{1,2}$ and q^2 are constants of integration and $2\epsilon = (p^1)^2 + (p^2)^2 > 0$.

We shall analyze a dynamical behaviour of the model both in Brans–Dicke and Einstein conformal frames where metrics are connected as follows ([5])

$$\begin{aligned}
g &= -e^{2\gamma}d\tau \otimes d\tau + a_e^2 g^{(e)} + a_i^2 g^{(i)}\\
&= -dt \otimes dt + a_e^2 g^{(e)} + a_i^2 g^{(i)}\\
&= \Omega^2(-d\tilde{t} \otimes d\tilde{t} + \tilde{a}_e^2 g^{(e)}) + a_i^2 g^{(i)}\,.
\end{aligned} \tag{5}$$

Here, t $(\tilde{t})$ is a physical/synchronous time and a_e, a_i $(\tilde{a}_e, \tilde{a}_i)$ are scale-factors of the external and internal spaces in the Brans–Dicke (Einstein) frame. The conformal factor $\Omega = a_i^{-d_i}$ (in the case of 4-D external spacetime).

2 Dynamical Behaviour in the Brans–Dicke Frame

Here, the Hubble and the deceleration parameters can be expressed via harmonic time as follows:

$$H_0 = \frac{1}{a_0}\frac{da_0}{dt} = -\frac{\xi_1 + \xi_2\tanh(\xi_2\tau)}{f(\tau)(d_0-1)}, \quad -q_0 = \frac{1}{a_0}\frac{d^2a_0}{dt^2} = -\xi_2\frac{\xi_2 + \xi_1\tanh(\xi_2\tau)}{f^2(\tau)(d_0-1)},$$

$$H_1 = \frac{1}{a_1}\frac{da_1}{dt} = \frac{\xi_1}{f(\tau)d_1}, \quad -q_1 = \frac{1}{a_1}\frac{d^2a_1}{dt^2} = \frac{\xi_1((D-2)\xi_1 + d_0d_1\xi_2\tanh(\xi_2\tau))}{f^2(\tau)d_1^2(d_0-1)},$$

where $f(\tau) = \frac{dt}{d\tau}$, $\xi_1 = \sqrt{\frac{d_1(d_0-1)}{D-2}}p^1$, $\xi_2 = \sqrt{\frac{(d_0-1)2\varepsilon}{d_0}}$,. We consider the expanding space as an external one. Thus, we arrive to two separate cases:

i. M_0 is the external space ($\xi_1 < 0$):
In this case, we find that M_0 undergoes an accelerated expansion only if $|\xi_1| > \xi_2$ with simultaneous compactification of M_1. For the asymptotical behaviour $\tau \to \infty$ we have two particular cases where external space undergoes accelerated expansion:

1. $d_0\xi_2 > |\xi_1| > \xi_2$, *Big Rip scenario*:
In this case the scale-factor of M_0 is

$$a_0(t) \simeq a_{(c)0}\left[\frac{(d_0\xi_2 - |\xi_1|)(c-t)}{(d_0-1)2^{\frac{d_0}{d_0-1}}a_{(c)0}^{d_0}a_{(c)1}^{d_1}}\right]^{-\frac{|\xi_1|-\xi_2}{d_0\xi_2-|\xi_1|}}. \tag{6}$$

It can be easily seen that the Universe behaves according to the Big Rip scenario, i.e. the scale-factor a_0 achieves infinite value in a finite period of time.

2. $|\xi_1| > d_0\xi_2$:
Here, for the scale factor of the external space we obtain the asymptote:

$$a_0(t) \simeq a_{(c)0}\left[\frac{(|\xi_1| - d_0\xi_2)(t-c)}{(d_0-1)2^{\frac{d_0}{d_0-1}}a_{(c)0}^{d_0}a_{(c)1}^{d_1}}\right]^{\frac{|\xi_1|-\xi_2}{|\xi_1|-d_0\xi_2}}. \tag{7}$$

For these cases, the following relations between momentum p^1 and kinetic energy ε hold:

- case $d_0\xi_2 > |\xi_1| > \xi_2$: $\frac{d_0d_1(p^1)^2}{(D-2)}2\varepsilon > \frac{d_1(p^1)^2}{d_0(D-2)}$, i.e real scalar field and Big Rip scenario
- case $|\xi_1| > d_0\xi_2$: $2\varepsilon < \frac{d_1(p^1)^2}{d_0(D-2)}$, so $2\varepsilon < (p^1)^2$, i.e it is possible only if $(p^2)^2 = (\dot{\varphi}(\tau))^2 < 0$ and scalar field is imaginary

ii. M_1 is the external space ($\xi_1 > 0$):
In this case space M_1 undergoes an accelerating expansion (via *Big Rip scenario*) with simultaneous compactification of M_0. Such behaviour for M_1 can be easily traced from the asymptotical behaviour of $a_1(t)$:

$$a_1(t) \simeq a_{(c)1}\left[\frac{(\xi_1 + d_0\xi_2)(c-t)}{(d_0-1)2^{\frac{d_0}{d_0-1}}a_{(c)0}^{d_0}a_{(c)1}^{d_1}}\right]^{-\frac{\xi_1(d_0-1)}{(\xi_1+d_0\xi_2)d_1}}. \tag{8}$$

3 Dynamical Behaviour in the Einstein Frame

Now, we analyze the behaviour of the model in the Einstein frame. Here, we also consider two separate cases:

i. M_0 is the external space:
If M_0 is the external space, its deceleration parameter is

$$-\tilde{q}_0 = \frac{1}{\tilde{a}_0}\frac{d^2\tilde{a}_0}{d\tilde{t}^2} = -\frac{\xi_2^2}{f_{E0}^2(\tau)(d_0-1)} < 0\,, \tag{9}$$

where $f_{E0}(\tau) = \frac{d\tilde{t}}{d\tau}$. It can be easily seen that in this case an accelerated expansion of the external space is impossible.

ii. M_1 is the external space:
If M_1 is the external space, its deceleration parameter reads:

$$\begin{aligned} -\tilde{q}_1 &= \frac{1}{\tilde{a}_1}\frac{d^2\tilde{a}_1}{d\tilde{t}^2} = -\frac{1}{f_{E1}^2(\tau)d_1^2(d_0-1)^2(d_1-1)} \times \\ &\times \left[[(D-2)\xi_1 + d_1 d_0 \xi_2 \tanh(\xi_2\tau)]^2 + \frac{d_0 d_1^2(d_0-1)\xi_2^2}{\cosh^2(\xi_2\tau)}\right] < 0\,, \end{aligned} \tag{10}$$

where $f_{E1}(\tau) = \frac{d\tilde{t}}{d\tau}$. Accelerated expansion of the external space is impossible in this case too.

4 Summary

Briefly, we can summarize our results as follows:

In the Brans–Dicke frame:
i. accelerated expansion of the positive curvature space can be achieved either if it undergoes the Big Rip scenario or if it contains a scalar field with negative kinetic term.

ii. Ricci-flat space undergoes late time acceleration via Big Rip scenario and its qualitative behavior does not depends on a particular choice of parameters of the model.

In the Einstein frame:
it is impossible to achieve an accelerated expansion of the external space (either M_0 or M_1).

Acknowledgements

The author would like to thank Alexander Zhuk for guidance and assistance.

References

[1] S. Perlmutter, M.S. Turner, M.J. White, *Phys. Rev. Lett.*, **83**, 670 (1999).

[2] A.H. Jaffe et al., *Phys. Rev. Lett.*, **86**, 3475 (2001).

[3] U. Bleyer, A. Zhuk, *Gravitation & Cosmology*, **1**, 37 (1995).

[4] V.D. Ivashchuk, V.N. Melnikov, A.I. Zhuk, *Nuovo Cimento*, **B104**, 575 (1989).

[5] M. Rainer, A. Zhuk, *Gen. Rel. Grav.*, **32**, 79 (2000).

4

Ghost Scalar Field as a Dark Energy

S.V. Sushkov

Kazan State Pedagogical University, Kazan, Russia
e-mail: sergey_sushkov@mail.ru

Abstract

We analyze the non-static spherically symmetric exact solution of general relativity with the ghost scalar field found in [1]. This solution describes a wormhole joining two homogeneous spatially flat universes expanding with acceleration. We show that out of the throat the ghost scalar field behaves effectively as phantom energy with the equation of state $p = \omega\rho$, $\omega < -1$, providing the expansion of the Universe with infinitely growing acceleration (the Big Rip scenario).

Keywords: dark energy, phantom energy, ghost scalar field

1 Introduction

Recent observations related to distant supernovae, cosmic microwave background, and galaxy clustering all together strongly indicate in favor of the accelerating expansion of the Universe.[1] In the framework of general relativity it means that a considerable part of the Universe consists of an exotic matter with negative pressure $p<0$ dubbed *dark energy*. The dark energy is usually described by the "equation-of-state" parameter $\omega \equiv p/\rho$, the ratio of the spatially homogeneous dark-energy pressure p to its energy density ρ. To describe the cosmic acceleration we must require $\omega < -1/3$. It is important to note that the possibility $\omega < -1$ is also not excluded by observations. In this case the dark energy is called *phantom energy*. The simple way leading to phantom energy can be provided by a ghost scalar field, i.e., scalar field with negative kinetic energy. Recently, a cosmological model including a ghost scalar field with an exponential potential has been examined in paper [1], and as a result an exact solution describing a wormhole in a cosmological setting has been found there. In this paper we will analyze in detail this solution and show

---

[1] Because of limited space allotted for a contribution we here do not give many references. The reader can find the references of interest in [1].

that actually the ghost scalar field behaves effectively as phantom energy providing the accelerating expansion of the Universe.

2 Exact Solution

In this section we briefly discuss results obtained in [1]. The action for a ghost scalar field ϕ minimally coupled to general relativity is

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{8\pi G}R + (\nabla\phi)^2 - 2V(\phi)\right], \tag{1}$$

where $g_{\mu\nu}$ is a metric, $g = \det(g_{\mu\nu})$, R is the scalar curvature, and the potential $V(\phi)$ is given in the exponential form:

$$V = V_0\exp(-k\phi). \tag{2}$$

Varying the action (1) gives the Einstein equations

$$R_{\mu\nu} = 8\pi G\left[-\nabla_\mu\phi\nabla_\nu\phi + g_{\mu\nu}V_0\exp(-k\phi)\right], \tag{3}$$

and the equation of motion of the scalar field

$$\nabla^\alpha\nabla_\alpha\phi = kV_0\exp(-k\phi). \tag{4}$$

The general time-dependent spherically symmetric solution of the system (3), (4) reads

$$ds^2 = -e^{-2\alpha^2 at+2u}dt^2 + e^{2at-2u}\left[dr^2 + (r^2+r_0^2)d\Omega^2\right], \tag{5}$$

$$\phi(t,r) = (4\pi G\alpha^2)^{-1/2}\left[u - \alpha^2 at\right], \tag{6}$$

where

$$u(r) = \frac{m}{r_0}\arctan\frac{r}{r_0}, \quad \alpha^2 = \frac{m^2}{m^2+r_0^2}, \tag{7}$$

with m, r_0 and a being three free parameters related to the parameters of the potential as follows:

$$k = 4\alpha(\pi G)^{1/2}, \quad V_0 = \frac{a^2(3+\alpha^2)}{8\pi G}. \tag{8}$$

The spacetime described by the metric (5) represents two asymptotically homogeneous spatially flat universes connected by a throat or, in other words, a *wormhole in cosmological setting*. At each fixed moment of time t = const the wormhole's throat radius corresponds to the minimum of the radius of a two-dimensional sphere, $R^2(r) = e^{2at-2u(r)}(r^2+r_0^2)$. The minimum of $R(r)$ is achieved at $r = m$ and is equal to

$$R_0 = \exp\left(2at - \frac{m}{r_0}\arctan\frac{m}{r_0}\right)(m^2+r_0^2)^{1/2}. \tag{9}$$

Both the universes and the wormhole's throat are simultaneously expanding with acceleration. The character of the acceleration qualitatively depends on the wormhole's mass parameter m. In the case $m = 0$ the acceleration is constant, so that the corresponding spacetime configuration represents two de Sitter universes joining by the throat. In the case $m \neq 0$ the acceleration turns out to be infinitely growing, so that the metric (5) describes now the inflating wormhole connecting two homogeneous spatially flat universes expanding with infinitely growing acceleration into the final singularity (the Big Rip scenario).

3 Ghost Scalar Field as Dark Energy

The stress-energy tensor of the ghost scalar field corresponding to the action (1) is

$$T_{\mu\nu} = -\nabla_\mu\phi\nabla_\nu\phi + \frac{1}{2}g_{\mu\nu}(\nabla\phi)^2 - g_{\mu\nu}V(\phi). \tag{10}$$

In our case because of spherical symmetry the only nonzero components are T_0^0, T_1^0, T_1^1, and $T_2^2 = T_3^3$. Let us discuss the energy density $\rho = -T_0^0$, the radial pressure $p = T_1^1$, and the transverse pressure $T_2^2 = T_3^3 = p_{tr}$. Substituting the solution (5), (6) into (10) yields

$$8\pi G\,\rho = -\frac{m^2+r_0^2}{(r^2+r_0^2)^2}e^{-2at+2u} + 3a^2e^{2\alpha^2at-2u}, \tag{11}$$

$$8\pi G\,p = -\frac{m^2+r_0^2}{(r^2+r_0^2)^2}e^{-2at+2u} - (3a^2+2a^2\alpha^2)e^{2\alpha^2at-2u}, \tag{12}$$

$$8\pi G\,p_{tr} = \frac{m^2+r_0^2}{(r^2+r_0^2)^2}e^{-2at+2u} - (3a^2+2a^2\alpha^2)e^{2\alpha^2at-2u}. \tag{13}$$

First of all, analyzing the above expressions, we have to stress that the radial pressure p given by (12) is *negative*, $p < 0$. The spatial distribution of ρ, p and p_{tr} at a fixed moment of time is shown in Figure 1. It is seen that the ghost scalar field configuration is inhomogeneous in the wormhole throat vicinity since the transverse pressure p_{tr} is not equal to the radial pressure p near the point $r = m$. At $r \to \pm\infty$ the value of p_{tr} quickly tends to p. This means that out of the throat the ghost

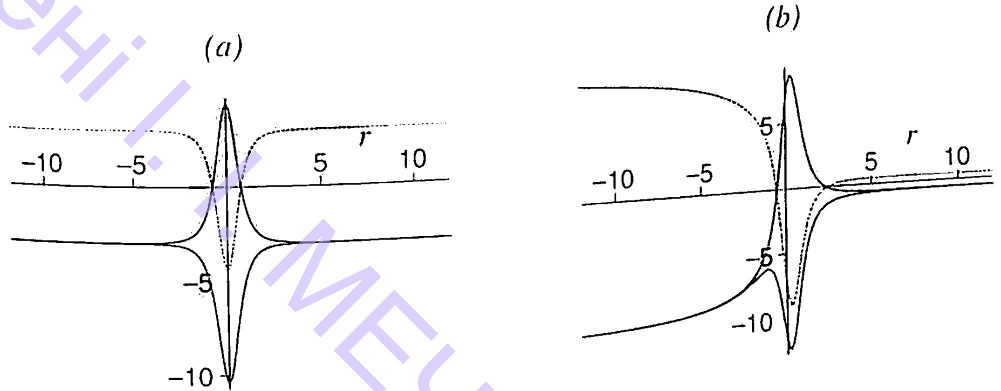


Figure 1: The spatial distribution of the energy density ρ (a dashed line), the radial pressure p (a lower solid line), and the transverse pressure p_{tr} (upper solid line). The plots (a) and (b) correspond to $m = 0$ and $m = 0.5$, respectively. The choice for the other parameters is $r_0 = 1$, $a = 1$.

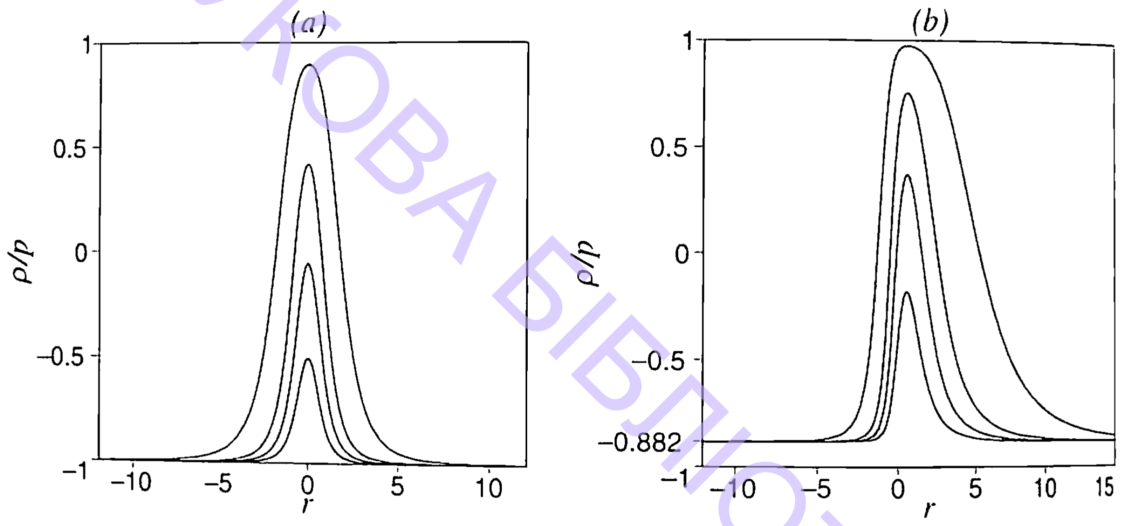


Figure 2: Spatial distribution of the ratio $\kappa = \rho/p$. Curves from top to bottom are given for various values of the time coordinate, $t = -2, -1, 0, 1$, respectively, and the plots (a) and (b) correspond to $m = 0$ and $m = 0.5$. The choice for other parameters is $r_0 = 1$, $a = 1$.

scalar field represents an homogeneous distribution of dark energy. To characterize the dark energy we usually consider the "equation-of-state" parameter $\omega = p/\rho$. In our case it is more convenient to analyze the inverse value $\kappa \equiv 1/\omega = \rho/p$. Curves describing the spatial distribution of κ are demonstrated in Figure 2. The curve's peaks are localized at the wormhole's throat. Out of the throat the value of κ tends to a negative constant, $-1 < \kappa < 0$, so that $p = \omega\rho$ with $\omega < -1$. This dependence means that the ghost scalar field out of the throat behaves like phantom energy.

4 Summary

In this paper we have analyzed the non-static spherically symmetric exact solution of general relativity with the ghost scalar field found in [1]. This solution describes a wormhole joining two homogeneous spatially flat universes expanding with acceleration. Our analysis has revealed that the radial pressure of the ghost scalar field is everywhere and always negative, and out of the wormhole's throat the radial and transverse pressures tend quickly to be equal. This feature indicates that out of the throat (i.e., in regions representing two homogeneous spatially flat universes expanding with acceleration) the ghost scalar field behaves effectively as dark energy providing the accelerated expansion of the Universe. Moreover, out of the throat the effective equation of state reads $p = \omega\rho$ with $\omega < -1$, i.e., the ghost scalar field represents the phantom energy providing the specific cosmological scenario called the Big Rip.

5 Acknowledgement

This work was supported by the Russian Foundation for Basic Research grant No 05-02-17344.

References

[1] S.V. Sushkov, S.-W. Kim, *Gen. Rel. Grav.*, **36**, 1671 (2004).

5

Dissolution of Black Holes in Phantom Energy

E.O. Babichev, V.I. Dokuchaev and Yu.N. Eroshenko

Institute for Nuclear Research, RAS, Moscow, Russia
e-mail: babichev@inr.npd.ac.ru

Abstract

A solution for a stationary spherically symmetric accretion of a relativistic perfect fluid with an equation of state $p(\rho)$ onto a Schwarzschild black hole is presented. This solution is a generalization of Michel solution and applicable to the problem of dark energy accretion. It is shown that accretion of phantom energy is accompanied with gradual decrease of the black hole mass. Masses of all black holes tend to zero in the phantom energy universe approaching the Big Rip.

Keywords: black holes, accretion, cosmology, dark energy

In recent years the concept of dark energy was accepted by cosmologists [1–3]. The energy with negative pressure was widely exploited for creating new interesting cosmological scenarios, including the exotic model of a universe which will die in a so-called Big Rip [4, 5]. In view of that the medii with negative pressure penetrate cosmology and particle physics and we need to revise the different phenomena of physics in which dark energy can take part. In particular, dark energy must participate in accretion onto black holes.

Usually the evolution of quintessence or k-essence is considered with the view of cosmological problems. However in the presence of compact objects such as black holes the evolution of dark energy should be sufficiently different from that in cosmological considerations. Indeed, what would be the fate of black holes in a universe filled with phantom energy and approaching Big Rip? In this paper (see [6] for more details) we considered the spherically symmetric stationary accretion of dark energy modelled by a perfect fluid into a black hole. The general equations describing the accretion onto a Schwarzschild black hole of a perfect fluid with equation of state $p = p(\rho)$ were obtained. In particular, these equations are applicable to radiation, dark energy and phantom energy. It has been found that the mass of a black hole is increasing as usually if $\rho + p > 0$. However a qualitatively new result has been obtained for the phantom energy, i. e. for the media with $\rho + p < 0$. We found that in this case the mass of the black hole decreases. It turned out that all black holes in a universe with Big Rip must lose their mass until they completely disappear at the moment of Big Rip.

In [7–9] the authors studied the accretion of scalar quintessence field into a black hole, using the specific quintessence potentials $V(\phi)$ for obtaining the analytical solution for the black hole mass evolution. We use essentially a different approach for the description of DE accretion into a black hole, namely, we model the DE by a perfect fluid with negative pressure.

We assume that the density of the dark energy is sufficiently low so that the metric can be described by Schwarzschild metric. We model the dark energy by a perfect fluid with energy-momentum tensor: $T_{\mu\nu} = (\rho + p)u_\mu u_\nu - p g_{\mu\nu}$, where ρ is the density and p is the pressure of the dark energy and the metric signature is so that for the four-velocity $u_\mu u^\mu = 1$. The integration of the time component of the energy-momentum conservation law $T^{\mu\nu}{}_{;\nu} = 0$ gives the first integral of motion

$$(\rho + p)\left(1 - \frac{2}{x} + u^2\right)^{1/2} x^2 u = C_1, \tag{1}$$

where $x = r/M$, $u = dr/ds$ and $C_1 = const$. Given the equation of state $p = p(\rho)$, we can introduce a function n by the relation $dn/n = d\rho/(\rho + p)$. The function n plays the role of concentration of the particles, though we can use n for the media without introducing any particles. From the conservation of energy-momentum along the velocity $u_\mu T^{\mu\nu}{}_{;\nu} = 0$ we obtain another first integral:

$$\frac{n(\rho)}{n_\infty} u x^2 = -A, \tag{2}$$

where n_∞ is the concentration of the dark energy at infinity. In the case of inflow $u = (dr/ds) < 0$ and the constant $A > 0$. The constant A which determines the flux is fixed for fluids with $\partial p/\partial\rho > 0$. This can be done through finding the critical point following Miclel [10]. For a linear model $p = \alpha(\rho - \rho_0)$ using the obtained integrals of motion we obtain

$$A = \frac{(1 + 3\alpha)^{(1+3\alpha)/2\alpha}}{4\alpha^{3/2}}. \tag{3}$$

For typical sound speeds the constant A has a value around unity.

The black hole mass changes at a rate $\dot{M} = -4\pi r^2 T_0{}^r$ due to the fluid accretion. With the help of (2) this can be expressed as

$$\dot{M} = 4\pi A M^2[\rho_\infty + p(\rho_\infty)]. \tag{4}$$

For phantom energy the relation (4) leads to the diminishing of the black hole mass. That means that in a universe filled with phantom energy, the black holes should melt away. This result is general, it does not depend on the equation of state $p = p(\rho)$, the only condition $p + \rho < 0$ is important.

Now we turn to the problem of the black hole evolution in a universe with Big Rip when a scale factor $a(t)$ diverges at finite time [4, 5]. For simplicity we will take into account only dark energy and will disregard all others forms of energy. The Big Rip solution is realized in the linear model $p = \alpha(\rho - \rho_0)$ for $\rho + p < 0$ and $\alpha < -1$. From the Friedman equations for the linear equation of state model we can obtain: $|\rho + p| \propto a^{-3(1+\alpha)}$. Taking for simplicity $\rho_0 = 0$ we find the evolution of the density of phantom energy in the universe:

$$\rho_\infty = \rho_{\infty,i}\left(1 - \frac{t}{\tau}\right)^{-2}, \tag{5}$$

where

$$\tau^{-1} = -\frac{3(1+\alpha)}{2}\left(\frac{8\pi}{3}\rho_{\infty,i}\right)^{1/2}, \tag{6}$$

$\rho_{\infty,i}$ is the initial density of the cosmological phantom energy and the initial moment of time is chosen so that the "doomsday" comes at time τ. From (5) and (6) it is easy to see that the Big Rip solutions are realized for $\alpha \equiv \partial p/\partial\rho < -1$. In general, the condition $\rho + p < 0$ is not sufficient for the Universe to come to the Big Rip.

We can find the black hole mass evolution in a universe approaching the Big Rip from equations (4) and (5)

$$M = M_i\left(1 + \frac{M_i}{\dot{M}_0\,\tau}\,\frac{t}{\tau - t}\right)^{-1}, \tag{7}$$

where

$$\dot{M}_0 = (3/2)\,A^{-1}|1 + \alpha|, \tag{8}$$

and M_i is the initial mass of the black hole. For $\alpha = -2$ and typical value of $A = 4$ (corresponding to $u_{\rm H} = -1$) we have $\dot{M}_0 = 3/8$. In the limit $t \to \tau$ (i.e. near the Big Rip) the dependence of black hole mass on t becomes linear, $M \simeq \dot{M}_0\,(\tau - t)$. While t approaches τ the rate of mass loss by black hole does not depend on the initial black hole mass and the density of phantom energy: $\dot{M} \simeq -\dot{M}_0$. In other words masses of all black holes in the universe tend to be equal near the Big Rip. This means that phantom energy accretion prevails over Hawking radiation until the mass of black hole is the Planck mass. However, formally all black holes in the universe evaporate completely at Planck time before the Big Rip due to Hawking radiation.

The dark energy is usually modelled by a scalar field ϕ with potential $V(\phi)$. The Lagrangian of a scalar field is $L = K - V$, where K is the kinetic term of a scalar field ϕ and V is the potential. For the standard choice of a kinetic term $K = \phi_{;\mu}\phi^{;\mu}/2$ the energy flux is $T_{0r} = \phi_{,t}\phi_{,r}$. Jacobson [7] found the scalar field solution in Schwarzschild metric for the case of zero potential $V = 0$: $\phi = \dot{\phi}_\infty[t + 2M\ln(1 - 2M/r)]$, where ϕ_∞ is the value of the scalar field at infinity. For this solution we have $T_0{}^r = -(2M)^2\dot{\phi}_\infty^2/r^2$ and correspondingly $\dot{M} = 4\pi(2M)^2\dot{\phi}_\infty^2$.

The energy-momentum tensor constructed from Jacobson solution completely coincides with one for a perfect fluid in the case of ultra-hard equation of state $p = \rho$ under the replacement $p_\infty \to \dot{\phi}_\infty^2/2$, $\rho_\infty \to \dot{\phi}_\infty^2/2$. It is easy to see the agreement of our result (4) for $\dot{M}$ in the case of $p = \rho$ and the corresponding result of [7, 9].

To describe phantom energy, the Lagrangian of a scalar field must have a negative kinetic term [4, 5], for example, $K = -\phi_{;\mu}\phi^{;\mu}/2$ (for the more general case of the negative kinetic term see [11]). In this case the phantom energy flux onto the black hole has the opposite sign, $T_{0r} = -\phi_{,t}\phi_{,r}$, where ϕ is the solution of the same Klein–Gordon equation as in the case of standard scalar field, however with the replacement $V \to -V$. For zero potential this solution coincides with that obtained by Jacobson [7] for a scalar field with positive kinetic term. A Lagrangian with negative kinetic term and $V(\phi) = 0$ does not describe, however, the phantom energy. At the same time, the solution for a scalar field with potential $V(\phi) = 0$ is the same as with a positive constant potential $V_0 = const$, which can be chosen so that $\rho = -\dot{\phi}^2/2 + V_0 > 0$. In this case the scalar field represents the required accreting phantom energy $\rho > 0$ and $p < -\rho$ and provides the decrease of black hole mass with the rate $\dot{M} = -4\pi(2M)^2\dot{\phi}_\infty^2$.

Acknowledgements

This work was supported in part by the Russian Foundation for Basic Research grants 02-02-16762-a, 03-02-16436-a and 04-02-16757-a and the Russian Ministry of Science grant 1782.2003.2.

References

[1] N. Bahcall, J.P. Ostriker, S. Perlmutter, P.J. Steinhardt, *Science*, 284, 1481 (1999).

[2] C.L. Bennett et al., *Astrophys. J. Suppl. Ser.*, 148, 1 (2003).

[3] U. Alam, V. Sahni, T.D. Saini, A.A. Starobinsky, preprint astro-ph/0311364 (2003).

[4] R.R. Caldwell, *Phys. Lett. B*, 545, 23 (2002).

[5] R.R. Caldwell, M. Kamionkowski, N.N. Weinberg, *Phys. Rev. Lett.*, 91, 071301 (2003).

[6] E.O. Babichev, V.I. Dokuchaev, Yu.N. Eroshenko, *Phys. Rev. Lett.*, 93 021102 (2004).

[7] T. Jacobson, *Phys. Rev. Lett.*, 83, 2699 (1999).

[8] R. Bean, J. Magueijo, *Phys. Rev. D*, 66, 063505 (2002).

[9] A. Frolov, L. Kofman, *J. Cosmology Astrop. Phys.*, 5, 9 (2003).

[10] F.C. Michel, *Ap. Sp. Sc.*, 15, 153 (1972).

[11] P.F. Gonzalez-Diaz, *Phys. Lett. B*, 586, 1 (2004).

6

Constraints on Relic Gravitational Waves from CMB and LSS Observations

S. Apunevych and B. Novosyadlyj

Astronomical Observatory of
Ivan Franko National University of Lviv, Lviv, Ukraine
e-mail: apus@astro.franko.lviv.ua, novos@astro.franko.lviv.ua

Abstract

An inflation theory naturally predicts the existence of relic gravitational waves. There are no direct detections of gravitational waves (GW) so far, but some general constraints on the amplitude of relic GWs could be established already. We have determined these constraints on the basis of a set of current observational data on large-scale structure of the Universe, anisotropies of cosmic microwave background and measurements of global parameters of the Universe. Using statistical methods we have investigated the most common inflationary model with 11 parameters, all of them evaluated simultaneously. By this approach we have obtained the statistical upper limit on the possible amplitude of relic GW and so on energy scale of inflation, building the criteria to distinguish the most plausible model among different models of inflation.

Keywords: cosmology, relic gravitational waves, large-scale structure of the Universe, cosmic microwave background

1 Introduction

The current cosmological observations need a quite complex inflationary model with about 11 parameters for their proper interpretation [1–6]. The cosmological parameters include: contents of baryons Ω_b, massive neutrinos Ω_ν and cold dark matter Ω_{cdm}; density parameter for dark energy Ω_{de}, curvature parameter Ω_k and Hubble constant H_0; power spectrum, in its most general form $P_s(k) = A_s k^{n_s}$ for scalar and $P_t(k) = A_t k^{n_t}$ for tensor mode of space-time metric perturbations; the late re-ionization parameter τ_c and biasing parameters for the spatial distribution of astrophysical objects.

By comparison of model predictions with observational data are searched for the best-fit values for the set of parameters. Complex dependence of predictions upon the parameters renders the separate determination of single parameter impossible. So we use statistics to find the best-fit values for the full set of parameters, along with errors. This paper is devoted to the analysis of constraints on power spectrum of tensor mode of space-time metric perturbations (relic GWs) in the framework of inflationary cosmological model for a number of inflation models.

2 Theoretical Framework

GWs are inherent in any physical system with changes in quadrupole distribution of stress-energy density. By now there are no direct successful detections of any kind of sources (see [7] for a review). Cosmological GWs appear to be the most ancient relic in our Universe. They have been emitted in the times of inflation as a result of quantum zero-point oscillations. The average amplitude of relic GWs is determined by energy scales of inflation. Further evolution of metric linear perturbations is described by gauge-invariant formalism wherein relic GWs are represented by tensor mode of space-time metric perturbations (see [2] for a review).

Both tensor and scalar modes of space-time metric perturbations share a common origin. The scalar mode is connected to the density perturbations and later produces the large-scale structure (LSS) of the Universe. The tensor mode does not lead to LSS formation, and cannot be disclosed by observations. The angular power spectrum of CMB temperature fluctuations (C_ℓ) at large scales is produced by both modes. At smaller scales the power spectrum is formed only due to the scalar mode of perturbations. Here we propose to extract the relic GWs contribution using LSS and CMB observations for the scalar mode.

We have developed an analytical approximation for tensor contribution to CMB power spectrum within low multipoles range ($2 < \ell < 20$) to reproduce the spectrum with high accuracy ($\leq 5\%$) for a wide range of parameters values. The form of this approximation is

$$\ell(\ell+1)C_\ell^T = \frac{A(n_t,\Omega_0)}{1+b(n_t,\Omega_0)}\cdot K_\ell(\Omega_{de})\times\exp(-C(n_t,\Omega_0)\cdot\ell^2 + D(n_t,\Omega_0)\cdot\ell), \qquad (1)$$

where the coefficients A, b, C and D are polynomial functions of n_t and $\Omega_0 = 1-\Omega_k$, $K_\ell(\Omega_{de})$ is the amplification factor for the CMB power spectrum caused by dark energy. The analytical form for these coefficients is presented in an Appendix of [8].

For the scalar contribution to CMB power spectrum for the multipoles range $2 < \ell < 20$ we use a semi-analytical approach elaborated in [4]. For the CMB power spectrum at higher multipoles we propose to approximate its whole shape by the positions and amplitudes of the acoustic peaks and dips. We use the analytical approximations for the height and position of the first three peaks and for the position of the first dip in CMB power spectrum in models under consideration [1, 4]. This analytical approach for the CMB power spectrum and for LSS characteristics from [1] allows accurate predictions at low computational cost.

3 Observational Data

The amplitude and position of the 1st and 2nd acoustic peaks and the position of the first dip have been taken from results of the WMAP mission [9], and for the third peak, from results of the BOOMERanG experiment [3]. At large angular scales (lower multipoles range) we use all datapoints from COBE [10] and WMAP [11], except the dipole.

The CMB data are complemented by LSS data: mass function and spatial distribution of rich galaxy clusters, temperature function of X-ray clusters, Ly-α forest absorption lines in distant quasars spectra, peculiar velocities of galaxies. These data can establish the amplitude and shape of the spatial power spectrum of matter density perturbations. Determination of parameters has a problem of degeneracies in the dependence of observational manifestations upon the parameters. This problem was partially eliminated by the addition of other measurements to the observational set: determinations of expansion dynamics by luminosity distance to Supernovae Ia, independent determinations of the Hubble constant, constraints on baryon content from Big Bang nucleosynthesis theory, and content of intergalactic deuterium ([1, 4]).

4 The Method

41 observational datapoints were utilized with 1σ statistical errors. We assume statistical independence of datapoints and normal probability distribution for their errors. We calculated the best-fit values for parameters Ω_{CDM}, Ω_{de}, Ω_ν, N_ν, Ω_b, h, A_s, n_s, A_t, n_t and τ_c, using the Levenberg–Marquardt algorithm for minimization of χ^2-function [12]. The number of neutrino species N_ν takes discrete values 1, 2, 3. The parameters of matter density perturbations power spectrum, A_s and n_s, were treated as free parameters, meaning that no special normalization procedures were applied.

Here we propose a new approximated but fast enough method of marginalization in multiparametric space affordable for common personal computers. To estimate the confidence ranges for parameters we have explored a likelihood function $\mathcal{L} \propto e^{-\frac{1}{2}\chi^2}$ for parameter x_k by the minimization of χ^2 in a subspace of $n-1$ parameters for each fixed value of x_k. This representation can be obtained from the general integral form and the estimated errors virtually coincide with values obtained by integration.

By tradition the amplitude of the tensor mode power spectrum is described by the ratio of tensor T and scalar S contributions to the quadrupole ($\ell = 2$) component of the CMB power spectrum. Here we have used $\mathcal{L}(A_t)$ for the determination of upper bounds and later recalculated it to T/S.

5 Results and Conclusions

We have analyzed the likelihood functions for the observational data and some generic inflation models: with free $n_t = 0$, with flat spectrum of tensor mode ($n_t = 0$), with $n_t = n_s - 1$ and $n_t = 0.5(n_s - 1)$. Our results show that the used data on LSS of the Universe and CMB power spectrum does not require the presence of a considerable amount of relic GWs, the best-fit values of A_t and T/S are close to zero. In the case of free n_t it drops to -1.

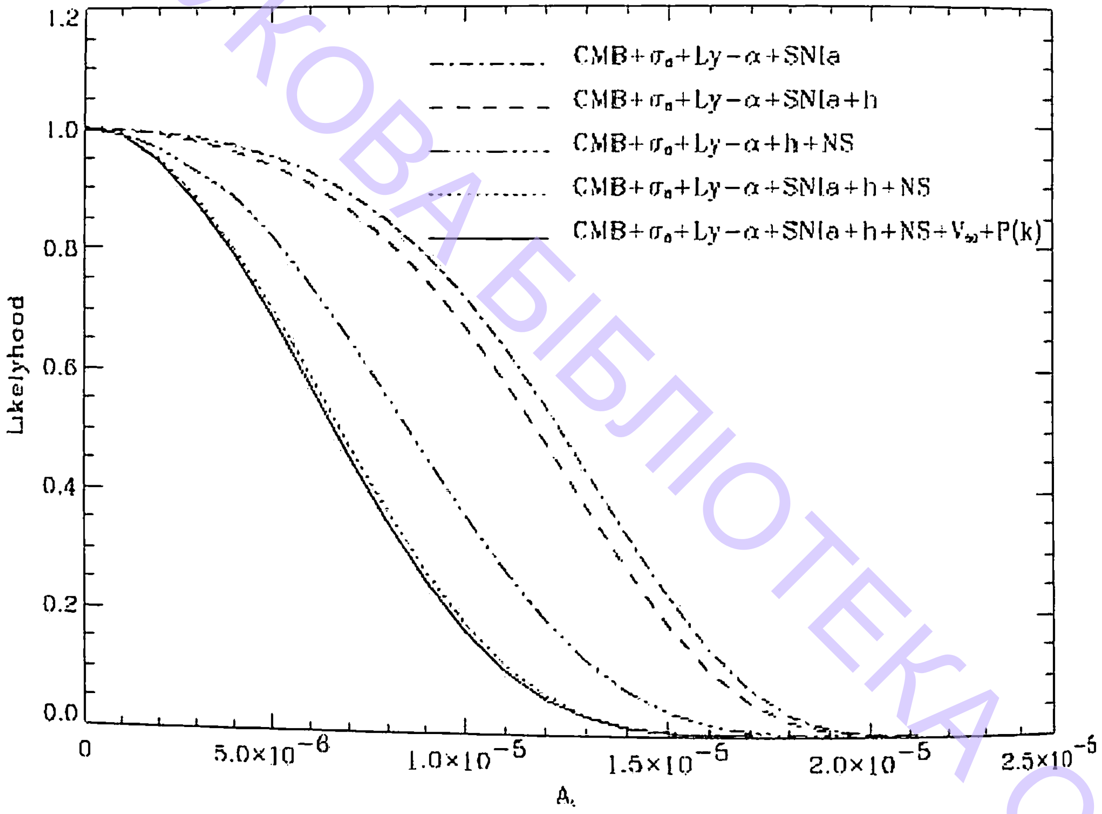


Figure 1: Likelihood functions $\mathcal{L}(A_t) = \exp[-\frac{1}{2}\chi(A_t)]$ for various sets of data.

The obtained upper 2σ constraints are: $A_t^{2\sigma} = 1.9 \cdot 10^{-5}$ for the free n_t and $A_t^{2\sigma} \approx .0 \cdot 10^{-5}$ for the rest cases. Corresponding values for them are $T/S = 0.6$ and 0.18. Figure 1 illustrates that adding the observational data with independent constraints on cosmological parameters, e.g. Hubble constant or LSS data, decreases the level of confidence for the models with high tensor amplitudes. In our previous estimations [4] based on BOOMERanG data this limitation was almost 4 times larger (at 1σ C.L. the ratio was estimated as $T/S = 1.7$ for a free n_t model). So we clearly see the achievements of WMAP.

The accordance of these best-fit values for other parameters with previous determinations [1, 4, 6] is quite satisfactory. The new precise measurements of CMB temperature anisotropies significantly tightened confidence ranges and lowered upper constraints for Ω_ν, τ_c and T/S. Our results agree with other calculations (see, for example, [5, 6]) based on the WMAP data.

References

[1] R. Durrer, B. Novosyadlyj, *Monthly Notic. Roy. Astron. Soc.*, **324**, 560 (2001).

[2] R. Durrer, *Journal of Phys. Studies*, **5**, N 2, 177 (2001).

[3] J.E. Ruhl et al., *Astrophys. J.*, **599**, 786 (2003).

[4] R. Durrer, B. Novosyadlyj, S. Apunevych, *Astroph. J.*, **583**,N 1, 33 (2003).

[5] D.N. Spergel et al., *Astrophys. J. Suppl.Ser.*, **148**, 175 (2003).

[6] M. Tegmark et al., *Phys. Rev. D.*, **69**, 103501 (2004).

[7] B.C. Barish, *American Astronomical Society Meeting*, **205**, 158.04 (2004).

[8] B. Novosyadlyj, S. Apunevych, *Journal of Phys. Studies*, in press, (2005).

[9] L. Page et al., *Astroph. J. Suppl Ser*, **148**, 233 (2003).

[10] C.L. Bennett et al., *Astrophys. J.*, **464**, L1 (1996).

[11] C.L. Bennett et al., *Astroph. J. Suppl. Ser.*, **148**, 97 (2003).

[12] W.H. Press et al., *Numerical recipes in FORTRAN*, Cambridge Univ. Press, Cambridge (1992).

7

Is the Universe Rotating?

P. Flin[1], W. Godlowski[2] and M.Szydlowski[2]

[1] *Pedagogical University, Institute of Physics,Poland*
[2] *Astronomical Observatory of the Jagiellonian University,*
Krakow, Poland

Abstract

George Gamow was the first one to express the idea of the global rotation of the Universe. This idea was developed by Li-Xin Li. In his model the angular momentum of the structure is due to global rotation of the Universe. We demonstrate the existence of a minimum in the angular momentum – mass of galaxy structure relation in this model. Our estimations show that the predicted null angular momentum $|J| = 0$ is for the masses of galaxy groups, while non-vanishing angular momenta for other structures are expected. We point out that this is in agreement with observations. We show that celestial bodies can acquire angular momenta during their formation from the global rotation of the Universe, in accordance with Gamow's suggestion.

Keywords: rotation of the Universe, galaxy clusters; properties

1 Introduction

The global rotation of the Universe is not a proved phenomenon. This possibility has been considered by Gamow [1]. In his pioneering paper he wrote: "We can ask ourselves whether it is not possible to assume that all matter in the visible universe is in a state of general rotation around some centre located beyond the reach of our telescopes?". The dynamical effects of a general rotation of the Universe are presently unimportant, contrary to the early Universe; now the period of rotation must be greater than the Hubble time, which is a simple consequence of the CMB isotropy [2]. The cosmic vorticity depends strongly on the cosmological models and assumptions connected with linearisation of homogeneous, anistropic cosmological models over the isotropic Friedmann Universe, for a flat universe $(\omega/H_0) = 2 \cdot 10^{-5}$ [3].

Li [4] explained the empirical relation between the angular momentum and mass of celestial bodies, usually presented as $J \sim M^{5/3}$, as caused by the global rotation of the Universe. We show [5] that this relation is more complicated and it has a minimum. In this paper we discuss the observational evidence in favour of this relation.

2 Theoretical Consideration

In homogeneous and isotropic models, the Universe with matter may not only expand, but also rotate relative to the local gyroscope. The motion of the matter can be described by the Raychaudhuri equation. This is a relation between the scalar equation Θ, the rotation tensor ω_{ab} and the shear tensor σ_{ab} [4]. The perfect fluid has the stress-energy tensor: $T_{ab} = (\rho + p)u_a u_b + pg_{ab}$, where ρ is the mass density and p is the pressure. The Raychaudhuri equation can be written as:

$$-\nabla_a A^a + \dot{\Theta} + \frac{1}{3}\Theta^2 + 2(\sigma^2 - \omega^2) = -4\pi G(\rho + 3p), \tag{1}$$

where: $A^a = u^b \nabla_b u^a$ is the acceleration vector, while $\omega^2 \equiv \omega_{ab}\omega^{ab}/2$ and $\sigma^2 \equiv \sigma^{ab}\sigma_{ab}/2$ are the scalars of rotation and sheer respectively. Because σ falls off more rapidly than the rotation ω as the Universe expands, it is reasonable to consider such a generalization of the Friedman equation in which only the "centrifugal" term is present, i.e.

$$\frac{\dot{a}^2}{2} + \frac{\omega^2 a^2}{2} - \frac{4\pi G a^2}{3c^2}\epsilon = -\frac{kc^2}{2}, \tag{2}$$

where $\epsilon = \rho c^2$, k is the curvature constant, a a scalar factor and $\dot{a} \equiv \frac{da}{dt}$.

Equation (2) should be completed with the principle of the conservation of the energy-momentum tensor and angular momentum. The conservation of angular momentum of a structure relative to the gyroscopic frames in dust epochs gives:

$$J = kM^{5/3} - lM, \tag{3}$$

where $k = \frac{2}{5}\left(\frac{3}{4\pi\rho_{d0}}\right)^{2/3}\omega_0$, ρ_{do} is the density of (dust) matter in the present epoch, and $l = \beta r_f^2(1 + z_f)^2\omega_0$, r_f is the radius of protostructure, z_f redshift formation, and β is a parameter determined by the distribution of the mass in it [4]. A suitable value for k is 0.4 (in CGS units) [4]. Taking $\rho_{d0} = 1.88 \cdot 10^{-29}\ \Omega\ \mathrm{h}^2\mathrm{g\ cm}^{-3}$ and $h = 0.75$, $\Omega = 0.01$ [6], we obtain $\omega_0 \simeq 6 \cdot 10^{-21}$ rad $\mathrm{s}^{-1} \simeq 2 \cdot 10^{-13}$ rad yr^{-1}.

We show in [5] that the relation (3) exhibits a clear minimum. The minimal mass of the structure:

$$M_{min} = \left(\frac{3l}{5k}\right)^{3/2} = 1.95 r_f^3(1 + z_f)^3\rho_{d0}, \tag{4}$$

does not depend on the present value of ω_0 rotation of the Universe. When $|J| = 0$, the corresponding mass $M_0 = \left(\frac{l}{k}\right)^{3/2} \approx 2.15 M_{min}$, and for less and more massive structures considered the model predicts $|J(M)| \neq 0$. We assumed that galaxy clusters are formed through the collapse of the protostructures. For a protostructure with a diameter of $60Mpc$ and $z_f = 6$, we obtain $M_0 \approx 5 \times 10^{13}\mathcal{M}_\odot$ as an estimate of dust mass for a formed structure with vanishing angular momentum. This gives the total mass of a structure of the order of $10^{14} - 10^{15}\mathcal{M}_\odot$, which is the mass of a small galaxy cluster.

3 Results and Discussion

We choose 18 Tully's galaxy groups, each containing more than 40 galaxies. We investigate the galaxy alignment in each group because the existence of alignment is interpreted as the existence of non-vanish angular momentum. For any galaxy we consider two parameters: the galactic position angle p and the inclination angle i [7]. With the use of these angles two orientation angles are determined: δ – the angle between the normal to the galaxy and the LSC plane, and η – the angle between the projection of this normal onto the LSC plane and the direction toward the LSC centre (seen from the Earth). The distributions of the "supergalatic position angles" P, as well as two angles δ and η were studied using a χ^2, Fourier and autocorrelation tests. From the performed statistical analyses it follows that no alignment of galaxy planes is observed. The analysis of the distribution of galaxy position angles shows non-randomness in one group only. So, for masses of structures close to M_0 we do not find evidence that angular momenta $J(M) \neq 0$, which is in agreement with predictions of the model considered.

Less massive structures are compact groups and galaxy pairs. In compact groups of galaxies, their members rotate along elongated orbits around the gravitational centre of the group [8], which obviously adds some angular momentum to the total angular momentum of the system. The study of paired galaxies showed that the angular momentum in these systems stays mainly in the orbital motion of galaxies [9-11]. These examples showed that for the structures with masses smaller than M_0 there are some indicators of the non-vanishing angular momentum. The study of galaxy orientation in Abell clusters yields different results. Some investigations did not find a galaxy alignment in clusters, disregarding the brightest cluster member [12], while in very rich clusters such as A754 [13] or A1656 [14, 15] non-random distribution of galaxies has been observed.

A lot of work has been done in the study of the alignment of galaxies in superclusters. The results are ambigous, but several independent investigations claimed the existence of a weak alignment of galaxies with respect to the supercluster's main plane. This was the case of the Local [7, 16–18], Hercules [19–21], Coma/A1367 [19-21] and Perseus superclusters [19-21]. From the above considerations it follows that the alignment of galaxies in the Abell clusters is much weaker than in the case of superclusters. There are no observational indications that groups, clusters and superclusters rotate. This means that the angular momentum of those structures depends on the spins of its components. Therefore, the larger, more massive a structure is, the greater is its angular momentum. Even if possible orbital motion is considered, the proportional relation between mass and angular momentum of a structure is still valid. This is in agreement with the fact that it is easier to find the evidence for an alignment for more massive structures.

4 Conclusions

Analysing Tully's groups of galaxies we did not find any evidence for galaxy alignments. Our results also showed that in our sample of galaxy groups, a sign of possible alignment of galaxy plane orientation was observed for the Virgo cluster itself, which is the most massive group studied by us. There is evidence of non-vanishing angular momentum in both smaller and larger structures, but not for galaxy groups. The results of galaxy alignment and spin investigations in pairs, yielded the existence of non-vanishing angular momenta in these structures. We also found evidence for an alignment of galaxies within clusters and superclusters of galaxies. These findings taken together were in agreement with the considered models, in which celestial bodies acquire angular momenta during their formation from the global rotation of the Universe. So, it is possible that George Gamow was also right in this case.

Acknowledgements

PF was partially supported by the Pedagogical University in Kielce grant No 052/*S* and MS by KBN Grant No 2*P*03*B* 00326.

References

[1] G. Gamow, *Nature*, **158**, 549 (1946).

[2] J. Silk, *MNRAS*, **147**, 13 (1970).

[3] J.D. Barrow, R. Juszkiewicz, D.H. Sonoda, *MNRAS*, **213**, 917 (1985).

[4] Li-Xin Li, *GRG*, **30**, 497 (1998).

[5] W. Godlowski, M. Szydlowski, P. Flin, M. Biernacka, *GRG*, **35**, 907 (2003).

[6] P.J.E. Peebles, B. Ratra, preprint astro-ph/0207347 (2002).

[7] P. Flin, W. Godlowski, *MNRAS*, **222**, 525 (1986).

