CONTENTS

FOREWORD

It gives me great pleasure to write these few words.

When Prof. Naresh Dadhich suggested the idea that the XVIII th conference of the IAGRG may be hosted by the Institute of Mathematical Sciences, I felt it was a welcome opportunity. There was a perception that while classical general relativity, gravitation, astrophysics, cosmology are active areas in their own right and as such have been discussed at the IAGRG meetings in the past, it is perhaps time now to expand the scope of these meetings to include the quantum gravity and particle physics aspects as well. Traditionally the general relativity community and the particle physics community have followed some what non overlapping developments. It would be mutually beneficial and healthier if both communities can interact more closely and share their experiences and perceptions. It is here that IMSc had a significant opportunity to play a role. Some of my colleagues concurred with this perception and we decided to host the XVIII the Conference of the IAGRG.

Just around the time we took the decision, Prof. S. Chandrasekhar passed away. In view of the phenomenal contributions of Chandra to General Relativity and Gravitation, it was but natural to dedicate this meeting to his memory. IMSc owes a special debt to Chandra as he played a crucial role in the foundation and the early stages of development of this Institute. Indeed the birth of the Institute of Mathematical Sciences was marked by the inaugural lecture by Prof. Chandrasekhar on January 3, 1962 in the lecture halls of Presidency College.

This IMSc report reflects the envisaged expanded scope of the IAGRG meeting and I hope that this trend will continue in the future IAGRG meetings as well. I also hope that these proceedings will make the frontline developments accessible to a larger body of researchers in the country particularly to those from the universities and colleges.

I may also take this opportunity to thank Drs G. Date and Bala Iyer for their efforts as Secretaries in the smooth conduct of the Conference and in putting together these proceedings.

R. Ramachandran

SCIENTIFIC ORGANISING COMMITTEE

1. Bala R Iyer, RRI, Bangalore (Chairman)
2. Asit Banerji, Jadavpur University, Jadavpur
3. G. Date, IMSc, Madras
4. N.D. Hari Dass, IMSc, Madras
5. Varun Sahni, IUCAA, Pune
6. T.P. Singh, TIFR, Bombay
7. D.C. Srivatsava, Gorakhpur University, Gorakhpur

LOCAL ORGANISING COMMITTEE

1. R. Ramachandran, IMSc, Madras
2. G. Rajasekaran, IMSc, Madras
3. N. D. Hari Dass, IMSc, Madras
4. G. Date, IMSc, Madras

PREFACE

This Institute of Mathematical Sciences Report contains the proceedings of the XVIII Conference of the Indian Association for General Relativity and Gravitation (IAGRG), held during February 15-17, 1996. The conference was attended by over 50 participants from all over the country, about half of them being from universities and colleges. The topics range over classical general relativity, astrophysics, gravity waves, cosmology and quantum aspects of gravity. The invited talks were intended to give an overview and current status of research in the respective areas. In addition, there were presentation of abstracts and theses which were collated in the form of a booklet available to the participants at the time of conference.

The conference was dedicated to the late Prof. S. Chandrasekhar. Prof. N. Panchpakesan kindly agreed to the difficult task of summarising some of Chandra's contributions to General Relativity. Prof. H. Dalitz who was visiting the Institute of Mathematical Sciences at the time of conference also kindly consented to speak on his personal encounters with Chandra. We are thankful to both of them.

Prof. C.V. Vishveswara delivered the traditional Vaidya-Raychaudhuri endowment lecture. We thank him for the very delightful talk. Prof. P.C. Vaidya, as is usual, made his presense felt throughout the conference.

Since publication of the proceedings is usually a long drawn process, we felt that we could bring out the proceedings much faster as an Institute of Mathematical Sciences Report. In this computer age, with authors doing most of the document preparation effort, it is relatively easy to put together the proceedings. Thanks to the various e-print archives available, these proceedings will also be made available to a much wider set of researchers. We would like to thank the speakers for giving their manuscripts in tex/latex formats which helped reduce the editorial work considerably.

It is a pleasure to acknowledge the encouragement and help we received at various stages. We thank Prof. R. Ramachandran, Director, Institute of Mathematical Sciences and Prof. N. Dadhich, President of the IAGRG, for their active encouragement and participation at all stages of organisation. We thank Dr. Parthasarthi Majumdar who was a member of the Local organising Committee in the beginning stages of the conference and who had pointed out that the conference may be dedicated to Prof. S. Chandrasekhar. Thanks are due to the members of the Local Organizing Committee and the members of the Scientific Organizing Committee. The editors thought it would be of some interest to include some quotes on Chandra and by Chandra. Many of the quotes by Chandra are from the wonderful biography 'CHANDRA' by Kamesh Wali.

We acknowledge the efficient assistance of Mr. Jayaraman, Mr. Sankaran and their colleagues from the administrative staff; Mr. Sampath and his colleagues from the guest house staff; the library staff; the computer network of IMSc and the volunteers from IMSc.

We thank G. Manjunath of Raman Research Institute for his patient help in the preparation of these procedings for final publication. We also thank Dr. Dipankar Bhattacharya of RRI for advise and assistance in resolving vexatious LaTeX'nical and Post-Script problems.

We thank the speakers and participants for their active participation. Last but not the least we thank the Department of Science and Technology, Govt. of India , the U.G.C and the Inter University Center for Astronomy and Astrophysics, Pune and the Institute of Mathematical Sciences, Madras for the financial support.

G. Date
B. R. Iyer

I practice style in a very deliberate way. I acquired my style from not only just reading, for instance, the essays of T.S.Eliot, Virginia Woolf, and Henry James, but also by paying attention to how they write–how they construct sentences and divide them into paragraphs. Do they make them short or long? For example, the idea of just using one sentence for a paragraph, or of a concluding sentence without subject or object, just a few words . . . 'so it is' . . . or some small phrase like that. I deliberately follow such devices. In fact there is one technique I started following when I was writing my book 'Radiative Transfer', and I have followed it since; that is, as you know, in music you repeat periodically the same phrase in exactly the same form. Very often in my books, when I have a key idea, and I have to restate it at a later stage, I don't leave it to chance. I go back and copy exactly what I had written before.

— S. CHANDRASEKHAR

SEEING BEAUTY

IN THE SIMPLE AND THE COMPLEX :

CHANDRASEKHAR AND GENERAL RELATIVITY

N.Panchapakesan[1]

Department of Physics and Astrophysics
University of Delhi
Delhi 110 007

Abstract

Some aspects of S.Chandrasekhar's contribution to General Relativity are reviewed. These cover the areas of post-Newtonian approximation and its application to radiation reaction, black hole theory, colliding gravitational waves and non-radial oscillations of a star. Some examples of his perception of beauty in these areas are given, as also the way symmetries seem to speak to him like to no one else. His attempt to find counterparts to space components of the metric in Newtonian theory in the context of non-radial oscillations is also presented.

[1]e-mail: panchu@ducos.ernet.in

Sometimes in the course of human interests things happen that truly lift the spirit. For me your winning the Nobel Prize is such an occasion.

I have always regarded my relationship with you as one of special inspiration for me. Your kindness, graciousness, absolutely uncompromising dedication to science, culture and integrity have really had a profound impact on me.

Every now and then the Nobel Committee does something truly great. This is one of those times. I cannot adequately express how happy I am for you and this is a feeling shared by all those who have been privileged to know you.

— MURPH GOLDBERGER.

His development is motivated by deep insight into the forms and symmetries of the differential equations with which he struggles. The equations speak to him in a tongue they speak to no one else leading him inexorably forward to new results. This is particularly so when the mathematics becomes horrendously complex as in the development of the 2PN approximation to general relativity and the analysis of pulsations of Kerr black holes.

In Mathematical Theory of Black Holes (MTBH) he concentrates on the last idea and the last word: He cleans up and completes in a thorough manner the body of incomplete theory that his younger colleagues left behind. And he does it without entering into controversy with them- at least not on the surface. However, if one knows something of the literature and reads beneath the surface one sees Chandrasekhar riding smoothshod over the works of his younger colleagues- smoothshod on an elegant steed with velvet covered hooves'

'Other insights tied up in incomplete non-Chandrasekhar versions of black hole mathematics may be lost to researchers..Chandrasekhar's method has been canonised and in ten or twenty years Chandra may truely have the last word..In return for losing other viewpoints we get from Chandrasekhar's book a monumental and almost complete body of mathematical theory presented in a totally coherent and aesthetically pleasing way. We are struck by the splendour of the theory, by the intricacies of its interconnections, by the mysterious amenability of black holes to totally analytical analysis.

One of those exceedingly rare books that will have a useful lifetime of fifty years. A book filled with new approaches to old subjects, old approaches to new subjects; completes unfinished researches of other physicists and maddeningly for the first time in Chandra'a career leaves unfinished his researches. It is filled with nuggets of mathematical insight.

— KIP THORNE

1 Introduction

Subrahmanyan Chandrasekhar started his work in General Relativity in the early 1960s. He was then past 50 years of age. In retrospect it seems historically inevitable that he should have moved over into the study of general relativity especially relativistic astrophysics. Discussing the maximum mass of white dwarfs that he had discovered in 1931, Chandrasekhar had said in 1934 'A star of large mass can not pass into the white dwarf stage and one is left speculating on other possibilities.' These other possibilities, it turned out later, were collapse to a neutron star or to a black hole. These were discussed by Oppenheimer and Volkov and Oppenheimer and Snyder in 1939. Chandrasekhar's studies of these possibilities, neutron stars and black holes thus seem a natural culmination of the ideas that led to the Chandrasekhar limit. But the way he entered the subject was unique and bears his own distinctive stamp.

In 1963 came his first major discovery in this area. He showed that general relativistic gravity creates a radial instability to gravitational collapse in stars with adiabatic index a little larger than 4/3. For the next twenty years he was one of the leaders in the field of relativistic astrophysics research. Research in this area, according to Kip Thorne [1], focussed on the structure, pulsations and stability of stars, star clusters and black holes, the gravitational collapse to form black holes and the generation of gravitational waves and the back reaction of waves on their sources. Chandrasekhar contributed to all these fields except star clusters and gravitational collapse.

The methods of research used for these studies are: Global Methods (Differential Topology), Discovery and Study of Exact Solutions (like Kerr metric), Perturbations of Exact Solutions and Post - Newtonian approximation. Chandrasekhar was a major figure in the study of perturbations and created the post- Newtonian approximation and was its master.

The major areas of his interest during the last 35 years of his life can be roughly chronologically listed.

1. 1961 -1975: Post Newtonian Approximation and its applications to general relativistic instability, radiation reaction and radiation reaction induced instability.
   (a) 1975-1982: Black Hole theory, perturbation and stability, separation of variables, transformation theory, reflection and transmission of waves.
   (b) 1983-1990: Colliding gravitational waves, relation to Kerr metric.
2. 1990 - 1992: Non - radial oscillations of neutron stars.
3. 1990-1995: Newton's Principia (outside the scope of the present lecture)

2 Post Newtonian Approximation

We first present the general relativistic instability discovered by Chandrasekhar using post - Newtonian approximation and then go on to present the general formalism of the approximation scheme.

In Newtonian theory the total energy of a star (sum of kinetic and potential energy) as a function of central density is given by

$$E = aKM{\rho_c}^{\Gamma-1} - bGM^{5/3}\rho_c^{1/3}$$

where M is the mass of the star, ρ_c is the central density and Γ is the adiabatic index. a and b are constants. Setting

$$\frac{dE}{d\rho_c} = 0,$$

keeping M constant, we find

$$M \propto \rho_c^{(\Gamma-\frac{4}{3})\frac{3}{2}}$$
$$\frac{dM}{d\rho_c} \propto (\Gamma - 4/3).$$

So $\Gamma > 4/3$ implies stabiliy.

In the case of general relativity we find the extra energy is given by (using post-Newtonian approximation)

$$\Delta E_{GTR} = -0.9 M^{7/3} \rho_c^{2/3}.$$

This leads to

$$\Gamma > 4/3 + k\frac{GM}{Rc^2}$$

$$\rho_c = 2.6\ 10^{10} (\frac{\mu_c}{2})^2\ g/cc$$

where R is the radius of the star.

In many cases, before this density is reached the constituents become neutrons. If the critical density ρ_c given above is smaller than the density at which neutronisation takes place then general relativity decides the density limit for stability.

Chandrasekhar was led to this discovery after his mathematical analysis based on post-Newtonian approximation. A physical insight following a mathematical analysis ! Let us compare this with the way Feynman and Fowler came to the same conclusion at Caltech around the same time.

Fowler was giving a seminar at Caltech and Feynman was in the audience. When supermassive stars were mentioned Feynman seems to have remarked that there may be an instability as gravity is stronger (in general relativity) and so collapse must be easier. Fowler calculated the effect and discovered the instability. Physical insight followed by mathematical confirmation! Did Chandra have no physical insight before calculating? I think there is some oversimplification here. Without insight one can not have the will or patience to do such complicated calculations. It was probably the desire for verification before announcement that played a role in Chandra's case.

2.1 Formalism: Conservation laws in Newtonian hydrodynamics [2]

The Energy-Momentum tensor is defined as

$$T^{00} = \rho c^2, T^{0\alpha} = \rho c v_\alpha$$
$$T^{\alpha\beta} = \rho v_\alpha v_\beta + p\delta_{\alpha\beta}$$

where $\alpha, \beta = 1, 2, 3.$ and $i, j = 0, 1, 2, 3.$

$$T^{ij}{}_{,j} \equiv \frac{d}{dx_j} T^{ij} = 0.$$

The momentum

$$p^i = \int_V T^{0i}\, d\overline{x} = \text{constant}$$

So is the angular momentum

$$L_\gamma = \epsilon_{\alpha\beta\gamma} \int_V \rho x_\alpha v_\beta d\overline{x} = \text{constant}$$

If we assume preservation of entropy by every fluid element we get another Energy integral. Change in thermodynamic energy (Π) per unit mass is equal to work done by the pressure in changing volume

$$\frac{d\Pi}{dt} = -p\frac{d}{dt}(1/\rho)$$

Then

$$E = \int_V \rho(\frac{1}{2}v^2 + \Pi)\, d\overline{x} = \text{constant}.$$

So far no external forces, not even its own gravitation is included.

When we include gravitation, if U is the gravitational potential given by distribution of ρ then

$$\nabla^2 U = -4\pi G\rho$$

and

$$T^{\alpha j},_j - \rho\frac{\partial U}{\partial x_\alpha} = 0.$$

Using the symmetric tensor defined by

$$t^{00} = 0, t^{0\alpha} = 0, t^{\alpha\beta} = \frac{1}{16\pi G}[4\frac{\partial U}{\partial x_\alpha}\frac{\partial U}{\partial x_\beta} - 2\delta_{\alpha\beta}(\frac{\partial U}{\partial x_\mu})^2]$$

and letting $\Theta^{ij} = T^{ij} + t^{ij}$ we can write

$$\Theta^{ij},_j = 0$$

which again leads to conservation laws. We also have

$$E = \int_V \rho(\frac{1}{2}v^2 + \Pi - \frac{1}{2}U)d\overline{x} = \text{constant}$$

2.2 General Relativity

In General Relativity the physical character of the system is completely specified by choice of T^{ij}. For a perfect fluid

$$T^{ij} = \rho(c^2 + \Pi + \frac{p}{\rho})u^i u^j - pg^{ij}$$

where $u^i = \frac{dx^i}{ds}$. Note that the metric is yet *unspecified.*
When T^{ij} is inserted in the field equation viz.

$$G^{ij} \equiv R^{ij} - \frac{1}{2}Rg^{ij} = -\frac{8\pi G}{c^4}T^{ij}$$

we do not have the choices we had in Newtonian Theory. $T^{ij}_{;j} = 0$ necessarily includes the effect of gravitational field on fluid motions. The covariant derivative has in addition to the ordinary derivative the effect of gravitational fields included in it. In Newtonian limit the additional term is the same as that encountered earlier, that is $-\rho\frac{\partial U}{\partial x_\alpha}$.

As T^{ij} has no dissipative mechanism the flow must preserve its entropy. We have

$$(\rho u^j)_{;j}(c^2 + \Pi + \frac{p}{\rho}) + \rho u^j[\Pi,_j + p(\frac{1}{\rho}),_j] = 0.$$

So conservation of mass $(\rho u^j)_{;j} = 0$ is compatible with the equation of motion only if

$$u^j[\Pi,_j + p(\frac{1}{\rho}),_j] = 0$$

which is the requirement that motion is *isentropic.* Thus in general relativity conservation of mass and conservation of entropy are not independent.

To get a conservation law we need $\Theta^{ik},_k = 0$, where we have an ordinary derivative. For this we have to add a 'pseudo-tensor' t^{ik} to the energy- momentum tensor T^{ik} and t^{ik} is symmetric but not a tensor. Defining the 'Energy-Momentum' complex $\Theta^{ik} = (-g)(T^{ik} + t^{ik})$ we have $\Theta^{ik},_k = 0$. This leads to the conservation law

$$E = \int_V (\Theta^{00} - c^2\rho u^0\sqrt{-g})d\overline{x} = \text{constant}$$

2.3 Basic Problem

In Newtonian framework given a well defined physical system (such as n - mass points under their mutual gravitational attraction or a mass of perfect fluid subject to internal stresses and its own gravitation) we can write down a set of *equations of motion* which govern all possible motions that can occur in the system.

The question arises : Can we write a similar set of equations in the framework of general relativity ? Chandrasekhar's response is : It appears that in general we can not do so. We are then led to a more modest inquiry : Can we write down an explicit set of equations of motion which govern departures from Newtonian motion due to effects of general relativity in a well defined scheme of successive post- Newtonian approximations. Further can we specify conserved quantities which are generalisations of the corresponding Newtonian quantities and which are constants of the post-Newtonian equations of motion ?

Around 1938 Einstein, Infeld and Hoffman (EIH) did pioneering investigations on the n-body problem. They wrote down the Lagrangian which differs from the Newtonian Lagrangian by terms of the $\sim \frac{v^2}{c^2}, \frac{U}{c^2}$ and comparable ones. The formalism gives the Keplarian orbit of two finite mass points about one another. However extensions to higher orders has not been possible. Chandrasekhar preferred the use of perfect fluid instead of mass points. He writes "mass points are not concepts that are strictly consistent with the spirit of general relativity. Hermann Bondi says 'General Relativity is a peculiarly complete theory and may not give sensible solutions for situations too far removed from what is physically reasonable'". Thus is Einstein suitably admonished ! According to Chandrasekhar, 'the concept of perfect fluid does not suffer from such limitations. In any event we confine ourselves to relativistic hydrodynamics of a perfect fluid.'

One starts with

$$T^{ij} = \rho(c^2 + \Pi + \frac{p}{\rho})u^i u^j - p g^{ij}$$

and the equation for conservation of rest mass $c^2(\rho u^i \sqrt{-g})_{,i} = 0$. The field equation

$$G^{ij} = -\frac{8\pi G}{c^4} T^{ij}$$

completes the set of equations.

The basic question is how is the field equation to be solved for g^{ij} so that the equation $T^{ij}_{;j} = 0$ when written out explicitly will provide the equation of motion as a power series in a suitable parameter.

2.4 Scheme of Approximation

First we identify the small parameters. The physical quantities of interest are the kinetic energy $K.E. = \frac{1}{2}\rho v^2$, the potential energy $P.E. = -\frac{1}{2}\rho U$, internal energy $= \rho\Pi$ and energy of molecular motion $= \frac{p}{\rho}$. The rest mass $= \rho c^2$ dominates over all these quantities. So $\frac{v^2}{c^2}, \frac{U}{c^2}, \frac{\Pi}{c^2}$ and $\frac{p}{\rho c^2}$ are the small parameters in what is usually called 'slow motion approximation'.

Secondly Equivalence principle, in its weak form, implies the following relation between rate of clock and gravitational potential.

$$g_{00} = 1 - \frac{2U}{c^2} + O(c^4)$$

So for Newtonian theory $g_{00} = 1 - \frac{2U}{c^2}, g_{0\alpha} = 0, g_{\alpha\beta} = -\delta_{\alpha\beta}$. These two considerations suffice to develop an entirely consistent scheme of successive post Newtonian approximation.

The first post-Newtonian is of order $O(c^{-2})$, the second is of order $O(c^{-4})$. These are orders of the equation of motion. For a given order of the equation of motion the different $g_{ij}s$ have to be known to different orders. If $T_{00} = \rho c^2 + O(1)$ then $(0,0)$ component of $R_{ij} = -\frac{8\pi G}{c^4}(T_{ij} - \frac{1}{2}T g_{ij})$ combined with $g_{00} = 1 - \frac{2U}{c^2} + O(c^{-4})$ reduces to Poisson's equation

$$\nabla^2 U = -4\pi G \rho$$

confirming that Newtonian equations are indeed 'zero order' solutions to Einstein's field equations.

To get to the first post-Newtonian approximation we proceed farther to the (α, β) component of the field equation (with a suitable coordinate condition) and find $g_{\alpha\beta} = -\delta_{\alpha\beta}(1 + \frac{2U}{c^2})$. Curvature of space implied by this is what is at the base of deflection of light by a gravitational field in general relativity. We next modify T^{ij}. We take

$$T^{00} = \rho c^2[1 + \frac{1}{c^2}(v^2 + 2U + \Pi)] + O(c^{-2})$$

$$T^{0\alpha} = ... + O(c{-}3)$$

$$T^{\alpha\beta} = ... + O(c^{-4}).$$

Used in the field equation they give

$$g_{00} = 1 - \frac{2U}{c^2} + \frac{2}{c^4}(U^2 - 2\phi)$$

$$g_{0\alpha} = \frac{P_\alpha}{c^3}$$

We can then write $T^{ij}_{;j} = 0$ to order $O(c^{-2})$ giving first order post Newtonian approximation.

The post- Newtonian series is even. The non trivial odd step is required in the imposition of the outgoing wave boundary condition. As $\frac{\partial}{\partial r}$ and $\frac{\partial}{c\partial t}$ are in different orders this is not possible. We are restricted to the near zone (r, ct). The way to match in the far zone was given by Trautman in 1958 who, unfortunately, used, wrongly, T^{ij} instead of Θ^{ij} for the Newtonian expression. When Chandra corrected this it gives the right result. Chandrasekhar and Esposito [3] found that $T^{\alpha j}_{;j}$ gives (in 2.5 post-Newtonian)

$$< \frac{d}{dt}\int E_{2.5}dx >= -\frac{G}{45c^5} < (\frac{d^3 D_{\alpha\beta}}{dt^3})^2 >$$

in exact agreement with rate of emission of gravitational radiation predicted by linear theory. Here $D_{\alpha\beta} = 3I_{\alpha\beta} - \delta_{\alpha\beta}I_{\sigma\sigma}$ is the quadrupole moment of the system.

It took 15 years (till mid 1980) for this result to be accepted by the scientific community [1]. There was controversy over handling of some divergent integrals and over the method of imposing the out going wave boundary condition.One should not be too far ahead of one's time, perhaps! Infeld, who was a collaborator of Einstein and Hoffman in the pioneering work mentioned earlier has written a book called 'Quest'. He discusses in that book the excitement and disappointment felt by them when they were trying to decide whether gravitational radiation is theoretically predicted in general relativity. Chandra's result finally showed the consistency of the formalism of gravitational radiation and provided for its theoretical acceptance. However it should be mentioned that work by Bondi and his collaborators, in the 1960s, had also convinced the community of the consistency of gravitational radiation though not in such a transparent manner.

The discussion of the radiation reaction led to the discovery of 'Radiation Reaction Induced Instability' in 1970. In the years from 1965 to 1968 Chandra was working on an area which seemed archaic to many of us, then, at Chicago. These were presented in a prestigious lecture series at Yale in 1968 and was published in a book form under the title 'Ellipsoidal figures of equilibrium'[4]. Chandra talks about the Maclaurin spheroid, the Jacobi ellipsoid and the Dedekind ellipsoid and how a rotating star can pass through these phases and meet bifurcation points when one form separates from the other. All these archaic concepts suddenly became very important in the context of the stability of rotating stars especially in the case of neutron stars which were discovered as pulsars in the year 1968. When dissipation is included, Chandrasekhar's study, in 1970, revealed the existence of an instability caused by radiation reaction due to emission of large amount of radiation. This has been confirmed in later studies.

3 Black Holes

Beginning in 1970 the study of black holes occupied Chandrasekhar for the next ten to fifteen years. Mathematical Theory of Black Holes [5] was published in 1982. The analysis of stability was one of the main considerations in the study. One perturbed a system and then studied whether the perturbation was damped returning the system to its original configuration or the perturbation tended to grow. To do this separation of variables was one of the important considerations.

The black hole metric has the well known form

$$ds^2 = (1 - \frac{2M}{r})dt^2 - (1 - \frac{2M}{r})^{-1} dr^2 + r^2[d\theta^2 + sin^2\theta d\phi^2].$$

When perturbed the general form is

$$ds^2 = \exp^{2\nu} dt^2 - \exp^{2\psi}(d\phi - \omega dt - q_2 dx^2 - q_3 dx^3)^2 - \exp^{2\mu_2}(dx^2)^2 - \exp^{2\mu_3} dx^{3^2}.$$

For the black hole $\omega = q_2 = q_3 = 0$. When perturbed we have the parameters $\omega, q_2, q_3, \delta\nu, \delta\mu_2$, $\delta\mu_3$, and $\delta\psi$. These are metric perturbations.

We can also use Newman- Penrose formalism and have perturbations of the scalars of Weyl tensors $\Psi_0, \Psi_1, \Psi_2, \Psi_3, \Psi_4$.

Chandra exploited both methods. He introduced transformation theory. This involves putting the equation for perturbation in the form

$$\frac{d^2Z}{dr_*{}^2} + \sigma^2 Z = VZ$$

This equation, like Schrodinger equation in a barrier potential, can be solved easily numerically. Chandra studied scattering matrix and its unitarity to discuss why different potentials give the same results. He also discussed how complex potentials give conservative scattering.

Chandrasekhar considered in detail the stability of the black hole, a problem first studied by Vishveshwara in 1970 (twentyfive years ago). Chandra considered various types of black holes:
Schwarzchild charactrised by the parameter M (mass)
Reissner-Nordstrom by parameters M(mass) and Q(charge)
Kerr by M (mass) and J (angular momentum)
Kerr - Newman by M (mass), Q (charge) and J (angular momentum)

Separability of the variables like r, θ and ϕ is crucial in all these discussions. Chandra's facility in these methods is well known. Over one weekend at Princeton he separated the variables of the spin $\frac{1}{2}$ particle equation in Kerr metric.

The extensive and powerful mathematical analysis applied to this problem, in the usual Chandrasekhar way brings us again to Kip Thorne's remarks [1]: 'Insight into physical origin comes after the mathematical analysis was complete.' He also goes on to remark 'It may be tempting to deprecate Chandra's more mathematical ways, were he not so spectacularly successful.' Thorne continues 'The symmetries of the equation speak to him in a manner that they speak to nobody else I know, leading him inexorably forward toward interesting results.'

We would like to present an example of his sensitiveness to such symmetries. While discussing Schwarzschild black holes Chandra had to work in a gauge, which he called phantom gauge.(Chap. 4, sec 29 of [5]). We have the following equations in which the operators $L_n, L_n{}^+, D_0, D_2{}^+$ operate on Φ_0, Φ_1, k, s (equation numbers are those in the book).

$$L_2\Phi_0 - (D_0 + \frac{3}{r})\Phi_1 = -6Mk - - - (237)$$

$$\Delta(D_2{}^+ - \frac{3}{r})\Phi_0 + L_{-1}{}^+\Phi_1 == 6Ms - - - (238)$$

$$(D_0 + \frac{3}{r})s - L_{-1}{}^+k = \frac{\Phi_0}{r}. - - - (239)$$

A choice of gauge can be made which brings to equations above a symmetry which they lack. In these equations, the symmetry of these equations in Φ_0, k and s is only partially present in Φ_1, k and s. Equation (239) is, for example, the right equation which allows us to obtain a decoupled equation for Φ_0 after the elimination of Φ_1 between equations (237) and (238). But a similar elimination of Φ_0 does not lead to a decoupled equation for Φ_1 since we do not have a 'right' fourth equation. However exercising the freedom we have to subject the tetrad frame to an infinitesimal rotation, we can rectify the situation by supplying (ad hoc?) the needed fourth equation. Thus with the additional equation

$$\Delta({D_2}^+ - \frac{3}{r})k + L_2 s = \frac{2}{r}\Phi_1$$

we can eliminate Φ_0.

Chandra adds, 'we shall find (chapter 5, sec 46) that the function Φ_1 defined in this way describe Maxwell's field in Schwarzschild geometry.' Thus we have derived Maxwell's equations (appropriate for photons with spin ± 1) by finding a gauge which rectifies the truncated symmetry of equations (237) - (239) in the quantities which occur in them.

To give an example where the absence of such symmetry leads to difficulties we consider the equations in Kerr- Newman metric, which defied even Chandra in his attempts at separating the variables.The equations, there are

$$(L_2 - \frac{3ia\sin\theta}{\overline{\rho}^*})\Phi_0 - (D_0 + \frac{3}{\overline{\overline{\rho}}^*})\Phi_1 = -2k[3(M - \frac{{Q_*}^2}{\rho}) + \frac{{Q_*}^2\overline{\rho^*}}{\overline{\rho}^2}]$$

$$\Delta({D_2}^+ - \frac{3}{\overline{\rho}^*})\Phi_0 + ({L_{-1}}^+ + \frac{3ia\sin\theta}{\overline{\rho}^*})\Phi_1 = +2s[3(M - \frac{{Q_*}^2}{\overline{\rho}}) - \frac{{Q_*}^2\overline{\rho^*}}{\overline{\rho}^2}].$$

Chandrasekhar remarks 'In contrast to the simplicity of terms in the earlier case on the right hand side we now have the ugly combination of $(M - \frac{{Q_*}^2}{\overline{\rho}})$ and $\frac{{Q_*}^2\overline{\rho^*}}{\overline{\rho}^2}$. A separation of variables will be possible, if at all, only by contemplating equations of order 4 or higher.' That is symmetry speaking to Chandra. No wonder Kip Thorne said, while reviewing Chandra's book on Mathematical Theory of Black holes, that no student at Caltech wanted to take up the problem of separating the variables of Kerr- Newman metric after Chandra's failure to separate them.

Chandrasekhar's presentation of beauty in these complicated expressions and equations, was the reason behind the title of this talk 'Seeing beauty in the simple and the complex' The following quotation may amplify this point:

> 'The treatment of perturbations of Kerr space-time has been prolixious in its complexity. Perhaps, at a later time, the complexity will be unravelled by deeper insights. But meantime the analysis has led into a realm of the rococo: splendorous, joyful and immensely ornate.'

I looked up the meaning of rococo in the dictionary. it says ornamental; a bit, perhaps, like the doors and the dieties at the entrances and walls in South Indian temples (like say the one at Kancheepuram). The simple may, probably, be compared to the lotus shaped Bahai temple at New Delhi. If one is attuned, there is beauty to be seen in both cases.

In his treatment of colliding gravitational waves, whose metric is closely related to the stationary Kerr metric, Chandra pushed these ideas even further (went really overboard !). He compares the beauty seen in these formulations to the beauty in the series of paintings of Claude Monet. Monet has painted the same subject or scene as seen at different times of the day or in different seasons. Chandra compares this to the different viewpoints of the same metric which manifests itself as Kerr metric or colliding gravitational waves.

4 Non Radial Oscillations

The work on this subject done by Chandrasekhar in 1991 (he was more than 80 years old then) with Valeria Ferrari applies the various techniques perfected by Chandra to neutron stars. The surprising results they obtained, according to Chandra should have counterparts in the Newtonian theory too. Quotations from his paper may be of interest.

First about the work: ' We develop ab initio a complete unified version of the theory of non-radial oscillations of a spherical distribution of matter (a star!) that provides not only a different physical base for the origin and nature of these oscillations but also simpler algorithms.... for numerical evaluations of quasi normal modes ($l \geq 2$) and the real frequencies of dipole oscillations ($l = 1$).'

The mode is found by calculating the scattering cross section of a gravitational wave as a function of energy and locating the peak (resonance). A process which is familiar to particle physicists. However Chandra finds a source of puzzlement and possibly a new insight at a deeper level. We again quote:

> 'On the relativistic theory, the frequencies of oscillations of the non- radial modes (as we have shown) depend only on the distribution of the energy-density and the pressure in the static configuration and the equation of state only to the extent of its adiabatic exponent. If this is a true representation of the physical situation, then it must be valid in the Newtonian theory as well: the true nature of an object can not change with the mode and manner of one's perception. In the relativistic picture, the independence of the frequencies of the non-radial oscillations of a star, on anything except its characterization in terms of its equilibrium structure, is to be understood in terms of the scattering of incident gravitational waves by the curvature of the static space time and its matter content acting as a potential. But what are the counterparts of these same concepts in the Newtonian framework ? Perhaps they lie concealed in the meanings that are to be attached, in the Newtonian theory, to the four metric functions (and their perturbations) that describe a spherically symmetric space-time (and their polar perturbations). It is known that the Newtonian gravitational potential, in some sense, replaces the metric function g_{tt}. Are there similar meanings to be attached to $g_{rr}, g_{\theta\theta}$ and $g_{\phi\phi}$? That is the predominant question to which the present investigation seems to lead.'