[8] H. Tiersch, O. Martinez, V. Chavushian, O. Yam, H. M. Tovmassian, in Galaxies: *The third dimension ASP Conf. Ser* (2001) (in press).

[9] I.D. Karachentsev, V.A. Mineva, *Sov. Astr. Lett.*, **10**, 105 (1984).

[10] I.D. Karachentsev, V.A. Mineva, *Sov. Astr. Lett.*, **10**, 235 (1984).

[11] V.A. Mineva, *Sov. Astr. Lett.*, **13**, 150 (1987).

[12] D. Trevese, G. Cirimele, P. Flin, *AJ*, **104**, 935 (1992).

[13] W. Godlowski, F. Baier, H.T. Mac Gillivray, *AA*, **339**, 709 (1998).

[14] S. Djorgovski, *ApJ*, **274**, L11 (1983).

[15] G.X. Wu, F.X. Hu, H.J. Su, Y.Z. Liu *AA*, **323**, 317 (1997).

[16] W. Godlowski, *MNRAS*, **265**, 874 (1993).

[17] W. Godlowski, *MNRAS*, **271**, 19 (1994).

[18] N. Kashikawa, S. Okamura, *PASJ*, **44**, 493 (1992).

[19] P. Flin, *MNRAS*, **235**, 857 (1988).

[20] P. Flin, in W.C. Seitter (ed) *Cosmological Aspects of X-Ray Clusters of Galaxies*, Kluwer Academic Press, Dordrecht, p.363, (1994).

[21] P. Flin, *MNRAS*, **325**, 49 (2001).

8

Distribution of Galaxies in Clusters A2333 and A3846

E. Panko[1] and P. Flin[2]

[1]*Nikolaev State University, Astronomical Observatory, Nikolaev, Ukraine*
[2]*Pedagogical University, Institute of Physics, Kielce, Poland*

Abstract

From the Muenster Red Sky Survey galaxy catalogue we extracted two fragments of areas equal to 1.5 square degrees each. We then applied the Voronoi tessellation technique to the galaxy positions. As a result, overdense regions were obtained. We study the influence on this result of the accepted value of the overdensity factor r, describing the contrast between the average density in the investigated field and the extracted overdense regions. The existence of subclustering in both clusters, showing a unimodal structure in A2333 and complex structure in A3846, is discussed as well.

Keywords: galaxy clusters; catalogues, properties

1 Introduction

The main goal of our project is to investigate the properties of galaxy clusters. For this purpose, a large sample of galaxy clusters is needed, extracted in a uniform manner from a homogeneous set of data. The features obtained in such way could help us to better understand nearby galaxy clusters. Moreover, the properties of nearby clusters can be compared with the rapidly enlarging set of clusters located far away from the observer. Such a comparison should allow us to discriminate between the primordial and the evolutionary properties of clusters. At the present stage of our investigation, we are in the process of constructing a catalogue of galaxy clusters. We have decided to use the Muenster Red Sky Survey (MRSS) as our observational basis and the Voronoi tessellation technique for cluster identification [2]. We are now trying to determine the optimal parameters of the procedure. In this paper, we discuss the method of choosing galaxy clusters from the catalogue as well as the difference in cluster identification for various overdensities. We show

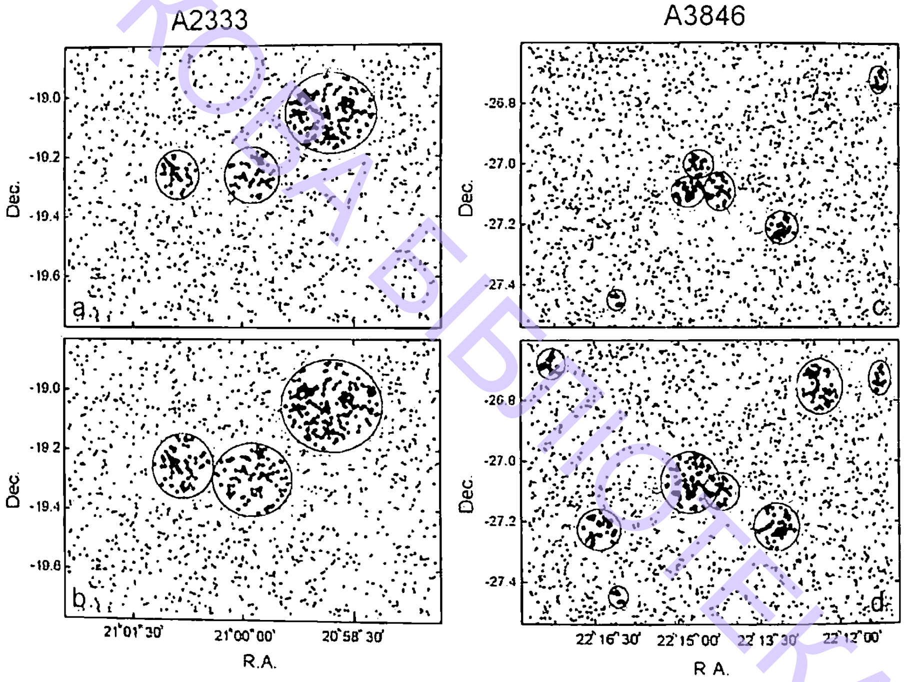


Figure 1: The structure detection in the investigated fields. Upper panels $r = 0.8$, lower panels $r = 0.7$.

how this works in the case of two clusters. The paper is organized as follows: in the next section, the observational data involved in our project are described, while the analysis constitutes the third section. The results, together with their discussion, are presented afterwards, with our conclusions ending the paper.

2 Observational Data

MRSS constitutes the observational basis of our work. It contains the scans of 217 adjoining plates of the ESO Southern Sky Atlas R, covering more than 5000 square degrees. The star/galaxy classification was performed with the help of an automatic procedure using the effective radius, central intensity and surface brightness profiles. The catalogue includes 5.5 million galaxies and is complete up to $r_F = 18^m.3$ [7]. We decided to use this catalogue of galaxies because it is one of the largest set of uniform data and the properties observed in red could be compared with those obtained in the B band such as, e.g., APM data.

3 Method of Analysis

There are three basic cluster detection algorithms available: the matched filter algorithm [3], the adaptive matched filter algorithm [1], and the Voronoi tessellation technique [4, 8]. The latter is completely non-parametric, and therefore can be applied correctly in the case of both spherically symmetric and non-symmetric (elongated and/or irregular) clusters [4, 5].

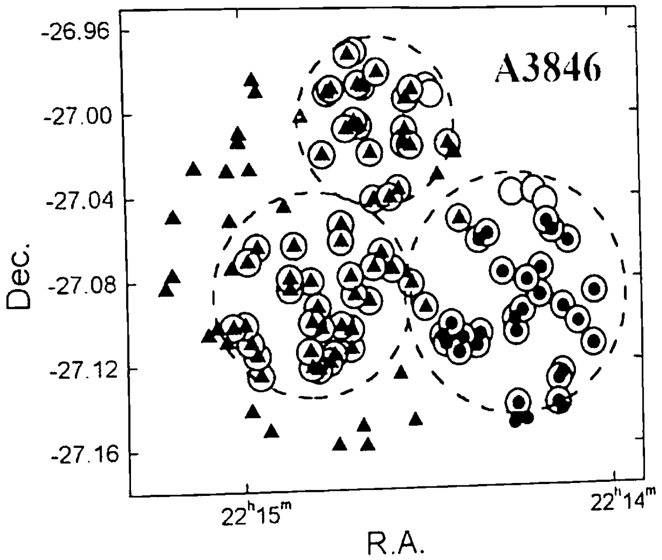


Figure 2: Detailed analysis of the galaxy cluster A3946.

Galaxy positions are input as the seeds for the 2D Voronoi tessellation, and the Voronoi cell around each galaxy is interpreted as the effective area that each galaxy occupies in the projection plane. The inverse value of these areas yield local densities for each galaxy position. This information is then used to define the threshold and select galaxy members that live in highly overdense regions, which we identify as clusters. We do this by calculating the density contrast at each galaxy position as: $\Delta = (\rho - <\rho>)/<\rho> = (<\sigma> - \sigma)/\sigma$, where σ is the area of the Voronoi cells and $<\sigma>$ is the mean area of all cells. We performed the galaxy cluster search using the VGCF (Voronoi Galaxy Cluster Finder) program, an automatic procedure for the identification of galaxy clusters in photometric galaxy catalogues [6]. The application of the procedure yields the following: the coordinates of the centre of the overdense region, the number of galaxies in the region, the estimated number of background galaxies, and the area of the structure.

4 Results and Discussion

For our analysis, we selected two different fragments from the catalogue, containing clusters A2333 and A3846. These are the regions centred at $\alpha = 21^h00^m50^s.40$, $\delta = -19°14'13''.2$ and $\alpha = 22^h14^m19^s.32$, $\delta = -27°04'05''.2$ respectively. In both cases, the size of the analysed field is 1.5 degrees in α and 1 degree in δ. In order to determine which objects are cluster members, the local overdensity r must be determined. We decided to compare the result for overdensity r equal to 0.7 and 0.8. Figure 1 presents the result of the application of the VGCF procedure for A2333 and A3846 respectively. The number of structures extracted from the data depends on the chosen value of the overdensity factor r. The existence of substructure in A3846 also depends on the value of r. Nevertheless, we can conclude that while A2333 is a unimodal structure, A3846 contains a substructure. Figure 2 presents a more detailed analysis of A3846. The galaxies seen in the two structures determined for $r = 0.7$ are denoted by dots and triangles. The galaxies constituting three substructures of the same cluster at $r = 0.8$ are denoted by open symbols. Surprisingly, the structures are not made up from the same galaxies! Some galaxies in the bi-modal structure of the cluster are not counted while determining three substructures denoted by great circles. Also, with a change in the value of r, some new galaxies are considered as clusters members. Moreover, some of these objects counted for larger r, are not cluster members for the smaller r. Thus, galaxies constituting one substructure for a given r can be split into different substructures with changing r.

5 Conclusions

We find the cluster-finding algorithm VGCF [6] to be a very useful and powerful package. It allows us to define structures to be analysed later. It tells us which galaxy belongs to the specific cluster. It is clear that the overdensity factor is of great importance in this study. When it is small, even scarce structures can be regarded as clusters. On the other hand, a large value of the factor r causes that only really large overdensities are regarded as clusters.

Similarly to [6] we think that $r = 0.8$ constitutes a good value, so it will be used in our further work. It can be easily seen that VGCF constitutes a very good tool for studying the existence of substructures. From our analysis it follows that A2333 is a single structure, while substructures exist in A3846.

Acknowledgements

We thank Dr. Renko Ungruhe for granting us the access to his catalogue before publication, and Dr. Dario Fadda for his kind help in finding the VGCF package.

References

[1] J. Kepner, X. Fan, N. Bahcall, J. Gunn, R. Lupton, G. Xu, *ApJ*, **517**, 78 (1999).

[2] E. Panko, P. Flin, in A. Diaferio (ed) *Outskirts of Galaxy Clusters, Proceed. IAU Coll. 195*, (in press).

[3] M. Postman, L. Lubin, J.E. Gunn, J.B. Oke, J.G. Hoessel, D.P. Schneider, J.A. Christensen, *AJ*, **111**, 615 (1996).

[4] M. Ramella, M. Nonino, W. Boschin, D. Fadda, in G. Giuricin, M. Mezzetti, P. Salucci (eds) *Observational Cosmology: The Development of Galaxy Systems, ASP Conf. Ser.*, **176**, ASP, San Francisco, 108 (1999).

[5] M. Ramella et al., *A&A*, **360**, 861 (2000).

[6] M. Ramella, W. Boschin, D. Fadda, M. Nonino, *A&A*, **368**, 776 (2001).

---

[7] R. Ungruhe, *Ph.D. thesis*, Astr. Inst. Univ. Muenster, Germany (1999).

[8] L. Zaninetti, *A&A*, **224**, 345 (1989).

9

Small-Scale Clumps in the Galactic Halo

V.S. Berezinsky[1], V.I. Dokuchaev[2] and Yu.N. Eroshenko[2]

[1] *Laboratori Nazionali del Gran Sasso, Assergi (AQ), Italy*
[2] *Institute for Nuclear Research, RAS, Moscow, Russia*
e-mail: erosh@inr.npd.ac.ru

Abstract

We study the cosmological origin of small-scale DM clumps in a hierarchical scenario with the most conservative assumption of adiabatic Gaussian fluctuations. The main included effect (tidal interaction) results in the formation of a large core in the center of a clump and in tidal destruction of large fraction of the clumps. The mass distribution of clumps has a cutoff at $M_{\rm min}$ due to diffusion of DM particles out of a fluctuation and free streaming at later stage. $M_{\rm min}$ is a model dependent quantity. The enhancement of annihilation signal due to DM clumpiness in the Galactic halo, valid for arbitrary DM particles, is calculated.

Keywords: dark matter, cosmology, clumps, gravitational dynamics

Large scale structures in the universe are developed from primordial density fluctuations $\delta(\vec{x},t) = \delta\rho/\rho$ of cold dark matter (CDM). Gravitationally bound objects are formed and detached from cosmological expansion when fluctuations enter the non-linear stage $\delta \geq 1$. This non-linear stage of fluctuation growth has been studied both by analytic calculations [1] and in numerical simulations [2–4] for Large Scale Structure (LSS). The density profile in the inner part of the formed objects is given by $\rho(r) \propto r^{-\beta}$, with $\beta \approx 1.7$–1.9 in analytic calculations [1], $\beta = 1$ in simulations of Navarro et al. [2] (hereafter NFW) and $\beta = 1.5$ in simulations of Moore et al. [3] and Jing and Suto [4]. In this work we apply this approach to the smallest DM objects in the universe, which we shall call *clumps*. The clumps, being the smallest structures, are produced first in the universe, and that makes our consideration different from LSS formation.

The small-scale clumps are formed only if the fluctuation amplitudes in the spectrum are large enough at the corresponding small scales. The inflation models predict the power-law primeval fluctuation spectrum. If the primeval power-law index $n_p \geq 1$, clumps are formed in a wide range of scales. During the universe expansion the small clumps are captured by the larger ones, and the larger clumps consist of the smaller ones and of continuously distributed DM. The analytic formalism which best describes statistically this hierarchical clustering, is the Press–Schechter theory and its extensions, in particular the "excursion set" formalism developed by Bond et al. However, this theory does not include the important process of the tidal destruction of small clumps inside the bigger ones. We take into account this process and obtain the mass function for the small-scale clumps in the Galactic halo. The theoretical result of this work is that the role of tidal interaction is crucial: a central core in the DM distribution is produced and some fluctuations can be fully disrupted. The LSS are produced much later, when the tidal interactions work in a different regime and their role is less essential. In the case of a power-law spectrum only a small fraction of the captured clumps survives, but even this small fraction is enough to dominate the total annihilation rate in the Galactic halo.

The destruction of clumps by tidal interaction occurs at the formation of the hierarchical structures, a long time before galaxy formation. This interaction arises when two clumps pass near each other and when a clump moves in the external gravitational field of the bigger host to which this clump belongs. In both cases a clump is excited by the external gravitational field, i. e. its constituent particles obtain additional velocities in the c. m. system. The clump is destroyed if its internal energy increase ΔE exceeds the corresponding total energy $|E| \sim GM^2/2R$. In [5] we have calculated the rate of excitation energy production by both aforementioned processes.

The differential fraction of mass, M, in the form of clumps which escape the tidal destruction in the hierarchical objects (survival probability) is found as

$$\xi(n,\nu) \simeq (2\pi)^{-1/2} e^{-\nu^2/2}(n+3)y(\nu), \tag{1}$$

where $\nu = \delta/\sigma(M)$ is the peak-height of a fluctuation with δ being the amplitude and σ variance (dispersion), $n \approx -3$ is the effective spectral index at $t \sim t_{\rm eq}$, and function $y(\nu)$ is given numerically in [5]. $\xi(n,\nu)$ depends weakly on the index of DM distribution in a clump β. Integrating over ν, we obtain $\xi_{\rm int} \simeq 0.01(n+3)$. Since n is close to -3, only a small fraction of clumps about 0.1–0.5% survive the stage of tidal destruction. However, this fraction is enough to dominate the total annihilation rate in the Galactic halo.

The core formation in a fluctuation begins at the linear stage of evolution and continues at the beginning of the non-linear stage. The tidal force diminishes with time as $t^{-4/3}$ (see [5]). Once the core is produced it is not destroyed in the evolution that follows. The stage of core formation continues approximately from $t_{\rm eq}$ to the time of maximal expansion t_s and a little above, when a clump decouples from the expansion of the universe and contracts in the non-linear regime. Soon after this period, a clump enters the hierarchical stage of evolution, when the tidal forces can destroy it, but surviving clumps retain their cores.

The calculations proceed in the following way (see [5] for details and references). The background gravitational field (including that of the host clumps) is expanded in series with respect to the distance from the point with maximum density in a fluctuation. The motion of a DM particle in this field is studied. The spherically symmetric term of the expansion causes radial motion of a particle in the oscillation regime. A spherically non-symmetric term describes the tidal interaction. It results in the deflection of a particle trajectory from the center (point with maximum density). The average deflection (over a statistical ensemble) gives the radius of the core R_c. After statistical averaging, R_c is expressed through the amplitude of the fluctuation $\delta_{\rm eq}$ and the variance $\sigma_{\rm eq}$ (or $\nu = \delta_{\rm eq}/\sigma_{\rm eq}$) as

$$x_c = R_c/R \approx 0.3\nu^{-2} f^2(\delta_{\rm eq}). \tag{2}$$

The fluctuations with ν ~0.5–0.6 have $x_c \sim 1$, i.e. they are practically destroyed by tidal interactions. Most of galactic clumps are formed from $\nu \sim 1$ peaks, but the main contribution to the annihilation signal is given by the clumps with $\nu \simeq 2.5$ for which $x_c \simeq 0.05$.

The mass spectrum of clumps has a low-mass cutoff at $M = M_{\rm min}$, which value is determined by a leakage of DM particles from the overdense fluctuations in the early universe.

CDM particles at high temperature $T > T_f \sim 0.05 m_\chi$ are in thermodynamical (chemical) equilibrium with the cosmic plasma. After freezing at $t > t_f$ and $T < T_f$, the DM particles remain for some time in *kinetic* equilibrium with the plasma, when the temperature of CDM particles T_χ is equal to the temperature of plasma T. At this stage the CDM particles are not perfectly coupled to the cosmic plasma. Collisions between a CDM particle and fast particles of ambient plasma result in exchange of momenta and a CDM particle diffuses in space. Due to diffusion the DM particles leak from the small-scale fluctuations and thus their distribution results in a cutoff at the minimal mass M_D.

When the energy relaxation time for DM particles $\tau_{\rm rel}$ becomes larger than the Hubble time $H^{-1}(t)$, the DM particles get out of kinetic equilibrium. This condition determines the time of kinetic decoupling t_d. At $t \geq t_d$ the CDM matter particles are moving in the free streaming regime and all fluctuations on the scale of free-streaming length λ_{fs} and smaller are washed away. The corresponding minimal mass

$$M_{\rm fs} = (4\pi/3)\rho_\chi(t_0)\lambda_{\rm fs}^3, \tag{3}$$

is much larger than M_D and therefore $M_{\rm min} = M_{\rm fs}$. In [5] we have performed the calculations using two methods: the transparent physical method, based on the description of diffusion and free streaming, and a more formal method based on the solution of the kinetic equation for DM particles starting from the period of chemical equilibrium. Both methods agree perfectly. The minimal mass in the mass distribution is given by the free-streaming mass and for the case of neutralino (bino) as DM particle it is equal to $M_{\rm min} = 1.5 \times 10^{-8} M_\odot$ fot the mass of neutralino $m_\chi =$ 100 GeV and the mass of selectron and sneutrino $\tilde{M} = 1$ TeV. Our calculations agree reasonably well with that of [7, 8], while M_{min} from [6] coincides with our value for M_D.

We calculate the enhancement of the annihilation signal due to the presence of small clumps in the Galactic halo as

$$\eta = (I_{\rm cl} + I_{\rm hom})/I_{\rm hom}, \tag{4}$$

where $I_{\rm hom}$ is the flux due to annihilation of unclumpy DM particles homogeneously distributed in the halo, and $I_{\rm cl}$ is the flux from the clumps. We used standard cosmology with WMAP parameters, different density profiles in the clumps, the distribution of DM clumps over their masses M and radii R, and the distribution of clumps in the galactic halo. The enhancement depends on the nature of the DM particles only through $M_{\rm min}$. The calculations details and corresponding plots can be found in [5].

The enhancements η for primeval power-law index $n_p = 1$ or less is not large: typically it is not larger than factor 2–5 for $M_{\rm min} \sim 10^{-8} M_\odot$. For example, $\eta = 5$ for $n_p = 1.0$ and $M_{\rm min} = 2 \cdot 10^{-8} M_\odot$. It strongly increases at smaller $M_{\rm min}$ and larger n_p. For example, for $n_p = 1.1$ and $M_{\rm min} = 2 \cdot 10^{-8} M_\odot$, enhancement becomes very large, $\eta = 130$ and $\eta = 4 \cdot 10^3$, respectively. Our approach is based on the hierarchical clustering model in which smaller mass objects are formed earlier than the larger ones. This condition is satisfied for objects with mass $M > M_{min} \simeq 2 \cdot 10^{-8} M_\odot$ only if the primordial power spectrum has the value of the power index $n_p > 0.84$.

Finally, we calculated the enhancement due to large clumps with masses in the range 10^8–$10^{10} M_\odot$ and number density distribution $\propto dM/M^2$, which are seen in numerical simulations (see e.g. [3]). This distribution of large clumps accidentally coincides with our derived distribution for small clumps (1). For a NFW profile of DM in the halo and mass fraction of large clumps 0.15 the calculated enhancement of the annihilation signal from large clumps is rather small, $\eta \simeq 1.07$.

Acknowledgements

This work was supported in part by the Russian Foundation for Basic Research grants 02-02-16762-a, 03-02-16436-a and 04-02-16757-a and by the Russian Ministry of Science grant 1782.2003.2.

References

[1] A. V. Gurevich, K. P. Zybin, *Sov. Phys. Usp.*, **165**, 723 (1995).

[2] J. F. Navarro, C. S. Frenk, S. D. M. White, *Astrophys. J.*, **462**, 563 (1996).

[3] B. Moore et al., *Astrophys. J. Lett.*, **524**, L19 (1999).

[4] Y. P. Jing, Y. Suto, *Astrophys. J. Lett.*, **529**, 69 (2000).

[5] V. S. Berezinsky, V. I. Dokuchaev, Yu. N. Eroshenko, *Phys. Rev.*, **D 68**, 103003 (2003).

[6] K. P. Zybin, M. I. Vysotsky, A. V. Gurevich, *Phys. Lett.*, **A 260**, 262 (1999).

[7] C. Schmid, D. J. Schwarz, P. Widerin, *Phys. Rev.*, **D 59**, 043517 (1999);

[8] D. J. Schwarz, S. Hofmann, H. Stocker, *Phys. Rev.*, **D 64**, 083507 (2001).

10

Quantum Scattering on a Cosmic String

Yu.A. Sitenko and N.D. Vlasii

Bogolyubov Institute for Theoretical Physics, Kyiv 03143, Ukraine
e-mail: yusitenko@bitp.kiev.ua, vlasii@bitp.kiev.ua

Abstract

Scattering of a quantum-mechanical particle on a cosmic string is considered. We discuss some peculiarities of the high-frequency scattering which are due to the nonvanishing string tension.

Keywords: cosmic strings, Bohm–Aharonov effect

1 Introduction

Cosmic strings are topological defects which are formed as a result of phase transitions with spontaneous breakdown of symmetries at early stages of evolution of the universe (see, for example, reviews [1, 2]). In general, a cosmic string is characterized by two quantities: flux

$$\Phi = \int\limits_{\rm core} d^2x\sqrt{g}B^3, \tag{1}$$

and tension

$$\mu = \frac{1}{16\pi G}\int\limits_{\rm core} d^2x\sqrt{g}R, \tag{2}$$

where the integration is over the transverse section of the core of the string, B^3 is the field strength which is directed along the string axis, R is the scalar curvature, and G is the gravitational constant. Space-time is conical outside the string core: a surface which is transverse to the axis of the string is isometric to the surface of a cone with a deficit angle equal to $8\pi G\mu$. Such space-times were known a long time ago (M. Fierz, unpublished, see footnote in [3]) and were studied in detail by Marder [4]. In the present context, as cosmological objects and under the name of cosmic strings, they were introduced in seminal works of Kibble [5] and Vilenkin [6]. A cosmic string resulting from a phase transition at the scale of the grand unification of all interactions is characterized by the values of tension $\mu \sim (10^{-7} \div 10^{-6})\,{\rm G}^{-1}$.

The nonvanishing of the string tension leads to various cosmological consequences and, among them, to a very distinctive gravitational lensing effect. A possible observation of such an effect has been reported recently [7], and this has revived an interest towards cosmic strings.

The flux parameter (1) is nonvanishing for the so-called gauge cosmic strings, i.e. strings corresponding to spontaneous breakdown of local symmetries. If tension vanishes ($\mu = 0$), then a gauge cosmic string becomes a magnetic string, i.e. a tube of the magnetic flux lines in flat space-time. If the tube is impenetrable for quantum-mechanical charged particles, then scattering of the latter on the magnetic string depends on flux Φ periodically with period $2\pi e^{-1}$ (e is the coupling constant - charge of the particle). This is known as the Bohm–Aharonov effect [8], which has no analogue in classical physics, since the classical motion of charged particles cannot be affected by the magnetic flux from the region inaccessible for the particles. The natural question is, how the nonvanishing string tension ($\mu \neq 0$) influences scattering of quantum-mechanical particles on the string. Thus, the subject of cosmic strings, in addition to tantalizing phenomenological applications, acquires a certain conceptual importance.

2 S-matrix and Scattering Amplitude

Due to nonvanishing flux Φ and tension μ, the quantum scattering of a test particle by a cosmic string is a highly nontrivial problem. It is impossible to choose a plane wave as the incident wave, because of the long-range nature of the interaction inherent in this problem. A general approach to quantum scattering in the case of long-range interactions was elaborated by Hormander [9]. This approach covers the cases of scattering on a Coulomb center and on a magnetic string ($\mu = 0$), but is not applicable to the case of scattering on a cosmic string ($\mu \neq 0$). Therefore the last case needs a special consideration and a thorough substantiation.

When the effects of the core structure of a cosmic string are neglected and the transverse size of the core is negligible, the field strength and the scalar curvature are presented by two-dimensional delta-functions. Scattering of a quantum-mechanical particle on an idealized (without structure) cosmic string was considered in [10-13]. A general theory of quantum-mechanical scattering on a cosmic string, permitting to take into account the effects of the core structure, was elaborated in [14]. According to this theory, the S-matrix in the momentum representation is

$$S(k,\varphi;\ k',\varphi') = \frac{1}{2}\frac{\delta(k-k')}{\sqrt{kk'}}\left\{\Delta(\varphi-\varphi'+\frac{4G\mu\pi}{1-4G\mu})\exp\left[-\frac{ie\Phi}{2(1-4G\mu)}\right]+\right.$$
$$\left.+\,\Delta\left(\varphi-\varphi'-\frac{4G\mu\pi}{1-4G\mu}\right)\exp\left[\frac{ie\Phi}{2(1-4G\mu)}\right]\right\}+\delta(k-k')\sqrt{\frac{i}{2\pi k}}f(k,\ \varphi-\varphi'), \quad (3)$$

where the initial (k) and final (k′) two-dimensional momenta of the particle are written in polar variables, $f(k,\ \varphi-\varphi')$ is the scattering amplitude, and $\Delta(\varphi) = \frac{1}{2\pi}\sum\limits_{n=-\infty}^{\infty} e^{in\varphi}$ is the angular part of the two-dimensional delta-function. Note that in the case of short-range interaction one has $2\Delta(\varphi-\varphi')$ instead of the figure brackets in Eq. (3). Thus, we can see that, due to the long-range nature of enteraction, even the conventional relation between S-matrix and scattering amplitude is changed, involving now a distorted unity matrix (first term in eq. (3)) instead of the usual one, $\delta(k-k')\Delta(\varphi-\varphi')(kk')^{-1/2}$.

In view of the comparison with the Bohm–Aharonov effect [8], we shall be interested in the situation when the string core is impenetrable for the particle. The scattering amplitude in this case takes the form:

$$f(k,\ \varphi) = f_0(k,\ \varphi) - \sqrt{\frac{2}{\pi ik}}\sum_{n=-\infty}^{\infty}\exp[in\varphi - i(\alpha_n - |n|)\pi]\frac{J_{\alpha_n}(kr_c)}{H^{(1)}_{\alpha_n}(kr_c)}, \quad (4)$$

where r_c is the radius of the string core, $J_\nu(u)$ and $H^{(1)}_\nu(u)$ are the Bessel and the first-kind Hankel functions of order ν,

$$\alpha_n = \left|n - \frac{e\Phi}{2\pi}\right|(1-4G\mu)^{-1}, \quad (5)$$

and

$$f_0(k,\varphi) = \frac{1}{\sqrt{2\pi ik}}\left\{\frac{\exp\left[i[[\frac{e\Phi}{2\pi}]](\varphi+\frac{4G\mu\pi}{1-4G_\mu})-\frac{ie\Phi}{2(1-4G\mu)}\right]}{1-\exp\left[-i\left(\varphi+\frac{4G\mu\pi}{1-4G\mu}\right)\right]}-\right.$$
$$\left.-\ \frac{\exp\left[i[[\frac{e\Phi}{2\pi}]]\left(\varphi-\frac{4G\mu\pi}{1-4G\mu}\right)+\frac{ie\Phi}{2(1-4G\mu)}\right]}{1-\exp\left[-i\left(\varphi-\frac{4G\mu\pi}{1-4G\mu}\right)\right]}\right\} \quad (6)$$

is the amplitude of scattering on an idealized (without structure) cosmic string, $[[u]]$ is the integer part of u. Sum over n in eq. (4) describes the core structure effects. In the low-frequency limit ($k \to 0$) these effects die out, and the differential cross section (i. e. the square of the absolute value of the amplitude) takes form

$$\frac{d\sigma}{d\varphi} = \frac{1}{4\pi k}\left\{\frac{1}{2\sin^2\left[\frac{1}{2}\left(\varphi+\frac{4G\mu\pi}{1-4G\mu}\right)\right]}+\frac{1}{2\sin^2\left[\frac{1}{2}\left(\varphi-\frac{4G\mu\pi}{1-4G\mu}\right)\right]}-\right.$$
$$\left.-\ \frac{\cos\left[\frac{e\Phi}{1-4G\mu}-\left(2[[\frac{e\Phi}{2\pi}]]+1\right)\frac{4G\mu\pi}{1-4G\mu}\right]}{\sin\left[\frac{1}{2}\left(\varphi+\frac{4G\mu\pi}{1-4G\mu}\right)\right]\sin\left[\frac{1}{2}\left(\varphi-\frac{4G\mu\pi}{1-4G\mu}\right)\right]}\right\}. \quad (7)$$

3 Peculiarities of the High-frequency Scattering

In the high-frequency limit ($k \to \infty$) the first term in eq. (4) dies out, and the differential cross section takes the form

$$\frac{d\sigma}{d\varphi} = \frac{1}{2}r_c(1-4G\mu)^2\left|\sum_l\sqrt{\cos[\frac{1}{2}(1-4G\mu)(\varphi-\pi+2l\pi)]}\times\right.$$
$$\left.\times\ \exp\{ie\Phi l - 2ikr_c\cos[\frac{1}{2}(1-4G\mu)(\varphi-\pi+2l\pi)]\}\right|^2, \quad (8)$$

where the summation is over integer l satisfying the condition

$$-\frac{\varphi}{2\pi}-\frac{2G\mu}{1-4G\mu}<l<-\frac{\varphi}{2\pi}+1+\frac{2G\mu}{1-4G\mu}. \tag{9}$$

Note that results (7) and (8) are periodic in the value of flux Φ with period equal to $2\pi e^{-1}$. This feature is common with the scattering on a purely magnetic string ($\mu=0$). The difference is that the Bohm–Aharonov differential cross section in the low frequency limit ($k\to 0$) diverges in the forward direction, $\varphi=0$, while eq. (7) diverges in two symmetric directions, $\varphi=\pm 4G\mu(1-4G\mu)^{-1}$. The difference becomes more crucial in the high-frequency limit ($k\to\infty$). In the $\mu=0$ case one gets

$$\frac{d\sigma}{d\varphi}=\frac{1}{2}r_c\sin\frac{\varphi}{2}, \tag{10}$$

which is the cross section for scattering of a classical pointlike particle by an impenetrable cylindrical shell of radius r_c; evidently, the dependence on fractional part of $e\Phi(2\pi)^{-1}$ disappears in this limit. In the $\mu\neq 0$ case the dependence survives, see eq. (8). In particular, if $0<\mu<(8G)^{-1}$, which is most interesting from the phenomenological point of view, then the cross section at $k\to\infty$ takes the following form in the region of the cosmic string shadow, $-\frac{4G\mu\pi}{1-4G\mu}<\varphi<\frac{4G\mu\pi}{1-4G\mu}$:

$$\begin{aligned}\frac{d\sigma}{d\varphi} &= r_c(1-4G\mu)^2\Bigg(\cos[\tfrac{1}{2}(1-4G\mu)\varphi]\sin(2G\mu\pi)+\\ &+\sqrt{\sin^2(2G\mu\pi)-\sin^2[\tfrac{1}{2}(1-4G\mu)\varphi]}\times\\ &\times\cos\left\{e\Phi+4kr_c\sin[\tfrac{1}{2}(1-4G\mu)\varphi]\cos(2G\mu\pi)\right\}\Bigg).\end{aligned} \tag{11}$$

Integrating eq. (11) over the region of the shadow and the appropriate expression (which is independent of Φ) over the region out of the shadow, we obtain the total cross section in the $k\to\infty$ limit:

$$\sigma_{\rm tot}=2r_c(1-4G\mu). \tag{12}$$

The high-frequency limit is usually identified with the quasiclassical limit. Although this identification is valid for the total cross section, it is found to be invalid for the differential cross section, see eqs.(8) and (11) revealing the periodic dependence on the flux which is a purely quantum effect.

These results are generalized to the case of scattering of a particle with spin.

References

[1] A. Vilenkin, E.P.S. Shellard, *Cosmic Strings and Other Topological Defects*, Cambridge Univ. Press, Cambridge (1994).

[2] M.B. Hindmarsh, T.W.B. Kibble, *Rep. Progr. Phys.*, **58**, 477 (1995).

[3] J. Weber, J.A. Wheeler, *Rev. Mod. Phys.*, **29**, 509 (1957).

[4] L. Marder, *Proc. Roy. Soc. London*, **A252**, 45 (1959).

[5] T.W.B. Kibble, *J. Phys.*, **A9**, 1387 (1976).

[6] A. Vilenkin, *Phys. Rev.*, **D23**, 852 (1981).

[7] M. Sazhin et al., *Mon. Not. Roy. Astron. Soc.*, **343**, 353 (2003).

[8] Y. Aharonov, D. Bohm, *Phys. Rev.*, **115**, 485 (1959).

[9] L. Hormander, *Analysis of Linear Partial Differential Operators IV*, Springer–Verlag, Berlin (1985).

[10] G.'t Hooft, *Commun. Math. Phys.*, **117**, 685 (1988).

[11] S. Deser, R. Jackiw, *Commun. Math. Phys.*, **118**, 495 (1988).

[12] P. de Sousa Gerbert, R. Jackiw, *Commun. Math. Phys.*, **124**, 229 (1989).

[13] Yu.A. Sitenko, *Nucl. Phys.*, **B372**, 622 (1992).

[14] Yu.A. Sitenko, A.V. Mishchenko, *JETP*, **81**, 831 (1995).

III Neutron Stars, Black Holes and High Energy Astrophysics

1

Accretion Disks with Optical Depth Transition and Advection

Yu.V. Artemova[1], G.S. Bisnovatyi-Kogan[1],
I.V. Igumenshchev[2] and I.D. Novikov[3]

[1] *Space Research Institute, Moscow, Russia*
e-mail:julia@mx.iki.rssi.ru,gkogan@mx.iki.rssi.ru
[2] *Laboratory for Laser Energetics, University of Rochester,*
Rochester, USA
e-mail: iigu@lle.rochester.edu
[3] *Niels Bohr Institute, Copenhagen, Denmark, and*
Astro Space Center, Lebedev Physical Institute, Moscow, Russia
e-mail:novikov@nbi.dk

Abstract

We consider the effects of advection and radial gradients of pressure and radial drift velocity on the structure of accretion disks around black holes. We concentrated our efforts on highly viscous disks with large accretion rate. Contrary to disk models neglecting advection, we find that continuous solutions extending from the outer disk regions to the inner edge exist for all accretion rates that we have considered. We show that the sonic point moves outwards with increasing accretion rate. Despite the importance of advection on its structure, the disk remains geometrically thin. Global solutions of advective accretion disks, which describe continuously the transition between optically thick and optically thin disk regions are constructed and analyzed.

Keywords: accretion, accretion disks – black hole physics

1 The Model and the Method of Solution

In this paper we will consider the full set of solutions to the disk structure equations with advection including the optically thick, optically thin and intermediate zones in accretion disks.

We use from now on geometric units with $G = 1$, $c = 1$, r as the radial coordinate scaled to $r_g = M$, and scale all velocities to c. We work with the pseudo-Newtonian potential proposed by Paczyński and Wiita (1980), $\Phi = -M/(r-2)$, that provides an accurate, yet simple approximation to the Schwarzschild geometry. We normalize the accretion rate as $\dot m = \dot M/\dot M_{\rm Edd}$, where $\dot M_{\rm Edd} = L_{\rm Edd} = 4\pi M m_p/\sigma_T$, in our units.

We use the same set of equations, ingredients and boundary conditions in our models as in Artemova et al. (2001), except for changes required by the formulae for radiative flux and radiation pressure to describe correctly the intermediate zone in the accretion disk at high accretion rates. We use the same numerical method (see Artemova et al. (2001)) to solve this modified system of algebraic and differential equations.

The following equations are therefore modified:

The vertically averaged energy conservation equation:

$$Q_{adv} = Q^+ - Q^-, \tag{1}$$

where

$$Q_{adv} = -\frac{\dot M}{4\pi r}\left[\frac{dE}{dr} + P\frac{d}{dr}\left(\frac{1}{\rho}\right)\right], \tag{2}$$

$$Q^+ = -\frac{\dot M}{4\pi} r\Omega\frac{d\Omega}{dr}\left(1 - \frac{l_{in}}{l}\right), \tag{3}$$

$$Q^- = \frac{2aT^4c}{3\kappa\rho h}\left(1 + \frac{4}{3\tau_0} + \frac{2}{3\tau_*^2}\right)^{-1}, \tag{4}$$

are the advective energy, the viscous dissipation rate and the cooling rate per unit surface, respectively, T is the midplane temperature, κ is the opacity, a is the radiation constant and τ_0 is the Thompson optical depth, $\tau_0 = \kappa\rho h$. Here we have introduced the total optical depth to absorption, $\tau_\alpha \ll \tau$,

$$\tau_\alpha = \frac{\epsilon_{ff} + \epsilon_{fb}}{aT_c^4 c}\rho h \simeq 5.2 * 10^{21}\frac{\rho^2 T^{1/2} h}{acT^4}, \tag{5}$$

and the effective optical depth

$$\tau_* = (\tau_0\tau_\alpha)^{1/2}.$$

Where ρ is the density and h is the half-thickness of the disk.

The equation of state for the matter consisting of a mixture of gas and radiation is

$$P_{\rm tot} = P_{\rm gas} + P_{\rm rad}, \tag{6}$$

where the gas pressure is given by

$$P_{\rm gas} = \rho\mathcal{R}T,$$

and $\mathcal{R}$ is the gas constant.

The expression for the radiation pressure is

$$P{\rm rad} = \frac{aT^4}{3}\left(1 + \frac{4}{3\tau_0}\right)\left(1 + \frac{4}{3\tau_0} + \frac{2}{3\tau_*^2}\right)^{-1}.$$

The specific energy of the mixture is

$$\rho E = \frac{3}{2}P_{\rm gas} + 3P_{\rm rad}. \tag{7}$$

Our method allows us to construct a self-consistent solution to the system of equations from very large radii, $r >> 100$, down to the innermost regions of the disk.

2 Results and Discussion

We will now compare the solutions with advection and without it. In the "standard model", for accretion rates $\dot m < \dot m_{cr} = 36$, (for α=0.5 and $M_{BH} = 10M_\odot$, where $\dot m_{cr}$ is the critical accretion rate) solutions always exist that extend continuously from large to small radii. When $\dot m > \dot m_{cr} = 36$ there are no solutions in a range of radii around $r \approx 13$, and therefore no continuous solutions extending from large radii to the innermost disk edge (see detailed discussion by Artemova et al. 1996, where however, a Newtonian potential was used resulting in $\dot m_{cr} = 9.4$ for α=1.0 and $M_{BH} = 10^8 M_\odot$).

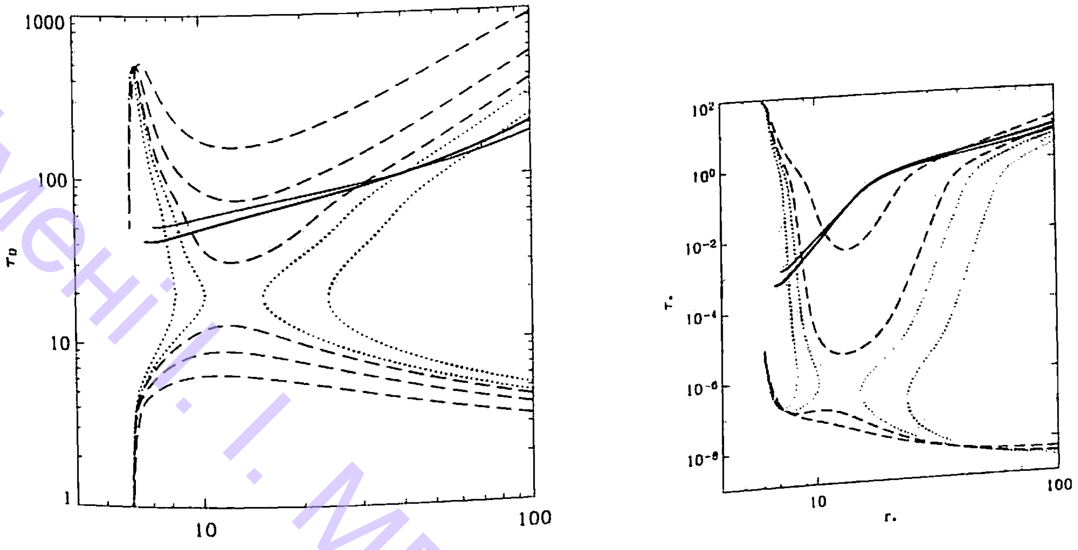


Figure 1: Dependence of the Thomson scattering depth (left panel) and the effective optical depth (right panel) on the radius. Dashed lines correspond to the solutions without advection and $\dot m < \dot m_{cr} = 36$. Dotted lines correspond to the non-physical solutions without advection for $\dot m = \dot m_{cr} = 36$ and $\dot m = 50$ (from the center to the edge of the picture respectively). Solid lines correspond to the solutions with advection and the mass accretion rate higher than the critical one ($\dot m = 36.0$, (upper curve) and $\dot m = 50.0$ (lower curve) from the left edge of the picture).

In Figure 1 we plot the Thomson scattering depth τ_0, (left panel) and the effective optical depth τ_* (right panel) as a function of radius, r, for different $\dot m$ (for $\alpha =$

0.5 and $M_{BH} = 10M_{\odot}$), clearly demonstrating that the solutions to the complete system of disk structure equations including advection and radial gradients have quite different properties at high $\dot{m}$ compared to the solutions of the standard disk model. For $\dot{m} < \dot{m}_{cr} = 36$ (dashed lines), including the gradient terms gives rather small corrections to the standard disk model. When $\dot{m} > 36$, advection becomes essential and for $\dot{m} > \dot{m}_{cr}$ it changes the picture qualitatively. When $\dot{m} > \dot{m}_{cr}$, solutions do exist extending continuously from large radii to the innermost disk region (the solid lines) where the solution passes through a "sonic point". However for small values of the mass accretion rate $\dot{m} < 0.1$ all types of solutions: solutions without advection, optically thick solutions with advection and solutions with advection and optical depth transition are very similar and have the same structure.

In Figure 2 we plot solutions with advection for $\dot{m} = 48 > \dot{m}_{cr}$, $\alpha = 0.5$, and $M_{BH} = 10M_{\odot}$. We compare the solution without optical depth transition (dashed line) and the solution with optical depth transition (solid line). In the standard model, no solution exists in the region around $r \approx 13$, but when advection is included, the structure of the solutions is completely different.

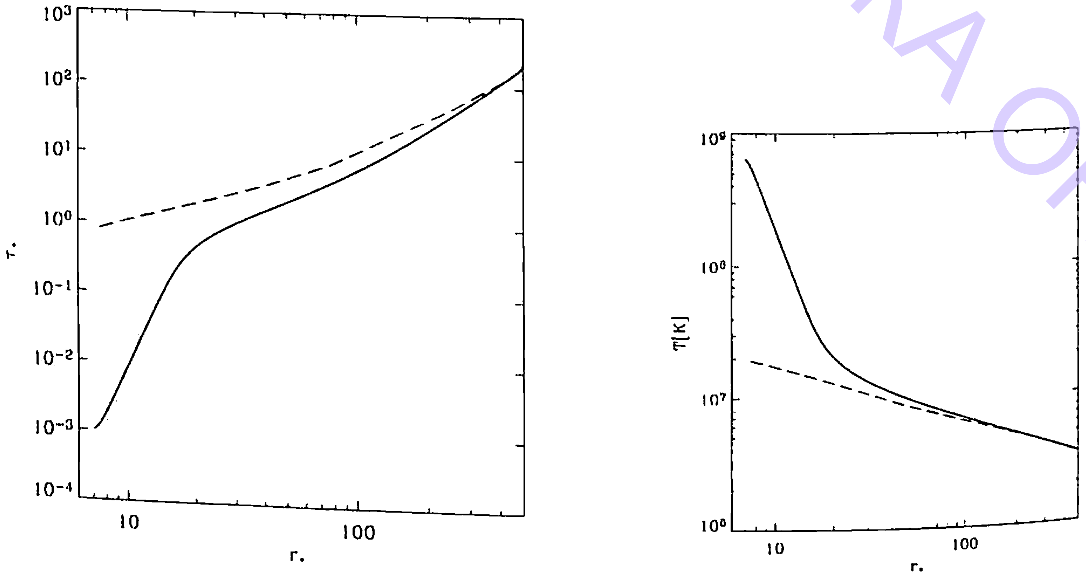


Figure 2: Dependence of the effective optical depth (left panel) and the temperature (right panel) on the radius for $M_{BH} = 10M_{\odot}$, $\dot{m} = 48$, $\alpha = 0.5$. Dashed line corresponds to the solution with advection and without optical depth transition, solid line corresponds to the solution with advection and optical depth transition.

Acknowledgements

Yu.A. and G.B.-K. thank RFFI Grant 02-02-16900 for partial support for this work.

References

[1] B. Paczyński, P.J. Wiita, *Astron.Astrophys*, 88, 23 (1980).

[2] I.V. Artemova, G. Bjornsson, G.S. Bisnovatyi-Kogan, I.D. Novikov, *ApJ*, 456, 119 (1996).

[3] I.V. Artemova, G.S. Bisnovatyi-Kogan, I.V. Igumenschev, I.D. Novikov, *ApJ*, 549, 1050 (2001).

2

Line-driven Winds Near Compact Objects

A.V. Dorodnitsyn

Space Research Institute, Moscow, Russia
e-mail: dora@mx.iki.rssi.ru

Abstract

A general physical mechanism for the formation of line-driven winds at the vicinity of the strong gravitational field of a Schwarzschild black hole is studied. It is demonstrated that if gravitational redshifting is taken into account, the radiation force becomes a function of the local velocity gradient (as in the standard line-driven wind theory) and of the gradient of g_{00}. It is shown that the proposed mechanism could have an important contribution to the formation of line-driven outflows from compact objects.

Keywords: radiation mechanisms: general – stars: mass loss – stars: winds, outflows – galaxies: active

1 Introduction

In [1] (hereafter D1) a mechanism was proposed when line-driven acceleration occurs in the vicinity of a compact object so that the gravitational redshifting can play an important role. The generalization of these studies in the frame of General Relativity (GR) is the problem that we address in this paper. The mechanism that we study is quite general and can be considered to work in any case when there is enough radiation to accelerate plasma and radiation driving occurs in a strong gravitational field. Particularly we discuss winds in active galactic nuclei as they manifest most important properties of the accretion disk plus wind systems keeping in mind however that our treatment allows us to consider their low mass counterparts.

In the standard line-driven wind theory a given parcel of gas sees the matter that is upstream redshifted because of the difference in velocities (assuming that a wind is accelerating gradually). This helps a line to shift from the shadow produced by the underlying matter and to expose itself to the unattenuated continuum. It was shown in D1, that together with Sobolev effect [4] the gravitational redshifting of the photon's frequency should be taken into account when calculating the radiation force. In the case of a strong gravitational field the gradient of the gravitational potential works in the same fashion as the velocity gradient when only Sobolev effect is taken into account, so that the radiation force becomes $g_l \sim (dv/dr + \frac{1}{c}d\phi/dr)$. As it was shown in D1 the gravitational field works in exposing the wind to the unattenuated radiation of the central source. We call such a flow "Gravitationally Exposed Flow" (GEF).

To compare the GEF regime with the standard line-driven wind (SLDW) we choose the simplest possible model: a plasma moving with non-relativistic velocities, spherical-symmetry, stationarity, wind is assumed isothermal, no ionization is treated and the radiation field is assumed to originate form a point source.

2 The Radiation-driven Wind

We consider wind that is accelerated by radiation and the gas pressure in the background gravitational field is described by the Schwarzschild line element: $ds^2 = -h\,dt^2 + \frac{dr^2}{h} + r^2 d\Omega^2$, where $h \equiv (1 - r_g/r)$, $r_g = 2M$ and M is the black hole mass.

The equations of hydrodynamics for the matter interacting with the radiation field read [3]: $T^{\alpha\beta}_{;\beta} - G^\alpha = 0$. The stress-energy tensor for the ideal gas reads: $T^{\alpha\beta} = (P+\rho)u^\alpha u^\beta + Pg^{\alpha\beta}$, where for the total mass-energy density we have $\rho = \rho_0(1+E_i)$, $\rho_0 = m_b n_b$ is the barionic rest-mass density and E_i is the internal energy of the ideal gas per unit mass.

The process of interaction of matter with radiation is described by the four-force density which is given by $G^\alpha = \int_0^\infty d\nu \oint d\Omega\,(\chi_\nu I_\nu - \eta_\nu)\,n^\alpha$, where χ_ν is the opacity, η_ν is the emissivity and I_ν is the specific intensity of photons of the frequency ν, propagating in the direction n^α.

The key ingredient of the CAK (Castor, Abbott and Klein, [2]) theory is that the optical depth (measured on the laboratory frequency of the line transition) can be expressed as a function of the velocity gradient in the flow. The same is true when a photon, emitted somewhere deeply in the potential well, becomes resonant with the line absorption both due to the velocity difference and to the GR effect of redshifting.