To sum up, to use Chandra's own words: *'One is left speculating'*.

References

[1] K.S.Thorne, 1990, Foreword in 'Selected Papers S.Chandrasekhar vol.5, University of Chicago Press.

[2] S.Chandrasekhar, 1970, 'Post-Newtonian methods and Conservation Laws' in 'Relativity', ed. M.Carmeli, S.Fickler and L.Witten, Plenum Press.

[3] S.Chandrasekhar and F.Paul Esposito, 1970, The Astrophysical Journal 160, 153-79.

[4] S.Chandrasekhar, 1968, Ellipsoidal figures of equilibrium, Yale University Press, USA

[5] S.Chandrasekhar, 1982, Mathematical Theory of Black Holes, Clarendon Press, Oxford.

ON THE BLACK HOLE TRAIL ... :

A PERSONAL JOURNEY

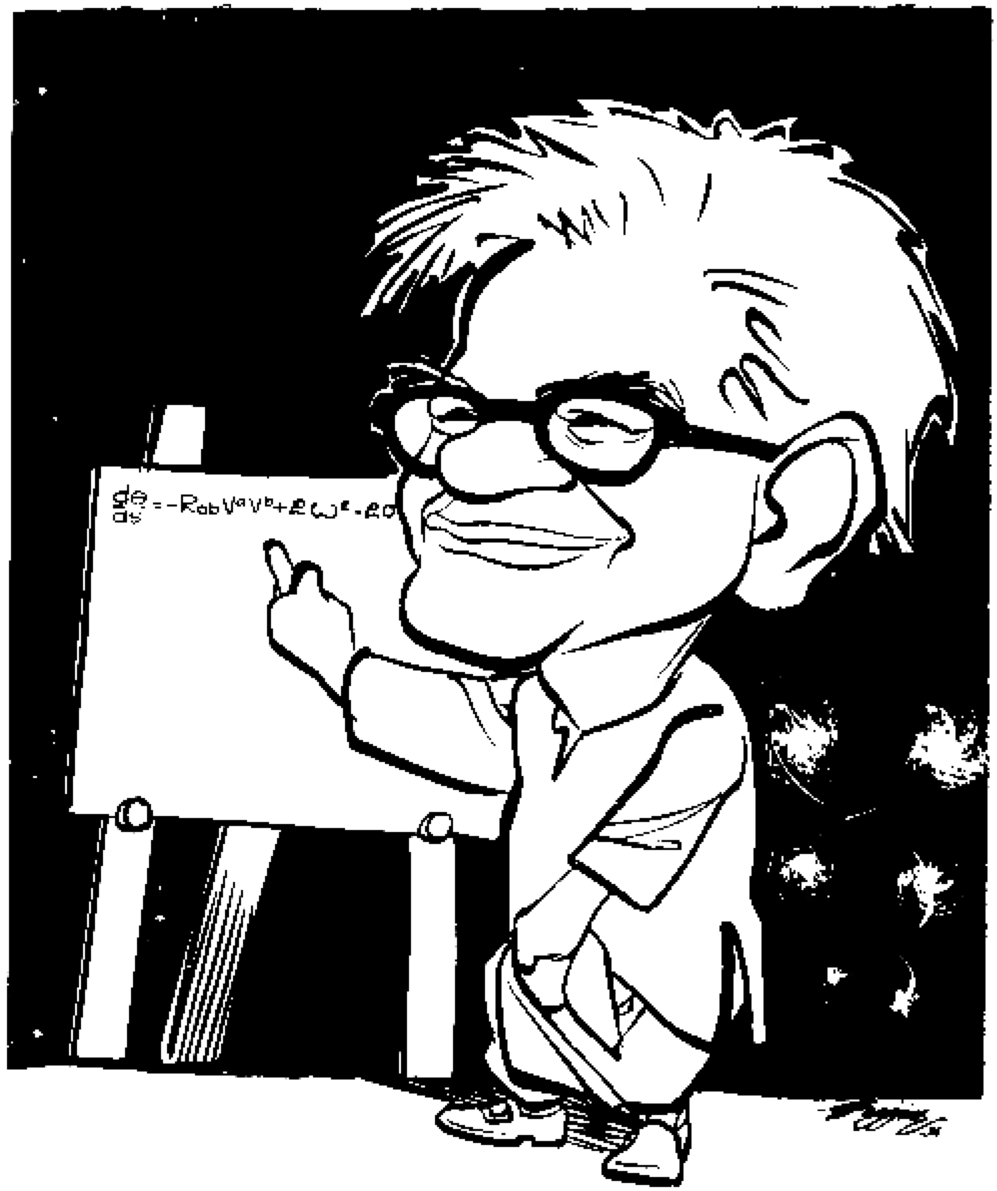
C V Vishveshwara[1]

Indian Institute of Astrophysics
Bangalore - 560 034

[1]e-mail: vishu@iiap.ernet.in

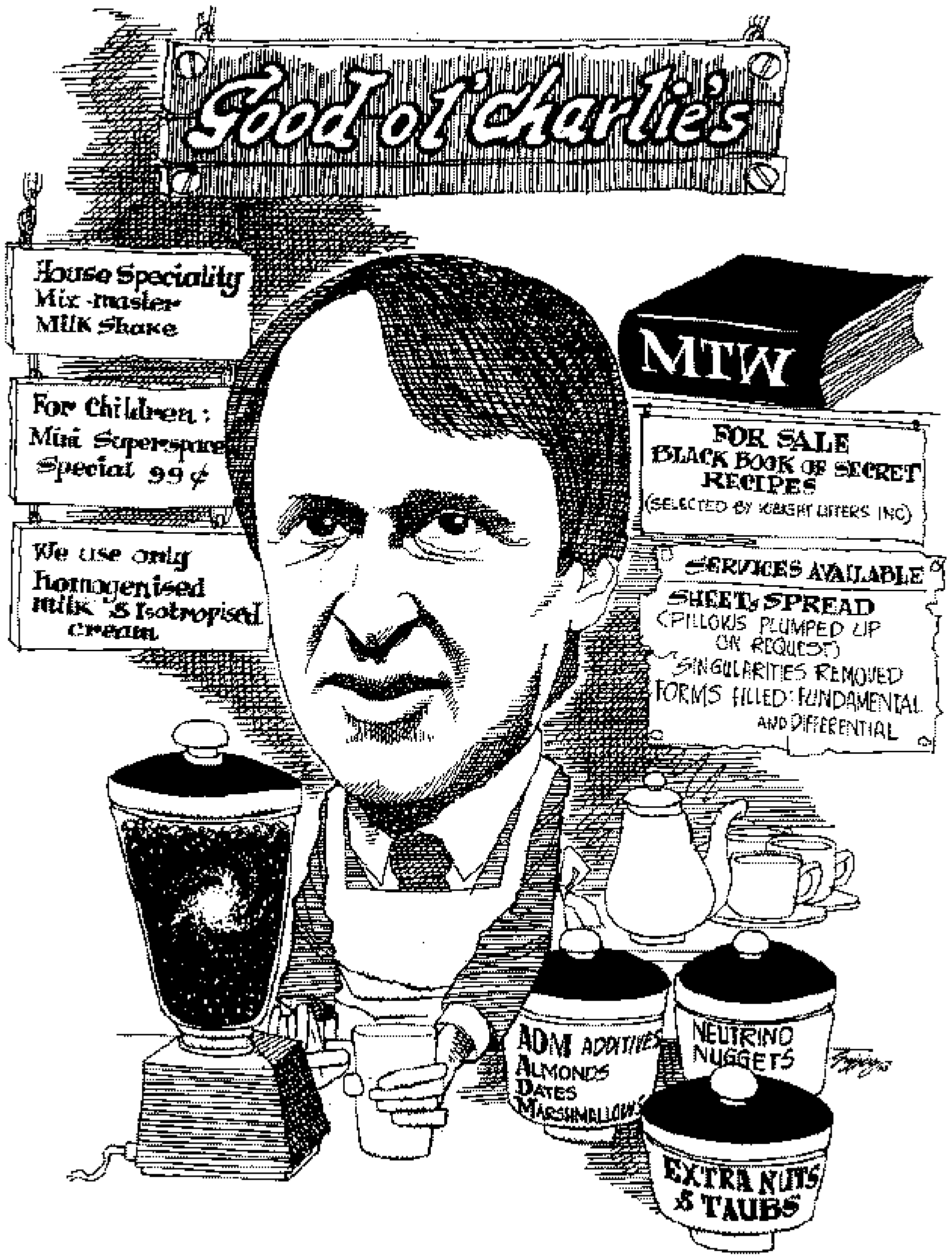
P. C. Vaidya

The Radiant Rider



A. K. Raychaudhuri

The Cosmic Converger



C. W. Misner

The Master Mixer

1 Beginning of the trail

It is a joy to give this talk as a tribute to Professors Vaidya and Raichaudhuri, the two father-figures of general relativity in India. If my talk is rather autobiographical in nature, the responsibility rests with Naresh Dadhich and Bala Iyer, respectively the President and the Secretary of the Indian Association for General Relativity and Gravitation, who persuaded me to make it so.

My personal journey along the black hole trail started in the sixties when I was a graduate student of Charles Misner at the University of Maryland. Also, that was when I first came to know about Vaidya and Raichaudhuri. Leepo Cheng was doing her master's thesis with Misner on the Vaidya metric. She was appalled by my ignorance when I told her that I did not know who Vaidya was. Later on, we were told that Raichaudhuri was coming as a Visiting Professor. Again, my colleagues were suitably impressed by my ignorance when I confessed that I did not know who this Raichaudhuri was either. I went on not only to take a course on cosmology from him but also to pass it with a little bit of honest cheating. Never did I dream that some day I would be delivering a lecture in honour of these two gentlemen.

Let me come back to black holes. That is not what they were called at that time. Schwarzschild singularity, which is a misnomer. Schwarzschild surface, which is better. The term 'Black Hole' was to be coined later on by John Wheeler. Perhaps, it was because of its intriguing name that so many people were enticed into working on the physics of the black hole. This is known as the Schicklgruber Effect. Scholars have speculated on how human history might have been different if Mr. Schicklgruber had not changed his name. But, he did change his name to Hitler. Misner proposed the following problem for my Ph.D. thesis. Take two of these entities that are now called black holes. Revolving around each other, they come closer as energy is radiated away in the form of gravitational waves. They coalesce into an ellipsoidal 'Schwarzschild surface' still rotating and radiating. Study the whole process, computing all the characteristics of the emitted gravitational radiation. Fine, I said, thy will be done! At the time I did not realize the magnitude of this problem. Had I pursued it, I might have entered the Guinness book of records as the oldest graduate student alive that too without financial support. Anyway, this proposed problem required the understanding of two aspects of black holes: the geometrical structure of a black hole and the perturbations of its spacetime.

2 Geometry of black holes

Those were the early days when very little was known about black holes. Wheeler was going around giving his talk "Gravitational Collapse: To what?" with missionary zeal. There was some vague notion of the metric component g_{00} of static spacetimes tending to zero on some surface. I distinctly remember the cold morning when, on the way to grab a sandwich at the little store run by the school of dairy research, Misner suggested that I look into this shady business. Fine, I said, thy will be done! There were false starts. I had this excruciating experience of translating to myself a lengthy paper in German by Ehlers and Sachs - or was it Ehlers, Kundt and Sachs, I forget - on ray optics and optical scalars in the hope that it would throw some light on the matter. It did not. Nevertheless, it was the article by Ehlers and Kundt in "Gravitation : An Introduction to Current Research", edited by Louis Witten[1] that gave the clue to the secret of the black hole structure. The covariant approach to the unravelling of the black hole geometry was via the spacetime symmetries or Killing vector fields. This way the fundamental properties of both Schwarzschild and Kerr black holes could be analysed, compared and contrasted. For instance, given a Killing vector field ξ one could derive the equation[2],

$$n^a n_a \ = \ \frac{1}{2}\left[\xi^a \xi_a (\xi_{b;c} \xi^{b;c}) \ - \ \omega^r \omega_r\right] \tag{1}$$

where ω^a is the vorticity of the Killing congruence and n^a is the normal to surfaces of constant Killing norm, ie.,

$$\Sigma : \xi^a \xi_a \;=\; \text{constant}. \tag{2}$$

This shows that the surface on which ξ^a becomes null ($\xi^a\xi_a = 0$), is itself a null surface, equivalently a one-way membrane or an event horizon ($n^a n_a = 0$), provided the vorticity also became null on the surface. This happens to be true for the global timelike Killing vector ξ^a in the Schwarzschild metric, which is rotation free. In the case of Kerr spacetime this is achieved for a suitable combination of the global timelike Killing vector ξ^a and the rotational Killing vector η^a. Consequently, the Schwarzschild black hole is both a one-way membrane and a static limit whereas in the Kerr spacetime these two surfaces are distinct thereby possessing the ergosphere between them. And this is where interesting phenomena like the Penrose process and the consequent energy extraction can occur.

Although the global timelike Killing vector field ξ^a of the Kerr spacetime possesses non-zero vorticity or rotation, Kerr spacetime admits an irrotational vector field

$$\chi^a = \xi^a - \frac{(\xi^b \eta_b)}{(\eta^c \eta_c)}\eta^a \tag{3}$$

which is timelike down to the black hole. This vector field defines the Locally Non-Rotating Frames, $LNRF$[3], or the Zero Angular Momentum Observers, $ZAMO$[4]. But this vector field exhibits much more interesting properties. These were investigated by Richard Greene, Englelbert Schücking and myself [5] around 1970. I had by then joined Schücking at New York University after a stint at the Institute for Space Studies of NASA in New York and a short period of unemployment. Perhaps it was too much to expect that black holes would be a source of income, since they were not sources of anything in the first place. To continue, we were able to generalize the irrotational vector field to arbitrary stationary, axisymmetric spacetimes with orthogonal transitivity. It was shown to be globally hypersurface orthogonal normal to t = constant surfaces. These are maximal surfaces. The vector field could become null on an event horizon. Some features of this study, such as the physical interpretation of the mathematical conditions necessary for these properties, are still open problems. Incidentally, B R Iyer and I [6] have recently renamed LNRF or ZAMO as GHOST - Globally Hypersurface Orthogonal Stationary Trajectories!

The geometry of Killing trajectories, i.e. the integral curves of Killing vector fields, that plays such a basic role in elucidating the black hole structure, sneaked into our investigations in an indirect manner. Eli Honig was studying the motion of charged particles in homogeneous electromagnetic fields using the Frenet - Serret (FS) formalism [7]. This formalism offers a geometric description of an arbitrary curve characterizing it by certain scalars and an orthonormal frame of reference at each point. In three dimensions these scalars are κ, the curvature and τ_1 the torsion. In four dimensional general relativity, we have an additional torsion τ_2. Furthermore, the derivatives are with respect to proper time and, as a consequence, κ turns out to be the magnitude of four-acceleration. Similarly, the precession rate of a gyroscope carried along the curve has components τ_1 and τ_2 with respect to two members of the Frenet- Serret tetrad at each point of the curve. Now the worldlines of charges moving in a constant electromagnetic field F_{ab} bear striking resemblance to Killing trajectories. In both cases, each member of the FS tetrad satisfies the Lorentz equation. For Killing trajectories

$$F_{ab} = e^{\psi}(\xi_{a;b}) \tag{4}$$

where the normalization factor $e^{\psi} = (\xi^a\xi_a)^{-\frac{1}{2}}$. In both cases, one can show κ, τ_1 and τ_2 are constants along the worldline and

$$\kappa^2 - \tau_1^2 - \tau_2^2 = \frac{1}{2} F_{ab}F^{ab} \tag{5}$$

In the case of Killing trajectories, τ_1 and τ_2 turn out be the components of vorticity. Further, acceleration is given by n_a, the gradient of equipotentials $\xi^a\xi_a = constant$, so that κ^2 is proportional to $n^a n_a$. With these substitutions, equation (5) reduces to equation (1). So, we have indirectly rederived the original black hole equation.

We shall return later to gyroscopic precession which we have mentioned here in passing.

3 Stability of the Schwarzschild black hole

The ultimate problem for my Ph.D. thesis, as mentioned earlier, was supposed to be the coalescence of two black holes. In order atleast to make a beginning on this problem, I had to study perturbations superposed on the Schwarzschild spacetime as the background. The canonical paper in this area was ofcourse the one by Regge and Wheeler[8]. To me this was a completely unknown territory. I remembered vaguely a remark in Wheeler's book "Geometro- dynamics" [9] to the effect that the stability of the Schwarzschild spacetime was a problem far from having been solved satisfactorily. In fact it was this book, totally incomprehensible to me when I was a first year graduate student in Columbia University and hence highly intriguing, that had drawn me towards general relativity. Another student of Misner, Lester Edelstein and I rederived the perturbation equations and published them [10] since these equations, as they appeared in the paper by Regge and Wheeler as well as at other places, contained errors. Lester was the first one to work out the radiation emitted by a particle falling into a black hole. He could not track down a factor of two that was missing when he compared his formula with the one of Landau and Lifshitz in the weak field limit. Unfortunately, as a result, he never completed his thesis and eventually switched from general relativity to actuaries.It was around this time that S. Chandrasekhar visited Maryland. It was a big event. Wheeler and his group, that included Uri Gerlach, Bob Geroch and Kip Thorne, drove down from Princeton. Chandra was getting interested in general relativity and in particular black hole perturbations. We gave him our newly derived, yet unpublished equations. I had met Chandra a couple of years earlier at the Boulder Summer School in Colarado when I was aspiring to be a particle physicist in Columbia University. Other participating Indian students and I discussed with him our research as well as a bit of Indian science. In later years, I was to have the great privilege of having many discussions with Chandra on black hole physics, which became his chosen territory, and Indian science, in which he was keenly interested.

Stability analysis, as Ed Salpeter once put it, consists in finding out whether a system breaks apart if an ant sneezed in its vicinity. In the case of the black hole the ant's sneeze is represented by metric perturbation which is a product of Fourier time mode exp $(i\omega t)$, angular function which is a suitable tensor spherical harmonic and a radial function. Assuming the radial function to be well behaved, one had to show that imaginary frequencies that would make the perturbations grow exponentially in time were not admitted. Good behaviour had to be tested in reference to Kruskal coordinates that are singularity free at the black hole. Moreover, the radial functions corresponding to real frequencies had to be shown to form a complete set so that wave packets could be built that did not blow up in time. All this could be done for odd parity perturbations for which the radial function was governed by a Schrödinger type equation with an equivalent potential. Frank Zerilli[11] would later derive a similar equation in the case of even parity perturbations. But, at the time, the even parity equation was a mess with frequency appearing all over the place. Stability analysis did not seem to go through. I was stuck, hopelessly stuck. Misner, who was going away to Cambridge for a year, suggested that I find a few simple, solvable problems and string them together into a thesis. My heart jumped into my mouth and my other organs rearranged themselves accordingly. I decided to devote another two weeks to the problem - body, mind and soul - and then quit if I did not make any progress. Those were the days when Joe Weber was setting up his gravitational wave detector. Weber and his group were observing a rather peculiar phenomenon. Regularly around midnight the detector would record a sharp, beautiful peak. And then again another peak after an interval of a few minutes. Joe Sinsky, a graduate student, stayed on in the laboratory one night to investigate this puzzling phenomenon. Around midnight the door opened, a security guard came in and banged the door shut - the first peak. After making sure everything was secure in the laboratory he went out banging the door shut again - the second peak. Probably it was the same security guard who used to visit me around one in the morning. My working hours used to be from nine in the night to two in the morning. He would remove his shoes and his belt heavy with holstered gun, put up his feet on the desk and rest for a while. He would tell me what was going on in the world, including the parking lot, appreciate my working hard all alone through the night, sympathise with my non-existent wife waiting for me and then move on. On the eighth day from my decision to give stability a last try, my friend found me in a state of absolute euphoria. I had solved the problem. It had taken quite a bit of complicated analysis of the messy equation. Misner did not believe at first that the stability problem had been solved. But, after being convinced, he pronounced that my thesis was in the bag[12] and went away to Cambridge. And I,on my part, goofed off for one whole year.

Apart from establishing the stability of the Schwarzschild black holes [13], the perturbation analysis had shown that the spacetime did not admit static perturbations that were regular at both the black hole and infinity. This was an indication that distorted static black holes could not exist in isolation. Nevertheless, it was startling to learn that Werner Israel had discovered the uniqueness of the Schwarzschild black hole [14]. There would be no potato shaped black holes for instance. Nature had been robbed of her infinite variety. On the other hand this clearly exhibited nature's simplicity. A static black hole could have only the shape of a sphere - the most perfect figure. After all, the philosopher Xenophanes, as early as in the sixth century B.C., had declared that even God, being perfect, had to be spherical in shape!

4 Quasinormal modes

Halfway through the defense of my Ph.D. thesis, the examiner from the mathematics department asked the question, probably in a rhetorical vein, why one should bother to prove the stability of an object that was impossible to observe and was of doubtful existence in the first place. My thesis advisor did not like the question in the least especially coming from a mathematician. The rest of the examination ended up as a verbal battle between the two which I watched with great satisfaction. But, the question remained: how do you observe a solitary black hole? To me the answer seemed obvious. It had to be through scattering of radiation, provided the black hole left its finger print on the scattered wave. I remembered from my first-year graduate course in quantum mechanics, how the reflection coefficient displayed maxima and minima in a wave scattered from a square barrier. In the case of the black hole also the scattering was from a barrier although of a different shape. So I thought I might discover maxima in reflection coefficient characteristic of the black hole. In order to carry out this calculation you needed a computer since the radial equation had to be numerically integrated. The days of the PCs were far in the future. I was working at the Institute for Space Studies in New York where we did enjoy some luxuries.One of them was chilled beer that was sold at a quarter a can during seminars. So much so, the listeners soaked up more alcohol than astrophysics. The other luxury was computer time which was quite dear and scarce at other places. In addition, we had the help of a numerical analyst and a computer programmer. The reflection coefficient did show maxima albeit extremely faint. I became highly excited. But, when the range of integration was increased, the maxima shifted to some other frequency region. After quite a bit of computer experimentation, I decided that these were spurious maxima produced by the abrupt cut off of the effective potential. My conjecture was that a completely smooth potential would not give rise to maxima in the scattering cross section. I consulted Regge and Wheeler when I gave a talk at Princeton in 1969 with the alliterative title 'Schwarzschild Surface as a Stable Scattering Centre'. It was just before my seminar that I heard for the first time the term 'black hole' newly coined by Wheeler, which he illustrated with a picture of automobile junkyard he drew on the black board. Regge and Wheeler both agreed that there was no theorem connecting the smoothness of the potential to the non-existence of maxima in the scattering cross section. I still do not know the answer.

Although the scattering of monochromatic waves did not show obvious characteristics of the black hole, I felt that scattering of wave packets might reveal the imprint of the black hole. So, I started pelting the black hole with Gausian wave packets. If the wave packet was spatially wide, the scattered one was affected very little. It was like a big wave washing over a small pebble. But when the Gaussian became sharper, maxima and minima started emerging, finally levelling off to a set pattern when the width of the Gaussian became comparable to or less than the size of the black hole. The final outcome was a very characteristic decaying mode, to be christened later as the quasinormal mode. The whole experiment was extraordinarily exciting.

By the time the above work was published in Nature [15], I had moved to New York University. Chandra made a visit and gave a talk on ellipsoidal figures of rotating fluids. He was very much interested in my work on scattering and in the phenomenon of decaying modes.Later on he was to compute the quasinormal mode frequencies with Detweiler[16]. Many calculations in this direction would follow finally culminating in the accurate determination of the frequencies by Nils Andersson[17].

Quasinormal modes are generated in astrophysical scenarios such as gravitational collapse and coalescence of black holes. Ed Seidel has shown how well the fundamental mode matches the outcoming wave during the coalescence of binary black holes[18]. Recently Aguirregabiria and I have studied the sensitivity of the quasinormal modes to the scattering potential[19]. The motivation is to understand how any perturbing influence, such as another gravitating source, that might alter the effective potential would thereby affect the quasinormal modes. Interestingly, we find that the fundamental mode is in general insensitive to a small changes in the potential, whereas the higher modes could alter drastically. The fundamental mode would therefore carry the imprint of the black hole, while higher modes might indicate the nature of the perturbing source.

Quasinormal modes are perhaps the rebuttal to the criticism of my thesis examiner regarding the nonobservability of black holes.

5 Ultracompact objects

One of the indirect offshoots of black hole research was the study of ultracompact objects or UCOs. While investigating the scattering of gravitational waves from the Schwarzschild black hole, I had noticed a peculiar phenomenon which I did not publish. Although, at the time, the radial equation for even parity perturbations was quite complicated and was not in the Schrödinger form, it yielded exactly the same reflection coefficient as the odd parity perturbation for a given angular parameter l. One day I got a very excited telephone call from Chandra enquiring whether I knew this fact. I answered, yes, I did. Did I know why this happened? No, I did not. He had found the reason, he said triumphantly. He asked me for the numbers I had computed which I sent him. He went on to publish his interesting conditions under which two potentials lead to identical scattering cross sections mentioning my foreknowledge of the fact but not the reason.

It is the same story with neutrinos as well. The two equivalent potentials corresponding to the two helicities are quite different from each other [20], but lead to identical reflection coefficients. One of them has the peculiar feature in that it has a potential well in the region $r < 3m$ attached to the usual potential barrier. This was terribly intriguing. Could there be a bound state in the potential well giving rise to some sort of neutrino trapping by the black hole? One can estimate the maximum number of bound states by integrating the potential over its spatial range. The well depth increases with the angular momentum quantum number and in the limit of its tending to infinity you get the answer one for the maximum number of possible bound states. In otherwords, there are no bound states.

I did not publish any of the above results. But out of it all another interesting question arose. Suppose you replaced the black hole by a spherical star of radius $r < 3m$. Then the potential well would not only exist, but would also be deepened by the enhanced gravitation of the matter. Could there then be bound states trapping neutrinos within the star? Ajit Kembhavi and I worked on this problem, found and computed the complex frequencies corresponding to the bound states of the neutrinos [21]. In a way these neutrino bound states with complex frequencies were forerunners of the quasinormal modes of ultracompact stars worked out by Chandrasekhar and Ferrari [22] as has been pointed out by Andersson[23]. It is a very happy feeling that some of the problems I had worked on interested Chandra also.

Ultracompact objects with radius $r < 3m$ are in fact quite interesting entities. In principle, trapping of massless particles in their potential well is possible. Or the object can oscillate in its quasinormal modes. Van Paradijs [24] has pointed out the peculiar behaviour of redshift for $r < 3m$. Recently Abramowicz and Prasanna [25] have discussed the reversal of centrifugal force at r = 3m for which you need a black hole or a UCO.

But, do such highly compact objects or stars with radius $r < 3m$ exist in nature? This question was considered by Dhuranhar, Iyer and myself [26,27] and we also coined the name 'Ultra-Compact Objects' or 'UCO's. By studying very carefully the general relativistic stellar models with different equations of state we established that, as a matter of fact, stable ultracompact objects can exist in nature.

6 Gyroscopic precession and inertial forces

Right in the beginning of this talk we discussed how the black hole structure can change dramatically when going from static to stationary spacetimes on account of the rotation inherent to the latter. The study of Killing trajectories in these spacetimes led to a covariant description of gyroscopic precession via the Frenet-Serret formalism. Precession is an important phenomenon. For instance, the Earth precesses. Ancient astronomers knew this. Astrologers did not, thereby making predictions that were doubly wrong. Or, can two wrongs add upto one right? In atomic physics Thomas precession, a manifestation of special theory of relativity, played a crucial role. Spacetime curvature gives rise to Fokker - De Sitter precession in the Schwarzschild spacetime. The spin of Kerr black hole contributes additional precessional effects. All these can be studied elegantly using the Frenet-Serret description[6].

Another related area concerns the general relativistic analogues of inertial forces as developed by Abramowicz and coworkers [28]. A particle at rest in a static spacetime experiences only the gravitational force, but is acted upon by the centrifugal force as well if it is moving uniformly in a circular orbit. In a stationary spacetime, there is an additional force, the Coriolis-Lense-Thirring force, which arises as a consequence of the metric components mixing space and time. In static spacetimes, such as the Schwarzschild spacetime, centrifugal force reverses at the circular photon orbit [25]. So does gyroscopic precession. The situation is far more complicated in stationary spacetimes[29,30]. My young colleague Rajesh Nayak and I have studied these effects and established covariant connections between gyroscopic precession on the one hand and inertial forces on the other [31,32,33]. These considerations should be of interest in black hole physics from a conceptual point of view as well as for astrophysical applications.

7 The trail goes on ...

I have tried to offer a glimpse, just a fleeting one at that, of my personal journey along the black hole trail. It has been a long journey spanning some three decades. There have been all sorts of ups and downs along the way. For instance, I have had my share of tussle with journals and referees. My very first paper [10], the one with Edelstein, was unceremoniously rejected as nothing more than a bunch of formulae. Misner had to write a strong letter pointing out that the same journal that had previously published the wrong equations was now rejecting the correct ones. The paper on the structure of black holes [2] was also rejected as it was considered to be just mathematics and had to be published in the Journal of Mathematical Physics. The stability paper [13] too had to cross some hurdles before seeing the light of the day. As with any important field, black hole physics has had its sociological factors sometimes leading to, among other things, inadequate recognition of significant contributions. All this becomes trivial in comparison to the exhilarating experience of exploration. It is a rare good fortune to have been trekking along the track right from the beginning. To have watched the seed germinate, the sapling sprout and the tree grow. It is also a good fortune to have had the company of congenial co-travellers on the journey - marvellous friends to work with and keen minds to lead the way. If sometimes you stray away from the road, you keep coming back. Even now my colleagues and I are working on different aspects of black hole physics, such as quasinormal modes, rotational effects and black holes in cosmological backgrounds.

What is the most important lesson I have learnt having traversed the trail for so long? Let me answer that question by quoting the Spanish poet Antonio Machado who wrote:

Caminante, no hay camino
Se hace camino al andar.

Traveller there is no path,
Paths are made by walking.

It is gratifying to feel that you have made a path however short, however narrow that has helped build a trail that was planned and paved by so many. It has been a joy to follow that trail. And I hope the trail will never end...

Acknowledgments

It is a pleasure to thank Sri. Gujjar for his delightful drawings of Professors Charles Misner, P. C. Vaidya and A. K. Raychaudhuri. Many thanks are due to Ms. G. K. Rajeshwari and Sri. Rajesh Nayak for their help in the preparation of the manuscript.

References

[1] J. Ehlers and W. Kundt, *Exact Solution of the Gravitational Field Equations* in *Gravitation: An Introduction to Current Research*, ed. Louis Witten, John Wiley (1962)

[2] C. V. Vishveshwara, *J. Math. Phys.*, **9**, 1319 (1968)

[3] J. Bardeen, *Ap. J.*, **162**, 71 (1970)

[4] *Black Holes - The Membrane Paradigm*, ed. K.S. Thorne, R.H. Price and D. A. Macdonald, Yale university Press (1986)

[5] R. D. Greene, E. L. Schücking and C.V. Vishveshwara, *J. Math. Phys.*, **16**, 153 (1975)

[6] B.R. Iyer and C.V. Vishveshwara, *Phys. Rev.*, **D48**, 5706 (1993)

[7] E. Honig, E.L. Schücking and C.V. Vishveshwara, *J. Math. Phys.*, **15**, 774 (1974)

[8] T. Regge and J. A. Wheeler, *Phys. Rev.*, **108**, 1063 (1957)

[9] J. A. Wheeler, *Geometrodymanics*, Academic Press (1962)

[10] L. A. Edelstein and C. V. Vishveshwara, *Phys. Rev.*, **D1**, 3514 (1970)

[11] F. J. Zerilli, *Phys. Rev. Lett.*, **24**, 737 (1970)

[12] C. V. Vishveshwara, *Ph.D. Thesis*, University of Maryland (1968)

[13] C. V. Vishveshwara, *Phys. Rev.*, **D1**, 2870 (1970)

[14] W. Israel, *Phys. Rev.*, **164**, 1776 (1967)

[15] C. V. Vishveshwara, *Nature*, **227**, 936 (1970)

[16] S. Chandrasekhar, *The Mathematical Theory of Black Holes*, Oxford University Press, New York (1983)

[17] N. Andersson, *Proc. R. Soc. Lond. A*, **439**, 47 (1992)

[18] E. Seidel, *Proc. of the Int. Conf. on Gen. Rel. and Grav. -1995*, Pune (1996)

[19] J. M. Aguirregabiria and C. V. Vishveshwara, *Scattering by Black Holes: A Simulated Potential Approach*, *Phys. Lett. A*, to be published (1996)

[20] D. Brill and J. A. Wheeler, *Rev. Mod. Phys.*, **29**, 165 (1957)

[21] A. K. Kembhavi and C. V. Vishveshwara, *Phys. Rev.*, **D22**, 2349 (1980)

[22] S. Chandrasekhar and V. Ferrari, *Proc. R. Soc. Lond. A*, **334**, 449 (1991)

[23] N. Andersson, *On the Oscillation Spectra of Ultracompact Stars: An Extensive Survey of Gravitational Wave Modes*, preprint (1996)

[24] J. Van Paradijs, *Ap.J.*, **234**, 609 (1939)

[25] M. A. Abramowicz and A.R. Prasanna, *Mon. Not. R. Astr. Soc*, **245**, 720 (1990)

[26] B. R. Iyer and C. V. Vishveshwara, *General Relativistic Aspects of Neutron Star Models* in *A Random Walk in Relativity and Cosmology*, ed. N. Dadhich, J. Krishna Rao, J. V. Narlikar and C. V. Vishveshwara, Wiley Eastern (1985)

[27] B.R. Iyer, C.V. Vishveshwara and S.V. Dhurandhar, *Class. Quantum. Grav,*, **2**, 219 (1985)

[28] M.A. Abramowicz, P.Nurowski and N. Wex, *Class Quantum. Grav.*, **10**, L183 (1993)

[29] S. K. Chakrabarti and A. R. Prasanna, *J. Astrophys. Astr.*, **11**, 29 (1990)

[30] S. Iyer and A. R. Prasanna, *Class. Quantum. Grav.*, **10**, L13 (1993)

[31] K. R. Nayak and C. V. Vishveshwara, *Gyroscopic Precession and Centrifugal Force in the Ernst Spacetime*, to be published

[32] K. R. Nayak and C. V. Vishveshwara, *Gyroscopic Precession and Inertial Forces in the Kerr-Newman Spacetime*, to be published

[33] K. R. Nayak and C. V. Vishveshwara, *Gyroscopic Precession and Inertial Forces in Axially Symmetric Stationary Spacetimes*, to be published

GRAVITATIONAL WAVES

FROM INSPIRALLING COMPACT BINARIES

Bala R. Iyer[1]

Raman Research Institute
Bangalore 560 080

Abstract

Inspiralling compact binaries are the most promising sources for gravitational wave detectors like LIGO and VIRGO. For detection one needs to know the gravitational waveform to accuracies much beyond quadrupole approximation. This requires a solution of the generation problem and radiation reaction problem in general relativity. This talk presents a broad brush overview of the techniques used to study these topics and the important nonlinear effects that obtain.