3 Optical Depth and Radiation Force in the Sobolev Approximation

A photon emitted at a given radius will suffer a continuous redshifting both gravitational and Doppler and may become resonant with a line transition at some point downstream. We restrict ourselves to the *radially streaming photons* only and assume that they are emitted from a point source. Sobolev optical depth between r_d and infinity is calculated for the field of *radially streaming photons* and can be represented as follows

$$t = \int_{r_d}^{\infty} \varphi(\tilde{\nu} - \nu_0)\chi_{l\nu}\frac{dr}{\sqrt{h}} = \frac{\chi_l v_{th}}{\left|\sqrt{h}\frac{dv}{dr} + c\mathrm{w}\right|}, \tag{1}$$

where $\chi_{l\nu}(\mathrm{Hz\cdot cm^{-1}}) = \Delta\nu_D\chi_l$, $h \equiv g_{00}$, and $\mathrm{w} = \frac{d}{dr}\sqrt{h}$. If no gravitational redshifting is taken into account then the Sobolev optical depth is obtained [2]: $\tau_{sob} = \frac{\chi_l v_{th}}{|dv/dr|} = \chi_l\, l_{sob}$ The radiation pressure can be calculated when summing over the ensemble of optically thin and optically thick lines. Using the parameterization law of CAK we obtain:

$$g_l = \sum_l G_l = \frac{F\sigma_e}{c} k \left(\frac{\sigma_e\,\rho_0 v_{th}}{\sqrt{h}\frac{dv}{dr} + c\mathrm{w}} \right)^{-\alpha}, \tag{2}$$

where σ_e is the electron scattering opacity per unit mass and F is the radiation flux.

4 Gravitationally Exposed Flow

The equation of motion for a stationary, spherically-symmetric, isothermal wind reads [5]:

$$\begin{aligned} &\frac{P+\rho}{h\rho_0}\left(vh\frac{dv}{dr} + \frac{GM}{r^2}\right) + \frac{1}{\rho_0}\frac{dP}{dr} - \frac{\sigma_e}{\sqrt{h}}\frac{L}{4\pi r^2 c} \\ &-\frac{\sigma_e}{\sqrt{h}}\frac{L}{4\pi r^2 c}k\left(\frac{4\pi}{\sigma v_{th}\dot{M}}\right)^{\alpha}\left\{\sqrt{h}\,vr^2\left[\sqrt{h}\frac{dv}{dr} + cw\right]\right\}^{\alpha} = 0. \end{aligned} \tag{3}$$

Here we retain only terms of the order $O(v/c)$. We adopt the equation of state for an ideal gas: $P = \rho_0\mathcal{R}T$, $E_i = 3/2\,\mathcal{R}T$, where $\mathcal{R} = k/m_p$ is the gas constant. For a given position of the critical point r_c we can calculate the value of the velocity and velocity gradient in the critical point. Adjusting the position of the critical point r_c we integrate the equation of motion inward, looking for the solution that satisfies the inner boundary condition. The qualitative picture that has been obtained in D1 is confirmed throughout our calculations. We detect a considerable gain in terminal velocity both in comparison with the CAK case and between the fully relativistic calculations presented here and the semi-classical treatment of D1.

Table 1: Comparison between the GEF and the SLDW solutions. $x_c = r_c/r_g$, $\Delta_c^{\mathrm{GEF}} = x_c^{CAK} - x_c^{\mathrm{GEF}}$. $\Delta_{\mathrm{CAK}}^{\infty} = (v_{\mathrm{GEF}}^{\infty} - v_{\mathrm{CAK}}^{\infty})/v_{\mathrm{CAK}}^{\infty}$.

Model	x_c^{GEF}	Δ_c^{CAK}	$\Delta_{\mathrm{CAK}}^{\infty}$
s_1	15	7.47	0.57
s_2	20	9.95	0.51
s_3	50	25	0.37
s_4	100	50	0.2

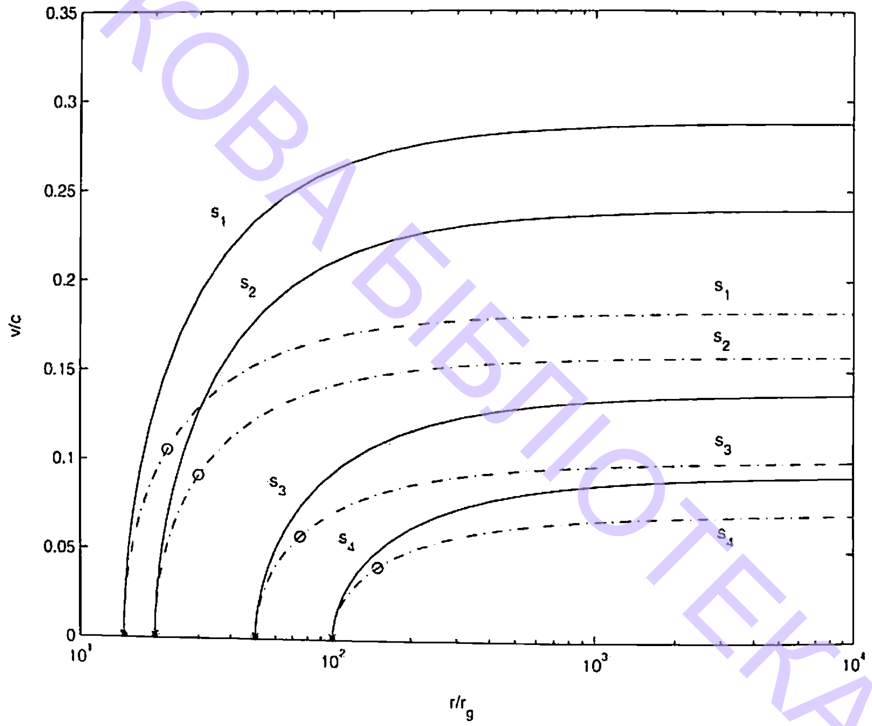


Figure 1: Solutions of the equation of motion. Solid line - GEF solution, dashed line - standard line-driven wind (SLDW) regime. Stars indicate GEF critical points, circles - SDLW critical points (cf. Figure 1 of Dorodnitsyn, 2003). s_1 - s_4 correspond to a different set of solutions.

The position of the GEF critical point is found to be closer to the BH than the SLDW critical point. From Table 1 we see that even when the solution originates (in fact it is determined by the position of the critical point) sufficiently far from the BH there exists a valuable gain in v^{GEF} in comparison with the CAK case. The results of the numerical integration of the equation of motion demonstrate that taking into account gravitational redshifting can result in a wind that is considerably faster than previously assumed on the ground of the CAK theory.

Acknowledgments

This work was supported in part by RFBR grant N 020216900, INTAS grant N 00491, and the Russian Federation President grant MK-1817.2003.02. The author thanks the Theoretical Astrophysics Center (Copenhagen) for hospitality during his visit.

References

[1] A. Dorodnitsyn, *MNRAS*, **339**, 569 (2003).

[2] J. Castor, D. Abbott, R. Klein, *ApJ*, **195**, 157 (1975).

[3] D. Mihalas, B. Mihalas, *Foundations of radiation hydrodynamics*, Oxford Univ. Press, New York (1984).

[4] V. Sobolev, *Moving envelopes of stars*, Harvard Univ. Press, Cambridge (1960).

[5] A. Dorodnitsyn, I. Novikov, *ApJ*, **621**, 932 (2005).

3

Appearance of Subsonic Propellers in a Strong Wind

N.R. Ikhsanov[1,2] and C.-S. Choi[1]

[1] *Korea Astronomy Observatory, Taejon, Korea*
e-mail: ikhsanov@kao.re.kr, cschoi@kao.re.kr
[2] *Pulkovo Observatory, St. Petersburg, Russia*

Abstract

The appearance of magnetized neutron stars in the state of a subsonic propeller, which are situated in a strong ($\dot{M}_c \sim 10^{17}\,\mathrm{g\,s^{-1}}$) wind, is similar to that of anomalous X-ray pulsars (AXPs) in several important aspects. Namely, the rotational periods of these objects are limited to the range of 3–16 s, and the spin-down rates are about a few $\times 10^{-11}\,\mathrm{s^{-1}}$. The emission of stars in this state is dominated by the process of accretion of material onto the stellar surface. The efficiency of the accretion process, which is governed by the reconnection of the magnetic field lines at the magnetospheric boundary, is high enough for the luminosity of the considered objects to be of order $10^{35}\,\mathrm{erg\,s^{-1}}$. Finally, the estimated ages of neutron stars under the conditions of interest are $< 10^6$ yr.

Keywords: accretion, propeller spin-down, supernovae, pulsars, magnetars

1 Introduction

As shown by Davies et al. [4], a neutron star in a massive close binary system can be observed in one of the following four states *ejector*, *supersonic propeller*, *subsonic propeller*, and *accretor*. This classification reflects four different evolutionary stages and four different mechanisms of energy release responsible for the neutron star emission. Stars in the state of *ejector* and *accretor* are identified with the spin-powered and accretion powered pulsars, respectively. Observational identification of stars in the state of *supersonic propeller* has been recently reported by Menou & McClintock [12] and Campana et al. [3]. The state of *subsonic propeller* remains unidentified so far.

Nevertheless, incorporation of the *subsonic propeller* state into the evolutionary tracks of neutron stars is still the only way to understand the origin of long-period

X-ray pulsars in massive close binaries containing main sequence normal companions [5]. Therefore, successful application of the complete spin-down track to a concrete source (see e.g., [6] and [8]) appears to be an indirect theoretical justification of the existence of stars in the state of *subsonic propeller*. A combination of this finding with the results of the spin evolution of neutron stars in the state of *accretor* reported by Bisnovatyi-Kogan [2], provides us with the basic solution of the problem of the origin of long-period pulsars and raises a question about the possibility of their observational identification.

The appearance of *subsonic propellers* in a weak wind, i.e. when the maximum possible mass capture rate by the star from the surrounding medium is $\dot{M}_c \ll 10^{17}\,\mathrm{g\,s^{-1}}$, has recently been discussed by Ikhsanov [9]. Here we envisage a situation in which a neutron star in the state of *subsonic propeller* is situated in a relatively strong wind, i.e. $\dot{M}_c \geq 10^{17}\,\mathrm{g\,s^{-1}}$.

2 The Definition of the Subsonic Propeller State

As shown by Davies & Pringle [5], the magnetosphere of a neutron star in the state of subsonic propeller is surrounded by a hot ($T \simeq T_{\rm ff}$) quasi-stationary plasma envelope, which is extended from the magnetospheric radius, $R_{\rm m}$ up to the Bondi (accretion) radius, R_α, of the star. The value of $R_{\rm m}$ is smaller than that of the corotational radius. The magnetospheric boundary is interchange stable and the process of plasma penetration into the magnetic field is governed by the Bohm diffusion or/and the reconnection of the magnetic field lines (for a discussion see [7]). The formation of the envelope in the first approximation, prevents the surrounding gas from penetrating to within the accretion radius of the star. As the neutron star moves through the interstellar medium with the velocity $V_{\rm rel}$, the interstellar gas overflows the outer edge of the envelope with a rate $\dot{M}_c$. The expected appearance of such an object is shown in Figure 1.

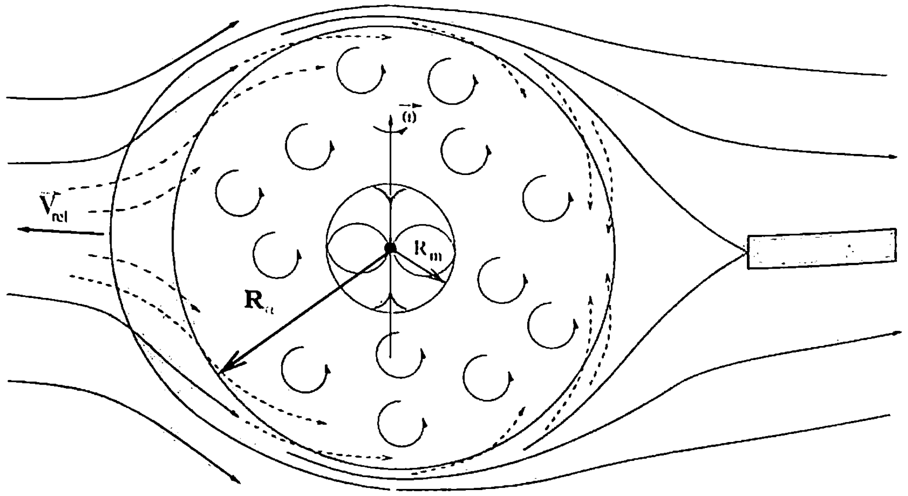


Figure 1: Schematic description of stars in the state of propeller.

3 Period Clustering

According to [6], the spin period of subsonic propellers satisfies the following condition: $P_{\rm cd} \leq P_{\rm s} \leq P_{\rm br}$, where

$$P_{\rm cd} \simeq 3.2\ \mu_{30}^{6/7}\ m^{-5/7}\ \dot{M}_{17}^{-3/7}\ \mathrm{s}, \tag{1}$$

is the period at which the star switches its state from *supersonic* to *subsonic* propeller, and

$$P_{\rm br} \simeq\ 16\ \mu_{30}^{16/21}\ \dot{M}_{17}^{-5/7}\ m^{-4/21}\ \mathrm{s}, \tag{2}$$

is the period at which the state of the star switches to *accretor*. Here m and μ_{30} are the mass and the dipole magnetic moment of the star expressed in units of $M_\odot$, and $10^{30}\,\mathrm{G\,cm^3}$, respectively. $\dot{M}_{17} = \dot{M}_c/10^{17}\,\mathrm{g\,s^{-1}}$.

4 Spin-down Rate

Following [5], the spin-down power of subsonic propellers can be evaluated as

$$L_{\rm ssp} = 8 \times 10^{33}\ \mu_{30}^2\ m^{-1}\ P_5^{-3}\ \mathrm{egr\,s^{-1}}, \tag{3}$$

where $P_5 = P_{\rm s}/5\,\mathrm{s}$. This means that

$$\dot{P} = \frac{P_{\rm s}^3 L_{\rm ssp}}{4\pi^2 I} \sim 2.5 \times 10^{-11}\ \mu_{30}^2\ m^{-1}\ I_{45}^{-1}\ \mathrm{s\,s^{-1}}, \tag{4}$$

where I_{45} is the moment of inertia of the neutron star expressed in units of $10^{45}\,\mathrm{g\,cm^2}$.

5 Accretion Power

Using eqs. (24) from [7] we can limit the accretion power of subsonic propellers to

$$L_{\rm acc} \leq 10^{35}\ m\ \dot{M}_{17}\ R_6^{-1} \left[\frac{\alpha_{\rm r}}{0.1}\right]\ \left[\frac{\lambda_{\rm m}}{0.1 R_{\rm m}}\right]\ \mathrm{erg\,s^{-1}}. \tag{5}$$

Here R_6 is the radius of the neutron star expressed in units of 10^6 cm. $\alpha_{\rm r}$ is the efficiency of the reconnection process, which is normalized according to [13] and $\lambda_{\rm m}$ is the average scale of plasma inhomogeneities at the base of the envelope, which is normalized following [1].

6 Age

The age of subsonic propellers can be expressed as the superposition of pulsar-like, $\tau_{\rm a}$, and supersonic propeller spin-down time scales, $\tau_{\rm c}$, where

$$\tau_{\rm a} \simeq 4 \times 10^5\ \mu_{30}^{-1}\ I_{45}\ \dot{M}_{17}^{-1/2}\ V_7^{-1/2}\ \mathrm{yr}, \tag{6}$$

and

$$\tau_c \simeq 2 \times 10^5 \, \mu_{30}^{-3/7} \, I_{45} \, \dot{M}_{17}^{-11/14} \, V_7^{1/2} \text{ yr.} \tag{7}$$

Here $V_7 = V_{\rm rel}/10^7 \, {\rm cm\,s^{-1}}$.

The life-time of a star in the *subsonic propeller* state is

$$\tau_{\rm d} \simeq 6 \times 10^3 \, \mu_{30}^{-2} \, I_{45} \, P_5 \, m \text{ yr.} \tag{8}$$

Hence, the mean age of subsonic propellers can be limited to $\leq 6.1 \times 10^5$ yr.

6.1 Discussion

What could be the origin of a strong wind? This question is easy to answer if a neutron star is the companion of a massive star in a close binary system. In this case the mass capture rate by the neutron star from the wind of its companion can be as high as $10^{18}\,{\rm g\,s^{-1}}$ (see e.g. [10]). In the opposite case the answer is not trivial. Nevertheless, a consideration of this case appears to be important taking into account that the appearance of subsonic propellers is very similar to that of the AXPs. In particular, a single neutron star can be situated in a strong wind if it is located inside a supernova remnant. As mentioned by Marsden et al. [11], the dense environment of the neutron star in this case can be explained in terms of a "pushed-back" disk. In this light, a discussion on the possibility to explain AXPs phenomenon in terms of the subsonic propeller model has certain theoretical and observational grounds.

Acknowledgements

The work was partly supported by the Korea Science and Engineering Foundation under the grant R01-2004-000-1005-0, the Russian Foundation of Basic Research under the grant 03-02-17223a and the State Scientific and Technical Program "Astronomy".

References

[1] J. Arons, S.M. Lea, *Astrophys. J.*, **210**, 792 (1976).

[2] G.S. Bisnovatyi-Kogan, *Astron. Astrophys.*, **245**, 528 (1991).

[3] S. Campana et al., *Astrophys. J.*, **580**, 387 (2002).

[4] R.E. Davies, A.C. Fabian, J.E. Pringle, *MNRAS* **186**, 779 (1979).

[5] R.E. Davies, J.E. Pringle, *MNRAS* **196**, 209 (1981).

[6] N.R. Ikhsanov, *Astron. Astrophys.*, **368**, L5 (2001).

[7] N.R. Ikhsanov, *Astron. Astrophys.*, **375**, 944 (2001).

[8] N.R. Ikhsanov, *Astron. Astrophys.*, **381**, L61 (2002).

[9] N.R. Ikhsanov, *Astron. Astrophys.*, **399**, 1147 (2003).

[10] V.M. Lipunov, *Astrophysics of neutron stars* Springer-Verlag, Heidelberg, (1992).

[11] D. Marsden, et al., *Astrophys. J.*, **550**, 397 (2001).

[12] K. Menou, J.E. McClintock, *Astrophys. J.*, **557**, 304 (2001).

[13] E.R. Priest, T.G. Forbes, *Magnetic reconnection: MHD theory and applications*, Cambridge University Press (2000).

4

Weak Magnetism Effects in the Direct URCA Processes in Cooling Neutron Stars

L.B. Leinson

Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation, RAS, 142190 Troitsk, Moscow Region, Russia
e-mail: leinson@izmiran.rssi.ru

Abstract

We study the direct URCA processes on relativistic nucleons in degenerate baryon matter under beta equilibrium. The analytic expression obtained for the neutrino energy losses incorporates the effects of nucleon recoil, parity violation, weak magnetism, and pseudoscalar interaction. For numerical testing of our formula, we use a self-consistent relativistic model of the multicomponent baryon matter. The relativistic emissivity is found to be substantially larger than that predicted in the nonrelativistic approach.

Keywords: neutron star, neutrino radiation, direct URCA processes

The most powerful neutrino-emitting processes by which neutron stars can lose their thermal energy are beta decay of neutrons, $n \rightarrow p + l + \bar{\nu}_l$, and the lepton capture on protons, $p + l \rightarrow n + \nu_l$. These reactions, where l is a lepton, either an electron or a muon, known as the direct URCA processes, are a central point of any modern scenarios of evolution of neutron stars because, if operative, they lead to a rapid cooling of the star core.

As applied to the degenerate nucleon matter under beta equilibrium, the relevant neutrino energy losses was derived by Lattimer et al. [1]:

$$Q_{nr} = \frac{457\pi}{10080} G_F^2 C^2 \left(C_V^2 + 3C_A^2\right) T^6 M_n^* M_p^* \mu_e \Theta\left(p_e + p_p - p_n\right), \tag{1}$$

Here[1] G_F - is the Fermi weak coupling constant; $C = \cos\theta_C = 0.973$ - is the

---

[1] We use the system of units $\hbar = c = 1$ and the Boltzmann constant $k_B = 1$.

Cabibbo factor; $C_V = 1$ and $C_A \simeq 1.26$ - are the nucleon weak coupling constants; $\Theta(x) = 1$ if $x \geq 0$ and zero otherwise. To satisfy the momentum conservation in the reaction, the Fermi momenta of the in-medium particles ($p_{\rm n}, p_{\rm p}$, and p_e) must obey the "triangle" condition $p_{\rm p} + p_e > p_{\rm n}$. Together with the charge neutrality of the medium, $n_{\rm p} = n_e$, the above condition gives the threshold proton fraction (about 11%) necessary to open the direct URCA reactions in the medium.

The above calculation has been performed in a nonrelativistic manner, while the central density of the star can be up to eight times larger than the nuclear saturation density, what implies a substantially relativistic motion of nucleons [2], and the relevant neutrino energy losses must be consistent with the used relativistic equation of state.

In the current report, we present the complete relativistic expression for neutrino energy losses caused by the direct URCA processes on nucleons in the strongly degenerate baryon matter under beta equilibrium (see [3], [4] and footnote on p.10 in [5]):

$$Q = \frac{457\pi}{10\,080} G_F^2 C^2 T^6 \Theta\left(p_l + p_{\rm p} - p_{\rm n}\right) F\left(n_b\right),$$

where

$$\begin{aligned} F\left(n_b\right) \quad = \quad & \left(C_A^2 - C_V^2\right) M^{*2} \mu_l \\ & + \frac{1}{2}\left(C_V^2 + C_A^2\right)\left[4\varepsilon_{\rm n}\varepsilon_{\rm p}\mu_l - \left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)\left(\left(\varepsilon_{\rm n} + \varepsilon_{\rm p}\right)^2 - p_l^2\right)\right] \\ & + C_V C_M \frac{M^*}{2M}\left[2\left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right) p_l^2 - \left(3\left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)^2 - p_l^2\right)\mu_l\right] \\ & + C_A\left(C_V + \frac{M^*}{M} C_M\right)\left(\varepsilon_{\rm n} + \varepsilon_{\rm p}\right)\left(p_l^2 - \left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)^2\right) \\ & + C_M^2 \frac{1}{16M^2}\left[8M^{*2}\left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)\left(p_l^2 - \left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)\mu_l\right)\right. \\ & + \left(p_l^2 - \left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)^2\right)\left(2\varepsilon_{\rm n}^2 + 2\varepsilon_{\rm p}^2 - p_l^2\right)\mu_l \\ & \left. - \left(p_l^2 - \left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)^2\right)\left(\varepsilon_{\rm n} + \varepsilon_{\rm p}\right)^2\left(2\varepsilon_{\rm n} - 2\varepsilon_{\rm p} - \mu_l\right)\right] \\ & - C_A^2 M^{*2}\Phi\left(1 + m_\pi^2\Phi\right)\left[\mu_l\left(\left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)^2 + p_l^2\right) - 2\left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right) p_l^2\right]. \end{aligned} \tag{2}$$

In the above, $\varepsilon_{\rm n} = \sqrt{M^{*2} + p_{\rm n}^2}$; $\varepsilon_{\rm n} = \sqrt{M^{*2} + p_{\rm n}^2}$ are kinetic energies of the relativistic nucleons near their individual Fermi surfaces. We replace the magnetic form-factors of the nucleons with anomalous magnetic moments of the proton, λ_p, and the neutron, λ_n, so that $C_M = \lambda_p - \lambda_n \simeq 3.7$. The last term, with

$$\Phi = \frac{1}{m_\pi^2 + p_l^2 - \left(\varepsilon_{\rm n} - \varepsilon_{\rm p}\right)^2}, \tag{3}$$

represents the contribution of the pseudoscalar interaction.

As well as in the nonrelativistic case, the "triangle" condition $p_{\rm p} + p_l > p_{\rm n}$, required by the step-function, is necessary for conservation of the total momentum in the reaction and exhibits the threshold dependence on the proton concentration. The temperature factor, T^6, is also the same as in the nonrelativistic case, while the dependence on the baryon density, $F\left(n_b\right)$, is much more complicated. In the nonrelativistic expression derived by Lattimer et al. [1] the only density dependence was implicitly included by means of the effective nucleon mass and the electron chemical potential. Our relativistic expression implicitly depends on the matter density by means of the effective mass and the Fermi momenta of the nucleons and the in-medium leptons.

To examine numerically the relativistic effects in the direct URCA processes, we employ the Walecka-type relativistic model of baryon matter [6], where the baryons interact via exchange of σ, ω, and ρ mesons. We calculate the matter composition in the mean field approximation, which allows us to calculate, in a self-consistent way, the composition of the matter together with the Fermi momenta, and the effective masses of the baryons [7, 8].

We consider first a simplified model for degenerate nuclear matter of the standard composition consisting of neutrons, protons, electrons, and muons within beta equilibrium. The left panel of Figure 1 shows the composition of neutrino-free matter in beta equilibrium among nucleons, electrons and muons. The fractions of the constituents are shown versus the baryon number density, in units of saturation density, $n_0 = 0.16\ fm^{-3}$. The right panel represents the emissivity of the URCA processes. The solid curve is the emissivity as given by our formula (2). The long-dashed curve demonstrates the energy losses obtained from the relativistic formula by formal setting C_M equal zero and Φ equal zero (see also [9]). This eliminates the weak magnetism and pseudoscalar contributions. The short-dashed curve is the nonrelativistic emissivity, as given by Lattimer et al. [1], see eq. (1). All the emissivities are given in units of $10^{27} T_9^6$ erg cm^{-3}s^{-1}, where the temperature $T_9 = T/10^9$ K. Thus the curves demonstrate how the emissivity varies with the matter density.

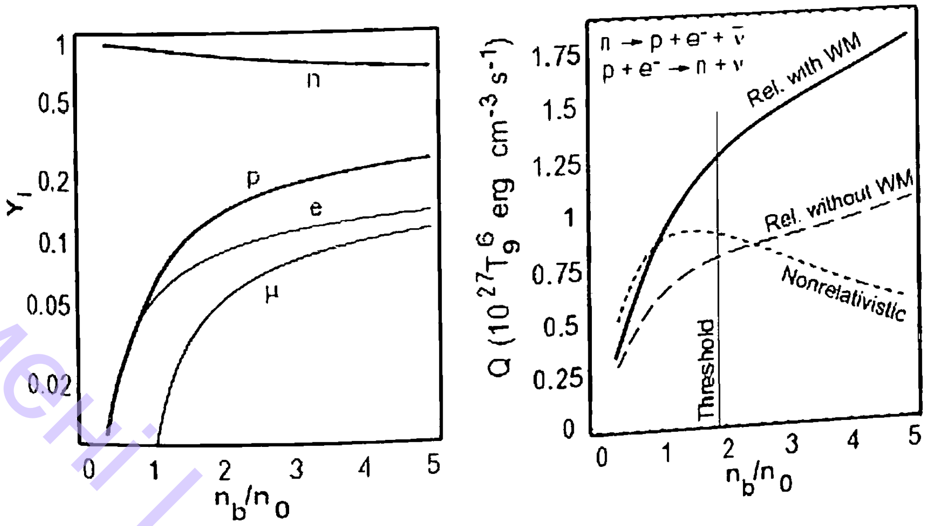


Figure 1: See explanations in the text below.

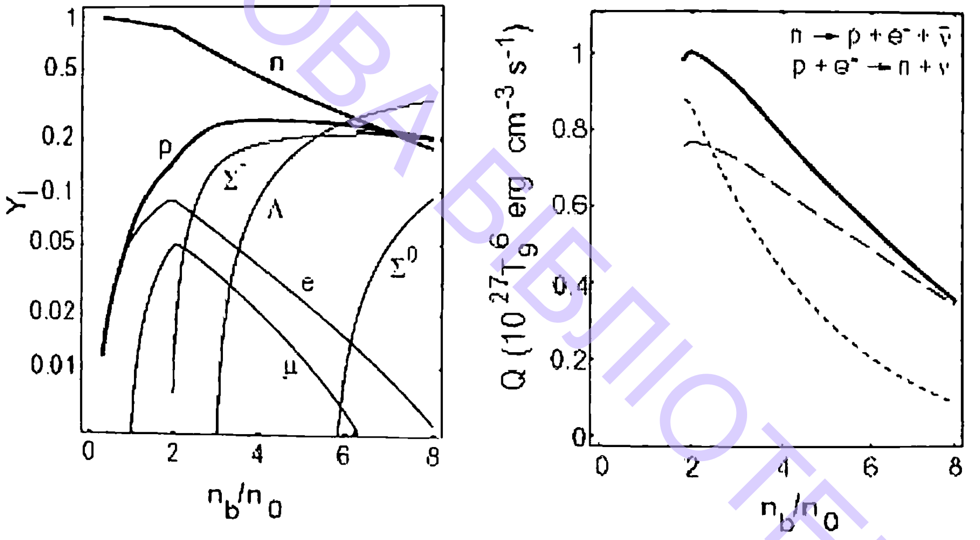


Figure 2: Same as Figure 1 but for the matter composition including hyperons.

We can observe that the relativistic effects dramatically modify the emissivity. The nonrelativistic approximation approaches our result only at densities much smaller than the threshold density, indicated by the vertical line. The direct URCA processes are kinematically forbidden in this density region. Above the threshold, the nonrelativistic formula predicts a decreasing of the emissivity due to decreasing of the effective nucleon mass. In contrast, our formula shows a substantial increase in the neutrino energy losses as we go to larger densities. A comparison of the long-dashed and solid curve demonstrates that due to weak magnetism effects, the relativistic emissivity increases by approximately 65–75%.

To examine the dependence of the energy losses on the matter composition we consider the model, which includes the nucleon and hyperon degrees of freedom. The composition of neutrino-free matter in beta equilibrium among nucleons, hyperons, electrons and muons is shown in the left panel of Figure 2 versus the baryon number density. The right panel represents the emissivity of the direct URCA processes, as given by our formula in comparison with the emissivity calculated without the effects of weak magnetism and pseudoscalar interaction. The short dashed curve is again the nonrelativistic emissivity. We see that the neutrino energy losses caused by the direct URCA processes on nucleons depend essentially on the composition of beta-stable nuclear matter. Appearance of hyperons in the system suppresses the fractions of nucleons and electrons in the medium. Therefore at densities where the number of hyperons is comparable with the number of protons, the relativistic emissivity reaches the maximum and then has a tendency to decrease. The relativistic emissivity, however, is found to be substantially larger than that predicted in the nonrelativistic approach. The weak magnetism effects again notably increase the emissivity.

References

[1] J.M. Lattimer et al., *Phys. Rev. Lett.*, **66**, 2701 (1991).

[2] M. Prakash et al., *Phys. Rep.*, **280**, 1 (1997).

[3] L.B. Leinson, *Phys. Lett.*, **B 532**, 267 (2002).

[4] L.B. Leinson, *Nucl, Phys.*, **A 707**, 543 (2002).

[5] L.B. Leinson, preprint hep-ph/0207116 (2002).

[6] B.D. Serot, *Rep. Prog. Phys.*, **55**, 1855 (1992).

[7] B.D. Serot, J.D. Walecka, *Adv. Nucl. Phys.*, **16**, 1 (1986).

[8] N.K. Glendenning, S.A. Moszkowski, *Phys. Rev. Lett.*, **67**, 2414 (1991).

[9] L.B. Leinson, A. Perez, *Phys. Lett.*, **B 518**, 15 (2001); Erratum: *Phys. Lett.*, **B 522**, 358 (2001).

5

On Hydrodynamic Shear Turbulence in Spectrally Stable Keplerian Disks

J.G. Lominadze[1], A.G. Tevzadze[1], G.D. Chagelishvili[1],
J.-P. Zahn[2] and R.G. Chanishvili[1]

[1] *Abastumani Astrophysical Observatory, Tbilisi, Georgia*
e-mail: j.lominadze@astro-ge.org
[2] *LUTH, Observatoire de Paris, Meudon, France*

Abstract

We report on the hydrodynamic shear turbulence in non-magnetized Keplerian disks drawing attention to a route to hydrodynamic turbulence, which seems to be little known by the astrophysical community, but which was worked out during the last decade by the hydrodynamic community for spectrally stable flows. In this so-called *bypass* concept for the onset of turbulence in shearing flows, vortical perturbations undergo transient growth by extracting energy from the shear (a linear process), thereby reaching an amplitude which is sufficient to allow for non-linear interactions which, by positive feedback, sustain turbulence. This transient growth (that is linear in nature) differs in principle from the well-known nonlinear instability. We outline the fundamental problems of fluid mechanics bearing on the description of plane-parallel smooth shear flows. We present the recent developments in shear flow analysis made by the hydrodynamic community, and the novel bypass concept for the onset of turbulence in shear flows. We describe non-axisymmetric vortex mode perturbations that according to the bypass concept are the most likely to lead to turbulence. We stress the similarity of the turbulent processes in Keplerian disks and in Cartesian flows and conclude that the prevalent scepticism of the astrophysical community about the occurrence of hydrodynamic shear turbulence in such disks is not founded. This report is based on [1, 2].

Keywords: accretion, accretion disks – hydrodynamics – instabilities – turbulence

1 Introduction

The emergence of X-ray astronomy exposed the importance of accretion phenomena in many astrophysical systems, such as binary stars, quasars and active galactic nuclei. The possible mechanisms governing accretion have been schematically understood during the very first years of investigations [3, 4]: the inward transport of matter and the outward transport of angular momentum in accretion disks was ascribed to a turbulent (anomalous) viscosity. The consequent development in understanding the physics of this phenomenon and the causes of that turbulence has been irregular and has taken considerable time. But substantial progress has been achieved in the 1990s with the discovery of a linear instability in magnetized disks [5, 6, 7]. In contrast, the solution of the accretion problem in the non-magnetized case has not yet reached sufficient maturity. Moreover, the very occurrence of turbulence in non-magnetized disks has been questioned by [7, 8].

The reason of this situation is that cylindrical flows with Keplerian profile belong to the class of smooth shear flows, *i.e.* which present no inflection point; it is well known that these flows are spectrally stable, although they may become turbulent in the laboratory. The explanation of this behavior remains one of the fundamental problems of fluid mechanics. But the situation has changed in recent years, with the recognition that the so-called nonmodal approach provides an adequate, and even optimal mathematical formalism for describing the dynamics of perturbations in plane-parallel shear flows. This led to the emergence of the *bypass concept* for the onset of turbulence in spectrally stable flows. This concept has triggered much interest among fluid dynamicists, but it had little impact so far on the astrophysical community. That is why we feel it necessary to present the salient results of that novel approach to the astrophysicists, to demonstrate its relevance for the dynamics of Keplerian disks, and to put together the bibliography on the subject.

Below we outline the fundamental problems of fluid mechanics bearing on the description of plane-parallel smooth shear flows. We present the recent developments in shear flow analysis made by the hydrodynamic community, and the novel bypass concept for the onset of turbulence in shear flows.

2 On the Onset of Turbulence in Spectrally Stable Shear Flows

Here we shall consider smooth shear flows (without inflection point in the velocity profile), as in Keplerian disks. These are linearly stable according to classical fluid mechanics. As shown by Rayleigh [9], the existence of an inflection point in the equilibrium velocity profile is a necessary condition for the occurrence of a linear (spectral) instability in hydrodynamic flows. Thus smooth shear flows (flows without a vorticity extremum) are "relaxed" in this context: they are spectrally stable, meaning that exponentially growing solutions are absent. However, it is well known from laboratory experiments and from numerical simulations that finite amplitude perturbations may cause a transition from laminar to turbulent state above some critical Reynolds number. This has led to the development of the concept of *nonlinear instability* in hydrodynamics (cf. [10–13]). Until about 10 years ago, the predominant view of this laminar-turbulent transition was centered around the slow linear amplification of exponentially growing perturbations (the familiar T-S waves), which modify the flow profile and thereby allow a secondary instability, further nonlinearity and finally a breakdown to turbulent flow. According to this concept of nonlinear instability, the perturbations (and the turbulent state itself) are energetically sustained by nonlinear processes. This concept of nonlinear instability has been borrowed by astrophysicists, who still use it to explain turbulent processes in smooth astrophysical flows, where no spectrally unstable solution is known, and in particular in Keplerian disks flows (see [7]).

However, there are subcritical transition phenomena that can not be attributed to the "T-S pass". During the last decade of the 20th century, another viewpoint emerged in the hydrodynamic community on understanding the onset of turbulence in spectrally stable shear flows, labeled as *bypass transition* (cf. [14–23]). Although the bypass transition scenario involves nonlinear interactions – which intervene once the perturbations have reached finite amplitude – the dominant mechanism leading to these large amplitudes appears to be linear. This concept of the onset of turbulence is based on the *linear transient growth of vortex mode (aperiodic) perturbations*. It implies that the perturbation energy extracted from the basic flow by linear transient mechanisms causes the increase of the total perturbation energy during the transition process. The nonlinear terms are conservative and only redistribute the energy produced by the linear mechanisms.

Thus, according to this concept, the transient growth of perturbations (*i.e.* the linear process) is the key element in the transition to turbulence in spectrally stable flows. The importance of this linear process has been stressed in the titles of several seminal papers:. [19] (*"On the role of linear mechanisms in transition to turbulence"*); [20] (*"A mostly linear model of transition to turbulence"*) and [22] (*"Transient growth: A factor in bypass transition"*).

3 On the Mechanism of Perturbations Amplification in Spectrally Stable Shear Flows

The scheme of investigation of shear flow dynamics implies the following steps: introduction of perturbations into a mean flow, linearization of the governing equations and description of the dynamics of perturbations and flow using the solutions of the initial value problem (*i.e.* following temporal balances in the flow). In principle this can be done, but in practice it is a formidable task. Therefore the mathematical approach was changed. This was done by assuming that the solution is separable in eigenmodes, and then establishing the existence of at least one unstable eigensolution. This approach became canonical in time, and resulted in a shift of ground: attention was directed to the asymptotic stability of the flow and *no attention* was paid to any particular initial value or to the finite time period of the dynamics. Indeed, this phase of the evolution was not thought to have any significance – it was left to speculation. But recently the early transient period for the perturbations has been shown to reveal "rich" and complicated behavior leading to different consequences.

It was found in the 1990s that smooth shear flows are crowded by intense processes of mean flow energy extraction by perturbations, energy exchange between perturbations, etc. even in the linear approximation, while following the classical theory they are spectrally stable and consequently "relaxed." Especially it has been shown that a superposition of decaying normal modes may grow initially, but will eventually decay as time goes on – a new *linear* transient channel of energy exchange between the mean flow and perturbations has appeared. Moreover, it has been shown that transient growth can be significant even for subcritical values of the Reynolds number and that its interplay with nonlinear processes can result in transition to turbulence without any "nonlinear instability" of the flows.

4 On the Nonorthogonality of Linear Operators and the Nonmodal Approach

In fact an exact resonance is not necessary to obtain the transient growth of perturbation. This is the consequence of the non-normal character of operators that describe the linear dynamics of perturbations in flows (see, *e. g.* [24]). The fact that the eigenfunctions of the linearized Navier-Stokes equations are not orthogonal (*i.e.* the operator is non-normal) is enough to allow for solutions that exhibit transient growth, depending on the initial conditions, before finally decaying (see [25, 26]). The operators are highly non-normal for large Reynolds numbers and the transient growth is asymptotically large in Re. (This fact is extremely important in the case of Keplerian accretion disks, where the Reynolds number is literally astronomical: $Re > 10^{10}$.)

These developments provoked a change of paradigm in the study of linear processes in the considered flows. It was the so-called *nonmodal approach* that became extensively used and even canonized in the 1990s on these grounds. As a result, substantial progress in the understanding of the shear flow phenomena has been achieved. The nonmodal analysis – some modification of the initial value problem – implicates the change of independent variables from a laboratory to a moving frame and the study of temporal evolution of spatial Fourier harmonics (SFH) of perturbations without any spectral expansion in time (see, *e. g.* [1, 2, 27, 28]).

5 Bypass Transition to Turbulence

The concept of nonlinear instability in smooth spectrally stable shear flows has undergone substantial revision in the hydrodynamic community. Nowadays the concept of *the bypass transition to turbulence* has become a favorite and is under intensive development.

Let us summarize the main features of this concept:

- the non-orthogonal nature of the linearized Navier–Stokes equations is the formal basis of the transient growth;
- the non-orthogonal nature increases with increasing Reynolds number; thus the operators are highly non-normal for the huge Reynolds numbers of Keplerian disks ($Re > 10^{10}$) and the transient growth is asymptotically large in Re;

- the onset of turbulence and the turbulent state itself in smooth spectrally stable shear flows is supported energetically by *the linear transient growth of vortex mode perturbations* – the key ingredient of the turbulence is the vortex mode (eddy) perturbation, and the key phenomenon is the linear transient growth of the perturbation (and not the nonlinear instability);
- nonlinear processes *do not contribute to any energy growth*, but regenerate vortex mode perturbations that are able to extract shear flow energy. Doing so, nonlinear processes only *indirectly* favor the energy extraction by the vortex mode perturbations.

In this report we wanted to contribute to the revival of hydrodynamic shear turbulence as a possible explanation for the "anomalous" viscosity in non-magnetized Keplerian accretion disks. Specifically, we wished to draw the attention of astrophysicists to the bypass concept. Detail description of the bypass concept for the Keplerian disks is presented in [1].

In papers [1, 2] it is shown that in disk flows there exist vortex mode perturbations that are similar to those which are held responsible for the onset of turbulence in the Cartesian shear flow. The key point of the analysis is the interpretation of the important role of the pressure perturbations in the dynamical processes. In fact, the kinematics and energetics of the vortex mode perturbations are identical in the rotating disk and the Cartesian shear flows in the 2D case (see [1] for the details). The Coriolis force only causes deviation of the pressure perturbations from the Cartesian case. By focusing on the epicyclic motions, and underestimating the action of the pressure terms, one is led to the false conclusion that the Coriolis force suppresses hydrodynamic turbulence in Keplerian flows [7, 8].

This property of the pressure perturbations has been established in [1] in the 2D case. In the more general case of 3D perturbations, the dynamics of vortex mode perturbations is somewhat more complicated, as it has been shown in [2]. But the role of the pressure perturbations is still to counteract the Coriolis force. In the 3D disk case, the wave-number domain where the perturbation undergoes transient growth is smaller in comparison with 3D Cartesian flows, but this is compensated by the very large Reynolds number characterizing astrophysical disks. In [29] revisited accretion disk instability it is also discussed the role of the bypass concept in the disk turbulence. Similar research is done in [30].

Decisive conclusions about the self-sustenance of the turbulent state need to be supported by numerical simulations. The simulations reported [7, 8] have failed to detect hydrodynamic shear turbulence in rotating flows with angular momentum increasing outwards. In [29–31] the possible explanations of this negative result have been discussed. The main reason is probably the lack of spatial resolution, which still prevents from reaching even the Reynolds numbers at which turbulence is detected in the laboratory. At this stage, we hope that the analysis presented here (see also [1, 2, 29–31]) will help to build a fruitful background for the nonlinear numerical analysis of rotating astrophysical shear flows.

Acknowledgements

This work is supported by ISTC grant G-553.

References

[1] G. Chagelishvili et al., *A&A*, **402**, 401 (2003).

[2] A. Tevzadze et al., *A&A*, **407**, 779 (2003).

[3] N. Shakura, R. Sunyaev, *A&A*, **24**, 337 (1973).

[4] J. Pringle, *Ann. Rev. Astron. Astrophys.*, **19**, 137 (1981).

[5] S. Balbus, J. Hawley, *ApJ*, **376**, 214 (1991).

[6] S. Balbus, J. Hawley, *ApJ*, **400**, 610 (1992).

[7] S. Balbus, J. Hawley, *Rev. Mod. Phys.*, **70**, 1 (1998).

[8] S. Balbus et al., *ApJ*, **467**, 76 (1996).

[9] Lord Rayleigh, *Scientific Papers* **1**, 474, Cambridge Univ. Press, Cambridge (1880).

[10] B. Bayly, *Phys. Rev. Lett.*, **57**, 2160 (1986).

[11] B. Bayly et al., *Ann. Rev. Fluid Mech.*, **20**, 359 (1988).

[12] S. Orszag, L. Kells, *J. Fluid Mech.*, **96**, 161 (1980).

[13] S. Orszag, A. Patera, *Phys. Rev. Lett.*, **45**, 989 (1980).

[14] L. Broberg, U. Brosa, *Naturforschung Teil*, **43a**, 697 (1988).

[15] K. Batler, B. Farrel, *Phys. Rev A*, **4**, 1637 (1992).

[16] B. Farrell, P. Ioanou, *Phys. Fluids A*, **5**, 1390 (1993).

[17] S. Reddy, D. Henningson, *J. Fluid Mech.*, **252**, 209, (1993).

[18] T. Gebhardt, S. Grossmann, *Phys. Rev. E*, **50**, 3705 (1994).

[19] D. Henningson, S. Reddy, *Phys. Fluids*, **6**, 1396 (1994).

[20] J. Baggett et al., *Phys. Fluids*, **7**, 833 (1995).

[21] S. Grossmann, *Rev. Mod. Phys.*, **72**, 603 (2000).

[22] E. Reshotko, *Phys. Fluids*, **13**, 1067 (2001).

[23] G. Chagelishvili et al., *JETP*, **94**, 434 (2002).

[24] S. Reddy et al., *SIAM J. Appl. Math.*, **53**, 15 (1993).

[25] W. Criminale, P. Drazin, *Studies in Appl. Math.*, **83**, 123 (1990).

[26] L. Trefethen et al., *Science*, **261**, 578 (1993).

[27] J. Lominadze et al., *Sov. Astr. Lett.*, **14**, 364 (1988).

[28] A. Fridman, *Sov. Astr. Letters*, **15**, 487 (1989).

[29] P. Yecko, *A&A*, **425**, 385 (2004).

[30] O. Umurhan, O. Regev, *A&A*, **427**, 855 (2004).

[31] P-Y. Longaretti, *ApJ*, **576**, 587 (2002).

6

Trans-sonic Propeller Stage

M.E. Prokhorov[1,*] and S.B. Popov[1,2,**]

[1] *SAI MSU, Moscow, Russia*
[2] *University of Padova, Italy*
*e-mail: *mike@sai.msu.ru, **polar@sai.msu.ru*

Abstract

We follow the approach used by Davies and Pringle [1] and discuss the trans-sonic substage of the propeller regime. This substage is intermediate between the supersonic and subsonic propeller substages. At the trans-sonic regime an envelope around a magnetosphere of a neutron star passes through a kind of reorganization process. The envelope in this regime consists of two parts. In the lower part turbulent motions are subsonic. Then at some distance, r_s, the turbulent velocity becomes equal to the sound velocity. During this substage the boundary r_s propagates outwards till it reaches the outer boundary, and so the subsonic regime starts.

We found that the trans-sonic substage is unstable, so the transition between supersonic and subsonic substages goes on at the dynamical time scale. For realistic parameters this time is in the range from weeks to years.

Keywords: compact objects, isolated neutron stars, evolution

1 Introduction

Observational appearances of neutron stars (NSs) are mainly determined by their interactions with the surrounding plasma. The main stages are: *Ejector*, *Propeller* and *Accretor* (see [2]). Normally a NS is born at the stage of ejection, then as the spin period increases the NS passes the propeller and accretor stages. For NSs with large spatial velocities another stage (*Georotator*) can appear.

In a simplified model it is possible to define transitions between stages by comparing external and internal pressure. The external one can be approximated as the ram pressure of a flow of the interstellar medium or as the pressure of matter falling down onto the NS in its gravitational field. The internal one inside r_ℓ (the light cylinder) can be estimated as a pressure of the magnetic field of a NS.

2 Propeller stage with an intermediate regime

In [3, 1] the authors distinguish three substages of the propeller regime. (1) *Very rapid rotator* ($c_s(r_{in}) \simeq r_{in}\Omega \gg v_{ff}$). (2) *Supersonic propeller* ($r_{in}\Omega \gg c_s(r_{in})$). (3) *Subsonic propeller* ($r_{in}\Omega \ll c_s(r_{in})$ and $v_t(r) < c_s(r)$ in $r_{in} < r < r_{out} = r_G$). Here r_{in} and r_{out} – are internal and external radii of the envelope, c_s – the sound velocity, v_t – the turbulent velocity, v_{ff} – the free-fall velocity, $r_G = 2GM/v^2$ – the gravitational capture radius.

We will not discuss the stage of the very rapid rotator here. In this note we focus on the super- and subsonic substages and on the transition between them. At the supersonic stage (a classical propeller) accretion is not possible due to a centrifugal barrier. At the subsonic stage the magnetospheric (Alfven) radius, r_M, is smaller than the corotation radius, r_c, but the accretion does not start because temperature in the envelope is too high. Why does an intermediate (transitional) regime between the supersonic and subsonic substages should exist?

In general supersonic and subsonic regimes do cover all possible values of the rotational velocity, $\Omega = 2\pi/P$. The supersonic propeller formally operates till $r_{in}^{super}\Omega \geq c_s(r_{in}^{super})$, where $r_{in}^{super} = r_G^{2/9} r_M^{7/9}$ [1]. The subsonic regime is on when

$$r_{in}^{sub}\Omega \leq c_s(r_{in}^{sub}) \text{ and } v_t(r) \leq c_s(r) \qquad \text{for} \quad r_{in}^{sub} \leq r \leq r_{out} = r_G\,, \tag{1}$$

here $r_{in}^{sub} = r_M$ [1]. As for the subsonic stage $v_t(r_{in}^{sub}) \simeq r_{in}^{sub}\Omega$ and $v_t/c_s \propto r^{1/3}$, then this regime is valid for $r_{in}^{sub}\Omega \leq c_s(r_{in}^{sub}) \left(r_{in}^{sub}/r_G\right)^{1/3}$.

It is easy to check, that the end of the supersonic substage and the beginning of the subsonic substage both correspond to:

$$\Omega_0 = \sqrt{2GM}\, r_M^{-7/6} r_G^{-1/3}. \tag{2}$$

However, the structure of the envelope on two substages is different, and an intermediate regime during which the structure of the envelope is reorganized is unavoidable. We call this intermediate stage *trans-sonic propeller*.

3 Trans-sonic propeller

We assume that for all substages we can write $c_s(r) \simeq v_{ff}$ [1]. At the trans-sonic substage the envelope is divided into two parts with a boundary at r_s. Processes in the lower part of the atmosphere are similar to those on the subsonic stage: $v_t(r_{in}) \simeq r_{in}\Omega < c_s$. In the outer part of the envelope the physical conditions are similar to those on the supersonic substage.

If the envelope is adiabatic, then for its lower part the polytropic index is equal to $n = 3/2$ and $\rho(r) \propto r^{-3/2}$, $p(r) \propto r^{-5/2}$. Following [1] we assume that the rotational energy of the NS is dissipated at the magnetospheric boundary, and that this energy is transported outwards by turbulence. For such assumptions we have: $v_t(r) \propto r^{-1/6}$ and $\mathcal{M}_t(r) \equiv v_t(r)/c_s(r) \propto r^{1/3}$.