[1]e-mail: bri@rri.ernet.in

What impressed me about your picture was the extremely striking manner in which you visually portray one's inner feeling towards one's efforts at accomplishments: one is half-way up the ladder, but the few glimmerings of structure which one sees and to which one aspires are totally inaccessible, even if one were to climb to the top of the ladder. The realization of the absolute impossibility of achieving one's goals is only enhanced by the shadow giving one an even lowlier feeling of one's position.

— S. CHANDRASEKHAR

1 Introduction

On 2 July 1974 Hulse and Taylor discovered the first signals from the binary pulsar 1913+16, a system of two neutron stars that orbit around each other. Since the two neutron stars are moving fast and relatively close together according to general relativity they should emit moderate amounts of gravitational radiation. This makes them lose energy. Their orbit should therefore shrink and their orbital period should shorten. Since its discovery this system has been monitored continuously and it is observed that the orbiting period has in fact decreased. Quantitatively the observations have yielded a verification of the quadrupole formula for radiation damping to better than 0.4 %[1]. All this is considered as sufficient evidence albeit indirect, that gravitational waves exist and the Nobel Prize in 1993 certifies the quality of the evidence. It is currently one of our strongest support for validity of general theory of relativity.

With an orbital period of eight hours, the frequency of gravitational waves from the binary pulsar today is very low: (10^{-4} Hz). However if we wait patiently for a short time of about hundred million (10^8) years the two stars will inspiral and the gravitational waves will sweep upward in frequency to about 10 Hz. In the following fifteen minutes before the neutron stars collide and coalesce the frequency will rise to about 1000 Hz with increasing amplitude producing a characterstic chirp waveform containing about 16000 cycles. This is the way its Life ends..Not with a whimper but a bang! The amplitude is still too small, however the large number of cycles brings the characterstic signal strength $h_c \sim h\sqrt{n}$ to the realm of the measurable. If we can build detectors that are sensitive to this bandwidth and if there are a sufficient number of such events in our observable universe we should be able to *directly* see gravitational waves on the earth. LIGO[2] and VIRGO[3] are two such experiments that will look for and hopefully see such events by 2001! Estimates of the rate of such coalescence events are about a few per year upto 200 Mpc. Advanced LIGO which would look upto cosmological distances [4, 8] would get to numbers of hundreds per year[9, 12].

The excitement over these experiments arises not just from the possibility of direct detection of a new radiation but from the hope of a new astronomy: Gravitational Wave Astronomy. The possibility of opening a new window to the universe is reflected in the name itself: A **Gravitational Wave Observatory** rather than a mere *gravitational wave detector*. What makes all this possible is the fact that though the gravitational wave signal is extremely weak and buried deep in the detector noise it consists of a large number of precisely predictable cycles in the detector bandwidth. This enables one to use the technique of matched filtering first for detection and later for estimation of parameters of the coalescing binary. For instance, the LIGO and VIRGO observations of inspiralling compact binaries should provide precise measurements of the masses of the objects, possibly of their spins and probably, in the case of neutron stars, of their radii [13, 20]. The absolute luminosity distance of the binary will be measured independently of any assumption concerning the nature (and masses) of the objects[21]. But if the intrinsic masses of the objects are known, the cosmological red-shift of the host galaxy where the event took place can be measured. There is hope to deduce in this way a measurement of the Hubble parameter from an assumption concerning the statistical distribution of masses of neutron stars [22, 23]. The puzzle of the origin of gamma-ray bursts could be solved by comparison of the times of arrival of the gravitational waves and of gamma ray bursts[11, 24, 25]. Furthermore new limits on the validity of alternative theories of gravity, notably scalar-tensor theories[26] and new tests of general relativity in the strong field regime[27] should be possible by monitoring very precisely the inspiralling signal.

The strength of the wave from a gravitational wave source can be estimated using the Newtonian or quadrupole approximation [28, 29]. One has

$$h \sim \frac{G}{c^4}\frac{\ddot{Q}}{r}$$

The strongest sources must be highly nonspherical $Q \sim ML^2, \ddot{Q} \sim 2Mv^2 \sim 4E_{\rm kin}^{ns}$ implying

$$h \sim \frac{1}{c^2}\frac{4G\left(\frac{E_{\rm kin}^{ns}}{c^2}\right)}{r}$$

A huge internal kinetic energy by virial theorem is generically accompanied by a high gravitational potential energy *i.e* a high compactness. For kinetic energies of the order of a solar mass one has $h \sim 10^{-21}$ at 200 Mpc. The thumb rule values are thus around $10^{-21} - 10^{-22}$. Laser interferometers are capable of measuring displacements of order 10^{-16} cm and recalling that $h \sim \Delta L/L$ one is led to the interferometer sizes of 1 km -10 kms. The gravitational wave source cannot emit strongly at periods smaller than the light travel time $4\pi GM/c^3$ and thus the frequencies at which these sources emit are

$$f \leq \frac{1}{4\pi GM/c^3} \sim 10^4 \frac{M_\odot}{M} \text{Hz}.$$

Gravitational waves from astronomical objects that exist today lie between 10^{-4}Hz $-$ 10^4Hz. Gravitational waves from the early universe on the other hand could range from 10^{-18}Hz $-$ 10^8Hz. The gravitational wave frequency bands that are actively being explored experimentally may be broadly divided as[28, 29]:

1. High Frequency (HF) band that lies between 10^4Hz $-$ 1Hz. These are explored by earth based gravitational wave detectors like laser interferometers (LIGO, VIRGO) and resonant mass antennas. The sources that radiate in this bandwidth include stellar collapse of neutron stars or black holes in our galaxy or external galaxies, the rotation and vibration of neutron stars in our galaxy and the coalescence of neutron stars and stellar mass black holes (upto 1000 $M_\odot$) in distant galaxies. The stochastic background of vibrating loops of cosmic strings, phase transitions in the early universe and gravitational waves from the big bang itself could radiate in this range. The bandwidth limitation comes from the noise sources. The 1 Hz cutoff is due to fluctuating gravity gradients and the earth's vibration (seismic cutoff).

2. Low Frequency (LF) band lying between 1Hz $-$ 10^{-4}Hz. These are the realm of Doppler tracking of spacecraft by microwaves and in the future optical tracking of spacecraft by each other (LISA). This band is due to waves from short period binary stars in our galaxy (main sequence binaries, cataclysmic variables, white dwarf and neutron star binaries), white dwarfs, neutron stars and small black holes spiralling into massive black holes ($3 \times 10^5 - 3 \times 10^7 M_\odot$) in distant galaxies, inspiral and coalescence of supermassive black hole binaries ($100 - 10^8 M_\odot$) and stochastic background from the early universe. The cutoff at the lower edge is due to difficulty in isolating the spacecraft from forces due to fluctuations in solar radiation pressure, solar wind and cosmic rays.

3. Very Low Frequency (VLF) band between 10^{-7}Hz $-$ 10^{-8}Hz. This band is explored by the timing of millisecond pulsars. At 95 per cent confidence limit the density in these gravitational waves $\Omega_g(4 \times 10^{-9}\text{Hz}) < 6 \times 10^{-8} H^{-2}$ where H is the Hubble constant relative to 100 km sec^{-1} Mpc^{-1}. This range is the stochastic background from early universe processes like cosmic strings, phase transitions and the big bang.

4. Extreme Low frequency (ELF) between 10^{-15}Hz $-$ 10^{-18}Hz. This radiation should produce anisotropies in the CMBR . If all the anisotropy in the COBE measurements were due to this gravitational radiation it would correspond to $\Omega_g(10^{-18}\text{Hz}) \sim 10^{-9}$.

To prepare for the analysis of such signals in the LIGO and VIRGO detectors one needs to compute the gravitational radiation field generated by a system of two point-masses moving on a circular orbit (the relevant case because the orbit will have been circularized by radiation reaction forces). Since inspiralling compact binaries are very relativistic, this problem is highly nontrivial and represents a challenge to relativity theorists[6, 8] as its resolution involves carrying out the calculation to a very high order in terms of a post-Newtonian expansion[30]. The problem can be decomposed into two different problems, which can be referred to respectively as the "wave generation problem" and the "radiation reaction problem" [31].

The wave generation problem deals with the computation of the gravitational waveforms generated by the binary (at the leading order in $1/r$, where r is the distance of the binary) when the orbital phase and frequency of the binary take some given values ϕ and ω. This problem involves computing the amplitude of each harmonic of the wave corresponding to frequencies which are multiples of the orbital frequency, with the predominant harmonic being at twice the orbital frequency.

The radiation reaction problem consists of determining the evolution of the orbital phase $\phi(t)$ itself as a function of time, from which one deduces the orbital frequency $\omega(t) = d\phi(t)/dt$. The actual time variation of $\phi(t)$ is nonlinear because the orbit evolves under the effects of gravitational radiation reaction forces. In principle it should be determined from the knowledge of the radiation reaction forces acting locally on the orbit. However these forces are at present not known with sufficient accuracy (only the relative first post-Newtonian corrections are known[32, 33]), so in practice the phase evolution is determined by equating the high-order post-Newtonian energy flux in the waves or energy loss (averaged over one orbit) and the decrease of the correspondingly accurate binding energy of the binary.

Estimates of the precision needed in the resolution of these two problems can be inferred from black-hole perturbation techniques in the special case where the mass of one object is very small as compared with the other mass[34, 39]. The required precision is reached when the systematic errors due to the neglect of some higher-order approximation become less than the statistical errors due to noise in the detector. It turns out that the post-Newtonian corrections in the time evolution of the phase (radiation reaction problem) will be measurable in advanced detectors, probably up to three post-Newtonian (3PN) orders beyond the (Newtonian) quadrupole radiation. This corresponds to relativistic corrections in the energy loss as high as order $\sim (v/c)^6$ where v is the orbital velocity. The possibility of measuring such high-order corrections can be understood crudely from the fact that, in order not to suffer a too severe reduction in signal-to-noise ratio, one will have to monitor the phase evolution with an accuracy of one tenth of a cycle over the tens of thousands of cycles during the entire passage through the frequency bandwidth of the detector.

None of the above problems can be solved exactly. They are treated by a combination of approximation methods like Post Minkowskian approximation, Post Newtonian approximation and Perturbations about a Curved Background. We list the main features of these schemes below:

1. Post Minkowskian Approximation: It is an expansion in $\gamma_i = GM/c^2R$ or $\gamma_e = GM/c^2D$ where M, L, D are the characterstic mass, size and separation respectively. Loosely it is an expansion in G and hence it is also called weak field, non linear or fast motion approximation. It makes crucial use of the geometry of Minkowski spacetime and its causality properties. The equations in this scheme reduce to a hierarchy of wave equations on Minkowski background which are solved by retarded potentials. The basic complication is in the nonlinear iteration.

2. Post Newtonian Approximation: It is an expansion in $\beta \sim v/c \sim L/\lambda \sim L/c/P$ where v, L, λ and P are the characterstic velocity, size, wavelength and period respectively. Loosely it is an expansion in $1/c$ and is also called slow motion expansion. It uses newtonian concepts like absolute space with an Euclidean metric and absolute time. It uses newtonian techniques and in this viewpoint Einstein theory provides small numerical corrections to Newtonian theory. The equations in this scheme are a hierarchy of Poisson equations which are solved by instantaneous potentials.

3. Perturbation about a Curved Background : This is an expansion about an arbitrary curved background rather than flat Minkowski spacetime. Unlike the PNA scheme which is also applicable to binary systems of comparable masses this scheme is applicable only if one of the bodies is much heavier than the other *i.e* the ratio of the masses is very small. However it takes into account the full background curvature and is applicable for fully relativistic situations $v \sim c$.

The use of multipole expansion methods in combination with the PMA scheme offers one of the most powerful techniques in gravitational radiation studies. Multipole expansion is most conveniently implemented by using symmetric trace free tensors rather than tensor spherical harmonics. The Multipole Post Minkowskian (MPM) method exploits the computational advantage of working in De Donder, Lorentz or harmonic gauge. In this gauge Einstein's equations can be conveniently written in terms of a wave operator of flat space with source terms that include the non linear gravitational stress energy. This permits a solution in terms of flat space Green functions and decomposition of the solution in terms of STF tensors which are eigen functions of the flat space D'Alembertian. For instance[40, 41] the multipole analysis of linearised gravity generated by a compact source $T_{\mu\nu}$

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \tag{1}$$

shows that the linearised metric may be represented as

$$\bar{h}^{00}_{can}(\vec{X},T) = +\frac{4}{c^2}\left[\sum_{l\geq 0}\frac{(-)^l}{l!}\partial_L\left(\frac{M_L(U)}{R}\right)\right] \tag{2}$$

$$\begin{aligned}\bar{h}^{0i}_{can}(\vec{X},T) &= -\frac{4}{c^3}\left[\sum_{l\geq 1}\frac{(-)^l}{l!}\left\{\partial_{L-1}\left(\frac{\dot{M}_{iL-1}(U)}{R}\right)+\right.\right.\\ &\left.\left.+\ \frac{l}{l+1}\varepsilon_{iab}\partial_{aL-1}\left(\frac{S_{bL-1}(U)}{R}\right)\right\}\right]\end{aligned} \tag{3}$$

$$\begin{aligned}\bar{h}^{ij}_{can}(\vec{X},T) &= +\frac{4}{c^4}\left[\sum_{l\geq 2}\frac{(-)^l}{l!}\left\{\partial_{L-2}\left(\frac{\ddot{M}_{ijL-2}(U)}{R}\right)+\right.\right.\\ &+\ \left.\left.\frac{2l}{l+1}\partial_{aL-2}\left(\frac{\varepsilon_{ab(i}\dot{S}_{j)bL-2}(U)}{R}\right)\right\}\right]\ ;\end{aligned} \tag{4}$$

in terms of two sets of STF time dependent moments: the Mass moments and Current moments analogous to the representation of the electromagnetic field in terms of electric and magnetic moments:

$$\begin{aligned}M_L(U) &= G\int d^3x\int_{-1}^{1}dz\left\{\delta_l\hat{x}_L\tilde{\sigma}-\frac{4(2l+1)}{c^2(l+1)(2l+3)}\delta_{l+1}\hat{x}_{aL}\frac{\partial\tilde{\sigma}^a}{\partial U}+\right.\\ &\left.+\frac{2(2l+1)}{c^4(l+1)(l+2)(2l+5)}\delta_{l+2}\hat{x}_{abL}\frac{\partial^2}{\partial U^2}\tilde{T}^{ab}\right\}\quad l\geq 0\ ;\end{aligned} \tag{5}$$

$$\begin{aligned}S_L(U) &= \underset{L}{\mathrm{STF}}\ G\int d^3x\int_{-1}^{1}dz\left\{\delta_l\hat{x}_{L-1}\varepsilon_{i_lab}x^a\tilde{\sigma}^b-\right.\\ &\left.-\frac{2l+1}{c^2(l+2)(2l+3)}\delta_{l+1}\varepsilon_{i_lab}\hat{x}_{acL-1}\frac{\partial\tilde{T}^{bc}}{\partial U}\right\}\quad l\geq 1;\end{aligned} \tag{6}$$

$$\text{where } \sigma = \frac{T^{00}+T^{ss}}{c^2} \text{ and } \sigma^a = \frac{T^{0a}}{c}\ . \tag{7}$$

$$\delta_l(z) = \frac{(2l+1)!!}{2^{l+1}l!}(1-z^2)^l \tag{8}$$

2 Model of the Source

The model mostly used is that of two point objects without structure moving under their mutual gravitational interaction. The extension to include effects due to intrinsic spin have been accomplished. There has been work on taking the tidal corrections into considerations. The model also needs the specification of the equations of motion. This is discussed in the next section. It should be noted that in this problem there are two time scales: the orbital motion and the radiation damping with the former being much smaller than the latter.