Till $\mathcal{M}_t < 1$, i.e. while $r < r_s$ ($r_{in} < r_s < r_G$) the structure of the envelope is not changed. For large radii turbulence becomes supersonic. Small-scale shock waves are formed, and they quickly dissipate part of the energy, so that the turbulent velocity decreases down to the sound velocity. In the range from r_s up to r_G the envelope structure is different from the lower part: $\mathcal{M}_t(r) \simeq 1$, and $\rho(r) \propto r^{-1/2}$, $p(r) \propto r^{-3/2}$.

To determine parameters of the whole atmosphere it is necessary to calculate the position of the boundary between two parts of the envelope, r_s, and the position of the inner boundary of the lower part, r_{in} (during the transition it decreases from $r_G^{2/9} r_M^{7/9}$ to r_M). For that is necessary to solve the following system of equations:

$$\begin{cases} \dfrac{\mu^2}{8\pi}\dfrac{1}{r_{in}^6} & = \dfrac{1}{2}\dfrac{\dot{M}v_\infty}{4\pi r_G^2}\left(\dfrac{r_G}{r_s}\right)^{3/2}\left(\dfrac{r_s}{r_{in}}\right)^{5/2} \\ \sqrt{\dfrac{2GM}{r_s}} & = \Omega\, r_{in}^{7/6} r_s^{-1/6}. \end{cases} \tag{3}$$

However, the system is degenerate, and each equation can be reduced to: $r_s \propto r_{in}^{-7/2}$. If the following equation is fulfilled:

$$\frac{\mu^2}{8\pi} = \left(\frac{1}{2}\frac{\dot{M}v_\infty}{4\pi r_G^{1/2}}\right)\frac{(2GM)^{3/2}}{\Omega^3} \tag{4}$$

then the system is compatible, i.e. for $\forall r_{in}$ in $r_M < r_{in} < r_M^{7/9} r_G^{2/9}$ $\exists r_s$, that is a solution of (3) (where $r_M = (\mu^2/\dot{M}\sqrt{2GM})^{2/7}$ is the Alfven radius).

The compatibility condition is fulfilled at the end of the supersonic substage ($\Omega = \Omega_0$). Later (during the transition) at any given moment (for any $\Omega < \Omega_0$) left-hand side of eq.(4) is smaller than the right-hand side. It means that the magnetospheric pressure and the envelope pressure have the same dependence on r_{in}, but the latter one is always larger (first equation of (3)).

The energy release during the transition stage is negligible. Estimates for realistic isolated NS parameters give a value $\Delta E \lesssim 10^{30}$ erg.

4 Conclusions

1. The intermediate trans-sonic propeller substage in unstable.
2. The duration of the transition can be roughly estimated as $\sim r_G/v_{ff}$ (from weeks to years for realistic isolated NSs).
3. The spin frequency is nearly unchanged during this transition.
4. The energy release during the transition is small.

Acknowledgements

This work was partially supported by grants RBRF 04-02-16720 and 03-02-16068.

References

[1] R.E. Davies, J.E. Pringle, *MNRAS*, **196**, 209 (1981).

[2] V.M. Lipunov, *Astrophysics of neutron stars*, Springer-Verlag, Berlin – New York (1992).

[3] R.E. Davies, A.C. Fabian, J.E. Pringle, *MNRAS*, **186**, 779 (1979).

[4] N.R. Ikhsanov, *Astron. Astrophys.*, 399, 1147 (2003).

7

Oscillating Neutron Star as a Pulsar

A.N. Timokhin

Sternberg Astronomical Institute,
Universitetskij pr. 13, 119992 Moscow, Russia
e-mail: atim@sai.msu.ru

Abstract

The pulsar "standard model" of a rotating magnetized conducting sphere surrounded by plasma is generalized in its essential parts for the case of an oscillating star. Goldreich-Julian charge density, electromagnetic energy losses as well as polar cap scenario of particle accelerations are considered. Despite similarities, there are substantial differences between magnetospheres of rotating and oscillating stars. Distortion of particle acceleration mechanism in radiopulsars by neutron star oscillations, excited by a strong glitch, for example, is discussed.

Keywords: stars: neutron – stars: oscillations – pulsars: general

1 Introduction

Neutron stars (NS) are amongst the most dense objects in the Universe. The study of their internal structure is of great interest for fundamental physics, because the matter inside the NS is under extreme physical conditions (huge density, pressure, magnetic fields), which can not be reproduced in the laboratory conditions, at least in the near future. The only way to get information about the internal structure of a celestial object is to study its proper oscillations, i.e. doing some kind of seismology. Such investigations would be in principle possible if two things are present: an effective mechanism for excitation of oscillations and a physical process causing changes in the radiation of the celestial object due to its oscillations. NS are observed both as isolated objects (radiopulsars) and as members of binary systems (X-ray binaries). In the second case it would be very difficult to connect some observational features with proper oscillations of the NS itself, because there is a bunch of non-stationary processes connected with accretion of matter onto the NS. Because of this, revealing NS oscillations in the radiation of a radiopulsar seems to be more promising. In the latter case glitches could play a role on the excitation mechanism for neutron star oscillations. Almost the whole pulsar radiation is formed in the magnetosphere, so in order to be able to reveal signatures of NS oscillation in the pulsar radiation it is necessary to know how the pulsar magnetosphere will be distorted by the star oscillations. In this work we consider the magnetosphere of an oscillating neutron star and try to reveal the physical mechanisms able to produce observable signatures of stellar oscillations. We consider the main properties of an oscillating NS magnetosphere and compare them with similar properties of the magnetosphere of a rotating star.

2 Goldreich–Julian Charge Density

As it was first shown by Goldreich and Julian [2], the magnetosphere of a NS is filled with charged particles. In the last thirty years the model of rotating magnetized conducting sphere surrounded by non-neutral plasma became a standard model for radiopulsars. Despite difficulties of this model [1] it is the only model describing, at least qualitatively, the main properties of the pulsar phenomena. According to this model, the charge density of the magnetospheric plasma compensates the longitudinal electric field caused by the NS rotation. The knowledge of this so-called Goldreich–Julian (GJ) charge density, allows us to study the particle acceleration and the emission mechanism. In the case of oscillations, the charge density in the magnetosphere should compensate the longitudinal electric field caused by a motion of the NS surface. So, in order to predict the observational signatures of stellar oscillations it is necessary to generalize the standard pulsar model for the case of arbitrary oscillations of the NS. A formalism of such generalization was proposed by Timokhin, Bisnovatyi-Kogan and Spruit [5].

We have looked for the configuration of the electric field $\vec{\mathcal{E}}$ in the magnetosphere of the oscillating NS that is orthogonal to the NS's magnetic field $\vec{B}$:

$$\vec{\mathcal{E}} \cdot \vec{B} \equiv 0 \tag{1}$$

everywhere in the magnetosphere. The charge density supporting this electric field is the GJ charge density connected with NS oscillations:

$$\rho_{\text{GJ}} = -\frac{1}{4\pi} \nabla \cdot \vec{\mathcal{E}}. \tag{2}$$

In [5] the solution for GJ charge density for the toroidal oscillation modes had been constructed. It turned out that for half of all oscillation modes a smooth charge density distribution in the whole magnetosphere could exist only if a strong current with the density

$$j \simeq \rho_{\text{GJ}} c \left(\frac{c}{\omega r} \right) \tag{3}$$

is flowing along the closed magnetic field lines (ρ_{GJ} is the GJ charge density associated with the *oscillations*, ω is the oscillation frequency, r is the radius in the spherical coordinate system, and c is the speed of light). In this work we have constructed the solution for the GJ charge density for the spheroidal oscillation modes, covering the most interesting physical oscillation modes (p-,g-,r-modes). Similar to the case of toroidal oscillations, only a half of all spheroidal modes have a smooth solution without the presence of a strong current flowing in the magnetosphere. This has an important implication for possible observations of the NS oscillations, because this (non-stationary) current will flow in the closed field lines regions, which are invisible in the radiopulsars. Hence, the radiation could be produced in the initially "dead" regions of the pulsar magnetosphere. On the other hand, this will cause a very rapid damping of the modes producing strong magnetospheric currents.

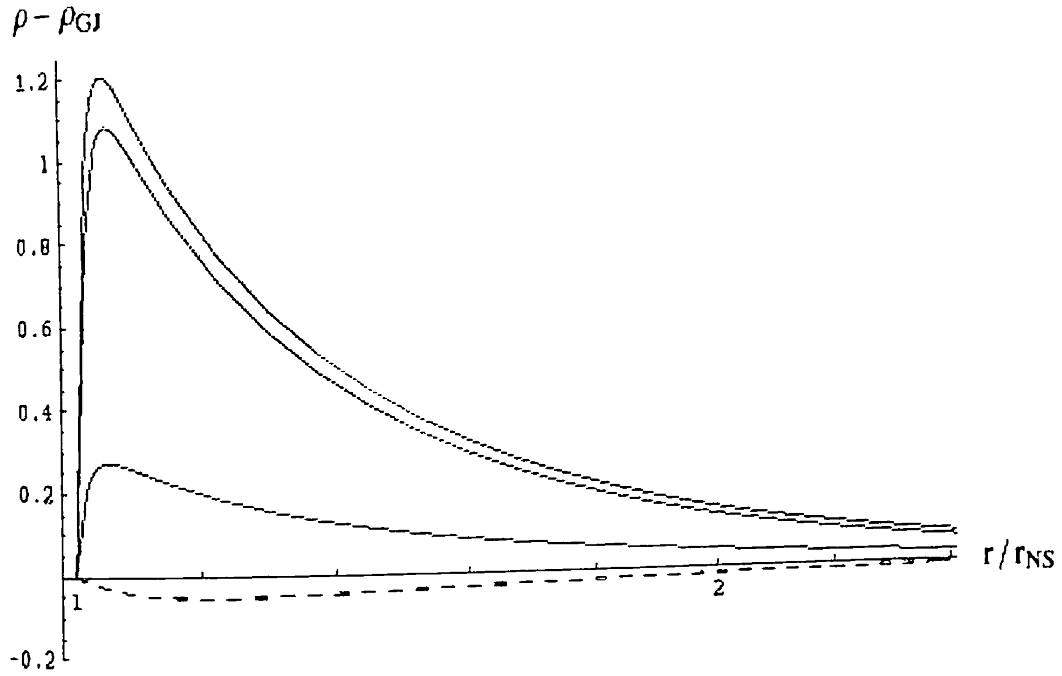


Figure 1: The difference between the charge density of the space charge limited flow and the local GJ charge density along the magnetic field lines in the polar cap region is shown (in arbitrary units) for three spheroidal modes with different (l,m): (64,14) – the upper solid line, (54,14) – the middle solid line, (54,2) – the lower solid line. The same quantity for the aligned rotator is shown by the dashed line. The x-axis shows the distance r from the center of the NS in units of the NS radius r_{NS}. For each oscillation mode this difference is shown along the field line where a local maximum of corresponding ρ_{GJ} is achieved. It was assumed that all modes have the same velocity amplitude which is equal to the linear velocity at the equator of the aligned rotator.

3 Distortion of the Particle Acceleration Mechanism in the Pulsar

In the open field line regions near the magnetic poles the GJ charge density associated with the oscillations for both toroidal and spheroidal modes with $l > 1$ decreases with r more rapidly than the density of the particles flowing away from the polar cap. Moreover, the difference between the charge density of the outflowing particles and the GJ charge density increases with increasing l and m (see Figure 1).

In the most popular model of the space charge limited flow, where particles freely escape the surface of the NS [4], this would lead to a formation of a *"decelerating"* electric field, accelerating particles back to the NS. If this field is of compatible strength with the *accelerating* field due to the pulsar rotation, then for half the oscillation period the resulting accelerating electric field will be stronger or weaker than in the unperturbed pulsar, when the rotational and oscillation GJ charge densities have different and the same signs respectively. This will lead to variations in both the spectrum and the intensity of the pulsar radiation after a strong glitch. For the oscillation modes with a large l the corresponding GJ charge density decreases more steeply with r than the GJ charge density of the rotating star [3]. Because of this and of an increase of $\rho_{\rm GJ}$ with (l, m), the decelerating electric field due to the star oscillations with l, m large enough, could be of a strength compatible with the rotational accelerating field even for the surface oscillation velocity much lower than the velocity of rotation. If the decelerating electric field is stronger than the accelerating field due to NS rotation (such a, rather artificial, case is shown in Figure 1), then on the field lines close to the local maxima of $\rho_{\rm GJ}$ the particle acceleration will be periodically suppressed. To be able to make quantitative predictions about the threshold of the oscillation velocity amplitude, above which the oscillation of the NS could be detected in the pulsar radiation pattern, we need to perform a detailed investigation of the electromagnetic cascade development in the polar cap region. This is outside the scope of the current work and will be the subject of further investigations.

4 Electromagnetic Energy Losses of an Oscillating Star

The oscillating star loses the energy through the electric current flowing across the magnetic field lines in the open field line regions. Accepting as an estimation for the current density $j \simeq \rho_{\rm GJ} c$ and considering self-consistently the position of the last closed field line we can estimate the electromagnetic energy losses of an oscillating star (see [5], where a detailed consideration for the toroidal modes is given). In contrast to the case of the rotating star, where electromagnetic energy losses due to the current flow are well described by the magnetodipolar formula, in the case of a star oscillating in the modes with $l > 1$, the energy losses have another dependence on the oscillation frequency than the energy losses in vacuum, and depend on the oscillation velocity magnitude. For a rotating oscillating star, where the position of the last closed field line is set by the rotation, the damping time of the oscillations also differs substantially from the estimations based on the vacuum formulas.

Acknowledgements

This work was partially supported by the Russian Federation President Grant Program, grants MK-895.2003.02, NSh-388.2003.2 and RFBR grant 04-02-16720.

References

[1] V. S. Beskin, *Physics Uspekhi*, **167**, 689 (1997).

[2] P. Goldreich, W. H. Julian, *ApJ*, **157**, 869 (1969).

[3] A. G. Muslimov, A. I. Tsygan, *MNRAS*, **255**, 61 (1992).

[4] E. T. Scharlemann, J. Arons, W. M. Fawley, *ApJ*, **222**, 297 (1978).

[5] A. N. Timokhin, G. S. Bisnovatyi-Kogan, H. C. Spruit, *MNRAS*, **316**, 734 (2000).

8

Spherical Accretion to a Magnetized Star in the "Propeller" Regime

O.D. Toropina[1], M.M. Romanova[2], Yu.M. Toropin[3] and
R.V.E. Lovelace[4]

[1] *Space Research Institute, Moscow, Russia*
e-mail: toropina@iki.rssi.ru
[2] *Department of Astronomy, Cornell University, USA*
e-mail: romanova@astro.cornell.edu
[3] *CQG International Ltd, Moscow, Russia*
e-mail: ytoropin@cqg.com
[4] *Department of Astronomy, Cornell University, USA*
e-mail: rvl1@cornell.edu

Abstract

This work investigates spherical accretion to a rotating magnetized star in the "propeller" regime using axisymmetric resistive magnetohydrodynamic simulations. In this regime the accreting matter tends to be expelled from the equatorial region of the magnetosphere where the centrifugal force in the matter rotating with the star exceeds the gravitational force. This regime is predicted to occur if the magnetospheric radius is larger than the corotation radius and smaller than the light cylinder radius.

Keywords: MHD, neutron stars, accretion, magnetic field, propeller regime

1 Introduction

Rotating magnetized neutron stars pass through different stages in their evolution [5, 7]. Initially, a rapidly rotating ($P \leq 1$s) magnetized neutron star is expected to be active as a radiopulsar. The star spins down owing to wind of the magnetic field and relativistic particles from the region of the light cylinder r_L [3]. However, after the neutron star spins-down sufficiently, the light cylinder radius becomes larger than the magnetospheric radius r_m where the ram pressure of the external matter equals the magnetic pressure in the neutron star's dipole field. The relativistic wind is then suppressed by the inflowing matter [8]. The external matter may come from the wind of a binary companion or from the interstellar medium for an isolated neutron star. The centrifugal force in the equatorial region at r_m is much larger than the gravitational force if r_m is much larger than the corotation radius r_{cor}. In this case the incoming matter tends to be flung away from the neutron star by its rotating magnetic field. This is the so called "propeller" stage of evolution [2, 4].

The "propeller" stage of evolution, though important, is still not well-understood theoretically. We discuss results of an axisymmetric two-dimensional resistive MHD simulation of the accretion onto a rotating magnetized star in the "propeller" regime. We treat the case when matter flows spherically with Bondi accretion rate [1]. Bondi accretion to a non-rotating and slowly rotating star was investigated by Toropin et al. [9] and Toropina et al. [10]. The investigation of the accretion onto a rotating star in the "propeller" regime was started by Romanova et al. [6].

2 The model

We simulate the plasma flow in the propeller regime using an axisymmetric, resistive MHD code. The code incorporates the methods of local iterations and flux-corrected-transport. The code is described in our earlier investigation of Bondi accretion to the non-rotating magnetized star [9, 10]. The equations for the resistive MHD are

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\ \mathbf{v}) = 0\ , \tag{1}$$

$$\rho\frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \frac{1}{c}\mathbf{J}\times\mathbf{B} + \mathbf{F}^g\ , \tag{2}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B}) + \frac{c^2}{4\pi\sigma}\nabla^2\mathbf{B}\ , \tag{3}$$

$$\frac{\partial(\rho\varepsilon)}{\partial t} + \nabla\cdot(\rho\varepsilon\mathbf{v}) = -p\nabla\cdot\mathbf{v} + \frac{\mathbf{J}^2}{\sigma}\ . \tag{4}$$

We assume axisymmetry ($\partial/\partial\phi = 0$), but calculate all three components of $\mathbf{v}$ and $\mathbf{B}$. The equation of state is taken to be that of an ideal gas, $p = (\gamma-1)\rho\varepsilon$, with specific heat ratio $\gamma = 7/5$. The equations incorporate Ohm's law $\mathbf{J} = \sigma(\mathbf{E}+\mathbf{v}\times\mathbf{B}/c)$, where σ is the electrical conductivity. The associated magnetic diffusivity, $\eta_m \equiv c^2/(4\pi\sigma)$, is considered to be a constant within the computational region. In the equation (2) the gravitational force $\mathbf{F}^g = -GM\rho\mathbf{R}/R^3$, is due to the central star.

We use a cylindrical inertial coordinate system (r, ϕ, z) with the z-axis parallel to the star dipole moment μ and the rotation axis $\mathbf{\Omega}$. The equatorial plane is treated as a symmetry plane. The vector potential $\mathbf{A}$ is calculated so that $\nabla\cdot\mathbf{B} = 0$ at all times. The star rotates with angular velocity $\mathbf{\Omega}_* = \Omega_*\ \hat{\mathbf{z}}$. The intrinsic magnetic field of the star is taken to be an aligned dipole, $\mathbf{B} = [3\mathbf{R}\,(\mu\cdot\mathbf{R}) - R^2\mu]/R^5$ with $\mu = \mu\ \hat{\mathbf{z}}$ and the vector-potential $\mathbf{A} = \mu\times\mathbf{R}/R^3$.

We measure the length in units of the Bondi radius $R_B \equiv GM/c_\infty^2$, with c_∞ as the sound speed at infinity, density in units of the density at infinity ρ_∞, and magnetic field strength in units of B_0 which is the field at the pole of the neutron star ($r = 0, z = Z_*$). Pressure is measured in units of $B_0^2/8\pi$. The magnetic moment is measured in units of $\mu_0 = B_0 R_B^3/2$.

Simulations have been done in the cylindrical region ($0 \le z \le Z_{max}$, $0 \le r \le R_{max}$). The size of the region is less than the sonic radius of the Bondi flow R_s: $R_{max} = Z_{max} = R_s/\sqrt{2}$. Thus matter inflows supersonically in the computational region. The inflow rate is taken to be the Bondi accretion rate, $\dot{M}_B = 4\pi\lambda(GM_*)^2\rho_\infty/c_\infty^3$, where $\lambda = 0.625$ for $\gamma = 7/5$. A uniform (r, z) grid with 513 x 513 cells was used.

Initially, the density $\rho(r,z)$ and the velocity $\mathbf{v}(r,z)$ are taken to be the values given by the Bondi [1] solution. Thus, initially $v_\phi = 0$. Also, initially the vector potential $\mathbf{A}$ was taken equal to that of a dipole so that $B_\phi = 0$. The star was initialized to be rotating at the rate Ω_*.

At the outer boundaries, $r = R_{max}$ and $z = Z_{max}$, the variables (ρ, v_r, v_z, ε) are fixed to the values of the Bondi flow which has $v_\phi = 0$. The inflowing matter is unmagnetized with $\mathbf{B} = 0$. At the inner boundary, the vector-potential $\mathbf{A} = \hat{\phi}\ A_\phi$ of the intrinsic dipole field of the star is determined on the surface of the model star. A more detailed description of the numerical model can be found in [6].

3 Results

We first discuss here the simulations for the case $\omega_* = \Omega_*/\Omega_{K*} = 0.5$, $\beta = 10^{-7}$, the magnetic diffusivity $\tilde{\eta}_m = 10^{-5}$. For the mentioned values of β and η_m, the magnetospheric radius r_{m0} and the corotation radius $r_{cor} = (GM/\Omega_*^2)^{1/3}$ are equal for $\omega_* \approx 0.16$. For smaller angular velocities $\omega_* < 0.16$, the matter flow around the star is similar to that of the non-rotating case. For $\omega_* > 0.16$ the flow exhibits the features expected in the "propeller" regime.

Figure 1 shows the general nature of the flow in the propeller regime. Two distinct regions separated by a shock wave are observed. One is the *external* region where the matter inflows with the Bondi rate, and the density and the velocity agree with the Bondi solution. The second one is the *internal* region, where the flow is strongly influenced by the stellar magnetic field and rotation. The shock wave, which divides these regions, is propagating outward as in the non-rotating case [9], [10]. For the rotating star in the propeller regime the shock wave has the shape of an ellipsoid flattened along the rotation axis of the star. The region of the flow well within the shock wave is approximately time-independent. The accretion rate onto the star becomes constant after about 1–2 rotation periods of the star.

A new regime of matter flow forms inside the expanding shock wave. The rapidly rotating magnetosphere expels the matter outward in the equatorial region. This matter flows radially outward forming a low-density rotating torus. The outflowing matter is decelerated when it reaches the shock wave. There, the flow changes direction and moves towards the rotation axis of the star. However, only a small fraction of this matter accretes onto the surface of the star. Most of the matter is expelled again in the equatorial direction by the rotating field of the star. Thus, most of the matter is circulating inside this inner region driven by the rapidly rotating magnetosphere.

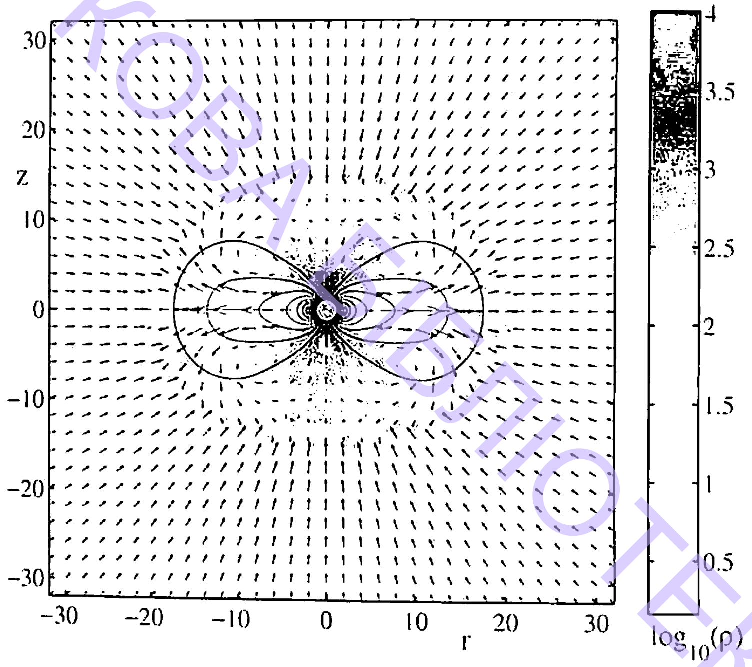


Figure 1: Matter flow in the "propeller" regime for a star rotating at $\Omega_* = 0.5\Omega_K$, after 6.9 rotation periods of the star. The axes are measured in units of the star's radius. The background represents the density, and the length of the arrows is proportional to the poloidal velocity. The thin solid lines are magnetic field lines.

The solid line in Figure 2 shows the Alfvén surface, where the magnetic energy-density equals the thermal plus kinetic energy-density, $\mathrm{B}^2/8\pi = \varepsilon + \rho \mathrm{v}^2/2$. In the equatorial plane, the Alfvén radius is $r_A \approx 4.5R_*$. For $r < r_A$, the magnetic pressure dominates, and the magnetic field is approximately that of the dipole. At larger distances the field is stretched by the outflowing plasma. Matter inside the magnetopause rotates with the angular velocity of the star.

Figure 3 shows the radial dependence of different radial forces in the equatorial plane. We can see that for $r > 3.5R_*$, the centrifugal force is becoming dominant in accelerating the matter outward. However, at larger distances, $r > 5.5R_*$, the pressure gradient force is becoming larger and determines the matter acceleration. Thus, the centrifugal and pressure gradient forces accelerate matter in the radial direction. Note, that in the majority of the region ($r > 4.3R_*$) the magnetic force is negative so that it opposes the matter outflow.

The results we have shown are for a relatively strong propeller. If the rotational velocity is smaller, the matter outflow becomes less intense. For $\Omega_* < 0.16$, no equatorial outflow is observed and the shock wave becomes nearly spherical as it is in the non-rotating case. On the other hand the shock wave becomes a more flattened ellipsoid for larger Ω_*. Matter rotates rigidly inside the Alfvén radius r_A while at large distances $\Omega = v_\phi/r$ decreases, see Figure 4.

Figure 5a shows that the accretion rate onto the star decreases as the angular velocity of the star increases, $\dot{M}/\dot{M}_B \propto \omega_*^{-1.0}$. Figure 5b shows that the accretion rate to the star decreases as the star magnetic moment increases, $\dot{M}/\dot{M}_B \propto \mu^{-2.1}$.

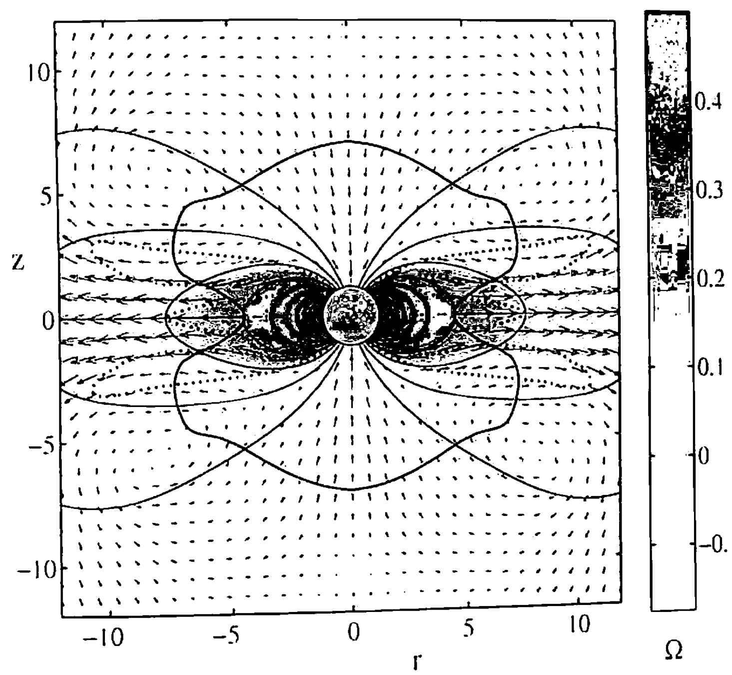


Figure 2: Same as Figure 1, but the background represents angular velocity $\Omega = v_\phi(r, z)/r$. The axes are measured in units of the stellar radius R_*.

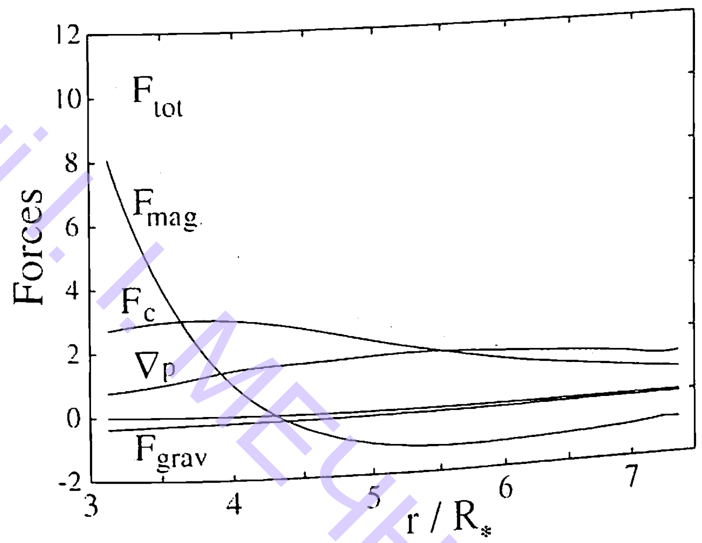


Figure 3: The radial dependence of different radial forces acting in the equatorial plane; $\mathrm{F_{mag}} = (\mathbf{J} \times \mathbf{B})_r/c$ is the magnetic force (per unit volume), $\mathrm{F_c} = \rho v_\phi^2/r$ is the centrifugal force, $\nabla p = (\nabla p)_r$ is minus the pressure gradient force, $\mathrm{F_{grav}} = -\rho(\nabla\Phi_{\mathrm{grav}})_r$ is the gravitational force, and $\mathrm{F_{tot}}$ is the total radial force.

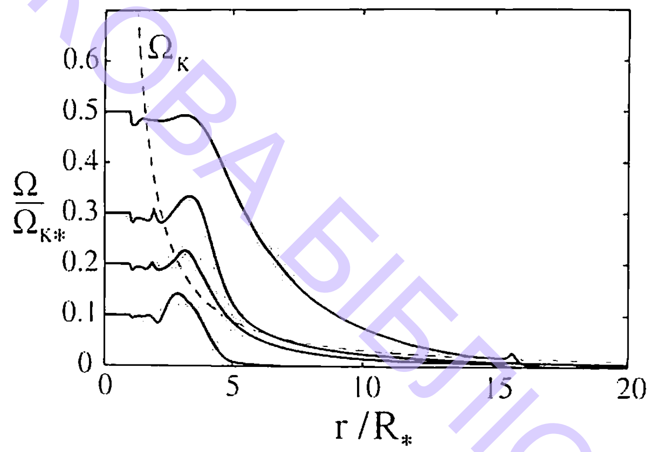


Figure 4: Angular velocity of the matter $\Omega = v_\phi/r$ versus r in the equatorial region for different angular velocities of the star. The dashed line represents the Keplerian angular velocity $\Omega_K = \sqrt{GM/r^3}$.

We have also done a number of simulation runs for different magnetic diffusivities in the range $\tilde{\eta}_m = 10^{-6} - 10^{-4.5}$, and from this we conclude that $\dot{M}/\dot{M}_B \propto (\eta_m)^{0.7}$.

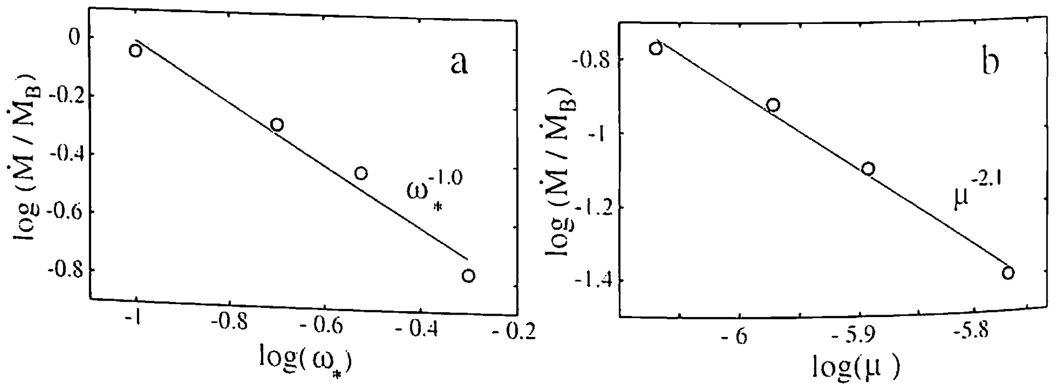


Figure 5: The left-hand panel (a) shows the fraction of Bondi accretion rate reaching the star as a function of the star's angular velocity $\omega_* = \Omega_*/\Omega_{K*}$. The right-hand panel ($b$) shows the dependence on the magnetic moment μ for $\Omega_* = 0.5\Omega_K$.

Acknowledgements

This work was supported by NSF grant AST 0307817, and by the Russian program "Astronomy." We thank Dr. V.V. Savelyev for the development of the original version of the MHD code used in this work.

References

[1] H. Bondi, *MNRAS*, **112**, 195 (1952).

[2] K. Davidson, J.P. Ostriker, *ApJ*, **179**, 585 (1973).

[3] P. Goldreich, W.H. Julian, *ApJ*, **157**, 869 (1969).

[4] A.F. Illarionov, R.A. Sunyaev, *Astron. Ap.*, **39**, 185 (1975).

[5] V.M. Lipunov, *Astrophysics of Neutron Stars*, Springer Verlag, Berlin (1992).

[6] M.M. Romanova, O.D. Toropina, Yu.M. Toropin, R.V.E. Lovelace, *ApJ*, **588**, 400 (2003).

[7] S.L. Shapiro, S.A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars*, Wiley-Interscience (1983).

[8] V.F. Shvartsman, *Radiofizika*, **13**, 1852 (1970).

[9] Yu.M. Toropin, O.D. Toropina, V.V. Savelyev, M.M. Romanova, V.M. Chechetkin, R.V.E. Lovelace, *ApJ*, **517**, 906 (1999).

[10] O.D. Toropina, M.M. Romanova, Yu.M. Toropin, R.V.E. Lovelace, *ApJ*, **593**, 472 (2003).

9

Radio Observations of X-ray Binaries

S.A. Trushkin

Special Astrophysical Observatory, RAS, Nizhnij Arkhyz, Russia
e-mail: satr@sao.ru

Abstract

We discuss recent results of radio monitoring and mapping of microquasars – X-ray binaries with relativistic jets. The sample of microquasars, defined amongst about 330 cataloged X-ray binaries in the Milky Way, consists of 15-20 radio emitting X-ray binaries, the well-known SS 433, GRS 1915+105, Cyg X-3, GRO 1655−40, and new ones G339−4, XTE 1118+480, XTE J1550−564, V4641 Sgr. For the last five years we have carried out sets of intensive monitoring of radio variability of microquasars with the RATAN-600 radio telescope, in the 1–22 GHz range. A dozen powerful radio flares have been detected. We discuss the spectral and temporal properties of the variability, and their relation with formation and evolution of jets. We also discuss different variations of finite segments of the models for the description of the synchrotron radio emission of jets in microquasars.

Keywords: stars, binaries, radio continuum, X-rays, microquasars

1 Introduction

X-ray binaries are divided into low-mass and high-mass binary stellar systems, according to the mass of the normal star, with a black hole or neutron star companion. Jets phenomena could appear in both types of X-ray binaries, and we see accretion disk-jets coupling ([1]). Black-hole X-ray binaries with relativistic jets mimic, on a much smaller scale, many of the phenomena seen in quasars and AGN [2]. The prominent radio emission from a star is a trace of high brightness temperature or of large angular size of the region. The former case corresponds to the synchrotron mechanism of emission and the last case is realized in different kinds of shells around stars. The persistent (Cyg X-1), transient (GRO 1655−40, V 4641), and typically highly variable microquasars satisfy both cases. In recent years intensive studies of the microquasars have been performed. Here we will only describe the most recent RATAN-600 monitoring results.

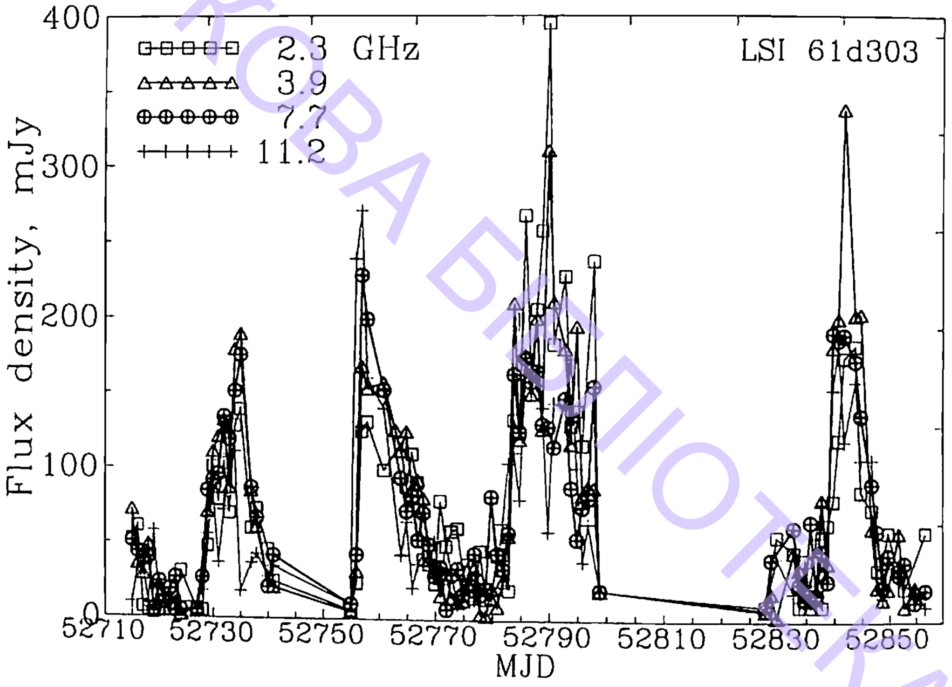


Figure 1: Light curves of LSI+61°303 at four frequencies, based on RATAN-600 observations in 2003

2 Recent Data

In a long term program to obtain daily observations of the microquasars SS433, Cyg X-3, GRS 1915+105, V4641 and LSI+61°303 with the RATAN-600 radio telescope, we have obtained unique multi-frequency light curves, which could be compared with Rossi XTE all-sky monitoring (ASM) data in the 2–12 keV band on scales from 1 to 100 days.

A 6.05-days modulation in the SS433 quiet radio emission was for the first time detected in the two sets, in the summers of 1997 and 1999 [3]. More than 1000 days of measurements of SS433 are discussed in [4]. These spectral data are accessible from the site: http://cats.sao.ru/cgi-bin/ss433.cgi.

The first periodically flaring radio-source GT0236+61, identified with an X-ray binary, containing the bright emissive star LSI+61°303 [5], was monitored during four orbital periods (26.5 days) at four frequencies (Figure 1). We have obtained the light curves, and four flaring events with different maxima have been detected near the orbital phase $\phi(P) = 0.55$. We confirm the increase of the maximum flux density during the events in accordance with the phases (from 0.62 to 0.69) of the 1667-day modulation. The decay of the flares follows a power law. Radio curves together with the optical light curve of LSI+61°303 definitely show that generation of the relativistic electrons is connected with the periastron moment (phase $\phi = 0.4$) of a binary with large orbit eccentricity ($e = 0.8$). While in this massive X-ray binary the compact component was considered as a neutron star with reduced magnetic field, the total data (and the detected jets) do not allow to reject a Kerr black hole companion.

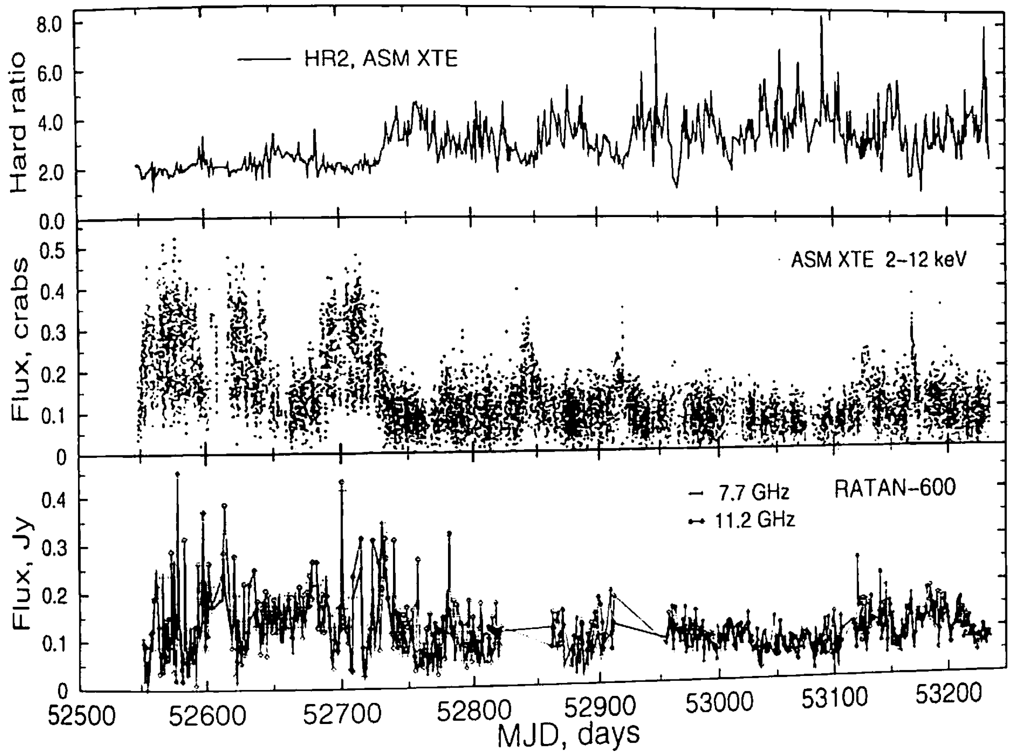


Figure 2: Light curves of Cygnus X-3 at two frequencies, based on monitoring measurements with RATAN-600 in 2003-2004, and 2-12 keV X-rays light curve, and the hard (hardness) ratio 2 (HR2) according to RXTE.

Figure 2 shows the light curves of Cygnus X-3 at 7.7 and 11.2 GHz, and RXTE ASM 2-12 keV flux and hardness ratio HR2 from October 2002 to August 2004. There are no powerful (>1 Jy) flares in Cyg X-3 since 14 September 2001, after the 15-Jy flare. Then after our alert announcement about this event, massive VLBA multi-frequency mapping, and flux density measurements with RATAN and VLA were carried out, which showed dramatic changes in the cock-sure jets structure of Cyg X-3. Miller-Jones et al. (2004) showed that Cyg X-3 jet suffers precession and there are intensive winds around the binary. In Figure 2 there is an obvious correlation of the soft X-rays with high frequency radio flux. The higher activity is seen at the beginning of the set, and in the last period the very low X-ray level accompanies low radio activity. From our former observations [7] we showed that during flares a clear soft X-ray/radio anti-correlation was detected. Probably such behavior is connected with a transient formation of the jets in Cyg X-3, contrary to SS433. There is no clear understanding of the physical origin of the jet, but obviously the process is triggered by the accretion of matter from the primary component Wolf–Rayet star, to the secondary one – a neutron star or a black hole.

In the multi-frequency light curves in Figure 3 there is a transition to "low/hard" X-ray state in GRS1915+105, followed by a relatively powerful flare.

In a 90-days set of monitoring the transient V4641 Sgr [8] we have detected two powerful radio flaring events (24 June and 4 July) with inverted spectra, and fluxes about 300 mJy at 11 GHz.

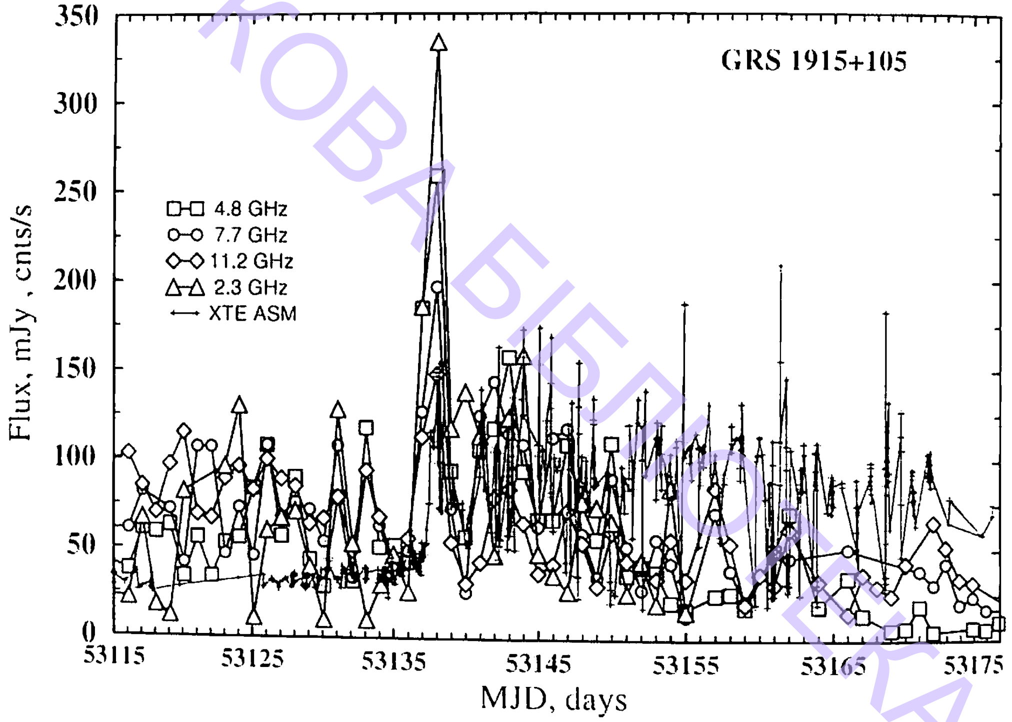


Figure 3: Light curves of GRS1915+105 at four frequencies, measured with RATAN-600 in 2004, and RXTE 2-12 keV X-rays light curve.

Acknowledgements

We are grateful to the Russian Basic Research Foundation for supporting of these studies through grant N 02-02-17439.

References

[1] R. Fender, preprint astro-ph/0303339 (2003).

[2] I.F. Mirabel, L.F. Rodrguez, *Ann. Rev. Astron. Astrophys.*, **37**, 409 (1999).

[3] S.A. Trushkin, N.N. Bursov, Yu.V. Smirnova, *Astron. Rep.*, **45**, 804 (2001).

[4] S.A. Trushkin, N.N. Bursov, N.A. Nizhelskij *Bulletin of SAO RAS*, **56**, 57 (2003).

[5] P.C. Gregory, A.R. Taylor, *Nature*, **272**, 704 (1978).

[6] J.C.A. Miller-Jones et al., *Astrophys. J.*, **600**, 368 (2004).

[7] S.A. Trushkin, *Astron. Let.*, **24**, 16 (1998).

[8] R.M. Hjellming et al., *Astrophys. J.*, **544**, 977 (2000).

10

Phase-resolved Spectroscopy of 4U0115+63 Based on Beppo-Sax Observations

A.N. Baushev, A. La Barbera, A. Santangelo and A. Segreto

Instituto di Astrofisica Spaziale e' Fisica Cosmica,
Palermo section, Italy
e-mail: baushev@pa.iasf.cnr.it

Abstract

We present results of the phase-resolved spectroscopy of the X-ray pulsar 4U0115+63 based upon Beppo-Sax satellite observations. Light curves on different energy ranges and spectra on various phase intervals have been obtained. It was shown that on the pulsar's surface there are most likely two active regions with different physical conditions instead of one region with a complex emission diagram. Analysis of the obtained data suggests that the magnetic field of the pulsar differs noticeably from the dipole one. In accordance with previous observations a strong cyclotron feature in the spectrum has been observed. Second and third harmonics have been detected at all phases of the source; even a fourth harmonic was visible at some phases. An absorption model has been used to fit the peculiarity. The feature turned out to be strongly unequidistant and phase-dependent. Physical conditions under which this sort of cyclotron peculiarity could appear are considered in close detail. Some difficulties emerging under an absorption interpretation are discussed.

Keywords: X-ray pulsars, cyclotron lines, neutron stars, magnetic field, X-ray binaries

1 Introduction

4U0115+63 is a well-known transient X-ray binary source [1, 2]. The orbital period of the binary system is 24.3 days. The optical companion of the source is *V*635 Cassiopeiae, an *O*9 star [4]. Hence the system is a massive X-ray binary. In this article we tried to make the whole phase-resolved analysis of the spectrum of 4U0115+63

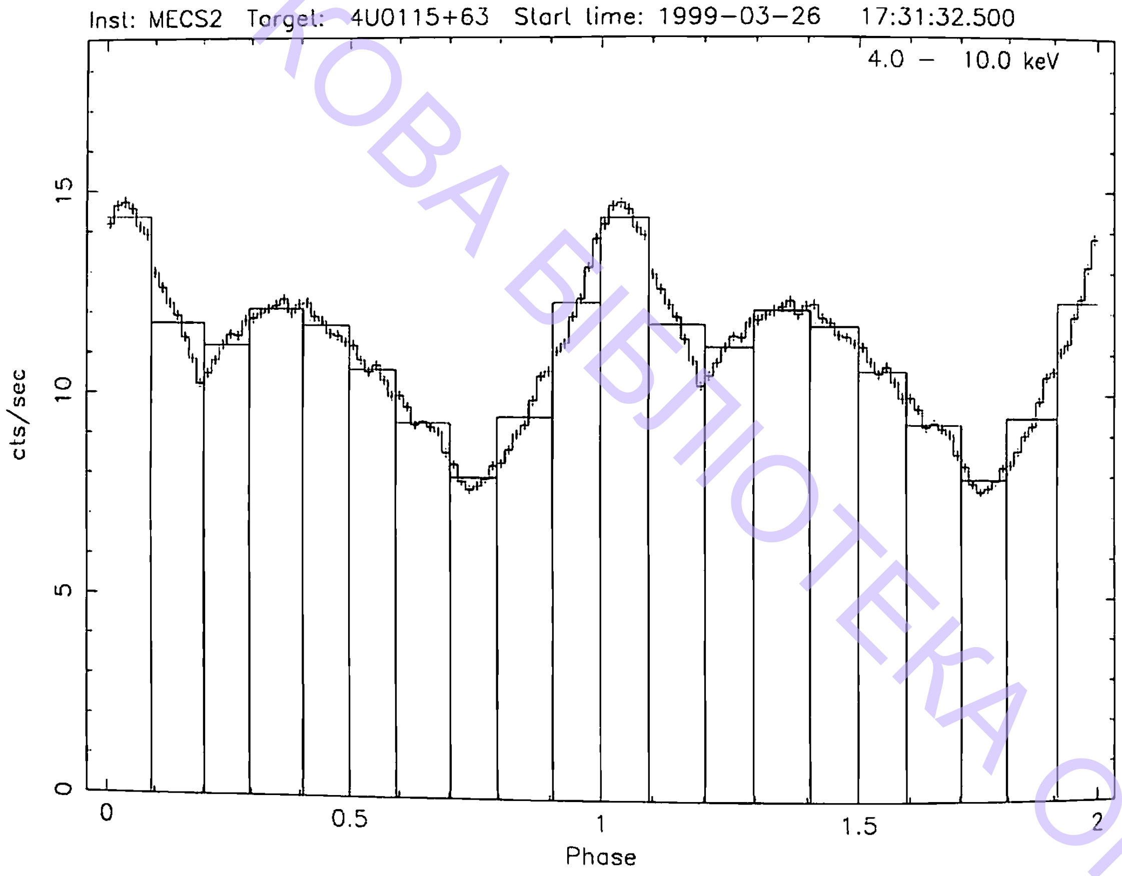


Figure 1: Pulse profile for the energy range 4–11 KeV. Phase region selection is shown.

based upon data of Beppo-Sax satellite, focusing our attention on the cyclotron feature's property.

2 Observations and Spectral Analysis

This work is based upon the observations which had been made by Beppo-Sax satellite on 1999 March, 26 (OP6714). Observation time was $17^h 12.5^s$. Since the first cyclotron harmonic in the spectrum of 4U0115+63 had been observed on ~ 12 KeV ([1, 3]), the data from the spectral interval from 4 KeV to 100 KeV have been used.

In order to perform a phase-resolved analysis of the spectrum of the pulsar all phase intervals have been divided in ten equal phase regions. The phase region selection is represented in Figure 1.

The source 4U0115+63 has a power-law spectrum with high-energy cutoff; this continuum is strongly distorted by cyclotron absorption [1, 3].

In order to represent the cyclotron feature continuum models have been multiplied by the so-called Gaussian filters, one for each line.

$$G_n(E) = 1 - D_n \exp\left(-\frac{(E - N_n \cdot E_1)^2}{2w_n^2}\right). \qquad (1)$$

Here n is the number of the harmonic, E_1 is the energy of the first harmonic. Parameters N define positions of the other harmonics. Initially they were taken equal to n, but they were considered as free parameters in order to represent unequidistant cyclotron spectra. Three cyclotron lines have been detected at all phases. The fourth harmonic turned out to be necessary only at the phases 0.9–1.2.

3 Discussion

The fundamental cyclotron energy is not stable: its variations are of the order of 10% and they are near correlating with the variations of luminosity. Harmonics are very unequidistant. This effect is the most distinct on the second harmonic. The ratio between energies of the second and the first harmonics is as small as 1.6 at some phases.

The situation with the theoretical interpretation of these effects is not so simple. These effects are usually interpreted as a result of complex geometry of the emission forming region and magnetic field variations in it. Let as analyze this possibility.

The neutron star will be considered to have a radius ~ 10 km, and a dipole magnetic field (for estimate this assumption is quite acceptable). The luminosity of the source 4U0115+63 is $\sim 10^{37}$ erg/sec. When determined from the cyclotron energy by the usual nonrelativistic formula the magnetic field of the source appears to be $1.1 \cdot 10^{12}$ Gs. There is a well-known formula for the angular size χ of the hot spot on the X-ray pulsar's surface if its magnetic field is a dipole.

$$\chi \simeq \sin\chi = 7.5 \cdot 10^{-4} H^{-2/7} L^{1/7}.$$

In the case under consideration we obtain $\chi \simeq 0.06$. Relative variation of the magnetic field (and hence of the cyclotron energy) within the boundaries of such small a spot is of the order of $\Delta H/H \approx 1.4 \times 10^{-3}$. Of course it is not enough to explain neither widths of lines nor their arrangement. Consequently all these effects should be considered as a result of variations of the magnetic field along the height of the accretion column.

On average the second harmonic is the strongest. The first harmonic is the most pronounced only in the region out of the major peak (phases 0.45–0.85). At the phase 1.15 where the cyclotron feature is the most distinct, the second and third harmonics have practically the same depths which is significantly higher than the depths of the first and the fourth one. These ratios between depths of different harmonics looks mysterious. As it was shown for example in [5], cyclotron cross sections averaged over directions and polarizations under the considered physical conditions are related as

$$\sigma_1 : \sigma_2 : \sigma_3 = 400 : 20 : 3. \qquad (2)$$

Even for the direction perpendicular to the magnetic field this ratio is

$$\sigma_1 : \sigma_2 : \sigma_3 = 200 : 18 : 3. \qquad (3)$$

We can see that in any case the first harmonic should be significantly stronger than the others. Meanwhile at the main peak the second and the third harmonics are stronger than the first one. The most believable explanation of this fact was given in the series of articles (see, for example, [6]). The authors explain the contradiction by the influence of Raman scattering. Under the physical conditions that exist in the accretion column the probability of the transition of an electron from the second Landau level through the branch ($2 \rightarrow 1 \rightarrow 0$) is approximately seven times higher [8] than through the branch ($2 \rightarrow 0$). An influence of the true absorption is negligible [9]. As a result a photon of the second harmonic being absorbed decays on two photons of the first one. This spawning leads to significant decreasing of the first harmonic's depth.

But this mechanism could be acceptable only for the second harmonic: for the ratio between third and second harmonics it fails. In fact, the integrated probability of the branches ($3 \rightarrow 1 \rightarrow 0$) and ($3 \rightarrow 2 \rightarrow 0$) is only 29%, and only one photon of the second harmonic appears instead of two photons in the previously described case. Furthermore even this rare photon has no chance to reach the distant observer. As it was shown in [7], Raman scattering of photons of the third harmonic can influence effectively the spectra of previous harmonics only if an optical depth of the scattering electrons on the third harmonic is at least comparable to 1. But in this case their optical depth on the second harmonic should be ~ 6. The spawned photon of the second harmonic is sure to experience the second scattering and decay on two photons of the first harmonic. So the efficiency of the spawning of photons of harmonics of the orders higher than first is negligible. But ratios of their depths are also very far from the theoretical relations between cyclotron cross-sections.

The nonequidistancy of harmonics is also very difficult for the theoretical explanation. Owing to the above-mentioned relations between the cyclotron cross-sections the magnetized substance is much more transparent on the energy of the second harmonic than on the energy of the first one. Consequently the second harmonic should appear in the deeper regions of the accretion column with the stronger magnetic field. We might have expected that the distance between first and second harmonics is bigger than the harmonic one. Meantime the distance is smaller than the harmonic: the ratio between energies of the second and the first harmonics sometimes is as small as 1.6. The same result has been already reported in [10].

References

[1] A. Santangelo et al., *Ap.J. Letters*, **523**, 85 (1999).

[2] S. Rappaport et al., *Ap.J. Letters*, **224**, 1 (1978).

[3] N.E. White, J.H. Swank, S.S. Holt, *ApJ*, **270**, 711 (1983).

[4] S. Unger et al., *A&A*, **336**, 960 (1998).

[5] A.K. Harding, J.K. Daugherty, *Ap.J.*, **374**, 687 (1991).

[6] J.C.L. Wang, et al., *Phys.Rev. Letters.*, **63**, 1550 (1989).

[7] D.Q. Lamb, J.C.L. Wang, I.M. Wasserman, *Ap.J.*, **363**, 670 (1990).

[8] J.K. Daugherty, J. Ventura, *A&A*, **61**, 723 (1977).

[9] S. Bonazzola, J. Heyvaerts, J.L. Puget, *A&A*, **78**, 53 (1979).

[10] W.A. Heindl et al., *Ap.J. Letters*, **521**, 49 (1999).

11

High-Energy Neutrinos from a Collapsing Supermassive Star

V.S. Berezinsky[1,2] and V.I. Dokuchaev[2]

[1]*Laboratori Nazionali del Gran Sasso, INFN, Assergi (AQ), Italy*
e-mail: berezinsky@lngs.infn.it
[2]*Institute for Nuclear Research, RAS, Moscow, Russia*
dokuchaev@inr.npd.ac.ru

Abstract

We describe the scenario of massive black hole (MBH) formation inside a supermassive star (SMS) in the galactic nucleus. A SMS is naturally formed in the dynamically evolving galactic nucleus due to destruction of normal stars in direct collisions. The surviving neutron stars (NSs) and stellar mass black holes form a compact self-gravitating subsystem deep inside the SMS. This subsystem is short-lived in comparison with a host SMS and collapses into the MBH. The frequent NS collisions in this nearly collapsing subsystem are accompanied by the generation of multiple ultra-relativistic fireballs which are the suitable place for particle acceleration. Only the secondary high-energy neutrino escape from the SMS interior. The resulting high-energy neutrino signal can be detected by underground neutrino telescope with an effective area $S \sim 1$ km^2 and can give the evidence for MBH formation in the distant galactic nucleus.

Keywords: black holes, supermassive stars, neutrino

1 Introduction

Recently we proposed [1] a model of a very powerful but short lived hidden high-energy (HE) neutrino source which originates in the compact galactic nucleus just prior to its collapse into a MBH. In this model it is supposed that a MBH is formed by the natural dynamical evolution of the central stellar cluster in the normal galactic nucleus. Dynamical evolution of dense central stellar clusters in galactic nuclei is accompanied by the growth of the velocity dispersion, v, of constituent stars or, equivalently, by the growth of the central gravitational potential $\phi \propto v^2$. This process is accompanied by the contraction of the stellar cluster and culminates with the formation of a MBH, when the velocity dispersion of stars grows up to about the speed of light, $v \sim c$, or equivalently, when the stellar cluster radius shrinks near to its gravitational radius (see for a review [2, 3] and references therein). In the following section we shall elaborate this hidden source model for the possible case of a SMS formation in the galactic nucleus. It will be shown below as in the scenario of NS cluster formation simultaneously with a massive envelope [1], a hidden source of HE neutrinos also arises in the case of SMS formation, but would be shorter-lived and more intense.

2 Supermassive Star Formation and Evolution

The SMS in the galactic nucleus may be formed from the gas liberated in normal star self-destructions in the rather evolved central stellar cluster with a velocity dispersion $v \geq v_{\rm esc} = (2Gm_*/r_*)^{1/2}$, where $v_{\rm esc}$ is the escape velocity from the surface of a star with mass m_* and radius r_*. For a solar type star the escape velocity is $v_{\rm esc} \simeq 620$ km s^{-1}. In a cluster with $v > v_{\rm esc}$ all normal stars are eventually disrupted in mutual collisions or in collisions with the extremely compact stellar remnants: NSs or stellar mass black holes (BHs). Only these compact stellar remnants survive the stellar-destruction stage of galactic nucleus evolution ($v \simeq v_{\rm esc}$) and may form the self-gravitating subsystem. The characteristic time-scale for stellar cluster dynamical evolution is the (two-body) relaxation time $t_{\rm r}$.

Only the relativistic compact stellar remnants such as NSs and stellar mass BHs can survive during the stellar-destruction phase of evolution of the galactic nucleus at $v \geq v_{\rm esc}$ and populate the interior of the newly formed SMS. An individual NS with a mass m and local velocity V spirals down to the center of the formed SMS under the influence of dissipative dynamical friction force [4, 5] $F_{\rm df} = I \times 4\pi(Gm)^2\rho_{\rm SMS}/V^2$ (where $\rho_{\rm SMS} = \bar{n}_{\rm SMS}m_{\rm p}$ is the SMS gas density), which is directed opposite to the local NS velocity V. The dimensionless function $I \simeq 1$ for the case of $V \simeq c_{\rm s}$, where $c_{\rm s}$ is the SMS sound velocity. The corresponding effective time of a NS dynamical friction drag toward the center of the SMS is $t_{\rm df} = V/\dot{V} = mV/F_{\rm df}$. It can be easily seen that at the moment of SMS formation (when $V \sim c_{\rm s} \sim v_{\rm esc}$) the effective time of a NS friction drag is of the same order as the time-scale of SMS formation, $t_{\rm df}(v_{\rm esc}) \sim t_{\rm coll}(v_{\rm esc}) \sim t_{\rm r}(v_{\rm esc})$. We assume for simplicity that the SMS contains identical NSs with a typical NS mass $m_{\rm NS} = 1.4{\rm M}_\odot$ and with a total NS cluster mass $M_{\rm NS} = f_{\rm NS}M_{\rm SMS}$. Here the NS mass fraction relative to the SMS mass is $f_{\rm NS} = 0.01f_{-2} \ll 1$ and the total NS number in the cluster is $N_{\rm NS} = f_{\rm NS}M_{\rm SMS}/m_{\rm NS} \simeq 7.1 \cdot 10^5 M_8 f_{-2}$. The value of $f_{\rm NS}$ is a free parameter of the model and we will put $f_{-2} = 1$ in numerical estimations. The dynamical friction time-scale of a NS with velocity $V = (GM_{\rm SMS}/2R)^{1/2}$ inside the host SMS diminishes as $t_{\rm df} \propto R^{3/2}$ during the contraction of the SMS. On the contrary the evolution time-scale of SMS grows with SMS contraction, $t_{\rm SMS} \propto R^{-1}$. The subsystem of NSs evolves faster than the host SMS after reaching the stage when $t_{\rm df} \sim t_{\rm SMS}$. At this time all NSs sink deep to the central part of SMS and forms there the self-gravitating cluster. From approximate equality $t_{\rm df}(R) \sim t_{\rm SMS}(R)$ we find [6] the SMS radius at the moment of NS cluster formation inside it, $R_{\rm SMS} \simeq 4.6 \cdot 10^{15} M_8^{1/5}$ cm, where $M_8 = M_{\rm SMS}/(10^8{\rm M}_\odot)$.

Under the dynamical friction the NSs are sinking deep to the SMS until they are concentrated in the central region with some radius $r = R_{\rm NS} \ll R_{\rm SMS}$. At this time the total mass of NSs becomes of the same order as the mass of the ambient gas and a self-gravitating NS cluster is formed. Using the above relation we find the initial parameters of the newly formed NS cluster: the radius $R_{\rm NS} = f_{\rm NS}^{1/3} R_{\rm SMS}$ and virial velocity $v_{\rm NS} = f_{\rm NS}^{1/3} v_{\rm SMS}$, where $v_{\rm SMS} = (GM_{\rm SMS}/2R_{\rm SMS})^{1/2}$ is a SMS virial velocity. At the moment of NS cluster formation the corresponding SMS evolution time $t_{\rm SMS} = t_{\rm SMS}(R_{\rm SMS}) \simeq 7.3 \cdot 10^5 M_8^{4/5}$ yrs, virial velocity $v_{\rm SMS} \simeq 0.04 M_8^{2/5} c$. Finally the dynamical evolution of the NS cluster is terminated by its dynamical collapse to the MBH.

We employ the simple analytical "evaporation" model [7] for the dynamical evolution of the star cluster to fix the time dependence of parameters of the NS cluster. During the process of evaporation of the fast NSs the total virial energy of the cluster remains constant $E = -Nmv^2/2 = const$. Accordingly, the NS velocity dispersion in the remaining cluster grows as $v \propto N^{-1/2}$, and the cluster contracts as $R \propto N^2$ with diminishing star number N. After reaching $v \simeq 0.3c$, which is the onset of the NS cluster global dynamical instability, the remaining NS cluster collapses to the MBH [8, 9]). The evaporation of fast NSs is accompanied by the cluster contraction and by the increasing of the rate of direct NS collisions in the cluster $\dot{N}_{\rm cap}$. These collisions proceed most probably through the capture of two NSs into the short-lived binary due to gravitational radiative losses during flyby with a final NS coalescence [9, 10].

3 Stationary Cavity Inside the SMS

The numerous NS collisions produce the coexisting relativistic fireballs with a large energy release. The total energy of a single fireball is $E_0 = E_{52}10^{52}$ ergs with $E_{52} \sim 1$. The repeating fireballs due to frequent collisions of NSs in the dense enough NS cluster can dig out the rarefied quasi-stationary cavity inside the SMS with radius $R_{\rm cav} \ll R_{\rm SMS}$, which we first describe qualitatively below.

We model the nonstationary stage of cumulative shock expansion by the self-similar spherical shock solution for a central energy source varying in time, $E = At^k$, with $A = const$ and $k = const$ [11, 12]. The particular case of $k = 0$ will correspond to the Sedov *instant shock* solution. The radius of the stationary cavity is determined from the energy flux balance on its boundary at $r = R_{\rm cav}$. Correspondingly just outside $R_{\rm cav}$ this energy flux is carried by the hydrodynamic flow. This energy flux balance determines [6] the radius of the stationary cavity $R_{\rm cav} \simeq 10^{12} M_8^{-2/5}$ cm. The cavity is supported in the stationary state only if there are simultaneously several fireballs inside it. The power of the central source $L = \dot{N}_{\rm cap}E_0 \propto v^{45/7}$ gradually grows due to the evolutionary growing NS collision rate. The maximum radius, $R_{\rm max}$, of a stationary cavity in the evolving NS cluster and the maximum rate of NS collisions are reached at the maximum luminosity of the central source. Collisions of NSs during the rather short lifetime supply the total energy which is far less than the SMS binding energy, and so the fireballs inside the SMS do not influence its state.

Inside the cavity numerous inward and outward colliding shocks from fireballs result in a highly turbulent gas motion. The particles are accelerated in such media by the Fermi II mechanism. Particles accelerated in the cavity interact with the gas in the SMS interior producing high-energy neutrino flux. We assume an existence of the equipartition magnetic field induced by the turbulence and dynamo mechanism. By following the general procedure, we calculate [6] the corresponding high-energy cosmic ray and neutrino flux from the cavity. Crossing the Earth, these neutrinos create deep underground the equilibrium flux of muons. The most effective energy of muon detection is $E_\mu \geq 1$ TeV [13]. The rate of muon events $\dot{N}(\nu_\mu) = F_\mu S$ in the detector with an effective area S at distance r from the source is given by

$$\dot{N}(\nu_\mu) \simeq 2\left(\frac{L}{10^{49}\ \mathrm{erg\ s}^{-1}}\right)\left(\frac{S}{1\ \mathrm{km}^2}\right)\left(\frac{r}{10^3\ \mathrm{Mpc}}\right)^{-2} \mathrm{day}^{-1}. \qquad (1)$$

Thus, it is expected a few muons per day in the neutrino telescope of the Ice Cube with scale $S \sim 1$ km from a single source at the distance $\sim 10^3$ Mpc and with a total luminosity of the cavity $L \simeq 2.8 \cdot 10^{49}$ erg/s. This hidden neutrino source inside the evolving SMS appears only once during a galaxy lifetime prior to the MBH formation in it.

Acknowledgements

This work was supported in part by the Russian Foundation for Basic Research grants 02-02-16762-a, 03-02-16436-a and 04-02-16757-a and by the Russian Ministry of Science grants 1782.2003.2.

References

[1] V. Berezinsky, V. Dokuchaev, *Astropart. Phys.*, **15**, 87 (2001).

[2] M. Rees, *Ann. Rev. Astron. Astropys*, **22**, 471 (1984).

[3] V. Dokuchaev, 1991, *MNRAS*, **251**, 564.

[4] S. Chandrasekhar S., *ApJ*, **97**, 255 (1943).

[5] E. Ostriker, *ApJ*, **513**, 252 (1999).

[6] V. Berezinsky, V. Dokuchaev, preprint astro-ph/0401310.

[7] L. Spitzer, R. Harm, *ApJ*, **127**, 544 (1958).

[8] Ya.B. Zel'dovich, M.A. Poduretz, *Soviet Astron*, **9**, 742 (1965).

[9] C. Quinlan, S. Shapiro, *ApJ*, **321**, 199 (1987).

[10] V. Dokuchaev, Yu. Eroshenko, L. Ozernoy, *ApJ*, **502**, 192 (1998).

[11] J. Ostriker, C. McKee, *Rev. Mod. Phys.*, **61**, 1 (1988).

[12] V. Dokuchaev, *Astron. Astrophys*, **395**, 1023 (2002).

[13] V. Berezinsky, *Nucl. Phys. B (Proc. Suppl.)*, **19**, 375 (1990).

12

X-ray/Gamma-ray Observations of BL Lac in High State in 2000

V.V. Fidelis

Crimean Astrophysical Observatory, Nauchny, Ukraine
e-mail: fedelis@crao.crimea.ua

Abstract

The retrospective analysis of Cherenkov data of BL Lacertae exposures in 2000 reveals the mean flux at $E > 1$ TeV on the level of $\approx$0.5 Crab at 4.65 σ. The source exhibited a regular trend in increasing intensity from a low state priory to September to a high luminosity state in November. During this period the flux enhancement came approximately to fourfold increasing of the intensity. This increasing tendency is also distinctly seen in the light curve extracted from the ASM/RXTE quasi-simultaneous observations in the 2–10 keV energy band.

Keywords: BL Lac objects: individual (BL Lac) – gamma-rays: observations

1 Introduction

BL Lacertae, the prototype of the BL Lac class of Active Galactic Nuclei, is one of the best studied objects. It was observed in X-ray by different space experiments [1] and in TeV energy band by the Whipple observatory [2] and CrAO [3]. In 2000 BL Lac was the subject of an extensive multi-wavelength monitoring campaign in radio band, optical frequencies, in X-rays (BeppoSAX and RXTE) and in Very High Energy (VHE) γ-rays (HEGRA) [4]. The observations provided evidence that the synchrotron spectrum of the object extends out to $\sim$10 keV.

The simultaneous observations in X-ray and the TeV energy band allowed us to collect data concerning correlated variability of the object on different time scales which may be used for determining the nature of the dominant radiation processes.

2 Observations and Gamma/Hadron Separation

The observations of BL Lacertae reported here were obtained in 2000 with air-Cherenkov telescope GT-48, covering the energy band > 1 TeV and an ASM detector onboard the RXTE Project in the energy band 2–10 keV. In the TeV range the object was observed in the ON/OFF modes. Both ON and OFF runs were 35 minutes long. OFF runs were shifted on right ascension by 40 minutes. After cleaning the data collected in ON/OFF pairs for weather conditions the total effective source exposure time was 1330 minutes, split between July and November.

The showers produced by VHE γ-rays were selected from a background of showers produced by cosmic rays (CR), mainly protons, using differences in their images and arrival directions. Shower images from VHE γ-rays, as they are more compact, have smaller angular dimensions and are aligned in the field of view of the camera to the source position. Showers from CR have more fragmented form of their images, higher angular dimensions and arrive from all directions.

From the calculated first and second moments of the brightness distribution we derived the coordinates of their centers, the effective dimensions and the orientation parameters of shower images. The flashes were also selected by the minimal amplitudes of the detected signals.

3 Results

After applying the ON/OFF data selection procedure we have obtained the following statistics (Table 1). ON data represent the number of γ-like events observed in the source direction and OFF data – the background. ON-OFF is their difference and equals the number of registered VHE γ-quanta from the point source with coordinates of the BL Lac. A filtering cut includes selection by form, amplitude and dimension. Selection by parameter *azwidth* was accomplished jointly with the parameter *dist* [5].

Table 1: Results of BL Lac observation data analysis.

Reduction method	ON	ON-OFF	Significance, st. dev.
Raw data	27821	502	2.14
Filtering	405	99	3.71
azwidth$< 0^\circ.235$	358	114	4.65

The average flux of VHE γ-quanta for BL Lac (expressed as a fraction of the Crab Nebula steady flux) equals ≈0.5 Crab.

4 X-ray/Gamma-ray Correlation

In the observing period the source exhibited a regular trend in increasing intensity at $E > 1$ TeV from a low state prior to September (MJD51753–51782) to a high luminosity state in October–November (MJD51820-51868) (Figure 1a). During this period the flux enhancement came approximately to a fourfold increase in intensity. This increasing tendency is also distinctly seen in the light curve extracted from ASM/RXTE quasi-simultaneous observations in the 2–10 keV energy band (Figure

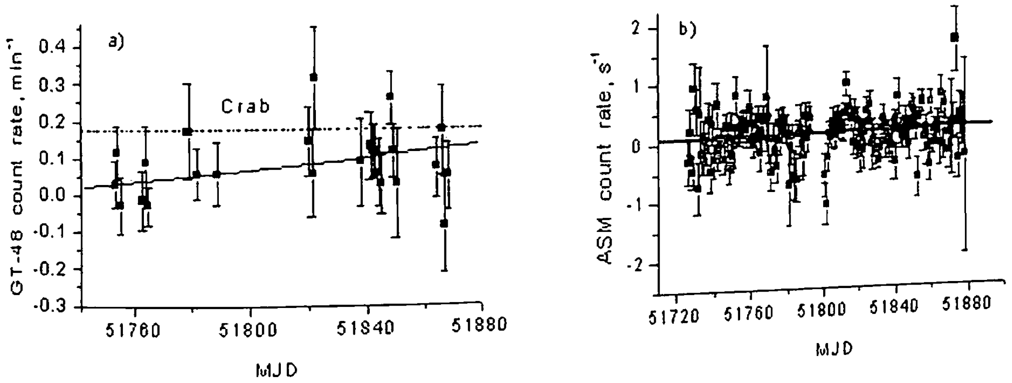


Figure 1: Time history of BL Lac emission a) at E>1 TeV and b) in the 2–10 keV energy band (quick-look results provided by the ASM/RXTE team). The GT-48 data are night average. The error bars are purely statistical. The dashed line on the left plot corresponds to the count rate of VHE γ-quanta from Crab Nebula.

From Figure 2 it may be seen that both light curves for three moonless periods have similar intensity features.

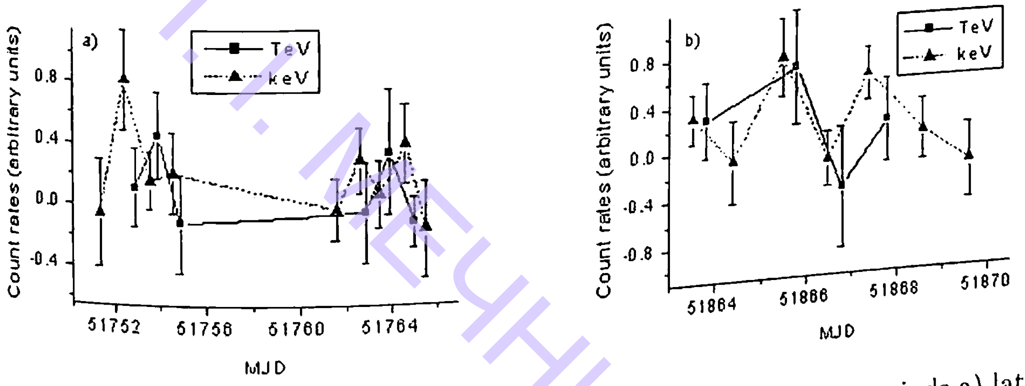


Figure 2: The comparison of TeV and keV light curves for moonless periods a) late August/early September, b) November.

Our estimation confirms the results of a multi-wavelength monitoring campaign of BL Lac in the second half of 2000 during which the object underwent a major transition from a rather quiescent state to a flaring state for the rest of the year [4].

References

[1] M. Ravasio et al., *Astron. Astrophys.*, **383**, 763 (2002).

[2] J. Quinn et al., *Astrophys. J.*, **456**, L86 (1996).

[3] Yu. Neshpor et al., *Astron. Reports.*, **78**, 291 (2001).

[4] M. Böttcher et al., *Astrophys. J.*, **596**, 847 (2003).

[5] A. Hillas, *NASA 19-th Intern. Cosm. Ray Conf.*, **3**, 445 (1985).

13

TeV/keV Correlated Activity from the Galaxy Mrk 501 in 2002

V.V. Fidelis

Crimean Astrophysical Observatory, Nauchny, Ukraine
e-mail: fedelis@crao.crimea.ua

Abstract

The atmospheric Cherenkov telescope GT-48 detected in 2002 the flux of Very High Energy (VHE) gamma-quanta (>1 TeV) from the extragalactic object Mrk 501 on the level of ≈ 45% of the steady flux in this energy band from the Crab Nebula with statistical significance 4.73 σ. The analysis of the data on daily basis revealed a correlation between the temporal properties of TeV emission from the object with the emission in 2–10 keV energy band according to ASM/RXTE data.

Keywords: BL Lacertae objects – gamma-rays: observations

1 Introduction

The extragalactic object Mrk 501, a BL Lac active galactic nucleus, is a point source of VHE γ-rays and it is associated with the elliptical galaxy at redshift 0.033. Its characteristic feature is the correlated variability in the keV and TeV energy bands on various time scales [1].

It is difficult to study BL Lacertae in the optical because of the absence of emission lines and problems with establishing distances. So that observations of BL Lac objects in X-ray and VHE γ-rays are of great interest.

The spectral energy distribution of BL Lacs is distinguished by two broad peaks, the first of them extends from radio to the UV/X-ray band and the second one may reach TeV energies. Observing BL Lacs nearly simultaneously in X-ray and the TeV band we may study the emission in two peaks, correlated in different ways. The resolution of the characteristic time-scales may be used to reconstruct the geometry and spatial structure of emitting regions on different scales.

Mrk 501 was detected by different ground-based experiments. In this paper we present the results of our observations and their comparison with X-ray data.

Table 1: Recorded and selected events.

Selection algorithm	Number of events on source	Number of events on background	Resulting events number	Significance, standard deviations
No selection	10196	10253	-57	-0.40
By form, dimensions and amplitude	211	170	41	2.10
By orientation ($\alpha < 36°$)	98	42	56	4.73

2 The Cherenkov Telescope and Observations

The Cherenkov telescope GT-48 is situated at an altitude of 600 m above sea level. It consists from two identical alt-azimuth mountings, or sections, spaced apart by 20 m. Each section is equipped with four cameras, operating in the visible band and consisting of 37 photomultiplier tubes (PMT) that organize 37 channels. Each camera, mounted in the focal plane of four mirrors with diameters of 1.2 m, records the images of atmospheric Cherenkov radiation from air showers produced by γ-rays and cosmic rays (CR). The field of view (FOV) of each PMT is $0°.4$, the total FOV of each camera is $2°.6$. The total area of mirrors on both sections is 36 m^2. The signals from the channels from the four cameras on each section are linearly added.

Flashes are registered only when the amplitude of the signals for any two of the 37 channels on each section exceed certain threshold. The time resolution for the coincidence scheme is 15 ns. The mountings are geared by a control system with a pointing accuracy of $0°.05$. Both sections of the telescope operate in coincidence mode. The effective threshold energy for the detection of the γ-rays is approximately 1 TeV. The telescope is described in detail in [2].

Mrk 501 was observed between April and July 2002 in the ON/OFF mode, when ON runs were followed by OFF runs with 30 minutes shift in right ascension. Both ON and OFF runs were 25 minutes in duration and were carried out under the same range of elevation. After cleaning the observing data for weather conditions a total of 700 minutes of source exposures and the same time of background observations were taken for processing the Cherenkov light images from extensive air showers (EAS) produced by γ-rays and CR.

3 Selection of EAS, Initiated by γ-Ray Primaries

The shower images from γ-rays have compact form and small effective dimensions and are aligned by their major axis in the direction of the source in the FOV of the camera, while the images from CR, mainly protons, are greater, have more fragmented form and isotropic orientation.

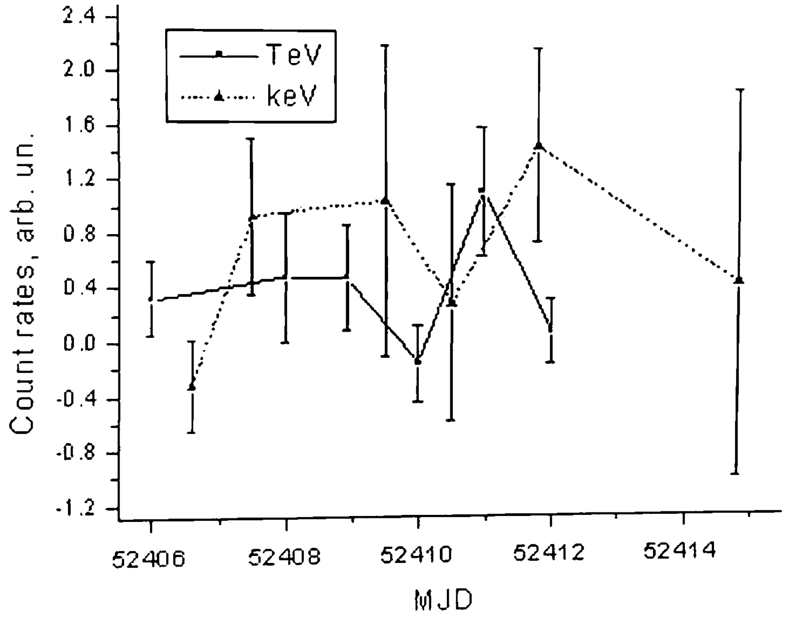


Figure 1: Distribution of TeV/keV emission from the object in May (quick-look results provided by the ASM/RXTE team).

To select the γ-like events we used: *length*, *width* (the effective dimensions of their images), orientation parameters α (the angle between the direction toward the source from the flash center and the direction of the major axis of the flash image) and *dist* (the distance of the image centroid from the source) [3]. Flashes were also selected by the minimal amplitudes of detected signals because the parameters of flashes with low energy are determined with high errors. The selection statistics is shown in Table 1.

4 Source Variability

During the observing period the source exhibited a rather low average flux in the VHE energy band $\approx 45\%$ of the steady flux from the Crab Nebula at the 4.73 σ level. Two flares were detected. The first flare, occurred in May 16, had an intensity close to 1.6 Crab and the second one, occurred in July, ≈ 1.8 Crab (Crab is the unit in terms of the steady flux of VHE gamma-quanta from the Crab Nebula). The statistical confidence of these events is 2.33 σ and 2.83 σ, respectively.

5 TeV/keV Correlated Activity

TeV/keV emission from the object in May and July correlated in different ways. Foe example, the intensity feature that includes a flare was also seen in the light curve at the 2–10 keV energy band in May 13–17 (Figure 1).

There is a leading role of the TeV component. In July 2–17 the source underwent a slow transition from a relatively high state to a low state in both the keV and the

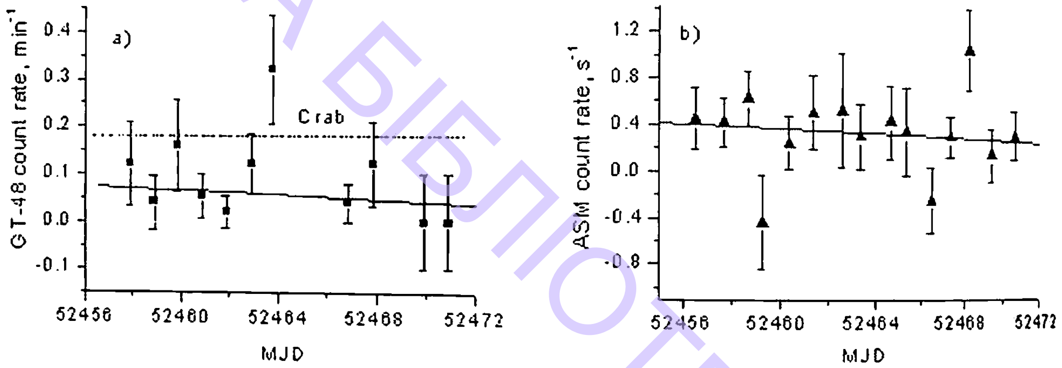


Figure 2: Distribution of TeV (a) and keV (b) emission from the object in July. The dashed line on a) corresponds to the count rate of VHE γ-quanta from the Crab Nebula.

The count rates of TeV/keV emission from the object during the observing period are slightly correlated (Figure 3).

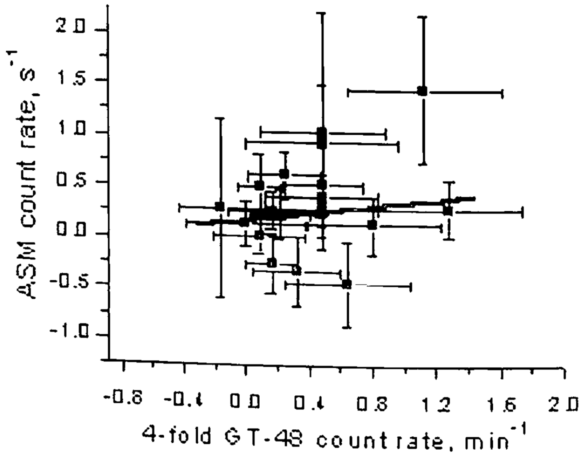


Figure 3: The correlation of ASM/GT-48 daily average count rates from Mrk 501.

6 Conclusion

Our observations of Mrk 501 in 2002 reveal emission from this object at E>1 TeV with a factor of half the steady γ-ray flux from the Crab Nebula and two flares with three–four-fold increasing of intensity relative to the base state.

The correlated emission in the X-ray and TeV bands from the object may be explained by co-spatial radiation in both bands produced by the same electrons in accordance with the most advanced to date synchrotron self-Compton model (SSC) [4].

References

[1] F. Aharonian et al., *Astron.Astrophys.*, **342**, 69 (1964).

[2] B. Vladimirsky et al., *Izv. Krim.Astrofiz.Observ.*, 91, 74 (1994).

[3] A. Hillas, *NASA 19-th Intern. Cosm. Ray Conf.*, 3, 445 (1985).

[4] L. Costamante, G. Ghisellini, *Astron.Astrophys.*, **384**, 56 (2002).

IV Galaxies, Stars and Interstellar Medium

1

Gamma-ray Burst Interaction with Dense Interstellar Medium

M.V. Barkov and G.S. Bisnovatyi-Kogan

Space Research Institute (IKI), RAS, Moscow, Russia
e-mail: barmv@sai.msu.ru, gkogan@iki.rssi.ru

Abstract

The interaction of cosmological gamma ray burst radiation with the dense interstellar medium of a host galaxy is considered. Gas dynamical motion of the interstellar medium driven by gamma ray bursts is investigated in a 2D approximation for different initial density distributions of host galaxy matter and different total energies of gamma ray burst. The maximum velocity of motion of the interstellar medium is 1.8×10^4 km s^{-1}. Light curves of gamma ray burst afterglow are calculated for a set of non homogeneous densities, a distribution of gamma ray burst total energies, and different viewing angles. Spectra of gamma ray burst afterglows are modeled taking into account conversion of hard photons (soft X-ray, hard UV) to soft UV and optical photons.

Keywords: gamma ray burst, optical afterglow

1 Introduction

Although gamma-ray bursts (GRBs) were discovered more than thirty years ago [1], their origin is still unclear. The most extensive data on the detection of GRBs have been obtained by the Compton Gamma Ray Observatory BATSE experiment [2, 3]. Analysis of the GRBs detected shows that their apparent distribution on the sky is isotropic, but that there is a significant departure of their (logN – logS) curve from the $N \sim S^{-3/2}$ in the sense that it is lower than that corresponding to a spatially uniform source distribution [2, 4].

Observation of optical afterglows of GRB, following after identification of GRB with a transient X-ray source by Beppo-SAX, and discovery of large (up to $z = 4.5$) redshifts in the spectra of optical transients had confirmed the cosmological origin of long GRB.

A powerful energy release should have a strong effect on a large volume of matter around the parent galaxy and should give rise to the formation of a GRB counterpart at other wavelengths. Without specifying the mechanism for GRB formation, we assume only the existence of a strong gamma radiation flux and consider the interaction of this radiation with the interstellar medium on large spatial and temporal scales.

We investigate the response of an interstellar medium of standard chemical composition to the passage of a short-term powerful pulse of gamma radiation (that is, the dynamical behaviour and radiative cooling after heating by the gamma rays).

The spherically symmetric model was investigated in [5]; a 2D model, allowing us to study different matter distribution and the reaction to anisotropic GRB has been studied in [6, 7] using numerical simulations by a PPM method. In the following sections we present the results from [6, 7].

2 The Main Equations

We solve the set of hydrodynamic equations describing the motion of matter, together with thermal processes in an axially symmetric case. The gamma ray pulse is considered as instantaneous having total energy Γ and luminosity $L = \Gamma\delta(t - r/c)$. We are interested in the behaviour of the gas heated by the GRB, which is taken as fully ionized. A flat spectrum for the GRB, $\frac{dL}{dE} = \frac{L}{E_{max}} e^{-E/E_{max}}$, is considered.

The main GRB photons have energies larger than the ionization energies of most electrons, so we consider the energy exchange of the GRB with the gas due to Compton and inverse Compton processes only. We have used GRB spectra with $E_{max} \geq 1/6$ MeV. We take into account deviations from Thomson cross section σ_T due to Klein–Nishina correction (σ_{KN}), which are taken from [8]. We also consider the radiation force due to electron scattering.

Cooling of the gas is due to different radiative processes (ff, fb, bb). For an optically thin plasma, which is used for describing the cooling, the function C_γ was calculated in [9, 10, 6] $C_\gamma = \Lambda(T)n^2/\rho$. We have neglected the heat conductivity (see, for example, [6]).

3 Numerical Results

More than 10 variants with different density distributions and GRB beaming have been calculated in [7]. Here we present the main results of these calculations.

The GRB heating of the cloud is more intensive in the central parts and leads to the formation of a shock wave propagating outwards. The speed of the shock is about 2×10^8 cm s^{-1} for the uniform spherical cloud, much less than the light speed. Therefore, the cloud is heated mainly by the light signal from the GRB, and effects connected with the formation of the central shock do not influence the integral light curve, except for the radiation in the hard X-ray band produced in the close vicinity of the GRB explosion.

When the GRB explosion takes place between two dense clouds, or in the cavity produced by a strong anisotropic stellar wind, the hydrodynamic effects may be much stronger than in the case of the uniform cloud, and the speed of the shock increases up to $\sim 2 \times 10^9$ cm s^{-1}.

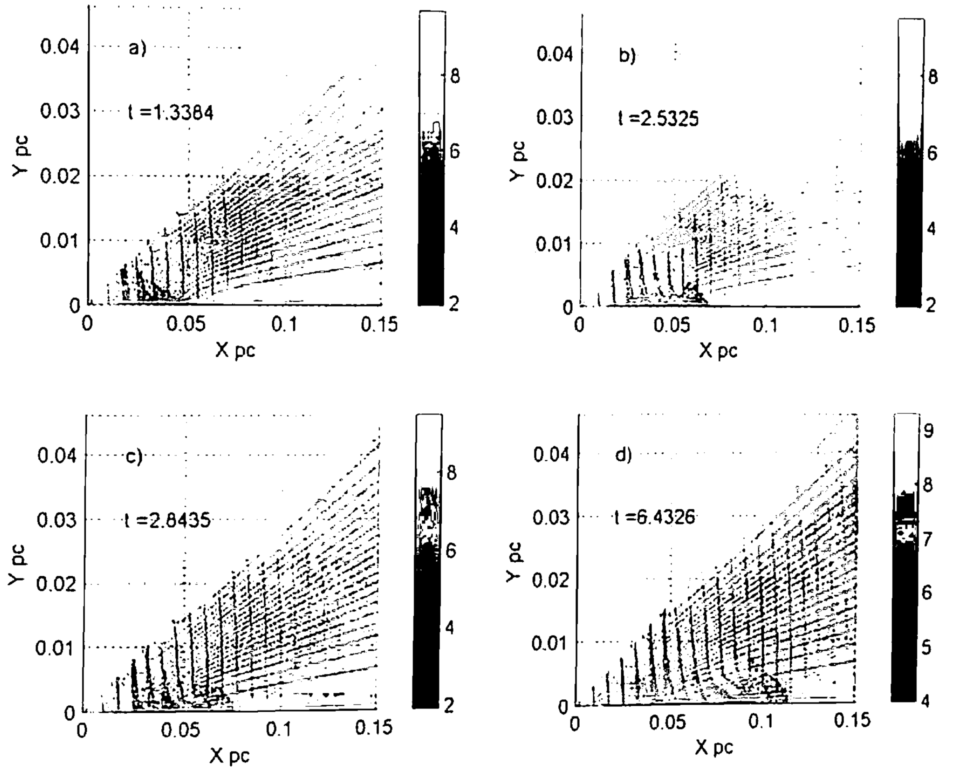


Figure 1: The evolution of the velocity field is represented for the conical density distribution in the cloud [6].

In Figure 1 (from [6]) the evolution of the velocity field is shown for a density distribution in the cloud, $n = 10^5 \times \exp[-2 - 2\cos(10\theta)]$ cm^{-3}, a radius of the cloud, $R = 1.5$ pc, energy of an isotropic GRB, $\Gamma = 1.6 \times 10^{53}$ erg, and $E_{max} \geq 1/6$ MeV. The calculations have been performed in the region $0 \leq \theta \leq \pi/10$, with the condition $v_\theta = 0$ on the outer boundary. The temperature of the heated gas depends only on the distance from the GRB, so a pressure gradient develops inside the cavity inducing the motion of matter to the axis of the cone. Collision of the flow at the cone axis produces a cumulative effect, leading to matter acceleration along the axis up to a velocity $\sim 2 \times 10^9$ cm s^{-1} [6]. The accelerated matter in this case forms bullets. If the explosion happens in the space between two spherical clouds, the ejected matter should have the form of an expanding ring.

The observed light curve of an anisotropic GRB exploding in a uniform gas cloud has different shapes for an observer at different angular distances from the symmetry axis. Such light curves are shown in Figure 2 [7]. Here an anisotropic GRB is considered with angular dependence of luminosity $\Gamma(\theta) = 10^{52} \times \exp[-(\vartheta/\vartheta_0)^2]$, $\theta = 0.1$ rad, the total energy of the GRB $\Gamma_{tot} = 2.5 \times 10^{49}$ erg, the density distribution of the cloud, $n = 10^5 \times \exp[-(r/R)^2]$ cm^{-3}, $R = 0.2$ pc, and the explosion located at 1 pc from the center of the cloud. The shortest optical burst (of a few days) with largest luminosity ($\sim 3 \times 10^{41}$ erg s^{-1}) is seen by an observer located at the symmetry axis which passes through the GRB and the center of the molecular cloud. The observer on the line which is perpendicular to the symmetry axis sees a much

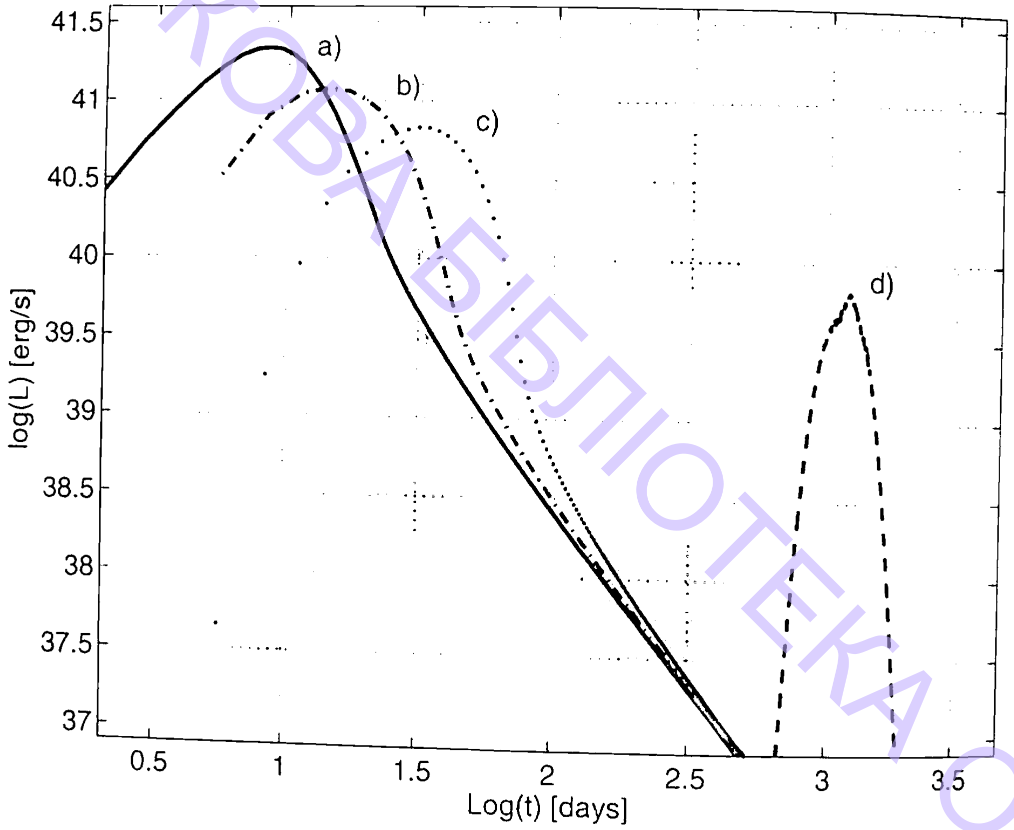


Figure 2: Light curves for a collimated GRB located at the distance of 1 pc from the center of the molecular cloud. Curves a), b), c), d) present the light curves for an observer who is located at: the line which passes through the GRB and the center of the molecular cloud (curve a), the lines which are shifted from this direction by $\alpha = 0.1$ radian (curve b), $\alpha = 0.2$ radian (curve c) and $\alpha = \pi/2$ radian (curve d).

longer optical afterglow (~ 1000 days), but with accordingly lower luminosity. All these differences are connected with the kinematics of the light propagation from the non-uniformly and non-simultaneously heated gas cloud.

An optical afterglow connected with the reradiation of the GRB by the dense molecular cloud could be observed as an optical transient. Indications that GRB explosions take place in regions of star formation filled with dense gas clouds [11, 12] support this possibility. Observation of a plateaux in the optical afterglow of GRB 030329 during a month between 64 and 94 days after the GRB detection [13] may be connected with such reradiation.

Acknowledgements

The authors would like to thank N.G. Bochkarev for useful discussions. This study was supported in part by the Russian Foundation of Basic Research (project no. 02-02-06598, 03-02-06786, 02-02-16900), and the INTAS (project no. 99-120, 00-491).

References

[1] R.W. Klebesadel et. al. *Astrophysical Jornal Letters*, **182**, 85 (1973).

[2] M.S. Briggs, *Astrophysics and Space Science*, **231**, 3 (1995).

[3] C.A. Meegan et al. *Nature*, **355**, 143 (1992).

[4] C. Kouveliotou, *The Astrophysical Journal Supplement Series*, **92**, 637 (1994).

[5] G.S. Bisnovatyi-Kogan, A.N. Timokhin, *Astronomy Reports*, **41** 423 (1997).

[6] M. V. Barkov, G.S. Bisnovatyi-Kogan, *Astronomy Reports*, **49**, 24 (2005).

[7] M.V. Barkov, G.S. Bisnovatyi-Kogan, *Astronomy Reports*, accepted (2005).

[8] A.M. Beloborodov, A.F. Illarionov, *Astrophysical Journal* **450**, 64 (1995).

[9] A.B. Kirienko, *Astronomy Letters* , **19**, 11 (1993).

[10] J. Raymond et al., *Astrophysical Journal*, **204** 290 (1976).

[11] V.V. Sokolov et. al. *Astronomy and Astrophysics*, **372**, 438 (2001).

[12] B. Paczynski, in M. Livio, N. Panagia, K. Sahu (eds) *Supernovae and Gamma Ray Bursts; The Largest Explosions Since the Big Bang, Proceedings of the Space Telescope Science Institute 1999 May Symposium 13*, Cambridge University Press, Cambridge, p.1 (2001).

[13] M.A. Ibrahimov et al. *GRB Coordinates Network*, **2191**, 1 (2003).

2

Cyg X-1 (V1357 Cyg) and Its Interstellar Environment

N.G. Bochkarev[1] and E.A. Karitskaya[2]

[1] *Sternberg Astronomical Institute, Moscow, Russia*
e-mail: boch@sai.msu.ru
[2] *Astronomical Institute, Russian Academy of Sciences, Moscow, Russia*
e-mail: karitsk@sai.msu.ru

Abstract

The first candidate to black hole Cyg X-1 is located in a sky area with very complicated structured interstellar matter (ISM) and significant interstellar absorption. We re-analyze available data on ISM absorption localization toward this object. As a result we renew Cyg X-1 distance limits. 3D-velocity vector and ISM spatial distribution reconstruction permit us to limit black hole age and birth place of this binary system.