3 PN Two body Equations of Motion

For the analysis of the timing of the binary pulsar the basic equations for two-body systems in the post-Newtonian approximation to general relativity were investigated in detail. Restricting attention to two-body systems containing objects that are sufficiently small that finite-size effects, such as spin-orbit, spin-spin, or tidal interactions can be ignored one can write the two-body equation of motion explicitly in the form

$$\mathbf{a} = \mathbf{a}_N + \mathbf{a}_{PN}^{(1)} + \mathbf{a}_{2PN}^{(2)} + \mathbf{a}_{RR}^{(5/2)} + \mathbf{a}_{3PN}^{(3)} + \mathbf{a}_{RR}^{(7/2)} + O(\epsilon^4)\,, \tag{9}$$

where the subscripts denote the nature of the term, post-Newtonian (PN), post-post-Newtonian (2PN), radiation reaction (RR), and so on; and the superscripts denote the order in ϵ. Through 2PN order, the individual terms are given by [42, 43, 15]

$$\mathbf{a}_N = -\frac{m}{r^2}\hat{\mathbf{n}}\,, \tag{10}$$

$$\mathbf{a}_{PN}^{(1)} = -\frac{m}{r^2}\left\{\hat{\mathbf{n}}\left[-2(2+\eta)\frac{m}{r} + (1+3\eta)v^2 - \frac{3}{2}\eta\dot{r}^2\right] - 2(2-\eta)\dot{r}\mathbf{v}\right\}, \tag{11}$$

$$\begin{aligned}\mathbf{a}_{2PN}^{(2)} = & -\frac{m}{r^2}\Big\{\hat{\mathbf{n}}\Big[\frac{3}{4}(12+29\eta)(\frac{m}{r})^2 + \eta(3-4\eta)v^4 + \frac{15}{8}\eta(1-3\eta)\dot{r}^4 \\ & -\frac{3}{2}\eta(3-4\eta)v^2\dot{r}^2 - \frac{1}{2}\eta(13-4\eta)\frac{m}{r}v^2 - (2+25\eta+2\eta^2)\frac{m}{r}\dot{r}^2\Big] \\ & -\frac{1}{2}\dot{r}\mathbf{v}\left[\eta(15+4\eta)v^2 - (4+41\eta+8\eta^2)\frac{m}{r} - 3\eta(3+2\eta)\dot{r}^2\right]\Big\}.\end{aligned} \tag{12}$$

Through 2PN order, the motion is conservative, that is, can be characterized by conserved total energy and angular momentum. At 5/2PN order, the first dissipative terms arise resulting from gravitational radiation reaction. At this order and beyond, conserved energy and angular momentum can no longer be defined. The 3PN terms in $\mathbf{a}$ are formally conservative, while the 7/2PN terms represent the post-Newtonian corrections to radiation reaction. Through 2PN order, the conserved energy and angular momentum are given by [42, 43, 44]

$$E = E_N + E_{PN} + E_{2PN}\,, \tag{13}$$

$$\mathbf{J} = \mathbf{J}_N + \mathbf{J}_{PN} + \mathbf{J}_{2PN}\,, \tag{14}$$

where

$$E_N = \mu\left(\frac{1}{2}v^2 - \frac{m}{r}\right), \tag{15}$$

$$E_{PN} = \mu\left\{\frac{3}{8}(1-3\eta)v^4 + \frac{1}{2}(3+\eta)v^2\frac{m}{r} + \frac{1}{2}\eta\frac{m}{r}\dot{r}^2 + \frac{1}{2}(\frac{m}{r})^2\right\}, \tag{16}$$

$$\begin{aligned}E_{2PN} = & \mu\Big\{\frac{5}{16}(1-7\eta+13\eta^2)v^6 + \frac{1}{8}(21-23\eta-27\eta^2)\frac{m}{r}v^4 \\ & +\frac{1}{4}\eta(1-15\eta)\frac{m}{r}v^2\dot{r}^2 - \frac{3}{8}\eta(1-3\eta)\frac{m}{r}\dot{r}^4 - \frac{1}{4}(2+15\eta)\left(\frac{m}{r}\right)^3 \\ & +\frac{1}{8}(14-55\eta+4\eta^2)\left(\frac{m}{r}\right)^2 v^2 + \frac{1}{8}(4+69\eta+12\eta^2)\left(\frac{m}{r}\right)^2\dot{r}^2\Big\},\end{aligned} \tag{17}$$

$$\mathbf{J}_N = \mathbf{L}_N\,, \tag{18}$$

$$\mathbf{J}_{PN} = \mathbf{L}_N\left\{\frac{1}{2}v^2(1-3\eta) + (3+\eta)\frac{m}{r}\right\}, \tag{19}$$

$$
\begin{aligned}
\mathbf{J}_{2PN} = \quad & \mathbf{L}_N \left\{ \frac{1}{2}(7-10\eta-9\eta^2)\frac{m}{r}v^2 - \frac{1}{2}\eta(2+5\eta)\frac{m}{r}\dot{r}^2 \right. \\
& \left. + \frac{1}{4}(14-41\eta+4\eta^2)\left(\frac{m}{r}\right)^2 + \frac{3}{8}(1-7\eta+13\eta^2)v^4 \right\}, \qquad (20)
\end{aligned}
$$

where $\mathbf{L}_N \equiv \mu \mathbf{x} \times \mathbf{v}$. These, together with the equations of motion, can be derived from a generalized Lagrangian that depends on the acceleration and Euler-Lagrange equations. The equations of motion were obtained by Damour by an Einstein- Infeld- Hoffmann type treatment complemented by analytic continuation techniques of Riesz. Unlike linear electromagnetism the non linear general relativity has the feature that its field equations contain the equations of motion. Damour deduced the n^{th} approximate equation of motion from integrability conditions on the $(n+1)^{th}$ approximated vacuum field equations[42]. The EOM have also been discussed by Grischuk and Kopeijkin [43] by more standard techniques of general relativity yielding equivalent results[42, 43].

4 The Quadrupole Wave Generation Formalism

The far field quadrupole equation is the solution of the generation problem at the lowest order. It asserts that for an isolated system S which is non-relativistic, i.e., slowly moving, weakly stressed and weakly self-gravitating, the gravitation radiation amplitude generated by the source in some Asymptotic coordinate system X^μ far from S is given by

$$
h_{ij}^{rad} = \frac{2G}{c^4 R} P_{ijkl}(\vec{N}) \left\{ \frac{d^2}{dt^2}\left[Q_{kl}\left(T - \frac{R}{c}\right)\right] + O(\frac{1}{c}) \right\} + O(\frac{1}{R^2}) \qquad (21)
$$

$$
\text{where} \quad Q_{ij}(t) = \int d^3x\, c^{-2}\, T^{00}\, \hat{x}^{ij} \qquad (22)
$$

$$
\hat{x}^{ij} = x_i x_j - \frac{1}{3}\delta_{ij}\vec{x}^2 \qquad (23)
$$

is the usual Newtonian quadrupole moment of mass density $c^{-2}T^{00}$ in S in some Cartesian coordinate system x^μ covering S but distinct from X^μ. P_{ijkl} is the transverse trace free projection operator onto plane perpendicular to $\vec{N}$ and $\hat{x}^{ij}$ the trace free part of $x^i x^j$. This equation is valid at leading order in R^{-1} and leading order in the small parameter c^{-1} describing the non-relativistic character of this system. The above standard quadrupole equation was first derived by Einstein[45] within the linearized approximation to general relativity and is applicable only to slowly moving sources with negligible self gravity. Consequently, it does not apply to any realistic astrophysical system since e.g., it cannot be applied to a binary system of two ordinary stars whose motion is governed by gravitational forces. Hence these equations need to be generalized to at least weakly self-gravitating systems. This was achieved along two very different lines by Landau-Lifshitz[46] and Fock[47] respectively. The associated total power or luminosity in the gravitational wave emission is

$$
\mathcal{L} = \int \frac{d\Omega}{4\pi}\frac{dE^{\text{wave}}}{dT d\Omega} = \frac{G}{c^5}\left\{ \frac{d^3 Q_{ij}}{dT^3} + O(\frac{1}{c^2}\right\} \qquad (24)
$$

This power can be expressed as the work done locally in the source by a radiation reaction force

$$
\mathcal{L} = -\int_{\text{source}} d^3x \vec{v}.\vec{F}_{\text{reac}} \qquad (25)
$$

$$
\vec{F}^{\text{reac}} = \rho \nabla V^{\text{reac}} \qquad (26)
$$

$$
\text{where} \quad V^{\text{reac}} = \frac{G}{c^5}\left\{ x_i x_j Q_{ij}^{(5)}(t) + O(\frac{1}{c^2}) \right\} \qquad (27)
$$

The above expression is valid only in a particular coordinate system covering the source (the Burke Thorne coordinate system [48, 49]).

The Landau-Lifshitz quadrupole equation is valid for slow motion sources with Newtonian internal gravity say a binary star or a star in oscillation. To obtain this extension one goes beyond linearized gravity by shuffling to the right hand side of Einstein's equation all terms non-linear in $\bar{h}^{\mu\nu} = \sqrt{g}\ \ g^{\mu\nu} - f^{\mu\nu}$ and combining it with the material tensor to make a total (matter plus gravitation) stress tensor of the type

$$\tau^{\mu\nu} = T^{\mu\nu} + t^{\mu\nu}\left(\bar{h}\right) \tag{28}$$

If one formally follows the standard derivation and *assumes* that lack of spatially compact support for $\tau^{\mu\nu}$ is irrelevant, one gets the new quadrupole equation where τ^{00} replaces T^{00}. For weakly self-gravitating systems, t^{00} is negligible relative to T^{00} and one recovers the standard quadrupole equation. The basic problem with this approach is that $\tau^{\mu\nu}$ does not have a spatially compact support and falls off rather slowly.

Fock's approach is conceptually different and lends itself to useful generalizations. The main idea is to solve the problem in two parts. First, compute the gravitational field in the near zone of the source where retardation is small compared to the characteristic period. Next, obtain the structure of the general radiative gravitational field in the wave zone. Finally, match the results of the two steps via an intermediate expression for the gravitational field that bridges the gap between the two zones.

5 The Generation Problem

The general approach to solve the generation problem may be broken up into the following steps[50, 58]:

1. Integrate the Einstein field equations in the vacuum exterior region D_e by means of a Multipolar Post Minkowskian series.
2. Integrate the non-vacuum field equations in the near zone D_i by means of a Post Newtonian expansion.
3. Match the two solutions in the exterior near zone $D_i \cap D_e$ to obtain the expression for the multipole moments of the source.
4. Expand the exterior solution in the wave zone D_w to obtain the observable moments of the radiative field.

5.1 The exterior Gravitational field

In harmonic coordinates

$$\partial_\nu h^{\mu\nu}_{\rm ext} = 0 \tag{29}$$

the vacuum Einstein equations read

$$\Box h^{\mu\nu}_{\rm ext} = \Lambda^{\mu\nu}(h_{\rm ext}) \tag{30}$$

Expanding the metric in terms of post Minkowskian series

$$h^{\mu\nu}_{\rm ext} = G h^{\mu\nu}_1 + G^2 h^{\mu\nu}_2 + \cdots + G^n h^{\mu\nu}_n \tag{31}$$

starting with the seed linearised metric in the form

$$h^{00}_1 = -\frac{4}{c^2} V^{\rm ext} \tag{32}$$

$$h^{0i}_1 = -\frac{4}{c^3} V^{\rm ext}_i \tag{33}$$

$$h^{ij}_1 = -\frac{4}{c^4} V^{\rm ext}_{ij} \tag{34}$$

where the external potentials $V^{\rm ext}, V_i^{\rm ext}, V_{ij}^{\rm ext}$ which are functionals of the Mass and Current moments, are of the form

$$\sum_{l\geq 0} \partial_L \left(\frac{1}{r} M_L(t-\frac{r}{c})\right) + \varepsilon \partial_L \left(\frac{1}{r} S_L(t-\frac{r}{c})\right) \tag{35}$$

the field equations are solved iteratively by means of the integral of the retarded potentials

$$(\Box_R^{-1} f)(\vec{x},t) = -\frac{1}{4*\pi} \int\int\int \frac{d^3x'}{|\vec{x}-\vec{x'}|} f(\vec{x'}, t - \frac{1}{c}|\vec{x}-\vec{x'}|) \tag{36}$$

For instance to order G^2 one has

$$h_{\rm ext}^{\mu\nu} = G h_1^{\mu\nu} + {\rm FP}_{B=0} \Box_R^{-1}[r^B \Lambda^{\mu\nu}(V_{\rm ext})] + G^2 q_2^{\mu\nu} + O(G^3) \tag{37}$$

5.2 The Inner Gravitational field

In harmonic coordinates the Einstein equations in the source read

$$\Box h_{\rm in}^{\mu\nu} = \frac{16\pi G}{c^4} T^{\mu\nu} + \Lambda^{\mu\nu}(h_{\rm in}) \tag{38}$$

$$\partial_\nu h_{\rm in}^{\mu\nu} = 0 \tag{39}$$

To the first post-newtonian iteration one finds

$$h_{\rm in}^{00} = -\frac{4}{c^2} V + \frac{4}{c^4}(W_{ii} - 2V^2) + O\left(\frac{1}{c^6}\right) \tag{40}$$

$$h_{\rm in}^{0i} = -\frac{4}{c^3} V_i + O\left(\frac{1}{c^5}\right) \tag{41}$$

$$h_{\rm in}^{ij} = -\frac{4}{c^4} W_{ij} + O\left(\frac{1}{c^6}\right) \tag{42}$$

where the inner potentials are

$$V = -4\pi G \Box_R^{-1} \sigma \tag{43}$$

$$V_i = -4\pi G \Box_R^{-1} \sigma_i \tag{44}$$

$$W_{ij} = -4\pi G \Box_R^{-1} \left[\sigma_{ij} + \frac{1}{4\pi G}\left(\partial_i V \partial_j V - \frac{1}{2}\delta_{ij}\partial_k V \partial_k V\right)\right] \tag{45}$$

5.3 Matching between the Inner and Outer fields

$h_{\rm in}^{\mu\nu}$ and $h_{\rm ext}^{\mu\nu}$ differ by a coordinate change in $D_i \cap D_e$

$$\delta x^\mu = \varphi(\vec{x}, t)$$

The inner and outer potentials differ by

$$V^{\rm ext} = \mathcal{M}(V) + c\partial_t \varphi^0 + O\left(\frac{1}{c^4}\right) \tag{46}$$

$$V_i^{\rm ext} = \mathcal{M}(V_i) - \frac{c^3}{4}\partial_i \varphi^0 + O\left(\frac{1}{c^2}\right) \tag{47}$$

$$W_{ij}^{\rm ext} = \mathcal{M}(W_{ij}) - \frac{c^4}{4}[\partial_i \varphi^j - \partial_j \varphi^i - \delta_{ij}(\partial_0 \varphi^0 + \partial_k \varphi^k)] + O\left(\frac{1}{c^2}\right) \tag{48}$$

(49)

where $\mathcal{M}()$ refers to the multipole expansion of the quantity in (). The matching equation between $h_{\rm in}^{\mu\nu}$ and $h_{\rm ext}^{\mu\nu}$ then implies to the relevant PN order

$$
\begin{aligned}
Gh_1^{\mu\nu} &= \Box_R^{-1}\left\{\frac{16\pi G}{c^4}T^{\mu\nu}+\Lambda^{\mu\nu}(V)\right\} \\
&\quad -FP_{B=0}\Box_R^{-1}\{r^B\Lambda^{\mu\nu}(\mathcal{M}(V))\}+\partial\varphi^{\mu\nu} \qquad (50)
\end{aligned}
$$

which can be written equivalently written in the form

$$
Gh_1^{\mu\nu} = FP_{B=0}\Box_R^{-1}\left\{r^B\left[\frac{16\pi G}{c^4}T^{\mu\nu}+\Lambda^{\mu\nu}(V)-\Lambda^{\mu\nu}(\mathcal{M}(V))\right]\right\}+\partial\varphi^{\mu\nu} \qquad (51)
$$

The external linear metric now appears to be the retarded integral of a compact supported source and our earlier result on the multipole decomposition of linearised gravity can be adapted to read off the required Mass and Current moments of this effective source $\tau^{\mu\nu}$ that includes the nonlinearities of the gravitational field.

$$
\tau^{\mu\nu}=T^{\mu\nu}+\frac{c^4}{16\pi G}\Lambda^{\mu\nu} \qquad (52)
$$

All reference to the multipole expansion $\mathcal{M}(V)$ of the potentials disappears in the final result.