Keywords: X-ray binary, black hole, interstellar matter, distance

1 Introduction

The X-ray source Cyg X-1 was discovered by UHURU about 35 years ago and was soon identified with the star HDE226868. This bright Northern sky source attracted much attention from observers. In spite of the mass function low value, absence of X- ray eclipses and small amplitude of optical variability, it was soon understood that the X-ray component of the binary system was historically the first candidate to be a black hole. Because of this fact the interest in this object remains very high up to now.

The system optical component is a supergiant 09.7Iab. Its radiation dominates in the optical spectral range. The input of the accretion structure is estimated as not exceeding 4%, which allows us to use the color properties of the star as a tool for determining the interstellar absorption along the way from the system. The color excess E(B-V) was found to be 1.09 ± 0.03^m.

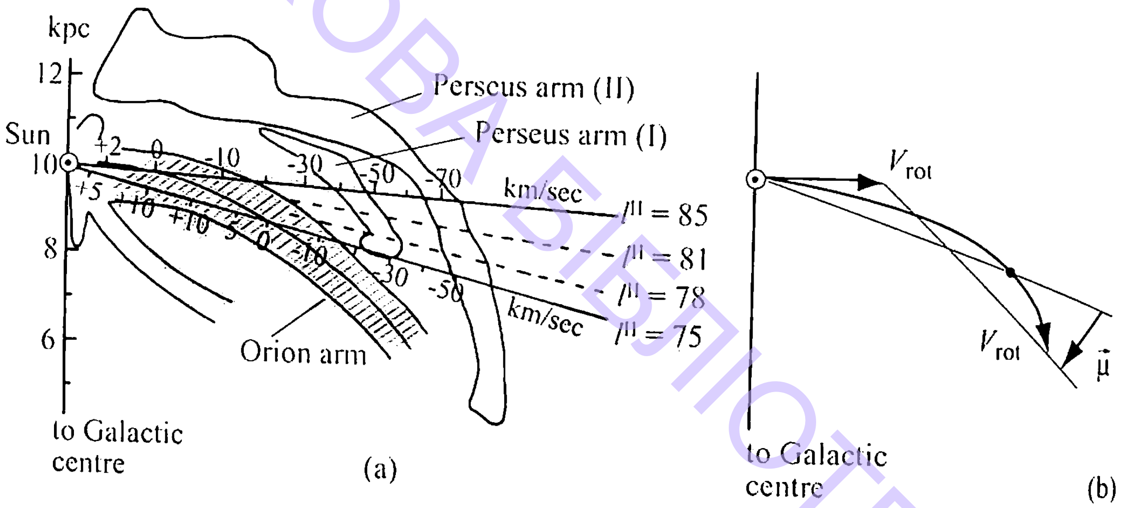


Figure 1: a) Position of the local (Orion) spiral arm derived from observations of the 21 cm HI line [6]. b) Cyg X-1 proper motion due to Galactic rotation.

The studies of interstellar matter distribution in the direction of Cyg X-1 yield a distance d =2–2.5 kpc, nevertheless recent work leads to a shorter one. The system parameters determined from photometric and spectral data [1, 2] correspond to the object luminosity consistent with a distance as short as about 1.5 kpc. This encouraged us to attempt to revise the distance to Cyg X-1 on the basis of the distribution of the interstellar matter in the direction to it.

2 Distance Limits for Cyg X-1

The Cygnus constellation has a galactic longitude of about 90 deg. So the kinematic methods for distance determination can not be used. The line of sight goes along the local (Orion) spiral arm (see Figure 1a).

Bochkarev and Sitnik [3] show that this area of the sky is rich with starformation centers, gas-and-dust structures etc. located at various distances d (Figure 2). We use Lucke [4] map of the distribution of absorbing matter at $d \leq 2$ kpc (Figure 3) to put a low limit d to Cyg X-1. Integration of E(B-V) over d in the direction to Cyg X-1 ($l = 71.34$, $b = 3.07°$ deg) shows that at $d \leq 1.8$ kpc the value of E(B-V) $\leq 0.55 \pm 0.05^m$.

The line of sight meets the star forming region Cyg OB3 where E(B-V) abruptly increases. The center of the region is located at a distance of 2.3 kpc [5], determined by the stellar luminosity method. Therefore the Galactic distances scale revision of the last decades cannot change such distance estimations.

So we conclude that Cyg X-1 distance $d > 1.8$ kpc. The most probable value equals 2.3 kpc. This implies that in a number of cases (e.g. see above mentioned papers) the spectrophotometric distance estimations based on the characteristics of the stellar atmosphere lead in the case of Cyg X-1 to significant errors.

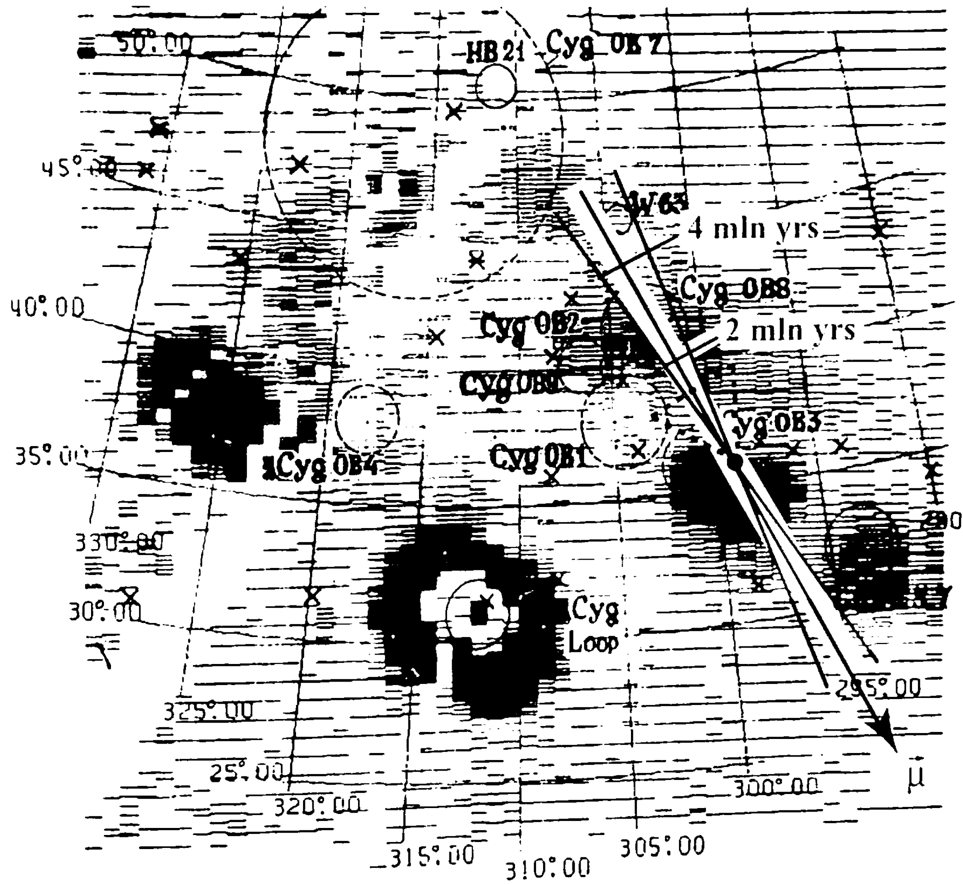


Figure 2: HEAO1 map of soft X-ray distribution in the Cygnus region (Figure from [3]). Solid circles on the X-ray picture mark supernova remnants, dash circles indicate the boundaries of stellar associations, and crosses mark the X-ray sources. Cyg X-1 proper motion μ is shown with $\pm\sigma$ uncertainty.

The upper limit for Cyg X-1 distance can be found by several methods. The low radial velocity – 2–10 km/s [2] is in agreement with the Galactic rotation curve up to about 4.5 kpc. However the Galactic latitude $b = 3.07°$ means that at this distance the altitude above the Galactic plane is $z = 240$ pc. There is no reason for it to be so high – massive stars are not born at such z. Cyg X-1 is not a "blue struggler", because its proper motion (see next section) has a very small component normal to the Galactic plane, and Cyg X-1 can not increase z more than 30–50 pc. The semi-thickness of massive star distribution does not exceed 120–150 pc. So the distance to Cyg X-1 can not be more than about 3.5 kpc. Moreover the luminosity is close to 10^6 $L_\odot$. It corresponds to the brightest stars in the Galaxy. So 3.5 kpc should be considered as the upper limit for the Cyg X-1 distance.

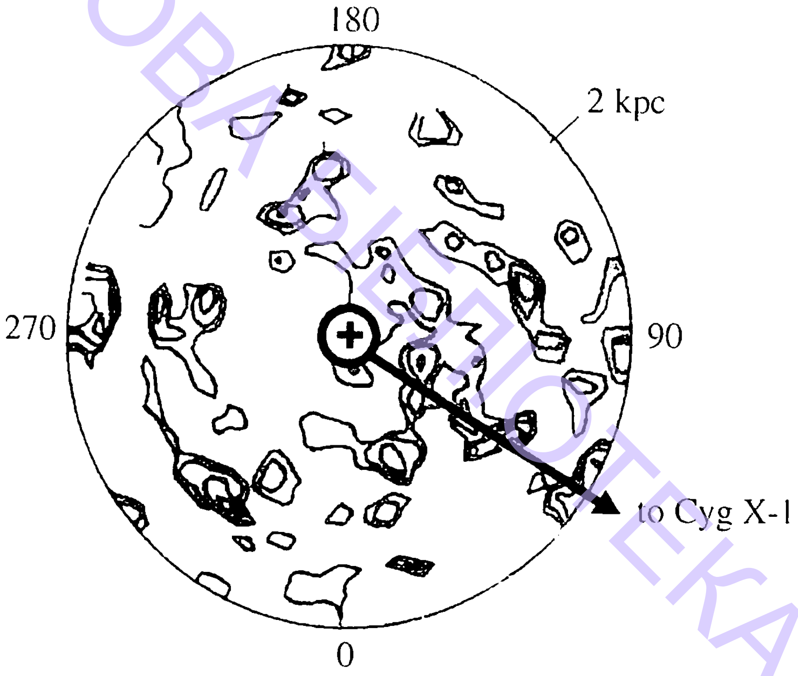


Figure 3: Distribution of obscuring matter in the galactic plane [4] in the region inside 2 kpc from the Sun. The contours represent lines of equal values $dE(B-V)/dr = 0.5, 1.0, 2.0$, and 4.0 mag · kpc^{-1}.

3 Proper Motion and Birth Place of Cyg X-1

HDE226868 proper motion $\vec{\mu}$ as measured by Hippacus is: $\mu_\alpha = -0.00382 \pm 0.00079''$; $\mu_\delta = -0.00762 \pm 0.00091''$. So Cyg X-1 $\vec{\mu}$ is determined at the level of 8 sigma. It is oriented almost exactly parallel to the Galactic plane and for a distance of about 2 kpc Cyg X-1 velocity is 100 km/s, which corresponds to the Galactic rotation. In contrast to the small radial velocity toward Cyg X-1 (Figure 1a) the Galactic rotation gives rather high tangential velocity (Figure 1b).

The average proper motion of CygOB3 stars coincide with that of Cyg X-1 within the accuracy of measurements. It is additional evidence that Cyg X-1 is probably a member of this stellar association. However taking into account the uncertainty of Cyg X-1 proper motion it is impossible to exclude the possibility that 5 million years ago Cyg X-1 black hole was born in CygOB1. Moreover in this case the distance to Cyg X-1 is the same as to CygOB1, namely 1.8 kpc. Because the tangential component of the Galactic rotation velocity at this distance is smaller by about 20% in comparison to CygOB3, Cyg X-1 has a tangential velocity relative to CygOB1 of about 20 km s^{-1}. This velocity is enough for traveling from the birth place to today's position on the sky in 5 Myr. According to [5] CygOB1 age is 7.5 Myr.

We can exclude any other close associations as its origin (Figure 2): Cyg X-1 travel time from any of them exceeds their ages (Table 1).

Table 1: Stellar associations in Cygnus region.

Association	*CygOB3*	*CygOB1*	*CygOB8*	*CygOB9*	*CygOB2*
Distance (kpc)	2.3	1.8	2.3	1.2	1.8
Age (x10^6 yr)	8.3	7.5	3.0	8.0	5.0

4 Conclusions

The most probable position of Cyg X-1 is CygOB3 association. It means that:

a) Cyg X-1 distance $d \approx 2.3$ kpc;

b) it was born in this association (not more than 8 Myr ago);

c) its velocity relative to CygOB3 is rather low, so there was no strong recoil effect during Cyg X-1 evolution.

Another less probable scenario is:

a) Cyg X-1 was born in CygOB1 association (no more than 7.5 Myr ago);

b) Cyg X-1 distance $d \approx 1.8$ kpc;

c) not less than about 5 Myr ago the system acquired a velocity of about 20 km · s^{-1} relative to CygOB1 (we assume that it was the recoil effect from the black hole creation).

Acknowledgements

This research was supported by grant RFBR 04-02-16924 and the Russian Federal Program "Astronomy". We are thankful to Prof. J. Zhilkowski for useful discussions.

References

[1] A. Herrero, R.P. Kudritski, R. Gabler, *et al.*, *Astron. Astrophys.*, 297, 556 (1995).

[2] M. Abubekerov, E. Antochina, A.M. Cherepashchuk, *Astron. Rep.*, 48, No. 7, 550 (2004).

[3] N.G. Bochkarev, T.G. Sitnik, *Astrophys. Space Sci.*, 108, 237 (1985).

[4] P.B. Lucke, *Astron. Astrophys.*, 64, 367 (1978).

[5] R.M. Humphreys, *Astrophys. J. Suppl. Ser.*, 38, 309 (1978).

[6] W.H. McCutcheon, W.L.H. Shutter, *Astron. J.*, 75, 910 (1970).

3

Star Complexes and Starburst Clumps in Spiral Galaxies

Yu.N. Efremov

Sternberg Astronomical Institute, Moscow, Russia
e-mail: efremov@sai.msu.ru

Abstract

Star complexes are the highest level groupings in the hierarchy of the embedded young stars, clusters and associations, which obey the size – age relation. Starburst clumps, superassociations, supergiant HII regions are different names for groupings of the same size as complexes but with active star formation over all the grouping. The coherent star formation in the regions of violent star formation was probably triggered by external pressure. The issue of gravitational boundness of complexes and superassociations is briefly discussed.

Keywords: large scale star formation, clusters, associations, star complexes, structure of galaxies

1 Introduction

The term “star complexes” was suggested in 1978 to designate the largest (up to 1 kpc) groupings of stars and clusters which unite objects with ages up to 100 million years; they were first identified in the solar neighbourhood; one of these was the Gould Belt system. It soon became evident that these groupings are the same objects which have been known for a long time as knots and clumps in the spiral arms of external galaxies (see [1] and references therein).

A mechanism for coherent star formation within such large groupings was suggested by the Elmegreens [2] who found that the regular spacing between “the giant HII regions” (complexes) along the spiral arms is explained by them being originated from HI/CO superclouds formed by the gravitational instability along the arm.

Star complexes are the largest and oldest objects in the hierarchy of the young embedded groupings, which starts from multiple stars. The younger and smaller clusters are always within older and larger ones. This hierarchy is similar to the one observed in the interstellar gas distribution and in fact is a result of the latter. This similarity as well as a correlation between the ages of the oldest stars and sizes of groupings point to turbulence as an important factor in the formation of star groupings [3, 4].

Sometimes a galaxy hosts one or two very bright complexes, which consist of OB-associations and HII regions. Long ago they were called superassociations and represent local starbursts. The coherent and effective star formation in such starburst clumps was evidently triggered by some event which led to high pressure in the ISM.

2 Star Complexes in the Spiral Arms

Chains of star complexes often delineate the spiral structure in grand design galaxies. This is observed in M31 and seemingly also in the Milky Way. The distribution of Cepheids around the Sun demonstrates a strong concentration in the local segment of the Sgr–Car arm, where they form a few vast groupings with separations of around 1 kpc [5]. The chain of HI/CO superclouds along the 40 kpc long Car–Sgr arm shows similar regular spacing [6].

Regular spacing of "giant HII regions" along the arm was noted by the Elmegreens [2] in a sample of 22 galaxies. Such a regularity (although less pronounced) is observed in many other galaxies, but mostly in one arm only, like NGC 5371, NGC 4321 (M 100) and NGC 4258.

In galaxies with a high rate of star formation, like NGC 6946 and M83, it is difficult to pick-up the individual complexes from the bright arm background; it is almost as if complexes fill all the arm without space between them. It is worth noting that long spurs start outwards from the largest complexes in M31 and M51.

The regular spacing of the complexes is assumed to be mainly determined by the joint action of magnetic and gravitational instabilities. In NGC 6946 the spiral arms of the regular magnetic field are located between the optical ones. This could be due to the high SFR that disorganize the magnetic lines within the arms, abundant in HII regions [7]. The leading role of the Rayleigh–Taylor instability in the formation of complexes is supported by the wave-like appearance of some arms (including the MW and M 31) in the Z-direction [8]. It is worth noting that regular field along the optical spiral arms is observed in M31 and MW – and both galaxies present regular spacing of complexes in the long segments of the arms.

At any rate, the regular spacing of complexes is difficult to reconcile with formation of superclouds resulting from random collisions of smaller clouds. Furthermore, the latter is a precess about 3 times slower than the action of gravitational instabilities [9].

The short segmented arms of flocculent spirals are star complexes themselves – sheared by the differential galactic rotation [10].

3 Star Complexes and Superassociations

Sometimes very bright and large blue complexes are observed in the arms. Many of them were noted as superassociations (SAs), or starburst clumps (see [11] and references therein). They do not seem to be just the extreme cases of the general distribution of stellar complexes in age and size (and therefore in luminosity). Most SAs are much brighter than the second complex in any individual galaxy, though in the general distribution over a number of galaxies SAs might not be separated by a gap from the lower luminosity complexes. This important issue is still open.

Anyway, in the nearest SAs, 30 Dor in the LMC and OB78 in M 31, Cepheids and red supergiants are almost absent. Thus there are no objects older than 30 Myrs. The present day star formation rate is low in 30 Dor (high density CO clouds are only found at the South-West of the SA) and more so in OB78, which is inside an HI superhole.

This means that the present day small range of ages will be preserved in such SAs for ever. We are probably witnessing an example of such aged SA at the East tip of the LMC bar, SE of the cluster NGC 2065, where within a region 300 pc in size about 180 Cepheids are observed, showing a small age range (~40 Myr) obtained from the period – age relation. The density of Cepheids there is 100 times higher than around the Sun, but there are no clusters! About 100 Myr ago the region might have looked like a small burst of star formation; a similar situation to that of OB78.

If so, it follows that during a time as long as half of the period of the Sun orbital motion a compact SA might not have dissipated yet. The strong concentration of complexes of about 100 Myr old within the Car–Sgr spiral arm near the Sun may imply that their destruction time (due to the shearing from the galactic rotation outside the arm) is longer. Otherwise the age of the oldest stars in these complexes may indicate that the local segment of the Car–Sgr arm is close to corotation and the time needed for crossing the arm is then quite large.

However, complexes formed from virialized superclouds must be initially bound and then we might expect to see older complexes outside the arms. After 100 Myr the brightest stars there are too dim to permit the detection of such complexes. Also, they must have been bound initially only with the parent gas, which since then might have been blown away. The gas superclouds and star complexes in an arm are usually close to each other, but do not coincide.

It looks like SAs are not just the young versions of common complexes. The small age range implies that star formation there was triggered by some source of external pressure, like a shock wave. The high pressure favours massive clusters to be formed bound [12]. However, the existence of SAs without clusters suggests another factor: difference in turbulence parameters at the scale of the whole complex might be acting too. All in all, it looks as if star groupings coherently formed from a supercloud might have different fate, depending on their location and on the efficiency of star formation.

4 Complexes of Star Clusters

Clumps of clusters have been long known in the LMC. The most impressive one is a concentration of a dozen clusters arond NGC 2164, four of these being within 200 pc, having masses around 10^5 $M_\odot$ each. If the velocity dispersion is ~10 km/s, the virial mass of this cluster of clusters should be compatible with the observed mass. Such a group might have been formed from a single gas supercloud, even more as the ages of the clusters are rather similar.

This cluster of clusters is at the outskirts of the LMC, but more often the massive bound clusters are formed within high pressure surroundings. They exist, for example, in the tight clump of three rich clusters found recently within the brightest HII clump in the young star complex NGC 5461 in M 101 [13].

The high pressure conditions explain the numerous occurences of bright compact and presumably bound clusters in interacting galaxies and around the centers of some spiral galaxies. Especially vast complexes of such clusters are known in NGC 4038/39 (Antennae) and have been more recently discovered in NGC 7673 (Mrk 325) and NGC 6621/22 (see references in [11]).

However a few such clusters are also known in normal spiral galaxies – but always within complexes. Two such complexes are known in M 51. The peculiar complex in NGC 6946 is a local starburst, it contains ~20 young clusters, one of them a real supercluster [14]. Signs of a HVC impact are observed near the cluster [15]. This enigmatic structure is under investigation. It looks quite similar to the BCD galaxy NGC 1705, which also hosts a supercluster.

The brightest giant HII region in M 101 (NGC 5471) is located outside the main body of the galaxy; it is round and has low abundance [13]. It also strongly resembles a small BCD galaxy. The similarities between these galaxies and supergiant extragalactic HII regions are well known and may be quite significant.

Acknowledgements

The support from grants RFBR 03-02-16288 and SciShc 389.2003.2 is appreciated.

References

[1] Yu.N. Efremov, *Astron. J.*, **110**, 2757 (1995).

[2] B.G. Elmegreen, D.M. Elmegreen, *MNRAS*, **203**, 31 (1983).

[3] B.G. Elmegreen, Yu.N. Efremov, *Astroph. J.*, **466**, 802 (1996).

[4] Yu.N. Efremov, B.G. Elmegreen, *MNRAS*, **299**, 588 (1998).

[5] L.N. Berdnikov, Yu.N. Efremov, *Astron. Lett.*, **19**, 389 (1993).

[6] Yu.N. Efremov, *Astron. Astroph. Trans.*, **15**, 3 (1998).

[7] P. Frick, R. Beck, A. Shukurov et al., *MNRAS*, **318**, 925 (2000).

[8] J. Franco, J. Kim, E. Alfaro et al., *Astroph. J.*, **570**, 647 (2002).

[9] B.G. Elmegreen, *Astroph. J.*, **357**, 125 (1990).

[10] B.G. Elmegreen, D.M. Elmegreen, S.N. Leitner, *Astroph. J.*, **590**, 271 (2003).

[11] Yu.N. Efremov, *Astrofisica*, **47**, 319 (2004).

[12] B.G. Elmegreen, Yu.N. Efremov, *Astroph. J.*, **480**, 235 (1997).

[13] C.H.R. Chen, Y.-H. Chu, K.E. Johnson, preprint astro-ph/0410240 (2004).

[14] S. Larsen, Yu.N. Efremov, B.G. Elmegreen et al., *Astroph. J.*, **567**, 896 (2002).

[15] Yu.N. Efremov, S.A. Pustilnik, A.Y. Kniazev et al., *Astron. Astrophys.*, **389**, 855 (2002).

4

Optimized Photoionization Models of the Planetary Nebula NGC 6720

N.V. Havrylova, V.V. Holovatyy and B.Ya. Melekh

Department of Astrophysics, Ivan Franko National
University of Lviv, 8 Kyryla and Mephodiya st., Lviv, 79005,
Ukraine
e-mail: havrylova@physics.wups.lviv.ua
gol@astro.franko.lviv.ua
melekh@physics.wups.lviv.ua

Abstract

Optimized photoionization models of the planetary nebula NGC 6720 taking into account inhomogeneities of the nebular gas density in its envelopes are calculated using Ferland's Cloudy96 beta5 code and van Hoof's PHYMIR. The algorithm of the optimized calculation is based on changing the value of the free parameters by a χ^2-function minimization. The chemical compositions obtained as a result of different types of optimized modelling are compared with previously calculated compositions.

Keywords: planetary nebulae, photoionization models, chemical composition

1 Introduction

The determination of planetary nebulae (PNe) chemical composition is important for studying chemical evolution of matter in the Galaxy. It is well known that PNe are formed from the late evolutionary stages of intermediate mass stars ($1 \leq M/M_{sun} \leq 8$), and their chemical composition reflects the nucleosynthesis processes that occur in these stars.

Some dozen photoionization codes exist to study the ionization and temperature structure of nebulae in order to obtain their physical parameters and chemical composition. The most popular code is Cloudy (see, for example, http://www.nublado.org). It has been developed by Ferland and co-workers and was tested for many different physical conditions. P. M. van Hoof [1] derives simultaneously and self-consistently some physical parameters of PNe from a set of observed quantities using the 84.06 version of Cloudy. In this method it is assumed that there exists a unique set of parameters, for which the resulting model predictions give the best fit to the observables. To compare the model predictions with the observed quantities, a goodness-of-fit estimator was calculated. This estimator is minimized by varying all the input parameters of the model, using an AMOEBA algorithm [2]. This method has been applied to a sample of five galactic bulge PNe.

We decided to modify this method, in order to take into account inhomogeneities of nebular gas density in the PN envelopes, and use it for accurate determination of PNe physical parameters and chemical composition.

2 Optimized Models Calculation: the Method and Main Results

The calculation of optimized photoionization models (OPhM) is based on finding optimal values of the free parameters which are selected from all input parameters of the model. The output model parameters in the case of our PNe OPhM are: selected relative lines intensities and luminosity of H_β. The criterion of the agreement between predicted and observed spectra is the χ^2-function, that can be written for any i observed and corresponding model quantity as:

$$\chi_i^2 = (\frac{Obs^i - Mod^i}{\sigma^i})^2, \tag{1}$$

where Obs^i and Mod^i are the observed and model values respectively and σ^i is the absolute error in the observed value. To evaluate the full value of χ^2 in the PN OPhMs we used such output parameters as luminosity in H_β, $L(H_\beta)$, and relative intensities of 14 lines. The logarithm of $L(H_\beta)$ absolute errors were in the range 0.05–0.35, the relative intensities absolute errors were around 3–15%. The algorithm of the optimization calculation is based on changing the free parameters value by a χ^2-function minimization. For our PNe OPhM calculations we used Cloudy96 beta5 (http://www.nublado.org) and PHYMIR [1].

The nebular shell in our models has been assumed to be spherically symmetrical. The radial distribution of the gas density was given by the empirical relation [3]:

$$n\left(r\right) = \frac{x^2\left(1 + 3e^{-1.2x}\right)}{\left(x^2 - 1\right)^2 + 0.36 r_c^{-0.43} x^2} \frac{A}{r_c^2} \text{ cm}^{-3}, \tag{2}$$

where $x = r/r_c$ and $r_c = \langle V_{\rm exp} \rangle t$. Here r is the distance from the central star to a particular point of the nebula, n is the gas density at this point, $\langle V_{\rm exp} \rangle$ is the mean expansion velocity of the shell (about 15 km/s), and t is the shell age. The parameter A is close to 1.6 for most of the PNe with low-mass central stars and increases with the masses of the stars.

The free parameters in our PNe OPhM were the following: 1) full number of ionizing photons Q_{tot} (47.66; 47.66; 47.66); 2) maximum hydrogen density (see eq. 2) in the PN envelope n(H) (3.5; 4.0; 4.5); 3) filling factor ff (−1.0, −1.0, 0.0); 4) chemical abundances of He/H (−1.09; −0.97; −0.90), 5)C/H (−3.34; −3.04; −3.04), 6) O/H (−3.47; −3.17; −3.09), 7) N/H (−3.82; −3.49; −3.01), 8) Ne/H (−4.21; −3.91; −3.44), 9) S/H (−5.49; −5.19; −4.57), 10) Ar/H (−5.92; −5.62; −5.15), 11) Si/H (−5.66; −5.36; −4.89). The logarithm of the input values of the free parameters for the "minimum", "average" and "maximum" initial sets are given in parentheses. The adopted ranges for changes of such free parameters as chemical abundances were three times smaller and three times larger from the input values; other free parameters had no limits. We constrain the change range of these values using real PNe data.

The values of the free parameters derived from the OPhMs for the PN NGC6720 are presented in Table 1. It is shown, that for different sets of initial values of free parameters (average, maximum and minimum initial values) we find the same χ^2-minimum. We assume that all spectra lines of the same ion in our models are not independent, a particular ionic abundance can be found using only one line of the ion. In order to check the validity of this assumption two models with 1 line per ion were calculated and it is shown that the freedom degree of our task is equal to 1 (12 really independent input (free) parameters minus 11 output parameters). Also we calculate 2 models with covering factor smaller than 1 (0.5 and 0.1). Only the PN model with quite small covering factor (0.1) has another χ^2-minimum.

Table 1: Values of the free parameters, derived from the optimized photoionization models for NGC6720. Here "cf" means covering factor.

Parameters	Optimized photoionization models for NGC 6720					
	average	max	min	line per ion	cf=0.1	cf=0.5
$n_H\,10^4$, cm^{-3}	$1.66^{+0.71}_{-0.52}$	$2.11^{+0.99}_{-0.73}$	$1.32^{+0.50}_{-0.33}$	$1.75^{+0.55}_{-0.40}$	$0.68^{+0.63}_{-0.15}$	$1.54^{+0.70}_{-0.33}$
$\log(Q_{tot})$	$45.7^{+0.02}_{-0.02}$	$45.7^{+0.02}_{-0.01}$	$45.7^{+0.02}_{-0.02}$	$45.7^{+0.02}_{-0.02}$	$45.8^{+0.01}_{-0.01}$	$45.7^{+0.02}_{-0.01}$
ff	$0.25^{+0.07}_{-0.05}$	$0.22^{+0.05}_{-0.04}$	$0.30^{+0.08}_{-0.07}$	$0.22^{+0.06}_{-0.04}$	$0.34^{+0.06}_{-0.00}$	$0.26^{+0.06}_{-0.06}$
He/H*10^2	$9.33^{+1.22}_{-1.21}$	$8.81^{+1.53}_{-0.79}$	$9.36^{+1.05}_{-1.22}$	$9.79^{+0.85}_{-1.55}$	$8.58^{+2.49}_{-0.57}$	$9.39^{+1.31}_{-1.31}$
O/H*10^4	$6.26^{+0.77}_{-0.63}$	$7.16^{+0.61}_{-0.82}$	$5.81^{+0.67}_{-0.59}$	$6.45^{+0.61}_{-0.62}$	$4.69^{+0.00}_{-0.04}$	$5.87^{+0.82}_{-0.59}$
N/H*10^4	$0.99^{+0.16}_{-0.13}$	$1.03^{+0.14}_{-0.20}$	$0.98^{+0.17}_{-0.15}$	$0.98^{+0.15}_{-0.07}$	$0.81^{+0.31}_{-0.03}$	$0.93^{+0.20}_{-0.13}$
Ne/H*10^5	$7.22^{+2.09}_{-2.05}$	$7.58^{+2.05}_{-2.45}$	$6.85^{+2.87}_{-2.10}$	$7.13^{+1.14}_{-1.81}$	$6.88^{+2.64}_{-1.91}$	$7.16^{+1.39}_{-2.20}$
S/H*10^6	$2.00^{+0.46}_{-0.44}$	$2.06^{+0.86}_{-0.50}$	$1.92^{+0.65}_{-0.57}$	$1.72^{+0.61}_{-0.18}$	$1.51^{+0.62}_{-0.42}$	$2.09^{+0.43}_{-0.49}$
Ar/H*10^6	$1.66^{+0.42}_{-0.48}$	$1.71^{+0.55}_{-0.55}$	$1.77^{+0.48}_{-0.74}$	$1.63^{+0.59}_{-0.50}$	$1.63^{+0.62}_{-0.58}$	$1.73^{+0.44}_{-0.44}$
C/H*10^4	$4.25^{+2.24}_{-0.97}$	$7.7^{+1.33}_{-3.19}$	$3.1^{+1.66}_{-1.05}$	$4.72^{+1.92}_{-0.67}$	$1.99^{+0.31}_{-0.85}$	$3.75^{+1.60}_{-0.84}$
Si/H*10^6	$1.71^{+0.65}_{-0.23}$	$2.28^{+0.57}_{-0.76}$	$1.55^{+0.41}_{-0.07}$	$1.81^{+0.23}_{-0.157}$	$1.69^{+0.23}_{-0.10}$	$1.74^{+0.38}_{-0.24}$

The observed and calculated spectra for NGC 6720 models are presented in Table 2. It is shown that the largest value of the χ^2-minimum occurs for small covering factor (0.1). So this result confirms that in reality NGC 6720 (the Ring

Nebula) has spherically structure. Also we conclude that the best set of initial values of the free parameters is the "average" set with χ^2-minimum value 0.930.

Our next task is to use this method for OPhMs calculation of a galactic PNe sample. Using this method, independent of ionization-correction factors, we can revise our previous results and accurately determine the main physical parameters and chemical composition of PNe.

Table 2: Relative intensities and luminosities $L(H_\beta)$ of observed and calculated spectra for NGC6720 OPhMs ($I(H_\beta) = 100$). Here "cf" means covering factor.

Lines	Observations	Optimized photoionization models					
		max	average	min	1 line	cf=0.1	cf=0.5
3727 [O II]	537.00 ± 26.9	534.13	537.47	548.31	538.87	611.54	534.22
3869 [Ne III]	138.00 ± 20.7	136.40	141.61	140.04	137.76	156.71	146.54
4363 [O III]	11.40 ± 1.71	12.49	13.84	14.83	–	14.68	14.41
4471 He I	4.31 ± 0.43	3.81	4.00	3.93	4.09	3.45	4.01
4686 He II	30.30 ± 3.03	33.53	34.58	35.01	35.01	34.72	35.10
5007 [O III]	1161.0 ± 34.8	1157.6	1152.9	1175.4	1154.3	1049.1	1139.5
5876 He I	11.10 ± 1.11	11.74	12.37	12.15	–	10.55	12.41
6678 He I	3.70 ± 0.37	3.01	3.15	3.10	–	2.71	3.15
6584 [NII]	383.0 ± 19.2	388.89	385.86	383.65	382.69	345.76	372.44
6716 [S II]	11.40 ± 3.06	13.53	14.70	15.09	–	15.36	16.26
6731 [S II]	20.40 ± 3.04	22.82	23.82	23.53	20.36	21.06	25.92
7135 [Ar III]	20.30 ± 3.58	24.11	24.41	26.46	23.91	25.79	26.08
1891 [Si III]	1.70 ± 0.25	1.67	1.64	1.71	1.68	2.27	1.86
1549 [C IV]	54.00 ± 8.10	53.74	53.70	58.16	52.91	63.13	60.09
$\log[L(H_\beta)]$	33.43; 32.43*; 33.13**	33.42	33.43	33.43	33.43	32.44	33.13
χ^2	–	1.099	0.930	1.147	0.274	3.669	1.243

* $\log[L(H_\beta)]$ for a model with covering factor 0.1;
** $\log[L(H_\beta)]$ for a model with covering factor 0.5.

References

[1] P.A.M. van Hoof, *Photo-Ionization Studies of Nebulae*, Ph.D. thesis, Rijksuniversiteit, Groningen (1997).

[2] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes*, Cambridge Univ. Press, Cambridge (1986).

[3] V.V. Golovatyi, V.L. Dmiterko, Yu.F. Mal'kov, O.V. Rokach, *Astron. Rep.*, **37**, 346 (1993).

5

Optimized Photoionization Modeling with Dust Grains of HII Regions in Blue Compact Dwarf Galaxies

B.Ya. Melekh and V.V. Holovatyy

Ivan Franko National University of Lviv, Ukraine
e-mail: melekh@physics.wups.lviv.ua

Abstract

A new method to determine the physical characteristics and chemical composition of HII regions is proposed. The Ly$_c$ spectrum is calculated independently of the Initial Mass Function method. Taking into account dust grain presence we used this method for the determination of the Ly$_c$ spectrum, physical characteristics and chemical composition of HII regions in the low metallicity blue compact dwarf galaxy SBS 0940+544. The influence of dust grains on Optimized Photoionization Model (OPhM) calculations low metallicity HII regions is analyzed based on the obtained results.

Keywords: blue compact galaxies: individual (SBS 0940+544), HII region, optimized photoionization modeling, dust grains

1 Introduction

We propose a new method for the determination of the chemical composition and physical characteristics of HII regions. This method is based on the calculation of OPhMs of HII regions and on the recalculation and optimization of the Ly_c spectrum using iterations. This method is independent of ionization correction factors and is oriented to reproducing HII regions observed spectra.

Photoionization models (PhM) for HII regions need to define the energy distribution in the ionizing spectra of nebular nuclei beyond $\lambda \leq 912$ Å (Ly$_c$ spectrum). Presently the Initial Mass Function (IMF) of the HII region nucleus in most cases is used for this purpose [1]. But the IMF parameters for most real HII regions are unknown. For the calculation of the Ly$_c$ spectrum we used instead a method described by Golovaty & Pronik [2] modified for the case of HII regions [3] by including four equations for the dependence on the effective temperature of the stars [4] of the Ly$_c$

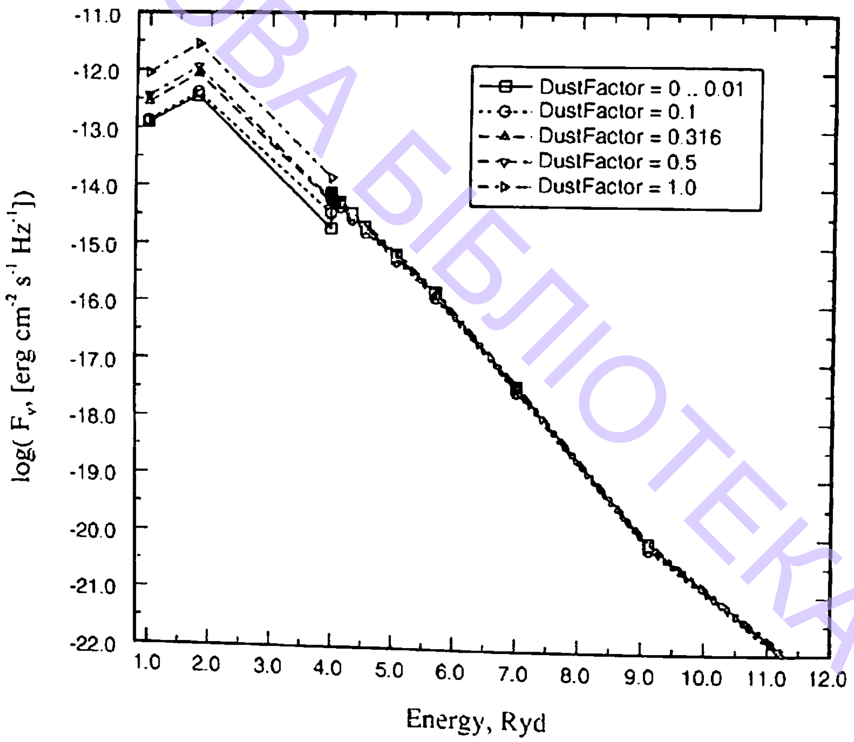


Figure 1: Ly$_c$-spectra obtained for different Grain Factor values.

spectrum shape in the ranges λ912–504 Å and λ228–22.8 Å derived from stellar atmosphere models [5] for O-B stars (SdK) with metallicity $Z = 0.004$. A jump at λ 504Å and the shape of the Ly_c spectrum in the wavelength interval λ228–22.8 Å can be determined using the SdK model that is closest to their temperature.

2 Description of the Calculation Method

We have written a code NLEHII (Nebula Light Excitation: HII regions) in C++ programming language for fast calculation of the Ly$_c$ spectrum. Every spectrum obtained is optimized for the number of ionizing photons beyond $\lambda \leq 228$ Å in order to reproduce the ratio of observed line intensities Iλ4686HeII/I(H$_\beta$).

We use a χ^2-function for determining the agreement between OPhM predicted and observed emission lines intensity. For our PhM calculation we used G. Ferland's code CLOUDY 96 beta 5 [6] with optimization function PHYMIR [7].

The free parameters in our OPhMs were the following (brackets give the allowed ranges for optimization): 1) the inner radius R_{in} of the HII region $\log R_{in} \in [15..20.98]$; 2) the number of ionizing photons $Q_{tot} \in [47..58]$; 3) the hydrogen density at the inner radius of the HII region $n_H(R_{in}) \in [50..1000]$; 4) the exponent for the hydrogen distribution over the HII region radius (a power-law was adopted) $\gamma \in [-2..0]$; 5) the filling factor $\epsilon \in [0.00001..1]$ 6) the chemical abundances relative to H of $He \in [0.05..0.11]$ and of the following elements, 7) $O \in [(0.8..14)10^{-5}]$, 8) $N \in [(0.3..6.5)10^{-6}]$, 9) $Ne \in [(2..40)10^{-6}]$, 10) $S \in [(0.6..4)10^{-6}]$, 11) $Ar \in [(0.8..8)10^{-7}]$, 12) $Fe \in [(0.0001..15.8)10^{-6}]$.

The output parameters used for the calculation of χ^2-function are H$_\beta$ luminosity, the relative intensities of 21 lines and the outer radius of the HII region obtained in [9].

For calculating the underlying stellar absorption in the OPhM we used recent data [8] for HeI recombination lines in which collisional enhancement and radiative transfer in the HeI lines are taken into account. For the analysis we adopted the intensities of 6 HeI lines: λ3889Å, λ4026Å, λ4471Å, λ5876Å, λ6678Å and λ7065Å.

The input parameters for the Ly_c-spectrum were obtained from the PhM results, via an iterative process using OPhMs.

3 Main Results and Conclusions

We applied the method to the observed spectrum of the northwest brightest HII region in BCDG SBS0940+544 [9]. The observed value of the equivalent width of H$_\beta$ in emission, EW(H$_\beta$), was used for determining the covering factor value.

In order to investigate the effect of dust grains we calculated the Ly$_c$ spectrum, physical parameters and chemical composition in the above HII region using different values characterizing the grains. We adopted as basic model, the InterStellar Medium (ISM) grain abundance described in [6]. Different grain abundances for the OPhMs were achieved by multiplying the ISM grain abundance by different Grain Factor values.

In Figure 1 the Ly$_c$-spectra obtained for different values of Grain Factors are shown. As we can see from the figure all Ly$_c$-spectra are similar beyond $\lambda < 228$Å where they start to differ.

Figure 2 shows the dependence of two free parameters on the chosen Grain Factor value. The horizontal lines correspond to the minimal (lowest dotted line), maximal (highest dotted line) and optimal value (solid line) of Q_{tot} and O/H obtained without grains. As we can see from these Figures the values of most free parameters at Grain Factors ≥ 0.01 differ from those from models without grains. As O/H abundance is in both gas and grain phases the corresponding abundance in gas phase is indicated by filled circles and full (gas + grain abundance) is indicated by triangles. To answer the question about the real grain abundance in HII regions of BCDGs we used the relationship from [10]:

$$n_g/n_H = 7.36 * 10^{-10} Z/Z_{Sun}, \quad (1)$$

where n_g/n_H is the relative abundance of dust to hydrogen in the HII region (by number); Z and Z_{Sun} is the abundance by mass of heavy elements in the HII region and the Sun respectively. The ratio n_g/n_H can vary for BCDGs from $3.68 \cdot 10^{-12}$ to $1.10 \cdot 10^{-10}$. We adopted a grain particles to hydrogen atoms number ratio based on the ISM element abundance ratios (a minimum of the real n_g/n_H ratio). The n_g/n_H values in most of our models are higher. Therefore, we conclude that the effect of grain presence on the determination of the Ly$_c$ spectrum, physical parameters and chemical composition of HII regions in BCDGs is very small and can be neglected. It therefore seems that grain abundance is not important in calculations of OPhMs of HII regions in BCDGs.

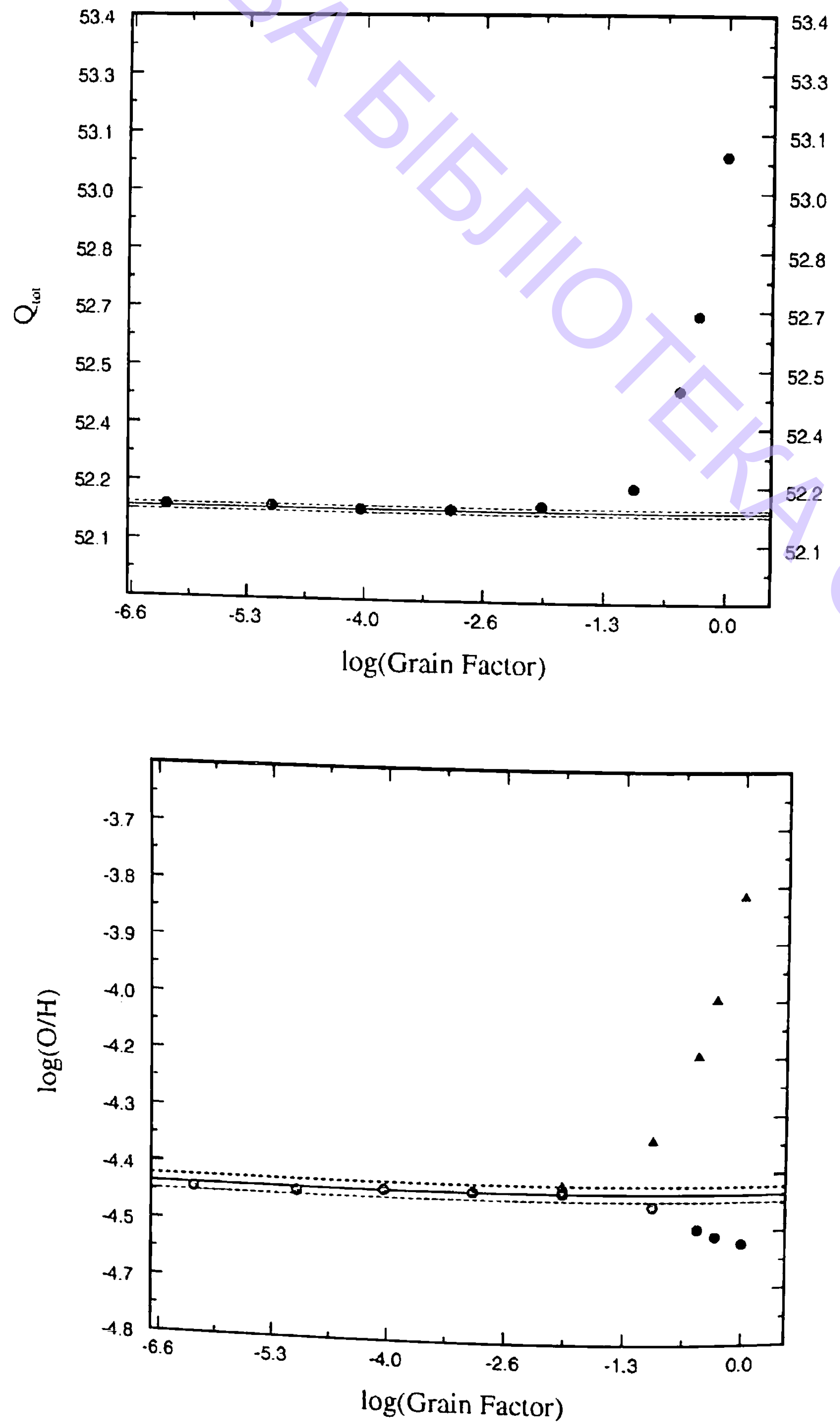


Figure 2: Dependence of Q_{tot} and O/H optimal values as a function of grain abundance.

References

[1] C. Leitherer et al., *ApJ Supl. Ser.*, **123**, 3 (1999).

[2] V.V. Holovatyy, V.I. Pronik, *Astrofizika*, **32**, 99 (1990).

[3] B.Ya. Melekh, *Journal of Physical Study*, 4, 225 (2000).

[4] B.Ya. Melekh 2003, *Investigation of physical characteristics and chemical composition of HII regions in blue compact dwarf galaxies*, Ph.D. thesis, Main Astronomical Observatory, Kyiv (2003).

[5] D. Schaerer et al., *Astronomy and Astrophysics*, **310**, 837 (1996).

[6] G.J. Ferland G.J., *Hazy, a Brief Introduction to Cloudy 96*, University of Kentucky, Department of Physics and Astronomy: Internal Report (2002).

[7] P.A.M. van Hoof, *Photo-Ionization Studies of Nebulae*, Ph.D. thesis, Rijskuniversiteit Groningen (1997).

[8] R.A. Benjamin et al., *ApJ*, **569**, 228 (2002).

[9] N.G. Guseva et al., *Astronomy and Astrophysics*, **378**, 756 (2001).

[10] I.Yu. Izotov, *Investigation of radiation of galaxies with active starburst processes in far infrared diapason*, Ph.D. thesis, Kyiv National University of Taras Shevchenko (2002).

6

Interstellar Matter and Star Formation

I.I. Pronik and L.M. Sharipova

Crimean Astrophysical Observatory, Nauchny, Ukraine
e-mail: ipronik@crao.crimea.ua, shali@crao.crimea.ua

Abstract

It is shown that observational evidence of a genetic link between gas nebulae, dust clouds and clusters of young stars was obtained in the 1950s by academician G.A. Shajn and his followers. Their results testified to the hypothesis of star formation from a gaseous and dust medium more than a decade before the current models that are based on the discovery of cold molecular clouds in the 1970s.

Keywords: interstellar matter, star formation

1 Introduction

This paper is dedicated to the 110 birthday of academician Gregory Abramovich Shajn, the first director of the Crimean Astrophysical Observatory (CrAO). G. Shajn and V. Gase actively studied the interstellar medium of the Galaxy in the 50s and made a H_α survey of the Milky Way from Crimea [1]. Only about 80 emission nebulae were known before their work. The Simeis catalogue by V. Gase and G. Shajn have listed about 301 objects,the majority of which were new diffuse emission nebulae. Their "Atlas of Diffuse Gaseous Nebulae" was much used at the time before the Palomar Sky Survey.

Shajn and Gase [2] estimated the masses of diffuse gaseous nebulae to be equal to hundreds and thousands of solar masses and showed that nebulae are not the result of ejection from stars. They showed observationally that interstellar matter is an important population of the Galaxy consisting of gas and dust with frozen magnetic fields.

The wealth of observational data on the emission nebulae permits to develop the idea of evolution of gaseous nebulae. Shajn and Gase [3, 4] suggested that there is a genetic connection between gaseous nebulae, dust clouds and groups of young stars and that their evolution happens in discrete spatial scales: about 10 pc, 100 pc and several hundred pc [5]. Shajn [6] investigated a giant feature of the spiral branch near the Sun of about 1000 pc, containing diffuse matter with a frozen magnetic field, which might represent a clump of future star formation.

For the investigation of emission nebulae new catalogues of early spectral type stars were needed to search for exciting stars. Shajn [7] designed a research program to look for early type stars: "Partly for that purpose observations are being carried out at the Simeis observatory of spectra of stars up to 12.^{m}2, by means of a 400mm camera supplied with a 410mm prism (6.°9). Also magnitudes, colour indices and colour excesses are being determined, mainly in the region of emission nebulae." This program was later transformed into an investigation of the structure of the Galaxy.

Shajn Plan was followed by 12 scientists: A. Alksnis, E. Brodskaja, N. Grigoreva, R. Ikhsanov, I. Kopilov, L. Metik, A. Numerova, I. Pronik, L. Raznik, P. Shajn, G. Sharapova, V. Strazys. Stars of different spectral types were investigated in 13 regions observed in Crimea (each one covering 10× 10 square degrees) along the Milky Way.