5.4 Expansion in the wave zone

The harmonic coordinates $(\vec{x},t)$ are not the simplest coordinates in the far zone. The good radiative coordinates are obtained by correcting for the deviations of true light cones from the flat light cones.

$$
T-\frac{R}{c}=t-\frac{r}{c}-\frac{2GM}{c^3}\ln\left(\frac{r}{cb}\right)+O\left(\frac{1}{c^5}\right) \qquad (53)
$$

This coordinate transformation can be implemented to all post minkowskian orders and brings the metric to Bondi type form. From the multipole decomposition

$$
h_{ij}^{\rm TT}=\frac{4G}{c^2R}\sum_{l\le 2}\frac{1}{c^l}\{N_{L-2}U_{ijL-2}+\varepsilon_{ab(i}N_{aL-2}V_{j)bL-2}\} \qquad (54)
$$

one can then compute the observable moments U_L and V_L in terms of the Thorne moments M_L and S_L and there by in terms of the source's parameters.

5.5 1PN Wave generation

This formalism is valid under the same conditions as the quadrupole formalism but $\varepsilon = v/c$ is not very small. (For $\varepsilon \sim 0.2$ the formalism is expected to be accurate to within one percent). The power in the waves is given by

$$
\mathcal{L}=\frac{G}{c^5}\left\{\frac{1}{5}I_{ij}^{(3)}I_{ij}^{(3)}+\frac{1}{c^2}\left[\frac{1}{189}I_{ijk}^{(4)}I_{ijk}^{(4)}+\frac{16}{45}J_{ij}^{(3)}J_{ij}^{(3)}\right]+O\left(\frac{1}{c^4}\right)\right\} \qquad (55)
$$

where I_{ij} is the 1PN relativistic mass quadrupole moment of the source, I_{ijk} and J_{ij} the newtonian mass octupole and current quadrupole moments.

$$
I_{ij}(t)=\int d^3x\left\{\hat{x}_{ij}\sigma+\frac{1}{14c^2}\hat{x}^{ij}\vec{x}^2\partial_t^2\sigma-\frac{20}{21c^2}\hat{x}^{ijk}\partial_t\sigma_k\right\}(\vec{x},t). \qquad (56)
$$

In a generalised Burke Thorne gauge the radiation reaction forces can be derived from the following reaction potential[56]

$$
\begin{aligned}
V^{\rm reac} &= \frac{G}{c^5}\left\{-\frac{1}{5}x_{ij}I^{(5)}_{ij} + \frac{1}{c^2}\left[\frac{1}{189}x_{ijk}I^{(7)}_{ijk} - \frac{1}{70}\vec{x}^2 x_{ij}I^{(7)}_{ij}\right]\right\} &(57)\\
V_i^{\rm reac} &= \frac{G}{c^5}\left\{\frac{1}{21}\hat{x}^{ijk}I^{(6)}_{jk} - \frac{4}{45}\varepsilon_{ijk}x_{jm}J^{(5)}_{km}\right\} &(58)
\end{aligned}
$$

The PN radiation reaction for the specific case of binary systems has been also obtained in generic coordinate systems starting from energy and angular balance arguments[32]. The results of both these schemes are consistent with each other.

6 2PN Generation

The detailed analysis along the lines indicated above leads us to the 2PN accurate mass quadrupole which contains compact support terms, quadratically nonlinear terms and a cubic nonlinear term. For illustration we quote the expression[58, 54]:

$$
\begin{aligned}
I_{ij}(t) &= FP_{B=0}\int d^3\vec{x}|\vec{x}|^B\Big\{\hat{x}_{ij}\Big[\sigma + \frac{4}{c^4}(\sigma_{kk}U - \sigma P)\Big]\\
&\quad \frac{1}{14c^2}|\vec{x}|^2\hat{x}_{ij}\partial_t^2\sigma - \frac{20}{21c^2}x_{ijk}\partial_t\Big[\Big(1+\frac{4U}{c^2}\Big)\sigma_k\Big]\\
&\quad +\frac{1}{\pi G c^2}\Big(\partial_m U[\partial_k U_m - \partial_m U_k] + \frac{3}{4}\partial_t U\partial_k U\Big)\Big]\\
&\quad +\frac{1}{504c^4}|\vec{x}|^4\hat{x}_{ij}\partial_t^4\sigma - \frac{10}{189c^4}|\vec{x}|^2\hat{x}_{ijk}\partial_t^3\sigma_k\\
&\quad +\frac{5}{54c^4}\hat{x}_{ijkm}\partial_t^2\Big[\sigma_{km} + \frac{1}{4\pi G}\partial_k U\partial_m U\Big]\\
&\quad \frac{1}{\pi G c^4}\hat{x}_{ij}\Big[-P_{km}\partial^2_{km}U - 2U_k\partial_t\partial_k U\\
&\quad +2\partial_k U_m\partial_m U_k - \frac{3}{2}(\partial_t U)^2 - U\partial_t^2 U\Big]\Big\}(\vec{x},t)\\
&\quad +O(\frac{1}{c^5})\\
P_{ij}(\vec{x},t) &= G\int\frac{d^3\vec{x}'}{|\vec{x}-\vec{x}'|}[\sigma_{ij}\\
&\quad +\frac{1}{4\pi G}\Big(\partial_i U\partial_j U - \frac{1}{2}\delta_{ij}\partial_k U\partial_k U\Big)\Big](\mathbf{x}',t)\ .
\end{aligned}
$$

$$
\begin{aligned}
J_{ij} &= \underset{ij}{\rm STF}\Big[\epsilon_{jab}\int d^3x\ \Big\{\sigma^b\ x^{ai} + \frac{1}{c^2}\Big(\frac{1}{2}\sigma^b\ x^{ai}\ P_{(-1)} -\\
&- \frac{1}{2}\sigma^s\ x^{ai}\ P^{bs}_{(-1)} + \frac{7}{4}\sigma^b\ x^a\ P^i_{(0)}\ + \frac{1}{4}\sigma^s\ x^a\ P^{bsi}_{(0)} +\\
&+\frac{7}{2}\sigma^b\ x^i\ P^a_{(0)} + \frac{11}{4}\sigma^a\ P^{bi}_{(1)} - \frac{7}{4}\sigma^i\ x^a\ P^b_{(0)}\Big) +\\
&+\frac{1}{c^2}\frac{d}{dt}\Big(\frac{5}{56}\sigma\ x^{ais}\ P^{sb}_{(-1)} - \frac{1}{56}\sigma\ x^{ass}\ P^{ib}_{(-1)} - \frac{3}{112}\sigma\ x^{as}\ P^{sbi}_{(0)} +\\
&+\frac{9}{112}\sigma\ x^{ai}P^b_{(0)} - \frac{9}{112}\sigma\ x^a\ P^{bi}_{(1)} - \frac{1}{14}\sigma\ x^{aiss}\ P^b_{(-2)} +\\
&+\ \frac{3}{28}x^{ass}\ T^{bi} - \frac{1}{28}x^{asi}\ T^{bs}\Big)\Big\}\Big] \qquad (59)
\end{aligned}
$$

$$\text{where, } P^{i_1 i_2 \dots i_l}_{(\alpha)} = G \int d^3y\, \sigma(\vec{y})\, r^{\alpha}_{xy}\, n^{i_1}_{xy} n^{i_2}_{xy} \cdots n^{i_l}_{xy}. \tag{60}$$

Once the mass quadrupole is explicitly obtained for the inspiralling binary system the waveform in the far zone can be computed using:

$$h^{TT}_{km} = (h^{TT}_{km})_{\rm inst} + (h^{TT}_{km})_{\rm tail} \;, \tag{61}$$

where the "instantaneous" contribution is defined by

$$\begin{aligned}(h^{TT}_{km})_{\rm inst} = \frac{2G}{c^4 R}\mathcal{P}_{ijkm}\Big\{ I^{(2)}_{ij} &+ \frac{1}{c}\left[\frac{1}{3}N_a I^{(3)}_{ija} + \frac{4}{3}\varepsilon_{ab(i}J^{(2)}_{j)a}N_b\right] \\ &+ \frac{1}{c^2}\left[\frac{1}{12}N_{ab}I^{(4)}_{ijab} + \frac{1}{2}\varepsilon_{ab(i}J^{(3)}_{j)ac}N_{bc}\right] \\ &+ \frac{1}{c^3}\left[\frac{1}{60}N_{abc}I^{(5)}_{ijabc} + \frac{2}{15}\varepsilon_{ab(i}J^{(4)}_{j)acd}N_{bcd}\right] \\ &+ \frac{1}{c^4}\left[\frac{1}{360}N_{abcd}I^{(6)}_{ijabcd} + \frac{1}{36}\varepsilon_{ab(i}J^{(5)}_{j)acde}N_{bcde}\right]\Big\} \;, \end{aligned} \tag{62}$$

and where the "tail" contribution reads

$$\begin{aligned}(h^{TT}_{km})_{\rm tail} = \frac{2G}{c^4 R}\frac{2Gm}{c^3}\mathcal{P}_{ijkm}\int_0^{+\infty} d\tau \Big\{ &\ln\left(\frac{\tau}{2b_1}\right) I^{(4)}_{ij}(T_R-\tau) \\ &+\frac{1}{3c}\ln\left(\frac{\tau}{2b_2}\right) N_a I^{(5)}_{ija}(T_R-\tau) \\ &+\frac{4}{3c}\ln\left(\frac{\tau}{2b_3}\right)\varepsilon_{ab(i}N_b J^{(4)}_{j)a}(T_R-\tau)\Big\} \;. \end{aligned} \tag{63}$$

The radiation reaction is simply computed by evaluating the flux of energy in the far zone. Upto 2PN accuracy this is given by:

$$\begin{aligned}\left(\frac{dE_B}{dT_R}\right)_{\rm inst} = -\frac{G}{c^5}\Big\{\frac{1}{5}I^{(3)}_{ij}I^{(3)}_{ij} &+ \frac{1}{c^2}\left[\frac{1}{189}I^{(4)}_{ijk}I^{(4)}_{ijk} + \frac{16}{45}J^{(3)}_{ij}J^{(3)}_{ij}\right] \\ &+ \frac{1}{c^4}\left[\frac{1}{9072}I^{(5)}_{ijkm}I^{(5)}_{ijkm} + \frac{1}{84}J^{(4)}_{ijk}J^{(4)}_{ijk}\right]\Big\} \;. \end{aligned} \tag{64}$$

The results are best represented in terms of gauge invariant variables like x defined by

$$x \equiv (Gm\omega_{\rm 2PN}/c^3)^{2/3}$$

In terms of x after a few 100 pages of calculation we obtain:

$$\begin{aligned}\frac{dE_B}{dT_R} = &-\frac{32c^5}{5G}\nu^2 x^5\left\{1 + \left(-\frac{1247}{336} - \frac{35}{12}\nu\right)x + 4\pi\, x^{3/2}\right. \\ &+ \left(-\frac{44711}{9072} + \frac{9271}{504}\nu + \frac{65}{18}\nu^2\right)x^2 \\ &\left. - \left(\frac{8191}{672} + \frac{535}{24}\nu\right)\pi x^{5/2} + O(x^3)\right\}. \end{aligned} \tag{65}$$

If ω is the orbital frequency then

$$\phi = \int \omega dt \tag{66}$$

From the energy balance law one deduces

$$\phi = -\int \frac{\omega dE}{\mathcal{L}} \tag{67}$$

$$\theta \equiv \frac{c^3\nu}{5Gm} T_R \ . \tag{68}$$

$$\begin{aligned} x &= \frac{1}{4}\Theta^{-1/4}\Bigg\{1 + \left(\frac{743}{4032} + \frac{11}{48}\nu\right)\Theta^{-1/4} - \frac{\pi}{5}\Theta^{-3/8} \\ &+ \left(\frac{19563}{254016} + \frac{24401}{193536}\nu + \frac{31}{288}\nu^2\right)\Theta^{-1/2} \\ &+ \left(-\frac{11891}{53760} + \frac{29}{1920}\nu\right)\pi\Theta^{-5/8} + O(\theta^{-3/4})\Bigg\} \end{aligned} \tag{69}$$

$$\begin{aligned} \phi &= -\frac{1}{32\nu}\Bigg\{x^{-5/2} + \left(\frac{3715}{1008} + \frac{55}{12}\nu\right)x^{-3/2} - 10\pi x^{-1} \\ &+ \left(\frac{15293365}{1016064} + \frac{27145}{1008}\nu + \frac{3085}{144}\nu^2\right)x^{-1/2} \\ &+ \left(\frac{38645}{1344} + \frac{15}{16}\nu\right)\pi \ln\left(\frac{x}{x_0}\right) + O(x^{1/2})\Bigg\} \end{aligned} \tag{70}$$

Table 1: Contributions to the accumulated number $\mathcal{N}$ of gravitational-wave cycles in a LIGO/VIRGO-type detector. Frequency entering the bandwidth is 10 Hz (seismic limit); frequency leaving the detector is 1000 Hz for 2 neutron stars (photon shot noise), and $\sim$ 360 Hz and $\sim$ 190 Hz for the two cases involving black-holes (innermost stable orbit). β and σ are spin parameters. Numbers in parentheses indicate contribution of finite-mass (η) effects.

$$\beta = \frac{1}{12}\Sigma_i(113m_i^2/m^2 + 75\eta)\hat{\mathbf{L}}\cdot\boldsymbol{\chi}_i$$

$$\sigma = (\eta/48)(-247\boldsymbol{\chi}_1\cdot\boldsymbol{\chi}_2 + 721\hat{\mathbf{L}}\cdot\boldsymbol{\chi}_1\hat{\mathbf{L}}\cdot\boldsymbol{\chi}_2).$$

	$2\times 1.4M_\odot$	$10M_\odot + 1.4M_\odot$	$2\times 10M_\odot$
Newtonian	16,050	3580	600
First PN	439(104)	212(26)	59(14)
Tail	−208	−180	−51
Spin-orbit	17β	14β	4β
Second PN	9(3)	10(2)	4(1)
Spin-spin	-2σ	-3σ	$-\sigma$
2.5PN	−11		

7 A Sampler of Nonlinear effects

The first interesting and dominant nonlinear effect we encounter is the Tail term. It arises at order 1.5PN and results from the nonlinear interaction between the varying quadrupole moment I_{ij} and the static mass monopole M of the source. The effect of this interaction is to modify the quadrupole I_{ij} by the tail induced correction

$$\delta I_{ij}^{\rm tail} = 2\frac{GM}{c^3}\int_0^{+\infty} d\tau I_{ij}^{(2)}(u-\tau)\left[\ln\frac{\tau}{2}+\frac{11}{12}\right] \tag{71}$$

It has important observational consequences in the dynamics of coalescing binaries.

The next sub-dominant nonlinear effect is the nonlinear memory term or the Christodoulou effect. It is dominantly 2.5PN. It corresponds to the re-radiation of gravitational waves by the stress energy distribution of the linear waves.

$$\delta_{ij}^{\rm memory}(U) = \frac{G}{c^5}\left\{-\frac{2}{7}\int_{-\infty}^{u} dv I_{k<i}^{(3)}(u)I_{j>k}^{(3)}(v) + \sum_{p=0}^{5}\alpha_p I_{k<i}^{(p)}(u)I_{j>k}^{(5-p)}(u)\right\} \tag{72}$$

In the limit $u \to +\infty$ the Tail term vanishes but the memory term tends to a finite limit.

$$\delta_{ij}^{\rm memory}(U) = -\frac{2G}{7c^5}\int_{-\infty}^{+\infty} dv I_{k<i}^{(3)}(u)I_{j>k}^{(3)}(v) \tag{73}$$

The memory effect does not contribute to the energy loss and hence has poor observable consequences.