More than 35,000 stars were studied and more than 4,000 new O–B stars were found among them. The data obtained allowed the investigation of the structure of the spiral branches closest to the Sun and the specification of the structure of dark matter, the composition of clusters and associations of O–B stars in the spiral arms of Orionis, Persei and Sagittarii. Regions were revealed with optical absorption up to 25^m per kpc.

I. Pronik [8] reviews the results of galactic structure investigations by the Shajn Program.

2 The Problem of Star Formation

The Shajn Program included the problem of star formation too. Shajn and Gase [1] wrote: "It is very much possible that the role of interstellar matter will appear very considerable also in just outlining hypothesis on the so-called protostars."

Interest in star formation increased in 1940–1950. The question was whether stars are formed in recent epochs or were they formed simultaneously with the Universe? If they are still being formed, then how and is it possible to look for protostars. Ambartsumyan [9] showed that star associations are unstable, star formation is taking place in them continuously "almost before our eyes". It was shown that the age of luminous stars is a factor of 10 less than the Universe age.

Bok and Reilly proposed that small dark objects (globulae) can represent evolutionary stages preceding star formation [10]. The minimum absorption in dark objects was estimated as 2^m–5^m.

The Shajn Program was followed in the 50s and 60s. 49 papers were published, 47 of them before 1966. One of the important results of this work was to link early type stars to dark matter. The complexes of emission nebulae, clusters of young stars and stars of high brightness were found to show heightened dust absorption. The two first papers of the series published in 1953 by Brodskaja and Kopilov were devoted to investigate the distribution of O-B5 stars and dark clouds in the Galaxy. Raznik [11] observed a striking example of enhanced dust absorption inside the O-B association in the region $\alpha = 0^h\ 00'$, $\delta = +66.°5$ (1950). She divided the region of 10×10 square degrees in the sections of equal dust absorption. Figure 1 shows the dependence of absorption A_v on the distance modulus v_0–M_v for 5 sections of the region. We can see that inside the O-B association (section 11) $\Delta A_v = 5^m$. The density of the dust cloud surrounding the O-B association is about 4 times higher than the density of clouds outside the association (sections 5, 6, 13, 14).

Ikhsanov and Raznik devoted several papers to this subject. Ikhsanov [12] considered 22 and Raznik [13] 31 clusters of young stars connected with emission nebulae and showed that the presence of large amounts of dust is one of the important characteristics of very young star clusters. According to data from the 60s, absorption in the Orion nebula is $A_v \sim 9^m$–10^m.

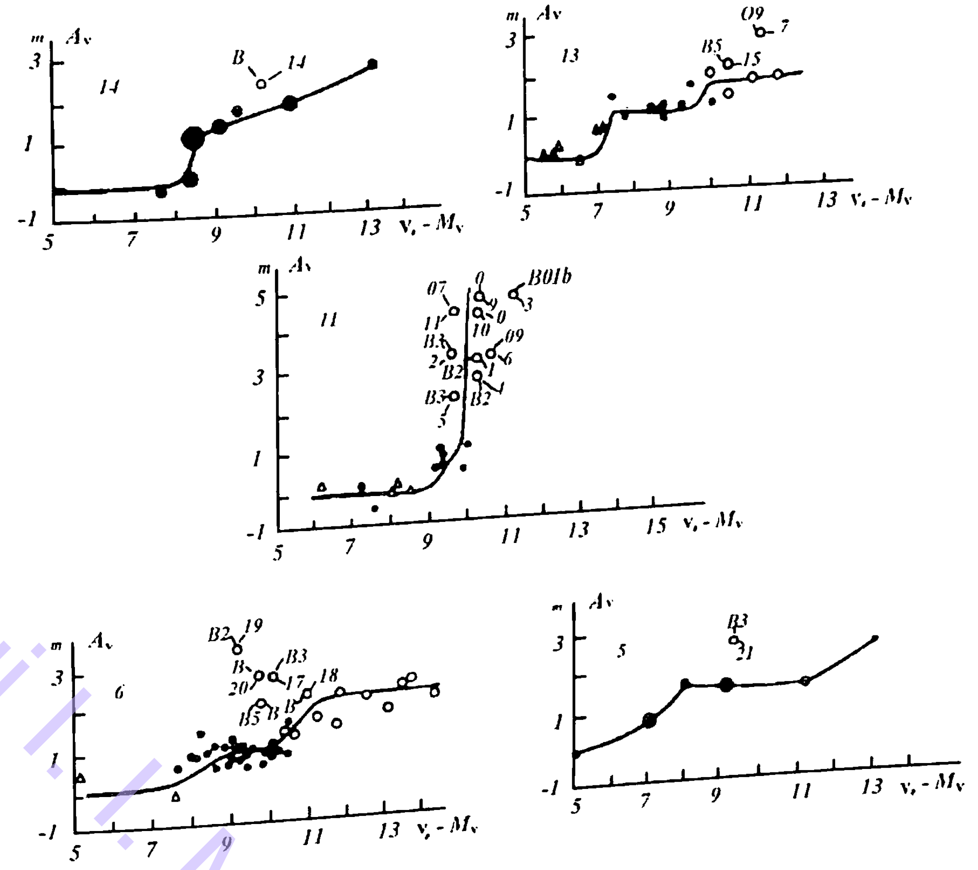


Figure 1: Dependence of absorption A_v on distance modulus v_0–M_v for 5 sections of the region investigated by Raznik [11]. The number at the top left represents the section number. Section 11 contains an O–B association.

The genetic link of gas nebulae with dust clouds and clusters of young stars was established by data from Shajn, Gase and collaborators. These results obtained during more than a decade allowed substantial advance in the understanding of the processes of star formation from interstellar matter in the 1970s.

A major break through was the discovery in 1963 of radioemission of hydroxyl at wavelength 18 cm, followed by observations on emission by other molecules. "Only in 1970 it turned out finally, that cold molecular gas is collected in massive clouds which serve as the "maternity hospitals of stars" [14]. Many molecules have been

Table 1: Hierarchy of sizes of regions of star formation and complexes of young stars and emission nebulae.

regions of star formation by Solomon et al., 1987 [16]		emission regions by Shajn and Gase	
object	size, pc	object	size, pc
1	2	3	4
clusters	10–15	separate nebulae	8–50
associations	30–200	groups of nebulae	100
superassociations	500	large groups of nebulae	to 300
complexes of nests star formation	1000	detail of regular magnetic field near the Sun	1000

detected since that time in the dense clouds which contain dust that protects them from the interstellar ultraviolet radiation.

Many of the basic topics of the present observational approach to understanding star formation are already proposed in Shajn and Gase work published between 1952–1955 and in papers published by the Shajn collaborators at the next decade, 1956–1966.

1. Stars are formed in groups in the places of accumulation of gas nebulae [3].

2. Clouds from which stars are formed are essentially more massive, not only than individual stars, but also than star clusters [2, 15].

3. Table 1 shows a comparison of sizes of different types of regions of star formation independently published in the literature ([16]), with those previously reported by Shajn [5].

4. Stars are formed inside the cocoons formed by very dense interstellar matter. It was shown that O–B stars in young star clusters are often surrounded by denser dust clouds than stars of later spectral types [11–13].

5. Clusters of giant H_α emitting clouds containing young stars are located in the spiral arms of galaxies [17].

I. Pronik and L. Sharipova [18] reviewed the genetic link of young star clusters with the interstellar dust and gas clouds covering the work by Shajn and collaborators.

References

[1] G. Shajn, V. Gase, *Izv. Crimean Astrophys. Obs.*, **6**, 3 (1951).

[2] G. Shajn, V. Gase, *Izv. Crimean Astrophys. Obs*, **9**, 13 (1952).

[3] G. Shajn, V. Gaze, *Izv. Crimean Astrophys. Obs.*, **10**, 152 (1953).

[4] G. Shajn, V. Gase, S. Pikelner, *Izv. Crimean Astrophys. Obs.*, **12**, 64 (1954).

[5] G. Shajn, *Astron. Zh.*, **31**, 217 (1954).

[6] G. Shajn, *Astron. Zh.*, **34**, 3 (1957).

[7] G. Shajn, in P. Oosterhoff (ed), *IAU Draft Reports of ninth General Assembly*, Cambridge Univ. Press, Cambridge p.305 (1955).

[8] I. Pronik, *Izv. Crimean Astrophys. Obs.*, 94, 14 (1998).

[9] V. Ambartsumyan, *Reports of Armenian Academy of Sci.*, p.35 (1947).

[10] B. Bok, E. Reilly, *Astrophys. J.*, **105**, N2, 255 (1947).

[11] R. Raznik, *Izv. Crimean Astrophys. Obs.*, **32**, 165 (1964).

[12] R. Ikhsanov, *Astron. Zh*, **37**, N4, 642 (1960).

[13] R. Raznik, *Izv. Crimean Astrophys. Obs.*, **36**, 264 (1967).

[14] V. Surdin, in *Birth of Stars*, URSS Press, Moscow, p.43 (2001).

[15] G. Shajn, V. Gaze, *Izv. Crimean Astrophys. Obs.*, **8**, 80 (1952).

[16] P. Solomon et al., in C. Persson (ed) *Starformation in galaxies, Proceedings of a NASA conference*, NASA Conf. Publ., **CP-2466**, p.37 (1987).

[17] G. Shajn, *Izv. Crimean Astrophys. Obs.*, **11**, 3 (1954).

[18] I. Pronik, L. Sharipova, *Izv. Crimean Astrophys. Obs.*, 99, 5 (2003).

7

Lithium in the Atmospheres of Slowly Rotating roAp Stars: γ Equ and 33 Lib

A. Shavrina[1], N. Polosukhina[2], S. Khan[3,4], Ya. Pavlenko[1],
V. Khalack[1], Yu. Lyubchik[1], A. Yushchenko[5,6], V. Gopka[5]
and D. Kudryavtsev[7]

[1] *Main Astronomical Observatory, NAS of Ukraine, Kiev, Ukraine*
e-mail: shavrina@mao.kiev.ua, yp@mao.kiev.ua, lyu@mao.kiev.ua
[2] *Crimean Observatory, Nauchny, Ukraine*
e-mail: polo@crao.crimea.ua
[3] *Institut für Astronomie, Universität Wien, Türkenschanzstraße 17, 1180 Wien, Austria*
[4] *Tavrian National University, Simpheropol, Crimea, Ukraine*
e-mail: serg@starsp.org
[5] *Observatory of Odesa University, Odesa, Ukraine*
e-mail: yua@odessa.net, gopka@arktur.tenet.odessa.ua
[6] *Astrophysical Research Center of the Structure and Evolution of the Cosmos (ARCSEC), Sejong University, Seoul 143-747, Korea*
[7] *Special Astrophysical observatory of the Russian Academy of Sciences, Nizhnij Arkhyz, Russia*

Abstract

The lines of lithium at 6708 Å and 6103 Å were analyzed in high resolution spectra for two very slowly rotating (P > 70 years) roAp stars γ Equ and 33 Lib. The spectral synthesis code SYNTHM was used. New lines of the rare earth elements from the DREAM database, and lines calculated on the basis of the NIST energy levels were included in synthetic spectra calculations. Magnetic splitting and broadening due to pulsation processes were taken into account. Enhanced abundances of lithium in the atmospheres of the stars and high ^{6}Li/^{7}Li ratio (0.5 and 0.2) can be produced by spallation reactions and be preserved in the strong magnetic fields.

Keywords: lithium isotopic ratio, roAp stars

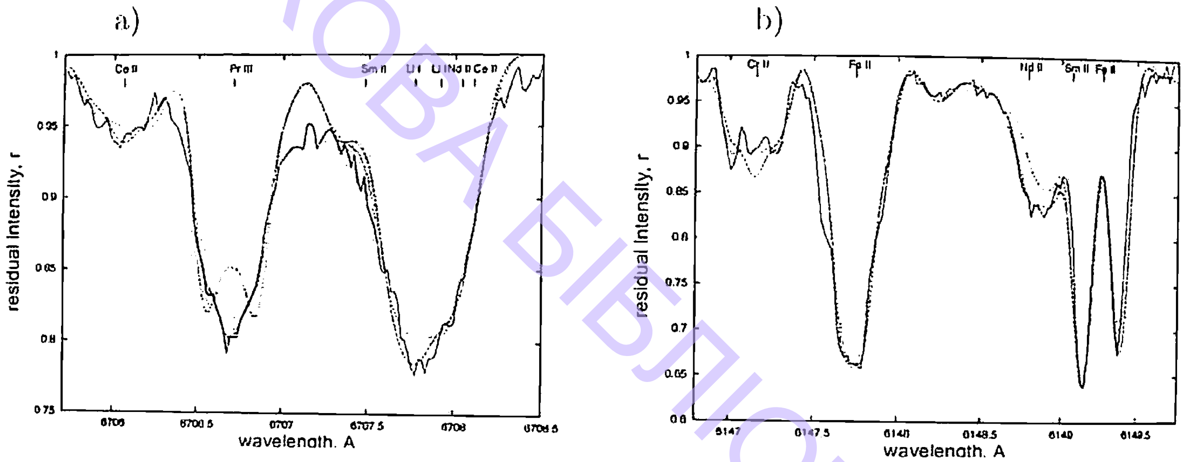


Figure 1: a) The estimation of field parameters from the lines CeII 6706.051Å and PrIII 6706.705Å for 33 Lib, long dash: $B_r = 2$kG, $B_m = 5$kG, $B_l = 0$kG and short dash: $B_r = 5$kG, $B_m = 2$kG, $B_l = 0$kG; b) Fe II 6147Å and 6149Å , long dash: $B_r = 2$kG, $B_m = 5$kG and $B_l = 0$kG like for Pr III 6706.7Å; short dash: $B_r = 4$kG, $B_m = 3$kG and $B_l = 0$kG.

1 Introduction

This work was carried out in the framework of the project "Lithium in CP stars" [6]. HD 201601 (γEqu) and HD 137949 (33 Lib), both roAp stars, are characterized by sharp lines in their spectra, by the strong overabundance of rare earth elements, and by strong magnetic fields (about 4 kG). The sharp lines in the spectra of both stars result from their slow rotation: their periods are approximately 77 years for γ Equ and more than 70 years for 33 Lib. The spectral observations of both stars show very strong resonance doublets of Li I at 6708Å.

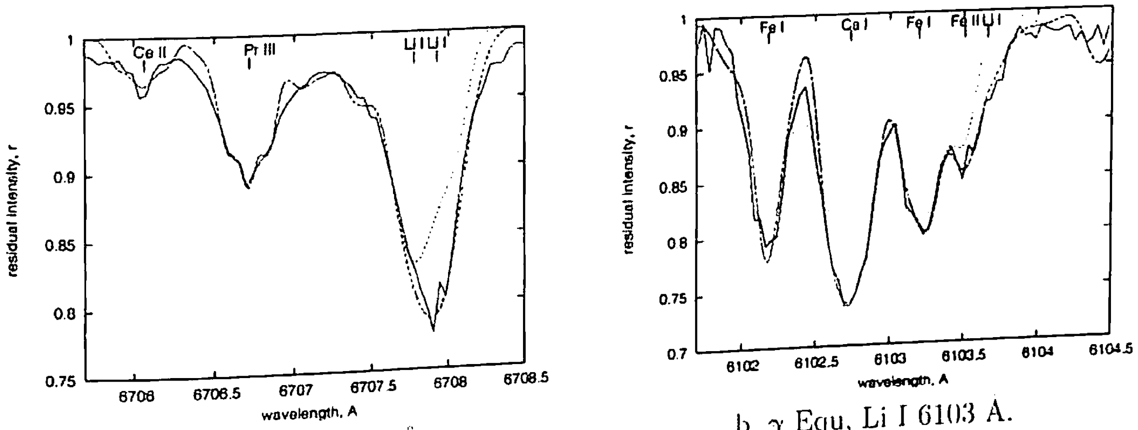
2 Observed Spectra

The spectra of 33 Lib were obtained by P. North in 1996, 8–14 March with the 1.4m Coude Auxiliary Telescope (CAT) at ESO, La Silla and the Coude Echelle Spectrometer (CES). The detector was ESO CCD N34 with 2048 pixels along the dispersion direction, which provides resolving power $R = 100\,000$, wavelength range 58.2 Å for central wavelengths 6705, 6645 and 6150 Å.

The observations of γ Equ were obtained during 9–10 of March 2004 by D. Kudryavtsev with the 6m BTA telescope and the Nasmyth Echelle Spectrometer (NES, Panchuk et al. 1999) of the Russian Special Astrophysical Observatory. They consist of continuous sequences of 25 and 16 spectra for 9 and 10 of March with exposure times about 3 min, that allow us to analyze pulsational variability of the spectral lines. The spectra with signal-to-noise ratio above 80 (15 and 4 spectra for 9 and 10 of March respectively) did not show any significant variability of the Li I 6708 Å line profiles and were selected for analysis. In Figure 2 we show the averaged spectrum for March 9 (covering 4 pulsation periods), which almost coincides with the similar one for March 10.

3 Synthetic Spectra

The lithium doublet lines 6708Å and 6103Å are very poorly studied for roAp stars. We study their spectra in detail in narrow ranges near 6708Å and 6103Å using synthetic spectra, taking into account Zeeman magnetic splitting and blending by REE lines. The additional broadening, most likely pulsational is described by the parameter $v \sin i$. Spectral calculations for γ Equ and 33 Lib were carried out using the model atmospheres of [5] with parameters from papers [7], [8]. For synthetic spectra calculations we applied the magnetic spectrum synthesis code SYNTHM ([2]), which is similar to Piskunov code SYNTHMAG and was tested in accordance with paper [12].



a. γ Equ, Li I 6708 Å

b. γ Equ, Li I 6103 Å.

Figure 2: For γ Equ a) Li I 6708 Å, lower dashed line: log N(Li) =–7.95, ^{6}Li/^{7}Li =0.5; upper dashed line: log N(Li)= –7.80, only ^{7}Li. b) Li I 6103 Å, $\log N(Li)/N(H) = -8.0$ and $\log N(Li)/N(H) = -8.3$.

The simplified model of the magnetic field is characterized by radial (along the line of sight), meridional and longitudinal components of the field B_r, B_m, B_l ($B_l = 0$ always, as it is justified for the plane-parallel model atmospheres), which were primarily determined from Fe II lines 6147Å, 6149Å, Ce II 6706.05 and PrIII 6706.70Å (see Table 1).

We used the VALD ([4]) and DREAM[1] databases for atomic spectral lines. These data do not in fact allow us to fit synthetic spectra to the observed ones for all stars studied. We therefore calculated additional REE II–III lines using NIST energy levels and estimated their "astrophysical" gf-values from the spectra of HD 101065 using element abundances from [1]. Also, the theoretical gf-values for important (for lithium abundance determination) blending lines were especially computed by P. Quinet with Cowan's code ([10]).

---

[1] [*http://www.umh.ac.be/~astro/dream.shtml*]

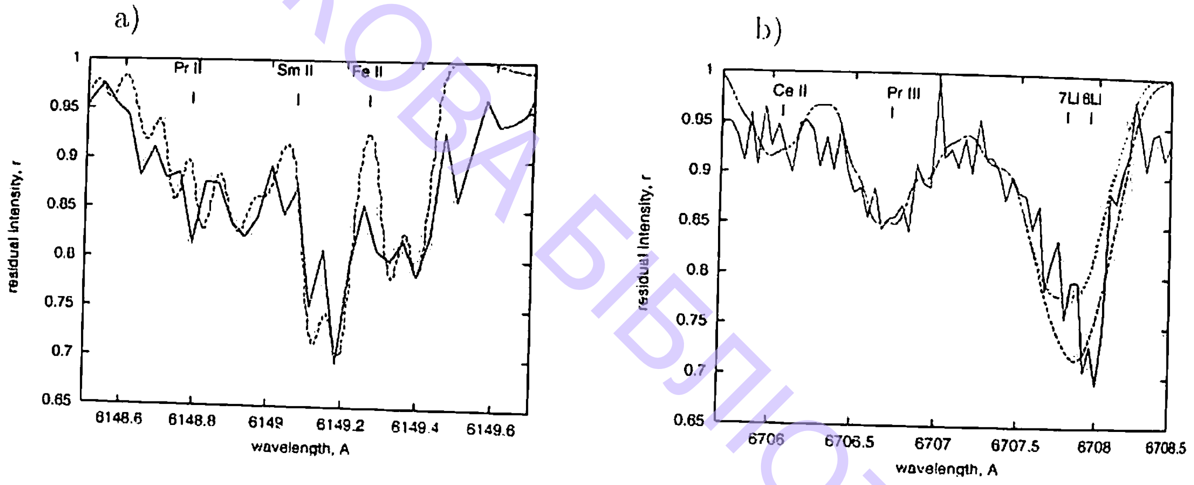


Figure 3: Fe II line 6149Å for spectrum N78 of γEqu with magnetic field components a) B_r = 3.1kG, B_m = 3.1kG, B_l = 0kG; the doubling was desribed by two layers with vsini=± 3.5 km/s; b) B_r = 1.5kG, B_m = 4.2kG and B_l = 0kG like for Pr III 6708Å N(Li)= 4.dex, $^6Li/^7Li$=0.5(lower dashed line) and only ^{7}Li - upper dashed line; wide profiles of Pr III and Li I were described by two layers with V_{macro}= 16 km/s and –19 km/s.

4 Results

Results are presented in the Table. The first line shows the HD numbers and the second line, the parameters of used model atmospheres. The calculations for star HD 137949 were carried out for two model atmospheres in a possible effective temperature range – 7750/4.5/0 and 7250/4.5/0. For each star (model) we give the abundances of Fe I and Fe II in the scale of log N(H)=12.0, derived from a group of Fe I (6102–6103 Å) and Fe II 6149 Å lines.

The abundances of lithium determined from both 6708 Å and 6103 Å lines and isotopic ratio from 6708 Å line are shown in the table. Under the solid line we give the parameters of the magnetic field and *vsini* found from the fitting of Fe II 6149 Å, Pr III 6706.7 and Ca I 6102.7 lines. The last value of *vsini* was used for spectra calculations in both lithium lines ranges. Magnetic field parameters from Ca I 6102.7 were used in the 6103 Å range and those from Pr III 6707.6 Å for the 6708 Å range. The results for γ Equ are given for averaged spectra (see section 'Observation'). We observed also the doubling of Fe II 6149 Å line, similar to 'the beating' in δ Sct stars, and extreme widening of the Pr III line 6706.7 Å and lithium line 6708 Å in two spectra – one on each night of observation. The calculations for one spectrum – N78 of April 10 are shown in Figure 3. In Figure 4 we show the averaged spectrum of γ Equ and dispersions for all spectra of the two nights (on the left) and selected spectra without abnormal line broadening (on the right). The last set included four full pulsation cycles in the first night and one in the second night.

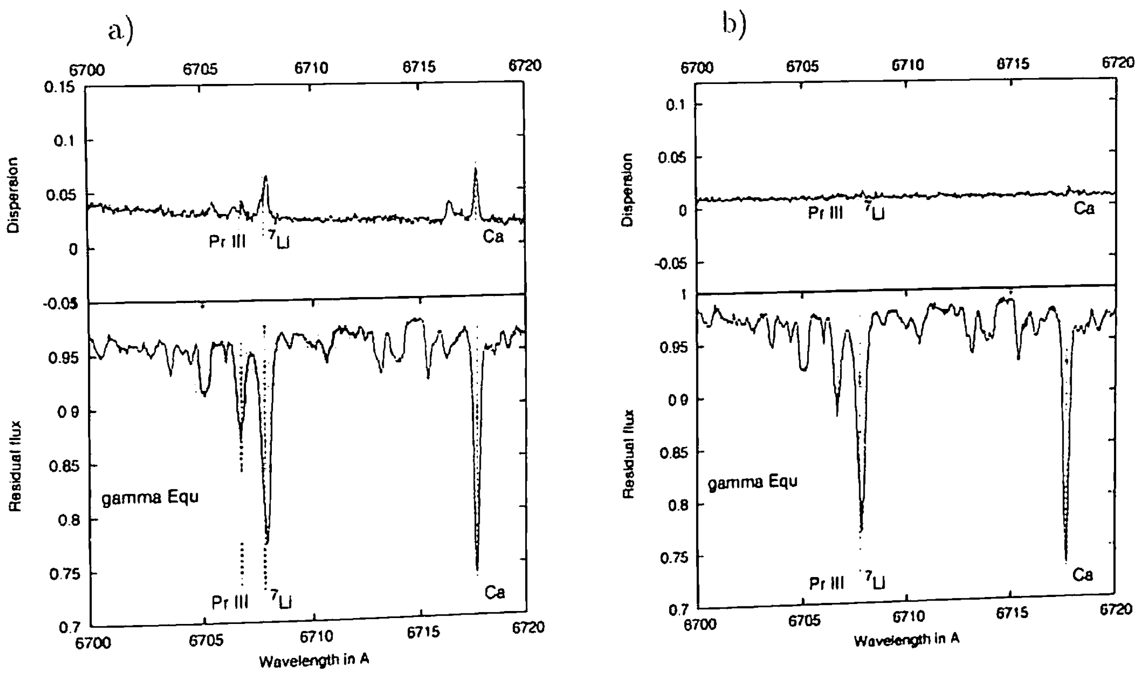


Figure 4: For γ Equ a) the averaged spectra and dispersions for all spectra of two nights and b) the same as a) for selected spectra without abnormal line broadening.

5 Conclusions

- The lithium abundance for all stars determined from the Li I 6103Å line is higher than the abundance determined from the 6708Å line. This may indicate vertical lithium stratification, an abnormal temperature distribution, or an unidentified blending with the 6103Å line.
- Our work on two roAp stars, HD 83368 and HD 60435 ([9]) provides evidence of an enhanced lithium abundance near the magnetic field poles. We can expect similar effects for γ Equ and 33 Lib. The high lithium abundance for all stars determined from the Li I lines and the estimates of the ^{6}li/^{7}Li ratio (0.2 ÷ 0.5) can be explained by the Galactic Cosmic Ray (GCR) production due to spallation reactions and the preservation of both ^{6}Li and ^{7}Li by the strong magnetic fields of these stars. The values of the ^{6}Li/^{7}Li ratio expected from GCR production are about 0.5 ÷ 0.8 ([3], [13]).
- The splitting of Fe II and the abnormal broadening of REE and lithium lines in two short exposure spectra of γ Equ could be explained by impact of shock waves in the line forming regions of the atmospheres of these slowly rotating stars (see [11]).
- New laboratory and theoretical gf-values for REE lines are necessary in order to refine our estimates of lithium abundances and the isotopic ratio.

Table 1: Results of Li abundances, ^{6}Li/^{7}Li-ratio, magnetic field parameters and *vsini* determination

	HD 137949	HD 137949	HD 201601
$T_{eff}/\log g/[m]$	7750/4.5/0	7250/4.5/0	7750/4.0/0
N(Fe I) 6103 Å	8.00	7.80	7.80
N(Fe II) 6149 Å	7.70	7.80	7.50
N(Li) 6708 Å	4.1	3.6	3.8
N(Li) 6103 Å	4.4	4.4	4.0
^{6}Li/^{7}Li, 6708 Å	0.2:	0.3:	0.5:
$B_r/B_m/B_l/$ (kG)			
Fe II 6149Å	4.1,4.1,0	4.2,3.3,0	3.5/2.6/0.8
Pr III 6706.7Å	2.0,5.0,0	1.5,5.0,0	2.7/3.5/0
CaI 6102.7Å	3.0,4.0,0	3.5,4.0,0	0/4.0/0
vsini(km s^{-1})(Fe II)	2.5	2.5	0.5
vsini(km s^{-1})(Pr III)	4.0	4.0	2.5

Acknowledgments

The authors are grateful to Dr. North for the ESO spectra. This work was supported by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (P-14984) and the BM:BWK (project COROT) to SK.

References

[1] C. R. Cowley et al., *MNRAS*, **317**, 299 (2000).

[2] S. Khan, *JQSRT*, **88**, 71 (2004).

[3] D. Knauth et al., *Astrophys. J.*, **586**, 268 (2003).

[4] F. Kupka F. et al., *Astron. and Astropys.*, **138**, 119 (1999).

[5] R. Kurucz, R., CDR(1-23), 1994

[6] N. Polosukhina et al., *Astron. and Astropys.*, **351**, 283 (1999).

[7] T. Ryabchikova et al., *Astron. and Astropys.*, **327**, 1137 (1997).

[8] T. Ryabchikova et al., *Astron. and Astropys.*, **343**, 229 (1999).

[9] A. Shavrina et al., *Astron. and Astropys.*, **372**, 571 (2001).

[10] A. Shavrina et al., *Astron. and Astropys.*, **409**, 707 (2003).

[11] H. Shibahashi et al., in D. W. Kurtz, K. R. Pollard (eds) *Variable Stars in the Local Group, IAU Col. 193*, ASP Conf. Proc., **310**, 287 (2004).

[12] G. Wade et al., *Astron. and Astropys.*, **374**, 265 (2001).

[13] W. Webber et al., *Astrophys. J.*, **568**, 210 (2002).

8

A Model of Nonthermal Radiation of a Shell-type Supernova Remnant

A.M. Bykov and Y.A. Uvarov

Ioffe Physico-Technical Institute, St. Petersburg, Russia
e-mail: byk@astro.ioffe.ru, uv@astro.ioffe.ru

Abstract

High energy nonthermal radiation detected from some SNRs could be direct evidence of an effective acceleration of protons and electrons up to energies ~ 10 TeV in these remnants. Modeling of the nonthermal radiation is very important because it could give a valuable information about SNR and the surrounding interstellar medium.

Based on a kinetic model of an electron acceleration by collisionless shock waves we studied electron acceleration in the vicinity of a shell-type SNR with applications to IC 443 and RX J1713.7-3946. We calculated the radiation spectrum of the remnants in radio, X-ray and gamma energy ranges taking into account bremsstrahlung, synchrotron and inverse Compton radiation processes. We showed that the nonthermal radiation spectrum of IC 443 is consistent with just nonthermal electron emission in all the energy ranges. Our model also can describe the nonthermal emission from the shell type SNR RX J1713.7-3946.

We studied possible consequences of the second order Fermi acceleration of electrons on magneto-hydrodynamic (MHD) fluctuations in the shell. This acceleration produces a spatially inhomogeneous electron distribution function that differs from a pure power law distribution. In case of IC 443 this difference could result in a special spectral feature in the energy band ~10–100 MeV that can be detected by upcoming γ-ray telescopes. Future observations in this energy range could provide an opportunity to study strength and spatial distribution of the MHD fluctuations in SNR IC 443.

Keywords: particle acceleration, nonthermal radiation, supernova remnants, IC 443, cosmic rays

1 Introduction

It is assumed that supernova remnants are the most plausible candidates for the role of the galactic Cosmic Ray source. A possible confirmation of this hypothesis is the fact that nonthermal radiation was detected from various SNRs. Among them there are remnants with emission spectrum that can be described as electron synchrotron radiation (see, for example, [1]). For others, the spectrum cannot be fitted by pure synchrotron radiation. There were attempts to find a nucleon radiation component in the spectrum of such remnants (see, for example, [2, 3]). Finding a nucleon component would be direct evidence of the hypothesis that SNRs really are a source of galactic cosmic rays. But in all these attempts the possibility that the emission spectrum has an electron origin could not be excluded. That is why a modeling of electron acceleration in SNRs is very important.

2 A Model of Electron Acceleration

Shock waves generated during the remnant expansion into the interstellar medium are effective particle accelerators due to a first order Fermi acceleration process. However in order to efficiently involve thermal electrons into Fermi acceleration an additional injection is needed. The process of injection was widely discussed in the literature but the complete theory of it is not yet constructed.

To resolve the problem of electron injection we developed in paper [4] the kinetic model that describes the generation of an electron distribution function in the vicinity of a collisionless quasiperpendicular shock wave. The electron distribution function[1] $f_e(p,z)$ (which is normalized to concentration $n(z)$ as $n(z) = \int f_e(p,z)\cdot p^2 dp$) in the stationary case must be a solution of the following kinetic equation that takes into account electron diffusion in the coordinate and momentum spaces due to their scattering on magnetic field fluctuations produced by MHD turbulence, an electron convection in plasma flow and electron energy losses:

$$\begin{aligned} & k_i(p)\frac{\partial^2 f_e(z,p)}{\partial z^2} - u_i(z)\frac{\partial f_e(z,p)}{\partial z} + \frac{p}{3}\frac{\partial f_e(z,p)}{\partial p}\frac{\partial u(z)}{\partial z} \\ & + \frac{1}{p^2}\frac{\partial}{\partial p}\left(p^2 D_i(p)\frac{\partial f_e(z,p)}{\partial p}\right) - \frac{1}{p^2}\frac{\partial}{\partial p}\left(p^2 F_i(p) f_e(z,p)\right) = 0\,. \end{aligned} \tag{1}$$

Index i in (1) describes a spatial region and is equal to 1,2,3 for upstream, shock transition region and downstream respectively. Function $F(p) = \frac{dp}{dt}(p)$ describes an electron energy loses. The diffusion coefficient in momentum space $D_i(p) = \alpha_i \frac{u_1^2}{9k_i(p)}$ depends on the MHD turbulence spectrum. The dimensionless parameter $\alpha_i \sim \frac{\delta u_i^2}{u_1^2}$ is a model parameter that describes the unknown spectrum of a MHD turbulence.

3 Supernova Remnant IC 443

We applied our model to calculate the nonthermal electron distribution and their emission for the well known SNR IC 443. The following processes influence the formation of the electron distribution function in the downstream: first and second order Fermi acceleration, electron energy losses and their escape from the system. In [5] we studied the electron distribution function taking into consideration first order Fermi acceleration, energy losses and electron escape. In this work we also include into consideration a second order Fermi acceleration which is important if MHD turbulence in the remnant is strong enough. We show that in this case a nonhomogeneous distribution function emerges in the remnant that differs from the pure power law distribution (Figure 1). In the emission of IC 443 this difference could result in an excess of radiation in the energy band 10–100 MeV that can be detected by upcoming γ-telescopes. On Figure 1 we show our model electron distribution functions and bremsstrahlung radiation spectra calculated for SNR IC 443 with different assumptions on the MHD turbulence level in the remnant.

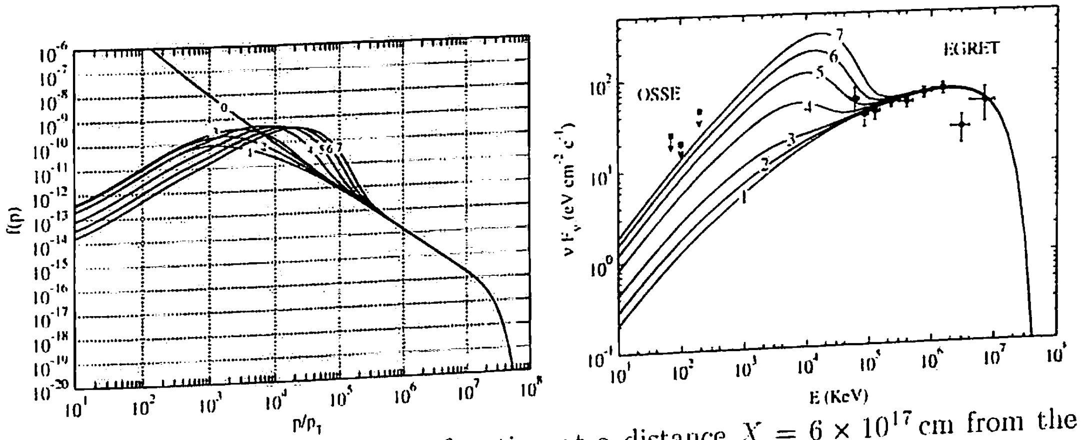


Figure 1: Electron distribution function at a distance $X = 6 \times 10^{17}$ cm from the shock front (on the left graph) and a nonthermal electron bremsstrahlung radiation from the remnant (on the right graph) calculated for different parameters α_3 in the shell. The curves {1, 2, 3, 4, 5, 6, 7} are calculated for $\alpha_3 = \{0; 5\times10^{-5}; 1\times 10^{-4}; 2\times10^{-4}; 3\times10^{-4}; 4\times10^{-4}; 5\times10^{-4}\}$. The curve 0 shows the distribution function just after the shock. P_T here is the thermal electron momentum for the temperature $T = 10^4$ K. The data points are EGRET observations from the 3rd EGRET catalogue [6] (circles) and OSSE observations [7] (squares).

On Figure 2 we show the total spectrum of IC 443 calculated according to our model. In this calculation we used almost the same parameters as we used in [5] but we also took into account a second order Fermi acceleration in the shell where we described the MHD fluctuations by $\alpha_3 \approx 3\times10^{-4}$. We also show by the long dashed curve on Figure 2 a possible emission due to secondary π_0 decay. For this calculation we approximated the flux of nonthermal protons by a power law and considered it to be equal to the electron flux on relativistic energies.

If we apply our model to SNR RX J1713.7-3946 assuming that the distance to it is also ~ 1.5 pc we obtain a similar shape of the electron distribution function and emission spectrum.

---

[1] We also use an electron distribution function $f(z,p) = p^2 f_e(z,p)$ which is normalized to concentration as $n(z) = \int f(z,p)dp$.

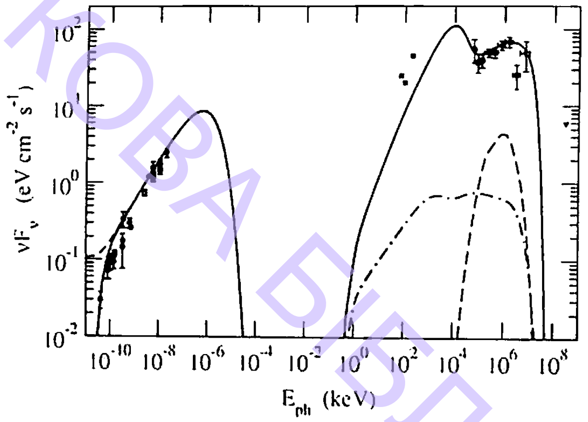


Figure 2: Our model emission spectrum of IC 443. Solid curves show synchrotron and bremsstrahlung radiation. The inverse Compton radiation is shown by dash-dotted curve. Short dashed curve shows the synchrotron radiation without taking into account a free–free absorption. Possible radiation due to secondary π_0 decay is shown by the long dashed curve. The data points of radio observations [8] (circles), EGRET observations from the 3rd EGRET catalogue [6] (rombs), OSSE observations [7] (squares) and WHIPPLE upper limit [9] (triangle) are also shown.

4 Conclusions

Using our model we studied the influence of the additional second order electron acceleration by MHD turbulence on the generation and evolution of the electron distribution function. We found that this acceleration could lead to substantial modification of the electron distribution function in the supernova remnant and such modification of the electron distribution function could lead to the excess in the remnant emission spectrum in the 10–100 Mev energy band that could be observed by the space telescope GLAST. So future GLAST observations of IC 443 will be very important. We also found that the nonthermal spectrum of SNR IC 443 (and RX J1713.7-3946) can be explained as pure electron radiation spectrum, but the possibility of some addition to the remnant emission from nucleons can not be excluded.

Acknowledgments

Our work was supported by RFBR grants 03-02-17433 and 03-07-90200, a President of the Russian Federation grant for young candidates of sciences 2642.2003.02 and a Russian science school grant 1115.2003.2.

References

[1] D.C. Ellison, F.C. Jones, S.P. Reynolds, *Astroph. J.*, **360**, 702 (1990).

[2] R.A. Chevalier, *Astroph. J.*, **213**, 52 (1977).

[3] L.O'C. Drury, F. Aharonyan, H. Volk, *Astron. Astroph.*, **287**, 959 (1994).

[4] A.M. Bykov, Y.A. Uvarov, *JETP*, **v.88**, N3, 465 (1999).

[5] A.M. Bykov, R.A. Chevalier, D.C. Ellison, Yu.A. Uvarov, *Astroph. J.*, **538**, 203 (2000).

[6] R.C. Hartman et al., *Astroph. J. Sup. Ser.*, **123**, 79 (1999).

[7] S.J. Sturner, J.G. Skibo, C.D. Dermer et al., *Astroph. J.*, **490**, 619 (1997).

[8] W.C. Erickson, M.J. Mahoney, *Astroph. J.*, **290**, 596 (1985).

[9] S. Fegan, in A. Carramiñana, O. Reimer, D.J. Thompson (eds) *The Nature of the Unidentified Galactic High-energy Gamma-ray Sources. Proceedings of the Workshop held at Tonantzintla, Puebla, México, 9-11 October 2000*, Astrophysics and Space Science Library, **267**, Kluwer Publ., Dordrecht, p.285 (2001).

9

Three-Dimensional Spectroscopy of a Starburst Region in the Spiral Galaxy NGC 6946

E.J. Alfaro[1], B. García-Lorenzo[2],
Yu.N. Efremov[3] and C. Muñoz-Tuñón[2]

[1]*Instituto de Astrofísica de Andalucía, CSIC, Granada Spain*
e-mail: emilio@iaa.es
[2]*Instituto de Astrofísica de Canarias, Tenerife, Spain*
e-mail: bgarcia@ll.iac.es, cmt@ll.iac.es
[3]*Sternberg Astronomical Institute, Moscow, Russia*
e-mail: efremov@sai.msu.ru

Abstract

3D spectroscopy of a massive star forming region in the starburst galaxy NGC 6946, has been performed. The data were taken at "Roque de los Muchachos" observatory with INTEGRAL IFU + WYFFOS camera attached to the 4.2 m telescope. The region is dominated by a massive and young star cluster (10^6 $M_\odot$ in mass and 15 My old) surrounded by half a circle of OB stars (~ 4 My old) and with a crown of Hα emission delimiting an outer ring ~ 600 pc in diameter. At least twenty other young clusters have been detected in the stellar complex, being their ages of a few tens My. The large number of young clusters makes this complex a natural lab for the study of the physical conditions driving two different modes of star formation: isolated stars or stellar clusters. In this contribution we analyze the 2D velocity field and the diagnostic diagrams of the ionized gas in order to discover the physical state of the gas.

Keywords: starburst, velocity field, emission gas, NGC 6946

1 Introduction

NGC 6946 is a starburst galaxy located at 6 Mpc from the Sun which shows some peculiar features. It displays a very high SFR over the entire disk (Larson & Richtler 1999) and, also shows a high density of star clusters, even though it could be contaminated by foreground stars due to the low galactic latitude of the object ($b \approx 12°$).

Eight supernovae have been detected in NGC 6946, being the galaxy with the highest number of observed SNe. In addition, many and very massive High Velocity Clouds (HVC) have been observed in its neighborhood which seem to be associated to the galaxy (Kamphuis 1993). This, together with the detection of some dwarf galaxies orbiting around it (Karachentsev et al. 2000), suggests some kind of material exchange between the galaxy ans its surroundings. However, the most conspicuous feature is the presence of a large, blue, and bright stellar complex located at the end of a sub-branch of one of the main spiral arms. This structure was first noted by Hodge (1967) and rediscovered by Larsen & Richtler (1999) twenty years later.

The complex has 600 pc in diameter, it is placed about 5 Kpc to the South-West of the galaxy center and shows a remarkable circular shape (see Figure 1 in Larsen et al. 2002). About twenty star clusters (younger than 30 My) have been detected inside the borders of the complex, one of them, located at the geometric center of its circular edge, has a mass of 10^6 $M_\odot$ and an estimated age of 15 My (Elmegreen et al. 2000). The SFH derived from HST stellar photometry shows that the current SFR is very high in this complex. Different stellar age groups present different locations inside the complex. Supergiants (25 My old) appear to be concentrated toward the center of the object, close to the SMYC, while the youngest OB stars (4 My old) form an arc shaped structure to the NW of the central cluster (Larsen et al. 2002). Long slit spectroscopy along three different directions shows that the interior of the complex is far from quiet, evidencing some small but very energetic structures (Efremov et al. 2002). This stellar complex is just an example of a peculiar class of objects whose study should shed light about: a) the physical variables, and their range, which control the star formation modes, e.g. isolated stars or star clusters, b) the formation of very massive clusters and, c) how the propagation of the star formation proceeds through the gas clouds. With this aim we have started the study of the ionized gas in this stellar complex.

2 Observations

A fiber device "INTEGRAL" (Arribas et al. 1998), connecting the output of the WYFFOS spectrograph to the CCD collector has been used. The instrument was attached to the WHT, a 4.2 m telescope, located at "Roque de los Muchachos Observatory" in La Palma (Canary Islands).

The bundle of fibers SB3 was used, where 135 fibers, 2.7 arseconds in diameter, cover a FOV of about 30×30 square arseconds. A linear scale of ~ 20 pc per arsecond corresponds to the distance of NGC 6946. Two spectra with R=1200, centered on 5200Å and 6565Å respectively, were taken for each field. More details on INTEGRAL & WYFFOS can be seen at the web page of the Observatory.

3 Results and Discussion

A visual inspection of the individual spectra showed that those fibers in the internal region, close to the massive cluster and the arc of the OB stars, present some contamination by stellar lines. Only the outer ring of the complex displays the typical spectrum of gas in emission, free of any trace of absorption lines. In this preliminary analysis we will focus on this outer ring, in particular the velocity field, and the ionization structure of the gas will be discussed in detail. Figure 1 shows the map of the Hα line in emission obtained from 3D spectroscopy. As commented before the regions of maximum emission form an almost circular ring surrounding the central cluster, which is also shown in Figure 1 as isophotes in V band. There is a drastic contrast between the emission coming from the outer ring and that generated in the NE region of the complex which presents a low content of ionized gas.

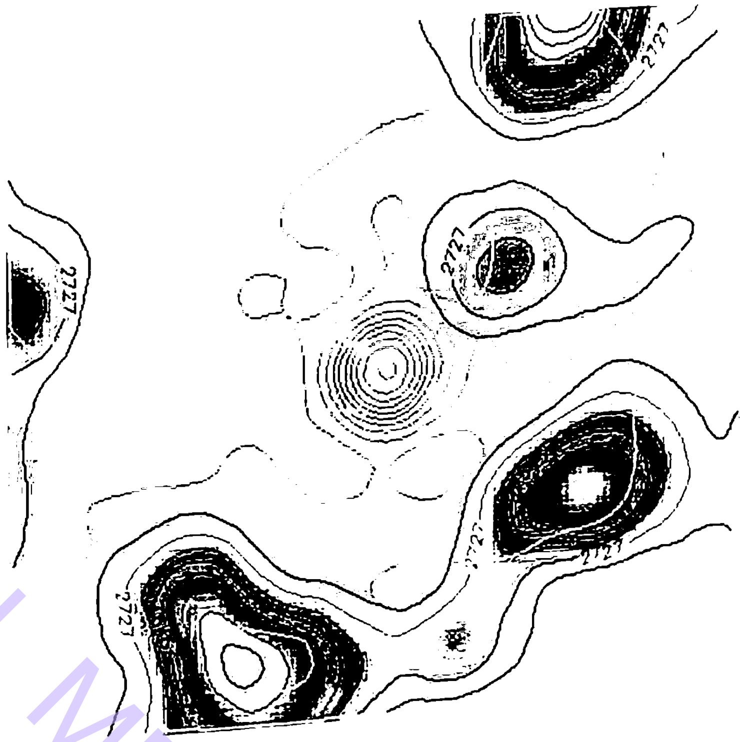


Figure 1: Hα map of the region showing the outer ring of ionized gas. Visual band isophotes (light lines) show the location of the young massive cluster. The size of the field is ~ 30×30 square arseconds. North at the top and East to the left.

One of the main objectives of this work is to analyze the 2D velocity field of the gas and compare it with the one expected for a pure rotation model. If the complex were either a dwarf galaxy interacting with the disk of NGC 6946

(Efremov et al. 2004) or a starburst triggered by the shock with a HVC, some kinematic perturbations, and/or traces of gas ionized by shock should be present.

The velocity map (Figure 2) has been obtained by averaging three radial velocity (RV) values per fiber estimated from three different emission lines (Hα, NII, SII). When the signal to noise ratio (S/N) of the Hα line was lower than 5 and the standard deviation of the mean velocity greater than 20 km s^{-1} the fiber was discarded. Excepting a velocity gradient ten seconds in extent and about 100 km s^{-1} in amplitude located to the West side of the massive cluster, no other special features are observed. The location of the lowest emissivity region is compatible with the previous long slit spectroscopy and the most kinematically disturbed regions were found inside this area (Efremov et al. 2004). The outer ring of ionized gas shows a RV field similar to the one expected for a pure rotation model of the galaxy (Carignan et al. 1990). Thus, the external ring does not show any evidence of peculiar motion and seems to be at rest with respect to the galaxy disk.

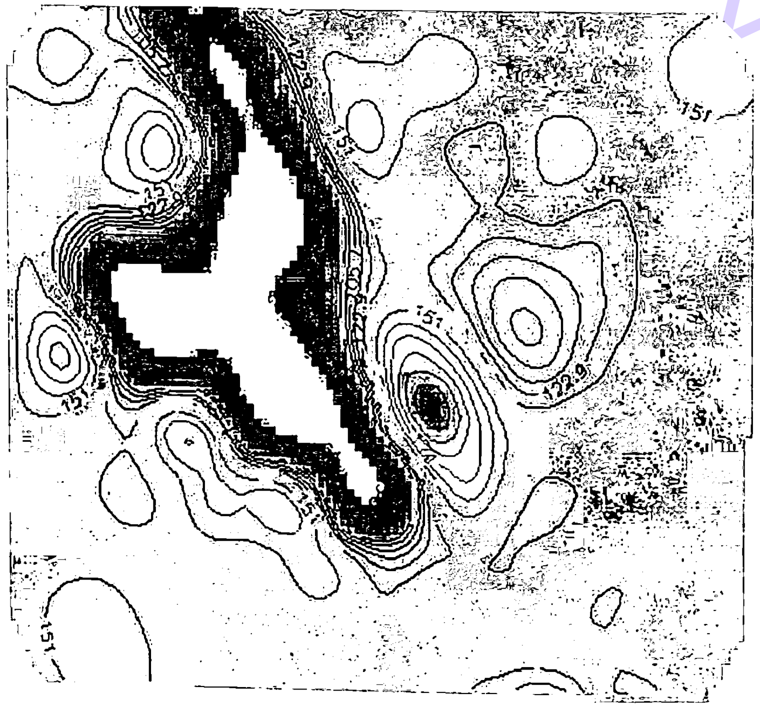


Figure 2: Radial velocity map of the ionized gas in the region. White area shows the places where the RV values have been discarded due to the low S/N ratio. Size and orientation as in Figure 1.

In order to elucidate the possible origin of this complex we need mode detailed information about the physical processes which have driven the current excitation of the gas. Thus, we would like to check whether the low ionization of the gas was originated by high energetic photons or by shocks. Several diagnostic tests have been proposed to this aim; here we will make use of the one based on the ratio of the lines (OIII/Hβ) and (NII/Hα). According to this test, all the fibers, concentrated along the region of high Hα emission, display line ratios compatible with emission originated by photoionization

Summarizing, our preliminary analysis of the 3D spectroscopy allows us to conclude that the rim of the complex shows neither evidence of kinematic perturbations nor signs of shocked gas.

Acknowledgements

E.J. A. acknowledges the financial support from MEyC (Spain) through grants AYA-2001-1696 and AYA-2001-3939-C03-01, and from Consejería de Educación y Ciencia (Junta de Andalucía) through PAI-TIC-101. Yu.N. E. thanks RFFI-03-0216288, SSP-389.2003.2 and AYA-2001-3939-C03-01 grants for partially supporting this work.

References

[1] S. Arribas et al., *SPIE*, **3355**, 821 (1998).

[2] C. Carignan, P. Charbonneau, F. Boulanger, F. Viallefond, *Astron.Astrophys*, **234**, 43 (1990).

[3] Yu.N. Efremov, S.A. Pustilnik, A.Y. Kniazev, B.G. Elmegreen, S.S. Larsen, E.J. Alfaro, P.W. Hodge, A.G. Pramsky, T. Richtler, *Astron.Astrophys*, **389**, 855 (2002).

[4] Yu.N. Efremov, E.J. Alfaro, P.W. Hodge, S.S. Larsen, C. Muñoz-Tuñón, *Astron. Astrophys. Trans.*, **23**, 95 (2004).

[5] B.G. Elmegreen, Yu.N. Efremov, S.S. Larsen, *Astrophys. J.*, **535**, 748 (2000).

[6] P.W. Hodge, *Publs. Astron. Soc. Pacific*, **79**, 297 (1967).

[7] J. Kamphuis, *PhD. Thesis*, Groningen University, The Netherlands (1993).

[8] I.D. Karachensev, M.E. Sharina, W.K. Hutchmeier, *Astron.Astrophys*, **362**, 544 (2000).

[9] S. Larsen, T. Richtler, *Astron.Astrophys*, **345**, 59 (1999).

[10] S.S. Larsen, Yu.N. Efremov, B.G. Elmegreen, E.J. Alfaro, P. Battinelli, P.W. Hodge, T. Richtler, *Astrophys J.*, **567**, 896 (2002).

10

Study of the Local Interstellar Medium and Galactic Magnetic Field by Multi-frequency Polarimetry of Synchrotron Radio Emission

V.A. Razin, E.N. Vinyajkin, A.M. Paseka and A.I. Teplykh

Radiophysical Research Institute (NIRFI), Nizhny Novgorod, Russia
e-mail: razin@nirfi.sci-nnov.ru

Abstract

Results of polarimetric measurements are presented of the diffuse Galactic radio emission at meter and decimeter wavelengths in a number of directions and regions of the sky. Spectra of brightness temperature of the linear polarized component, rotation measure map and the local structure of the Galactic magnetic field are discussed.

Keywords: interstellar medium, polarimetric measurements, the Galaxy

1 Introduction

The interstellar medium extending up to distances of a few hundred parsecs from the Sun (the local interstellar medium) is a subject of numerous experimental and theoretical studies (see [1] and references therein). The maximum distance from which the linearly polarized component of the diffuse Galactic radio emission comes to us (the polarization horizon) was estimated on the basis of the results of our polarimetric measurements in the area of the sky around $l = 144°$, $b = 8°$ [2–6] at 102, 210 and 290 MHz. For this estimation we used the data on the brightness temperature T_b^p of the diffuse Galactic radio emission linearly polarized component [2–6] and the interstellar medium synchrotron emissivity [7]. The distances turned out to be equal, respectively, to 10, 40 and 275 pc [2–6], i.e. they correspond to the local interstellar medium. There is very little information on the interstellar magnetic field at such distances from the Sun. This explains the substantial interest for special polarimetric studies of the diffuse Galactic radio emission at decimetre and metre wavebands aiming at getting information on the local interstellar medium and magnetic field.

2 Observations

Multi-frequency polarimetric observations of the diffuse Galactic radio emission in select directions of the sky are among basic observations carried out at the NIRFI Radio Astronomy Observatory "Staraya Pustyn′" near Nizhny Novgorod. In recent years multi-frequency polarimetric observations in five select directions were carried out at "Staraya Pustyn′" in the band of 151.5–1250 MHz. These observations were made by the tracking method using fully steerable radio telescopes with parabolic reflectors of 10 and 14 m in diameter and modulation radio polarimeters[1].

The table below shows the coordinates of the directions, the band and a number of frequencies, and type of the T_b^p spectrum[2]. We traditionally use here the "old" definition of T_b^p according to the following relation between polarization intensity I_p and T_b^p: $I_p = kT_b^p\nu^2/c^2$, where k is the Boltzmann constant, c is the speed of light, ν is the frequency ($I_p = 2kT_b^p\nu^2/c^2$ according to the "new" definition of T_b^p [9]).

Region, frequency band of the observations (MHz), and number of frequencies	North Celestial Pole, 151.5–1407, 14	Region of high meter wavelengths polarization, 151.5–1407, 26	Region of minimum radio brightness, 238–1407, 4	North Galactic Pole, 238–1407, 5	North Polar Spur, 238–1407, 6
α_{1950}, δ_{1950} l, b	$\delta_{1950} = 90°$ 123°, 27°24′	4^h30^m, 61° 146°47′, 9°03′	9^h40^m, 35° 190°10′, 49°15′	12^h49^m, 27°24′ $b = 90°$	14^h28^m, 14° 7°51′, 63°20′
Type of the T_b^p spectrum	Power (in the band 200–1407 MHz): $T_b^p(\mathrm{K}) = (1.95 \pm 0.05)\cdot\left(\frac{\nu}{300}\right)^{-1.87\pm0.05}$	Oscillating. The spectrum is fitted by a uniform synchrotron slab spectrum [2, 3, 5, 6]	Power (in the band 238–1407 MHz): $T_b^p(\mathrm{K}) = (1.03 \pm 0.10)\cdot\left(\frac{\nu}{300}\right)^{-1.55\pm0.34}$	Power (in the band 290–1407 MHz): $T_b^p(\mathrm{K}) = (2.34 \pm 0.72)\cdot\left(\frac{\nu}{300}\right)^{-1.82\pm0.72}$	Oscillating or power with maximum near 408 MHz

More extended areas in Loop III and Loop I were also observed [10–12]. In particular, the rotation measure (RM) map was obtained for the area of high linear polarization with Galactic coordinates (130° $< l <$ 155°, $-10° < b <$ 40°) by the results of our multi-frequency measurements of the position angle of the linear polarization plane taking into account the Faraday rotation in the ionosphere (see below).

---

[1] The method, calibration, data reduction and the radio telescopes are described in more detail in [2–6].

[2] We also used results of polarimetric observations at 1407 MHz [8].

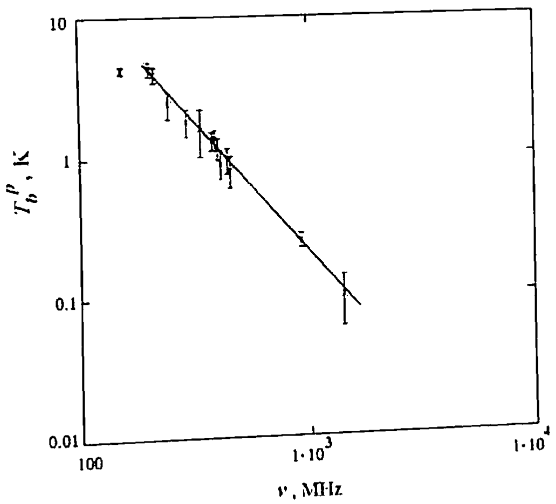


Figure 1: Spectrum of the polarization brightness temperature of the North Celestial Pole region and its approximation by a power spectrum with $\beta_p = 1.87 \pm 0.05$ in the band 200–1407 MHz.

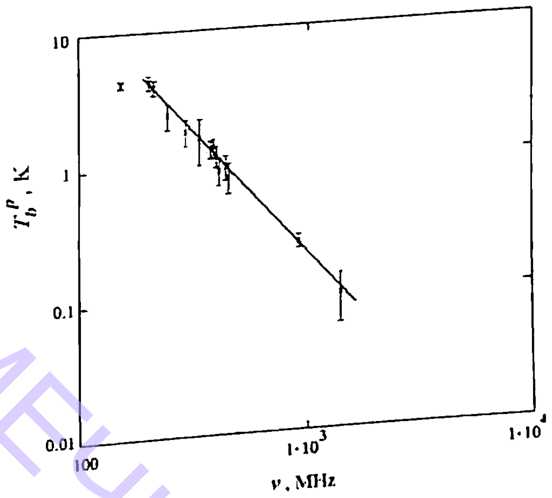


Figure 2: Spectrum of the polarization brightness temperature of the region of minimum radio brightness and its approximation by a power law with $\beta_p = 1.55 \pm 0.34$ in the band 238–1407 MHz.

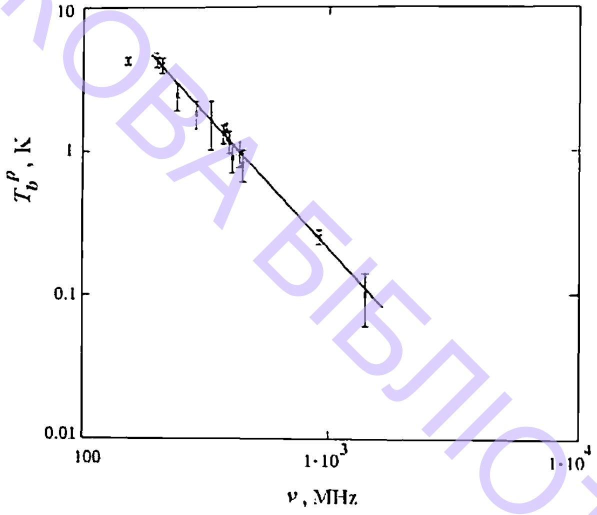


Figure 3: T_b^p spectrum of the region $\alpha_{1950} = 4^{\rm h}30^{\rm m}$, $\delta_{1950} = 61°$ in the band 151.5–448 MHz and its approximation by a uniform synchrotron slab spectrum with $RM = 0.86$ rad/m^2 and $\beta = 2.6$.

3 Discussion

Let us discuss the results of polarimetric measurements in the selected directions and the area of high linear polarization. The analysis has shown that there are two basic types of T_b^p spectrum in the band studied: the power spectrum (with or without the maximum) and the oscillating spectrum. Figures 1 and 2 show, respectively, the T_b^p spectra of the North Celestial Pole region and the region of minimum radio brightness. Both spectra are satisfactorily fitted by a power law ($T_b^p \propto \nu^{-\beta_p}$).

The temperature spectral index β of the minimum radio brightness region total Galactic radio emission is equal to 2.43 ± 0.02 in the band 17.5–1420 MHz. The degree of linear polarization is approximately proportional to λ^{-1} that testifies to the presence of a large number of "polarization" cells within the main antenna beam.

Another type of spectrum is revealed by the polarimetric observations of the regions in Loop III and Loop I. Figures 3 and 4 show, respectively, the T_b^p spectra of regions $\alpha_{1950} = 4^{\rm h}30^{\rm m}$, $\delta_{1950} = 61°$ and $\alpha_{1950} = 14^{\rm h}28^{\rm m}$, $\delta_{1950} = 14°$. It is clear from Figure 3 that the spectrum is non-monotone.

There is a maximum near 300 MHz and a deep minimum near 220 MHz. The simplest interpretation of this spectrum involves a homogeneous synchrotron slab model [2, 3] taking into account bandwidth depolarization [5, 6]. Formula (5) from [6] was used to fit the observed spectrum (see Figure 3).

As to the NPS region (Figure 4), an interpretation of its T_b^p spectrum by the same model seems to be the most probable because of high ordering of the magnetic field in the NPS upper part (see [12] and references in it). The model predicts a minimum at 482 MHz which we are planning to observe.

Figure 5 shows the RM map of the area of high linear polarization with Galactic coordinates ($130° < l < 155°$, $-10° < b < 40°$). It is seen that this map testifies to the fact that the structure of the magnetic field in the region has a loop character.

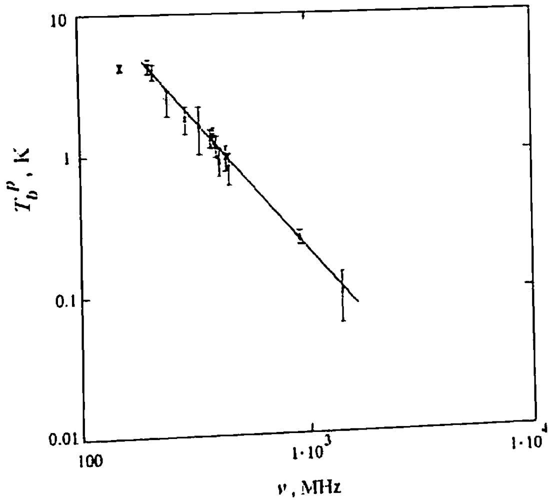


Figure 4: T_b^p spectrum of the region $\alpha_{1950} = 14^{\rm h}28^{\rm m}$, $\delta_{1950} = 14°$ in the band $238 \div 1407$ MHz and its approximation by a uniform synchrotron slab spectrum with $RM = 4.06$ rad/m^2 and $\beta = 3.3$.

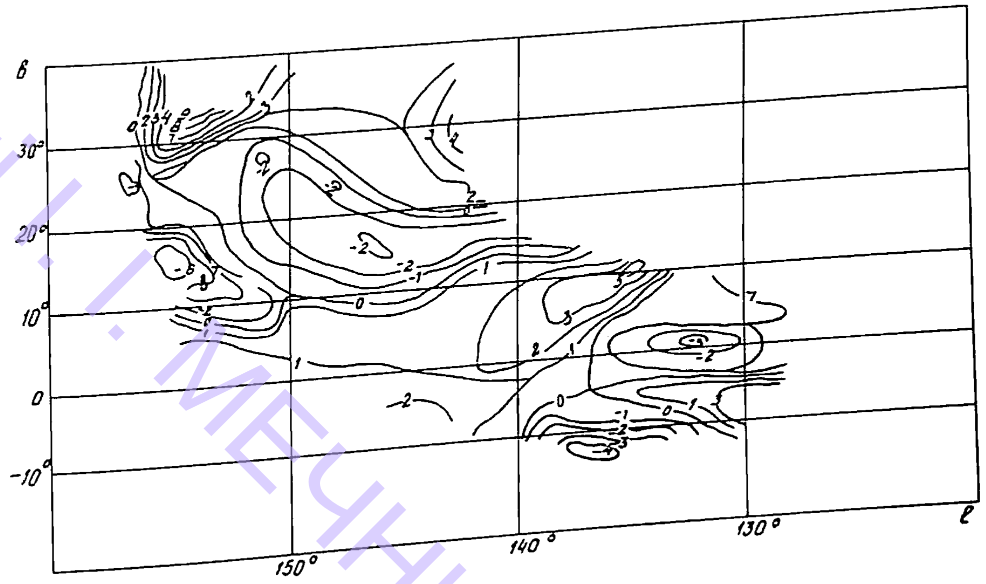


Figure 5: The RM map of the area with Galactic coordinates $130° < l < 155°$, $-10° < b < 40°$.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (project No 03-02-16685) and by grant No 1483.2003.2 of the leading scientific school.

References

[1] R. Ferlet, *The Astron. Astrophys. Rev.*, **9**, 153 (1999).

[2] V.A. Razin, *Radiotekhnika i Elektronika*, **1**, 846 (1956).

[3] V.A. Razin, *Soviet Astronomy*, **2**, 216 (1958).

[4] E.N. Vinyaikin et al., *Astronomy Letters*, **22**, 582 (1996).

[5] E.N. Vinyajkin, A.M. Paseka, A.I. Teplykh, *Radiophysics and Quantum Electronics*, **45**, 102 (2002).

[6] E.N. Vinyajkin, V.A. Razin, in S. Cecchini, S. Cortiglioni, R. Sault, C. Sbarra (eds) *Astrophysical polarized backgrounds, Proceedings of Workshop on Astrophysical polarized backgrounds*, American Institute of Physics, AIP, Melville, New York, **609**, p. 26 (2002).

[7] R.S. Roger, *Astrophys.J.*, **155**, 831 (1969).

[8] R.G. Bingham, *Mon. Not. R. Astronom. Soc.*, **134**, 327 (1966).

[9] E.M. Berkhuijsen, *Astron. and Astrophys.*, **40**, 311 (1975).

[10] V.A. Razin et al., *Radiophysics and Quantum Electronics*, **11**, 824 (1968).

[11] A.M. Paseka, *Soviet Astronomy*, **22**, 664 (1978).

[12] E.N. Vinyajkin, *Astronomy Reports*, **39**, 599 (1995).

V Galaxies, Stars and Interstellar Medium

1

Space Mission Rosetta to the Nucleus of Comet 67P/Churyumov–Gerasimenko: Goals, Successful Start and First Observations

K.I. Churyumov

Kyiv Shevchenko National University, Astronomical Observatory,
Kyiv, Ukraine
e-mail: klim.churyumov@observ.univ.kiev.ua

Abstract

The Rosetta space mission successfully started on March 2, 2004 from the Kourou cosmodrom to comet 67P/Churyumov–Gerasimenko. Upon entering its orbit around the nucleus observations will be made as the comet becomes more active as it journeys towards the Sun. A lander, named Philae, will be deployed and to attempt the first ever controlled landing on a comet. On its 10 year journey Rosetta will fly-by two asteroids – 2867 Steins in Sept. 2008 and 21 Lutethia in July 2010. Rosetta will be the first spacecraft to orbit a comet's nucleus. It will be the first spacecraft to fly alongside a comet as it heads towards the inner Solar System. Rosetta will be the first spacecraft to examine from close proximity how a frozen comet is transformed by the warmth of the Sun. Shortly after its arrival at Comet 67P/Churyumov–Gerasimenko, the Rosetta orbiter will despatch a robotic lander for the first controlled touchdown on a comet nucleus. The Rosetta lander's instruments will obtain the first images from a comet's surface and make the first in situ analysis of the relic matter of the Solar system.

Keywords: comet Churyumov–Gerasimenko, Rosetta, lander, nucleus

1 Introduction

The European Space Agency's unprecedented mission of cometary exploration is named after the famous "Rosetta Stone". This slab of volcanic basalt – now in the British Museum in London – was the key to unravelling the civilisation of ancient Egypt.

French soldiers discovered the unique stone in 1799, as they prepared to demolish a wall near the village of Rashid (Rosetta) in Egypt's Nile delta. The carved inscriptions on the Stone included hieroglyphics – the written language of ancient Egypt – and Greek, which was readily understood. After the French surrender in 1801, the 762-kilogram stone was handed over to the British.

By comparing the inscriptions on the stone, historians were able to begin deciphering the mysterious carved figures. Most of the pioneering work was carried out by the English physician and physicist Thomas Young, and the French scholar Jean François Champollion. As a result of their breakthrough, scholars were at last able to piece together the history of a long-lost culture.

Just as the Rosetta Stone provided the key to an ancient civilisation, so ESA's Rosetta spacecraft will unlock the mysteries of the oldest building blocks of our Solar System – the nucleus of comet 67P/Churyumov–Gerasimenko, which will play a role of "Rosetta Stone". As a worthy successor of Champollion and Young, the valuable data received by Rosetta will allow scientists to look back 4600 million years to an epoch when no planets existed and only a vast swarm of asteroids and comets surrounded the young Sun.

2 Comet 67P/Churyumov–Gerasimenko – Main Target of the Rosetta Mission

Comet 67P/Churyumov–Gerasimenko is a large dirty snowball that orbits the Sun once every 6.6 years. It was discovered October the 22, 1969, by Klim Churyumov from Shevchenko National University of Kyiv (Ukraine) and Svetlana Gerasimenko from the Institute of Astrophysics (Dushanbe, Tajikistan).

During this time, it commutes between the orbits of Jupiter and the Earth. However, little is known about it, despite its regular visits to the inner Solar System. Physical and orbital parameters of comet 67P/Churyumov–Gerasimenko are: diameter of the nucleus 4 km, orbital period 6.6 years, minimum distance from the Sun 186 million km, maximum distance from the Sun 857 million km, orbital eccentricity 0.6, orbital inclination 7.1°. Over an entire year, as it approaches the Sun, Rosetta will orbit the comet. As its ices evaporate, instruments on board the orbiter will study the gas and dust particles which surround the comet and trail behind it as a streaming system.

3 The Main Goals of the Rosetta Mission

The main goals of the Rosetta mission are [1]: to study the origin of comets, the relationship between cometary and interstellar material and its implications with regard to the origin of the Solar System.

The scientific results of the mission will be collected in several achievements. Among them:

- Global characterization of the nucleus, determination of dynamic properties, surface morphology and composition;

- Determination of the chemical, mineralogical and isotopic composition of volatiles and refractories in a cometary nucleus;
- Determination of the physical properties and interrelation of volatiles and refractories in a cometary nucleus;
- Study of the development of cometary activity and the processes in the surface layer of the nucleus and the inner coma (dust/gas interaction);
- Global characterization of asteroids, including the determination of dynamic properties, surface morphology and composition.

4 The Phases of the Rosetta Flight

The mission falls into several distinct phases: Launch - March 2, 2004. First Earth gravity assist – March 2005. Mars gravity assist – March 2007. Second Earth gravity assist – November 2007. Fly-by binary asteroid Steins (2867) – September 2008. Third Earth gravity assist – November 2009. Fly-by big asteroid Lutethia (21) – July 2010. Enter hibernation – July 2011. Exit hibernation – January 2014. Rendezvous manoeuvre – May 2014. Global Mapping - August 2014. Lander Delivery – November 2014. Perihelion Passage - August 2015. End of Mission – December 2015.

5 Rosetta's Scientific "First" Observations

The first phase of commissioning is close to completion and Rosetta has successfully performed its first scientific activity – observation of Comet C/2002 T7 (Linear).

On 30 April, the OSIRIS camera system, which was scheduled for commissioning on that date, took images of this unique cometary visitor – Comet C/2002 T7 (Linear). Later that day, three more instruments on board Rosetta (ALICE, MIRO and VIRTIS) were activated in parallel to take measurements of the comet.

The first data from the remote-sensing observations confirm the excellent performance of the instruments. The four instruments took images and spectra of Comet C/2002 T7 (Linear) to study its coma and tail in different wavelengths, from ultraviolet to microwave. Rosetta successfully measured the presence of water molecules in the tenuous atmosphere around the comet. The OSIRIS camera produced high-resolution images of Comet C/2002 T7 (Linear) from a distance of about 95 million kilometres. The image showing a pronounced nucleus and a section of the tenuous tail extending over about 2 million kilometres was obtained by OSIRIS in blue light.

An original image of the Earth–Moon system was taken by ESA's Rosetta comet-chaser spacecraft from a distance of 70 million kilometres. This image was taken by the Navigation Camera System (NAVCAM) on board the Rosetta spacecraft, activated for the first time on 25 July 2004. This system, comprising two separate independent camera units (for back-up), will help to navigate the spacecraft near the comet nucleus. The cameras perform both as imaging cameras and star sensors, and switch functions by means of a refocusing system in front of the first lens.

The successful observations of Comet Linear was a first positive test for Rosetta's ultimate goal, Comet 67P/Churyumov–Gerasimenko, which will be reached in 2014. Rosetta will be the first mission to undertake a long-term exploration of a comet at close quarters whilst accompanying it on its way towards the Sun. The unprecedented in-depth study conducted by the Rosetta orbiter and its Philae lander will help scientists decipher the formation of our Solar System around 4600 million years ago and provide them with clues of how comets may have contributed to the beginning of life on Earth.

References

[1] L. Colangelli et al. (eds) *The new Rosetta targets, Proceedings of the Capri-2003 conference*, Kluwer Publ., Dordrecht, pp. 1–392 (2004).

2

Detection and Properties of a Near-Earth Flux of Dark Electric Matter Objects (Daemons)

E.M. Drobyshevski

Ioffe Physico-Technical Institute, Russian Academy of Sciences
St. Petersburg, 194021 Russia
e-mail: emdrob@mail.ioffe.ru

Abstract

If the DM of galaxies consists of elementary Planckian black holes, their number should be fairly low, so that their flux should be small compared to that of hypothetical WIMPs. If, however, they carry an electric charge corresponding to their mass (up to $Ze \approx 10e$), such DArk Electric Matter Objects, daemons, should interact strongly with matter. They should be slowed down to some extent when crossing celestial bodies, build up in them, and in multiple systems, in close lying orbits too (e.g., in Earth-crossing orbits).

The scintillation detector consisting of two ZnS(Ag) screens stacked one upon the other (four modules 0.25 m^2 each) detects at a CL > 99% events with a time shift corresponding to velocities $V \sim$ 30-5 km/s (in both down- and upward crossings). Such velocities are typical for objects in helio- and geocentric orbits (with the latter crossing the surface of the Earth to become eventually confined to its interior). Of particular significance ($> 3\sigma$) is a group with $V \approx$10–15 km/s, which is characteristic of bodies falling from near-Earth almost circular heliocentric orbits. Their flux is in excess of $\sim 10^{-9}$ cm^{-2}s^{-1}. It varies with a period of 0.5 year. The data obtained in our experiments estimate the time of the daemon-stimulated proton decay as $\sim 10^{-6}$ sec.

Keywords: DM objects, discovery; Planckian scales; proton decay

1 The Crisis with Detection of DM Objects

The nature of the DM has remained unclear already for a few decades. The times had long passed when we believed it to be made up of conventional neutrinos with a mass >10 eV. The passionate enthusiasm for MACHOs has likewise waned [1].

Tremendous effort, staggering amounts of money, and extremely sophisticated techniques are concentrated presently on a search for the Weakly Interacting Massive Particles (WIMPs), i.e., neutralinos, heavy neutrinos, axions, and the like, which are predicted by some theories of elementary particles (see, e.g., [2]).

The only result that might be considered as an evidence for the existence of WIMPs is the 7% seasonal modulation of $<$6-keV signals, which was revealed by the DAMA collaboration at a level of 6.3σ during 7-year experiments [3]. These signals are believed to be caused by recoil nuclei knocked out by WIMPs which have a mass of 30–100 GeV and an elastic cross section $s \approx (2\text{–}10)\times10^{-42}$ cm^2/nucleon.

These data are not corroborated, however, by the recent CDMS-Soudan [4] and Edelweiss [5] experiments with cryogenic Ge detectors, which, by the way, lower s by one order of magnitude (down to $\sim5\times10^{-43}$ cm^2/nucleon). If we recall that the original estimates were $s \sim 10^{-38}$ cm^2 [2], straightforward reasoning brings us to the conclusion that the WIMPs as they were originally conceived simply do not exist; indeed, the cross section of interaction of nonexistent particles just has to continuously drop as the experiment is made increasingly refined.

Another controversy plaguing the WIMP ideology consists of the fact that the cross section of their mutual annihilation is estimated, based on the concepts of their formation in the epoch of freezeout thermodynamic equilibrium, as $\langle sv\rangle_A \sim 10^{-26}$ cm^3/s [2], i.e., as $s_A \sim 3\times10^{-34}$ cm^2. This value now exceeds already by $\sim$9 orders of magnitude the upper limit of s for elastic interaction with nucleons, which does not appear reasonable and makes one additionally assume the existence of some resonance (?) phenomena involved in the capture and annihilation of WIMPs with one another. The above reasoning forces an unbiased observer to draw the conclusion of a massive crisis in the WIMP ideology.

2 Our Ideology and Detection System

Our search for DM particles was based on the hypothesis that as long as the Universe started from Planckian scales, the main part of its mass should also have remained in Planckian objects. Their simplest version is represented by elementary black holes (BHs). We assumed also that such DArk Electric Matter Objects, daemons, carry an electric charge, whose repulsive action is counterbalanced by gravitation. Then $M_d = \mathrm{M}_{Planck} \approx 3\times10^{-5}$ g, $r_g \approx 2\times10^{-33}$ cm, and $Ze \approx 10e$. Such objects can be stable (they could be, say, remnants of evaporated more massive BHs, e.g. [6–11]). Apparently nobody attempted their search for a number of reasons.

First, the daemon having a giant mass, their flux from the Galactic halo can barely reach $\sim10^{-12}$ cm^{-2}s^{-1}, which makes their detection a difficult problem.

However, if we assume that the Galactic disk also has a DM population comparable to the mean density of the stellar population it contains [12], and that the velocity dispersion of these objects in it is 4–20 km/s [12], then the Sun should exert a strong focusing action on the daemons when moving through their background. We can readily show that when traversing the Sun, some of the disk daemons are decelerated to the extent where they can be captured by the Sun into elongated orbits with perihelia lying inside the Sun. In subsequent Sun crossings by the daemons these orbits gradually contract and lose orientation (remaining however in planes containing the Solar apex direction), with the daemons descending into the Sun [13]. If, however, in moving along such an orbit a daemon traverses the Earth's sphere of action, its orbit changes such that its perihelion will leave with a high probability the Sun's volume, to make the orbit stable until the next time it enters the Earth's sphere of action. In this way, daemons build up in heliocentric orbits crossing the Earth's orbit in about equinox zones mainly. Estimates show the daemon flux from such orbits onto the Earth may be as high as $\sim10^{-7}$ cm^{-2}s^{-1} [14].

Second, a daemon with a velocity $\leq$100 km/s cannot produce scintillations, that aroused scepticism concerning the possibility of their detection [6].

However, in their passage through matter *negative* daemons should capture nuclei with an energy release $W \approx 1.8ZZ_nA^{-1/3}$ MeV, which for Fe is $>$100 MeV. The excited nucleus ejects $\sim$10 nucleons (the first to be emitted are the atomic and refilling electrons). These particles are capable of producing a strong scintillation. For $Z_n \geq 2$, the ground state orbit of the daemon lies within the nucleus, and for $Z_n \geq 24/Z$, even inside a proton in the nucleus. It appears plausible that, on entering a proton, the daemon initiates its decay [13] (recall the monopole catalysis of proton decay [15]). Therefore, the charge of the nuclear remainder dragged along by the daemon should decrease continuously until it reaches the value $Z_n < Z$. After this, the daemon will again be capable of capturing another nucleus on getting rid of the neutron-excess remainder. Thus, daemon's traversal of condensed matter is accompanied by fast events of successive capture and evaporative de-excitation of a nucleus (at $V \sim$10 km/s, the capture of a nucleus occurs on a path $\sim$1 μm [16]) and slower processes of digestion of the nuclear remainder.

Thus, we invoke as a working hypothesis the existence of daemons. This assumption implies inevitably (*i*) their electromagnetic interaction with matter, including capture of nuclei with their strong excitation, (*ii*) proton decay catalysis, as well as (*iii*) advantages of looking for objects of a slow population in the Solar system accumulated as a result of their having been captured from the Galactic disk.

Based on this assumption, we develop a detector of four identical modules [16,17]. Each module represents a cubic case 51 cm on each side, made of tinned sheet iron. Its top side is a sheet of black paper. At the case's center, two transparent polystyrene screens 4 mm thick are arranged horizontally at a distance of 7 cm from one another, which are coated on the underside by a ZnS(Ag) scintillator layer $\sim$3.5 mg/cm^2 thick. The screens are separated by black paper. Each screen is viewed by a PM tube. The tube output signals are fed to a digital double-trace oscilloscope. The oscilloscope is triggered by signals from the top PM tube. If the second trace also carries a signal, such a paired event is sent to the computer memory.

We are interested in events with signals whose origins are displaced with respect to one another by Δt. The most frequent paired events are not displaced in time (their $\Delta t < 0.5$ μs). They are caused by cosmic rays. Such signals feature a short leading edge ($\leq$1 μs) and do not differ in shape from intrinsic PM tube noise. Sometimes, signals with a longer leading edge and a fairly flat, 2.5-3 μs long, maximum are observed. Such signals are characteristic of scintillations produced by α-particles. We called them Heavy Particle Scintillations (HPSs) [16-17].

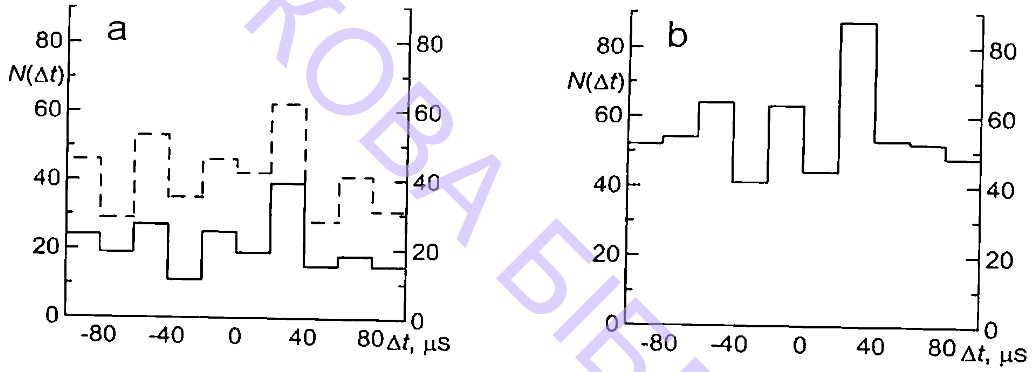


Figure 1: Distributions $N(\Delta t)$ of pair events on their time shift relative to the upper screen scintillations. (a) March, 2000; (b) Sum of March events of 2000, 2003, and 2004. (- - -) all type events in the top screen; (—) HPS events only in the top screen.

3 Main Results

Figure 1a displays the signal distribution $N(\Delta t)$ (with $\Delta t < 0.5$ μs events excluded) collected during March, 2000 [17]. A peak in the region $\Delta t \approx 30$ μs is clearly seen at a confidence level of 2.6σ. If we accept only events with HPSs in the top screen, their number drops (from 413 to 212), but the significance of the peak rises to 2.85σ (CL = 99.5%). This suggests that the peak is due to a physical cause, more specifically, to bombardment of the scintillator by heavy particles.

The sum distribution $N(\Delta t)$ of March events of three years is presented in Figure 1b. The significance of the 30-μs peak increased to 3.33σ (CL = 99.9%).

Our detector was designed assuming the lifetime of the daemon-containing proton to be $\Delta\tau_{ex} \sim 10^{-7}$ s. This accounts for the 7 cm separation between the scintillator screens. The time $\Delta t = 30$ μs yields, however, only 2.5–3 km/s for the velocity. It appears more reasonable [17] to choose for the base length the distance between the top screen and the bottom tinned-iron case lid, which is 29 cm. Then the transit velocity increases to $V \approx 10$–15 km/s, a figure corresponding to the fall of objects from near-Earth almost circular heliocentric orbits.

We may conceive now the following sequence of events [16–17]. A downward moving daemon captures a Zn or S nucleus in the top scintillator, where a HPS is created. While digesting the remainder of the nucleus, it crosses without interaction the bottom scintillator. Only when the system approaches the bottom lid of the case, the charge of the remainder drops to $Z_n < Z$, after which the daemon is capable of capturing a new Sn or Fe nucleus. Electrons (and, possibly, some protons) ejected in the process can traverse the 22-cm distance to the bottom scintillator and excite in it a scintillation in $\Delta t \approx 30$ μs. This is, in particular, what yields the estimate $\Delta\tau_{ex} \sim 10^{-6}$ s. It is also obvious that an upward moving daemon with $V \approx 10$–15 km/s is in an unfavorable position; indeed, it can excite the bottom scintillator even twice, but the probability for it to initiate the top scintillator located not far from the bottom one is slim. The number of events in the 30-μs peak corresponds to a ground-level flux of $\sim 10^{-9}$ cm^{-2}s^{-1}.

Interestingly, the April $N(\Delta t)$ distributions reveal maxima at $\Delta t = \pm(40$–$80)$ μs. They correspond to $V < 5$ km/s, i.e., to objects captured into Geocentric, Earth-Surface-Crossing Objects (GESCOs) and moving both down- and upward (see [16] for more details). These objects drift eventually beneath the Earth surface [18].

Figure 2 shows a seasonal variation of $1 - \alpha$, the confidence level that $N(\Delta t)$ cannot be represented by a constant, which was derived by the χ^2 criterion [19]. The totality of the data may be fitted by a sine curve with $P = 0.5$ yr. This is consistent with the above ideas of the daemon's build-up in orbits crossing the Earth's orbit in about equinox regions.

4 Some Conclusions and Implications

The above results characterize apparently the first attempt at a direct detection of massive (Planckian?) DM particles. We made use of the fairly obvious process of their buildup in Earth-crossing orbits due to a combined action of the Earth and the Sun. Nevertheless, gaining a satisfactory monthly statistics is a hard task. The smallest possible detector area is ~1 m^2, and seasonal variations permit us to collect reasonable data in a time of one year. Add here the difficulties involved in the the adjustment of a detector based on quite novel principles, where some of the parameters of its operation are poorly controllable (indeed, sometimes it is unclear to which side should the "tuning knob" be turned, but in either case we will have

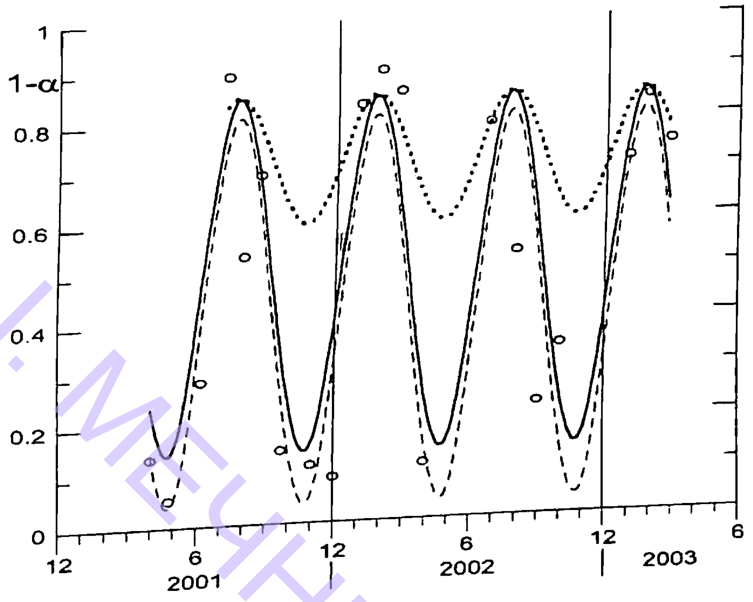


Figure 2: Seasonal variation of 1 - α (see text). (- - - -) Weights of all the points are equal, the correlation coefficient of the sine curve ($P = 0.5$ yr) with the points is $r = 0.86$, its CL > 99.9%. (——) Weights of the points are proportional to $1 - \alpha$; $r = 0.75$, its CL = 98.7%. (.........) Points with $1 - \alpha > 0.5$ are considered only; their weights are equal to $1 - 2\alpha$; $r = 0.36$, its CL ≈ 50%.

to wait a year to see whether the decision was right!). Nevertheless, the results thus far obtained are encouraging and have a high enough CL (99.9%!). It is not inconceivable that our data can be interpreted in some other way (for example, in terms of the "mirror matter"[20]).

For the daemon concept not to look as an *ad hoc* hypothesis, we may stress some of its implications. Our measurements of the rate of GESCO objects evolution suggest that up to 10^{24} daemons have accumulated inside the Earth, where they form a kernel of a few cm in size [18]. Proton decays in iron nuclei diffusing into the kernel are capable of accounting for the intriguing excess fluxes of heat, ^{3}He, and, possibly, even non-electron neutrinos out of the Earth's interior, the factors driving the convection and seismic anisotropy in the Earth's inner core, *etc.* [18–19].

The Sun's kernel, which contains up to $\sim 10^{30}$ daemons [21], may also account for a certain fraction of solar energetics (the possibility of the monopole-assisted proton decay as a source of stellar energy was pointed out in [22–23]), and for the flux of non-electron neutrinos in [13].

Acknowledgements

The author thanks N. N. Nikonov for processing the March, 2003 and 2004, data.

References

[1] M.W. Evans, V. Belokurov, preprint astro-ph/0411222 (2004).

[2] J.R. Primack, D. Seckel, B. Sadoulet, *Ann. Rev. Nucl. Part. Sci.*, **38**, 751 (1988).

[3] R. Bernabei et al., *Riv. Nuovo Cimento*, **26**, 1 (2003).

[4] D.S. Akerib et al., preprint astro-ph/0405033 (2004).

[5] V. Sanglard et al., preprint astro-ph/0406537 (2004).

[6] M.A. Markov, *ZhETF*, **51**, 878 (1966).

[7] K.P. Stanyukovich, *Doklady Acad. Scis. USSR*, **168**, 781 (1966).

[8] J.H. MacGibbon, *Nature*, **329**, 308 (1987).

[9] J.D. Barrow, E.J. Copeland, A.R. Liddle, *Phys. Rev.*, **D46**, 645 (1992).

[10] A.D. Dolgov, P.D. Naselsky, I.D. Novikov, preprint astro-ph/0009407 (2000).

[11] S. Alexeyev et al., *Class. Quantum Grav.*, **19**, 4431 (2002).

[12] J.H. Bahcall, C. Flynn, A. Gould, *Astrophys. J.*, **389**, 234 (1992).

[13] E.M. Drobyshevski, *A&A Trans.*, **23**, 173 (2004).

[14] E.M. Drobyshevski, in H.V. Klapdor-Kleingrothaus, Y. Ramachers (eds) *Dark Matter in Astro- and Particle Physics*, World Sci., p.417 (1997).

[15] V.A. Rubakov, *Pis'ma v ZhETF*, **33**, 658 (1981).

[16] E.M. Drobyshevski et al., *A&A Trans.*, **22**, 19 (2003).

[17] E.M. Drobyshevski, *A&A Trans.*, **21**, 65 (2002).

[18] E.M. Drobyshevski, *A&A Trans.*, **23**, 49 (2004).

[19] E.M. Drobyshevski, preprint astro-ph/0402367 (2004).

[20] R. Foot, S. Mitra, *Phys. Rev.*, **D68**, 071901 (2003).

[21] E.M. Drobyshevski, *Mon. Not. Royal Astron. Soc.*, **282**, 211 (1996).

[22] M. Izawa, *Progr. Theor. Phys.*, **75**, 556 (1986).

[23] G.S. Bisnovatyi-Kogan, *Sov. Astron. Lett.*, **13**, 358 (1987).

Afterword

I think the conference was successful, the quality of the papers presented was good and worthy of Gamow's memory.

The year 2004 is also the 90th year Jubilee of the outstanding Soviet scientist Yakov Borisovich Zeldovich. The fate of Ya. B. was quite opposite to that of Gamow, and there are some illustrations of this. While Gamow spent most of his life abroad, in the USA, Yakov Borisovich lived all the time in the Soviet Union, and it was only during the last five years of his life that he was able to go abroad. Gamow worked on the US H bomb; Ya.B. was the scientist behind the creation and development of the Soviet nuclear weapon. But as we know from mathematics and from life the extremes finally merge, and they both ended their lives as cosmologists, two of the greatest cosmologists of the 20th century. Gamow was the creator of the Hot Universe, and Ya. B. explained how the observed large structures are formed in this universe.

Igor' Dmitriyevich Novikov was, in the 1960s, the closest collaborator of Ya. B. In 1965, he met Gamow at a conference and told us later about the respect that Gamow expressed for Ya. B. For his part, Yakov Borisovich considered Gamow as the greatest physicist, making reference to his work in his papers and in his books with Novikov.

At that time, as a graduate student of Ya. B., I witnessed that respect. The two scientists never met, but their work had close connections, and as far as I could gather, they even had a similar scientific style based mainly on physical intuition, and only in a secondary way on mathematical calculations. Usually Ya. B. knew the answers before the calculations.

I would like to express our deep gratitude to all who participated in this successful conference, especially to the plenary speakers and conveners of the parallel sessions, who determined the high quality of the conference.

Special thanks should be given to the Local Team and its most efficient and active members: the LOC vice-chairman Mikhail Ivanovich Ryabov and the LOC secretary Irina Valer'evna Chernysheva. We are also grateful for the fantastic work of the SOC secretary Olga Dmitriyevna Toropina, who among many other obligations created and maintained the web pages of this conference and the previous Gamow's conference in 1999. Our greatest appreciation and gratitude are to Odessa University; its rector, Valentin Andreyevich Smyntyna, and also the director of Odessa

University, Observatory Valentin Grigiryevich Karetnikov who bravely took all of the financial burden to support the conference.

G.S. Bisnovatyi-Kogan

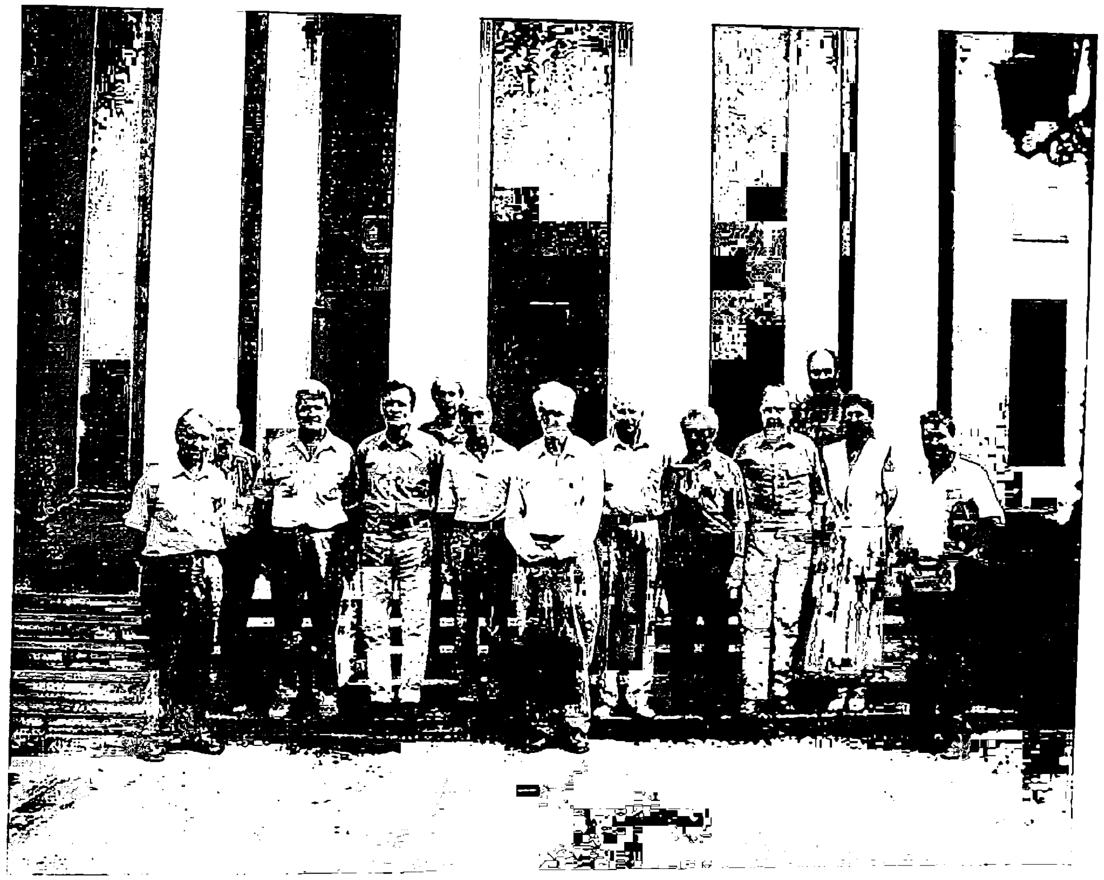


Plate 1 From left to right: Ryabov, Bochkarev, Cherepashchuk, Rudenko, Chernin, Khriplovich, Lominadze, Bisnovatyi-Kogan, Razin, Chechetkin, Ostrovski, unidentified, Karetnikov, standing in front of Odessa City Hall.

Plate 2 Conference photograph, Odessa National University.

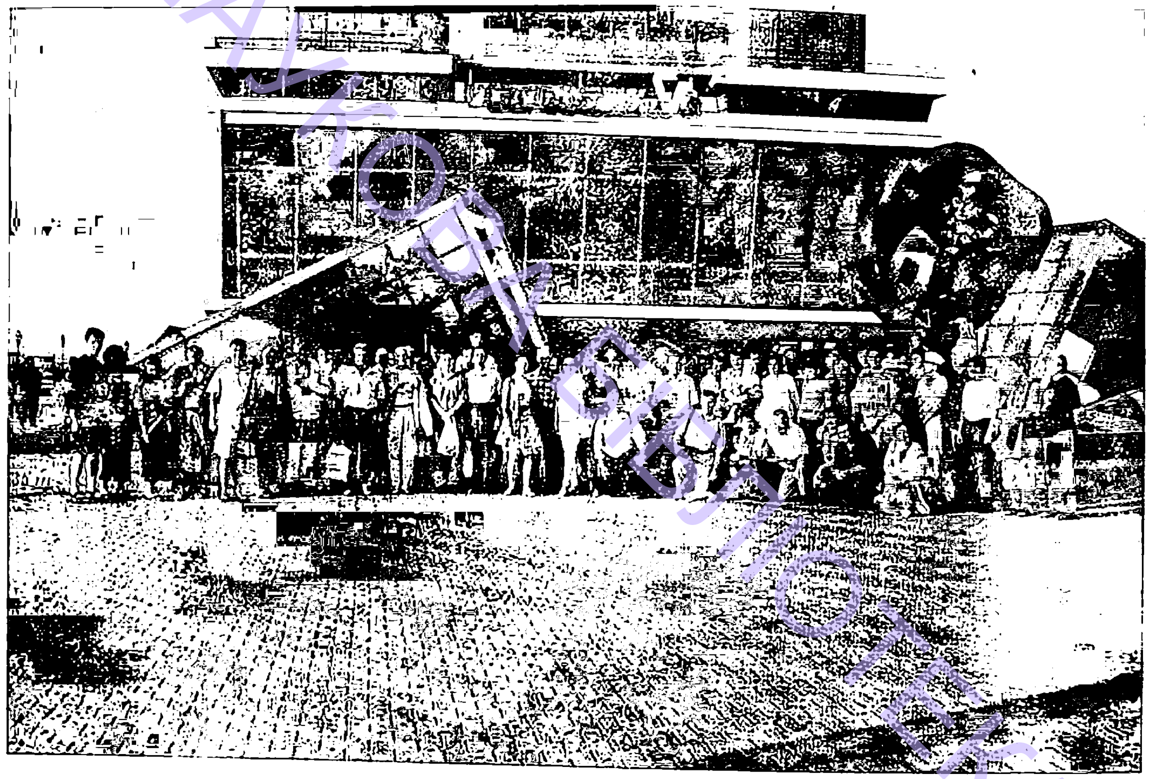


Plate 3 Conference participants at the Port in the background, a monument by E. Neizvestny.

Plate 4 The famous landmark: Potyomkin stairs.

Astrophysics and Cosmology After Gamow

Edited by

G.S.Bisnovaty-Kogan, IKI, Moscow, Russia
S.Silich, INAOE, Puebla, Mexico
E.Terlevich, INAOE, Puebla, Mexico
R.Terlevich, INAOE, Puebla, Mexico
A.Zhuk, Odessa National University, Ukraine

The Gamow Memorial International Conference was held in Odessa, Ukraine in August 2004. the conference was dedicated to 100th anniversary of George Gamow, one of the most prominent scientists of the twentieth century. His major scientific achievements include the development of the alpha and beta decay theory and the Big Bang theory of the expanding universe.

The conference covered key topics including:

- Cosmology and gravitation
- Large-scale structure of the universe
- Neutron stars, black holes and high energy astrophysics
- Stellar populations in galaxies and the interstellar medium

This volume provides a useful source of reference for graduates and researchers in astronomy, astrophysics and cosmology.

Cover illustration: © Prof. Artur Chernin.

ISBN 1-904868-38-X
9 781904 868385

CSP
CAMBRIDGE SCIENTIFIC PUBLISHERS