The next dominant cubically nonlinear effect is the tail of tail term. It is the nonlinear interaction between the tail of waves and the static mass monopole M. The effect is of order 3PN and should be detectable in the dynamics of coalescing compact binaries.

$$\delta I_{ij}^{\rm tail^2} = \int_0^{+\infty} d\tau I_{ij}^{(3)}(u-\tau)\left[\ln^2\frac{\tau}{2}+\frac{57}{70}\ln\frac{\tau}{2}+\frac{124627}{44100}\right] \tag{74}$$

The generalisation of the 2PN results for non circular orbits has been obtained by Gopakumar and Iyer[60]. They obtain the flux of energy, angular momentum and waveform for the general case and apply it to study the effect of this emission on the orbital period of the binary system.

It has been argued that most of the accessible information allowing accurate measurements of the binary's intrinsic parameters (such as the two masses) is in fact contained in the phase, because of the accumulation of cycles, and that rather less accurate information is available in the wave amplitude itself. To determine just what order is required in the waveform would necessitate a measurement-accuracy analysis not restrained to a model waveform taking only the higher-order phase evolution into account (namely the so-called restricted post-Newtonian waveform). For instance, it is conceivable that conclusions reached regarding the needed accuracy in the determination of the phase in the reaction problem are modified when the full amplitude evolution of the wave is taken into account. Some of these are under investigation by Kanti Jotania[61].

8 The Perturbation method

As mentioned earlier in this approach the Einstein's equations

$$G^{\mu\nu}[g] = 8\pi G T^{\mu\nu} \tag{75}$$

$$T^{\mu\nu} = \mu\int ds \frac{dz^\mu}{ds}\frac{dz^\nu}{ds}\delta^{(4)}(x^\alpha - z^\alpha(s)) \tag{76}$$

are expanded around the black hole background

$$g_{\mu\nu} = g_{\mu\nu}^{\rm BH} + h_{\mu\nu} \tag{77}$$

Linearising this equation with respect to μ/M *i.e* ignoring terms of $O(h^2)$ and $O(\mu^2)$ we have

$$\delta G^{\mu\nu}[h] = 8\pi G T^{\mu\nu} \tag{78}$$

The $h_{\mu\nu}$ is written using the curved space Green function

$$h_{\mu\nu}(x) = 16\pi G \int G^{BH}(x, x')_{\mu\nu,\alpha\beta} T^{\alpha\beta}(x'). \tag{79}$$

For the Kerr black hole the story is contained in the ψ_4, a Weyl tensor component describing outgoing gravitational waves,

$$\psi_4 \to \frac{1}{2}(\ddot{h}_+ - \imath \ddot{h}_\times) \tag{80}$$

Expanding the ψ_4 in terms of spheroidal harmonics of appropriate weight -2

$$\psi_4 = \frac{1}{(r - \imath a \cos\theta)^4} \sum_{lm\omega} R_{lm\omega}(r) {}_{-2}S^{a\omega}_{lm}(\theta) e^{\imath m\phi - \imath\omega t} \tag{81}$$

Einstein's equations and the Bianchi identities lead one to the Teukolsky equation

$$\left[\Delta^2 \frac{d}{dr}\frac{1}{\Delta}\frac{d}{dr} - U_{lm\omega}\right] R_{lm\omega} = T_{lm\omega} \tag{82}$$

where

$$\begin{aligned} \Delta &= r^2 + a^2 - 2Mr & (83)\\ U_{lm\omega} &= -\frac{K^2 + 4\imath(r - M)K}{\Delta} + 8\imath\omega r + \lambda^{a\omega}_{lm} & (84)\\ K &= (r^2 + a^2)\omega - ma & (85) \end{aligned}$$

Here $\lambda^{a\omega}_{lm}$ is the eigenvalue of ${}_{-2}S^{a\omega}_{lm}(\theta)$. It follows that

$$\begin{aligned} R_{lm\omega}(r) &= \int_{r_+}^{\infty} dr' G_{lm\omega}(r, r') T_{lm\omega}(r') & (86)\\ &\to Z_{lm\omega} r^3 e^{\imath\omega r} \quad (\Omega \sim 2\Omega_{\rm orbit}) & (87) \end{aligned}$$

Expanding $G_{lm\omega}(r, r')$ in powers of

$$\epsilon = 2M\omega = \frac{2GM}{c^2}\omega = O(v^3)$$

(similar to a post minkowskian expansion) and $T_{lm\omega}(r)$ in powers of

$$z = r\omega = O(v)$$

(similar to a post newtonian expansion) one can systematically study the gravitational wave luminosity for a particle in circular orbit around the black hole. The numerical discovery of the 4π term due to tail luminosity by Poisson[34], the $v^6 \ln v$ and $v^8 \ln v$ terms by Tagoshi and Nakamura[36] indicate the accuracy and efficacy of the method. The confirmation of these results analytically by Sasaki and Tagoshi[37, 38] as also the evaluation of the gravitational wave luminosity to 4PN are the impressive achievements of this programme. This provides one of the best guidelines to the more involved PN calculations. The extension to orbits with eccentricity as also particles with spin have been achieved. More difficult issues like radiation reaction as also the question of going beyond linearised perturbations are under study.

9 Chandrasekhar's contributions

Of his myriad contributions to astrophysics and in particular to general relativity this area is special in that not only did Chandrasekhar have the first word but he did not as was his wont write a monograph about it. The broad areas may be divided into,

- The influence of general relativity on pulsation and stability of stars.
- The back reaction of gravitational waves on their sources
- The post newtonian approximations to general relativity and its astrophysical applications

Each topic was pioneered by Chandrasekhar dominated by him and his papers deal simultaneously with aspects of two or three of the above topics. Chandra's discovery of the radiation reaction instability was an outgrowth of the general quest to understand gravitational radiation reaction[62, 64]. It was the first direct calculation à la Lorentz of a radiation reaction force which obtained a result in agreement with the quadrupole losses. Though in linearised gravity Einstein had discovered the existence of propagating wave like solutions people were still not clear about whether the waves were really physical or gauge? Do they really take away energy from the system? One of the first attempts was by Einstein Infeld and Hoffmann who tried to model the binary system by a system of two point masses but this failed. In analysing the result Trautman realised that the missing element was in the outgoing boundary condition but due to an algebraic error got a dissipation not in agreement with the radiation formula. This was the state of affairs when Chandra moved into the act. What should be realised is that partial results which concentrate only on the reactive terms are not good enough since one has to be sure that indeed to a lower order there are conserved quantities; extensions of the usual energy, linear momentum and angular momentum; which at the next half higher order are no longer conserved.That is why Chandra developed the 1PN and 2PN methods to iterate Einstein's equations to discover the correct conserved quantities. He had the fortitude to face the problem of developing a post newtonian approximation scheme that keeps ALL the terms including the large number that do not reverse sign but are of lower order. He realised the danger in the approximation scheme of EIH using delta functions in nonlinear theories. Hence he went back à la Fock to perfect fluid sources. However he was not happy with the choice of harmonic coordinates that Fock made and developed his own sytem where he could control everything at every order. He completed the 1PN work in 1965 and paused to apply it to ellipsoid figures of equilibrium and the onset of dynamical instability. Then with Nutku he constructed the 2PN equations and finally with Esposito the 2.5PN equations that were used to discuss the radiation reaction instability completely and rigorously for the first time. It was a tour de force and underappreciated now twenty years later...The results were under severe scrutiny when the now famous binary pulsar was discovered.. There are divergent integrals which make the results mathematically ill defined.. More care is needed to do better. Yet another issue was the applicability of these to the binary pulsar sytem where internal gravities are much stronger. This extension by T. Damour depends on subtle effacement of strong gravity effects in general relativity probably related to the strong principle of equivalence and once again for computational efficiency works with the delta functions à la EIH and harmonic coordinates à la Fock. But finally the formulas that emerge are the radiation reaction formulas equivalent to those of Chandrasekhar. It is curious that in connection with the high order perturbation calculations that are in progress in connection with LIGO one feels that the delta function approach has reached its end and that probably one may need to go back to the fluid treatment à la Chandrasekhar!

The PN and PPN have become standard working tools of physics and astrophysics. They are used in the studies of stars, star clusters, gravitational wave generation, motion of planets and moon, development of PPN formalism to compare general relativity with other theories and experiment. To assess Chandra's contributions to this area one can do no better than quote Kip Thorne:

> If Chandra had left us no other relativistic legacy beyond the PN and PPN formalisms he would still deserve a place among the great contributors to our subject.

References

[1] J. H. Taylor, A. Wolszczan, T. Damour, and J. M. Weisberg, *Nature*, **355**, 132 (1992)

[2] A. Abramovici, W. E. Althouse, R. W. P. Drever, Y. Gürsel, S. Kawamura, F. J. Raab, D. Shoemaker, L. Siewers, R. E. Spero, K. S. Thorne, R. E. Vogt, R. Weiss, S. E. Whitcomb and M. E. Zucker, *Science*, **256**, 32 (1992)

[3] C. Bradaschia, E. Calloni, M. Cobal, R. Del Fasbro, A. Di Virgilio, A. Giazotto, L. E. Holloway, H. Kautzky, B. Michelozzi, V. Montelatici, D. Pascuello and W. Velloso, *Proc. of the Banff Summer Inst. on Gravitation 1990, Banff, Alberta*, Eds. R. Mann and P. Wesson, (World Scientific, Singapore, 1991), p.499

[4] K. S. Thorne in *300 Years of Gravitation*, Eds. S. W. Hawking and W. Israel, (Cambridge Univ. Press, Cambridge, 1987), p.330

[5] B. F. Schutz, *Class. Quantum Grav.*, **6**, 1761 (1989)

[6] B. F. Schutz, *Class. Quantum Grav.*, **10**, S135 (1993)

[7] K. S. Thorne, *Proc. of the Fourth Rencontres de Blois*, Eds. G. Fontaine and J. Van Tran Thanh, (Editions Frontières, France, 1993)

[8] K. S. Thorne, *Proc. of the 8th Nishinomiya-Yukawa Symposium on Relativistic Cosmology*, Ed. Sasaki M, (Universal Acad. Press, Japan, 1994), p.67

[9] J. P. A. Clark, E. P. J. van den Heuvel and W. Sutantyo, *Astron. Astrophys.*, **72**, 120 (1979)

[10] E. S. Phinney, *Astrophys. J.*, **380**, L17 (1991)

[11] R. Narayan, T. Piran and A. Shemi, *Astrophys. J.*, **379**, L17 (1991)

[12] A. V. Tutukov and L. R. Yungelson, *Mon. Not. R. Astron. Soc.*, **260**, 675 (1993)

[13] A. Królak and B. F. Schutz, *Gen. Rel. Grav.*, **19**, 1163 (1987)

[14] A. Królak, *Proc. in Gravitational Wave Data Analysis*, Ed. B. F. Schutz, (Kluwer Academic Publishers, 1989)

[15] C. W. Lincoln and C. M. Will, *Phys. Rev. D*, **42**, 1123 (1990)

[16] C. Cutler, T. A. Apostolatos, L. Bildsten, L. S. Finn, E. E. Flanagan, D. Kennefick, D. M. Markovic, A. Ori, E. Poisson, G. J. Sussman and K. S. Thorne, *Phys. Rev. Lett.***70**, 2984, (1993)

[17] L. S. Finn and D. F. Chernoff, *Phys. Rev. D*, **47**, 2198 (1993)

[18] C. Cutler and E. Flanagan, *Phys. Rev. D*, **49**, 2658 (1994)

[19] E. Poisson and C. M. Will, *Phys. Rev. D*, **52**, 848 (1995)

[20] A. Królak, K. D. Kokkotas and G. Schäfer, *Phys. Rev. D*, **52**, 2089 (1995)

[21] B. F. Schutz, *Nature*, **323**, 310 (1986)

[22] D. Marković D, *Phys. Rev. D*, **48**, 4738 (1993)

[23] L. S. Finn, *Proc. of the 17th Texas Symposium on Relativistic Astrophysics*, (1996) (To appear)

[24] B. Paczynsky, *Acta Astron.*, **41**, 157 (1991)

[25] D. Eichler *et al.*, *Nature*, **340**, 126 (1989)

[26] C. M. Will, *Phys. Rev. D.*, **50**, 6058 (1994)

[27] L. Blanchet and B. S. Sathyaprakash, *Class. Quantum Grav.*, **11**, 2807 (1994); and *Phys. Rev. Letters*, **74**, 1067 (1995)

[28] K. S. Thorne, *Proc. of IAU symposium 165, Compact Stars in Binaries*, Eds. J. Van Paradijs, E. Van den Heuvel and E. Kuulkers, (Kluwer Academic Publishers, 1995)

[29] K. S. Thorne, *Proc. of the Snowmass 95 summer study on particle and nuclear astrophysics and cosmology*, Eds. E. W. Kolb and R. Peccei, (World Scientific, 1995)

[30] C. M. Will, *Proc. of the 8th Nishinomiya-Yukawa Symposium on Relativistic Cosmology*, Ed. M. Sasaki, (Universal Acad. Press, Japan, 1994), p.83

[31] T. Damour, *Gravitation in Astrophysics*, Eds. B. Carter and J. B. Hartle, (Plenum Press, New York and London, 1986), p.3

[32] B. R. Iyer and C. M. Will, *Phys. Rev. Lett.*, **70**, 13 (1993); and *Phys. Rev. D*, **52**, 6882 (1995)

[33] L. Blanchet, *Phys. Rev. D.*, **47**, 4392 (1993); and *Phys. Rev. D.*, (submitted) (1995)

[34] E. Poisson *Phys. Rev. D.*, **47**, 1497 (1993)

[35] C. Cutler, L. S. Finn, E. Poisson and G. J. Sussmann *Phys. Rev. D.*, **47**, 1511 (1993)

[36] H. Tagoshi and T. Nakamura, *Phys. Rev. D.*, **49**, 4016 (1994)

[37] M. Sasaki, *Prog. Theor. Phys.*, **92**, 17 (1994)

[38] H. Tagoshi and M. Sasaki, *Prog. Theor. Phys.*, **92**, 745 (1994)

[39] E. Poisson, *Phys. Rev. D.*, **52**, 5719 (1995)

[40] K.S. Thorne, *Rev. Mod. Phys.*, **52**, 299 (1980)

[41] T. Damour and B.R. Iyer, *Phys. Rev. D.*, **43**, 3259 (1991)

[42] T. Damour and N. Deruelle, *C. R. Acad. Sci. Paris*, **293**, 537 (1981); **293**, 877 (1981); T. Damour, *Gravitational Radiation*, Eds. N. Deruelle and T. Piran (North Holland, Amsterdam, 1983), p.59

[43] L. P. Grishchuk and S. M. Kopejkin, *Relativity in Celestial Mechanics and Astrometry*, Eds. J. Kovalevsky and V. A. Brumberg (Reidel, Dordrecht, 1986), p.19

[44] L. E. Kidder, A. G. Wiseman and C. M. Will, *Phys. Rev. D.*, **47**, 3281 (1993)

[45] A. Einstein, *Preuss. Akad. Wiss. Sitzber, Berlin*, p.154 (1918)

[46] L. D. Landau and E. M. Lifshitz, *Teoriya Polya*, (Nauka, Moscow, 1941)

[47] V. A. Fock, *Teoriya prostranstva vremeni i tyagoteniya*, (Fizmatgiz, Moscow, 1955)

[48] W. L. Burke, *J. Math. Phys.*, **12**, 401 (1971)

[49] K. S. Thorne, *Astrophys. J.*, **158**, 997 (1969)

[50] L. Blanchet and T. Damour, *Phil. Trans. R. Soc. Lon. A*, **320**, 379 (1986)

[51] L. Blanchet, *Proc. Roy. Soc. Lon. A.*, **409**, 383 (1987)

[52] L. Blanchet and T. Damour, *Phys. Rev. D.*, **37**, 1410 (1988)

[53] L. Blanchet and T. Damour, *Ann. Inst. Henri Poincaré, Phys.Théor.*, **50**, 377 (1989)

[54] T. Damour and B. R. Iyer, *Ann. Inst. Henri Poincaré, Phys.Théor.*, **54**, 115 (1991)

[55] L. Blanchet and T. Damour, *Phys. Rev. D.*, **46**, 4304 (1992)

[56] L. Blanchet, *Phys. Rev. D.*, **51**, 2559 (1995)

[57] L. Blanchet, T. Damour, B.R. Iyer, C.M. Will, and A.G. Wiseman, *Phys. Rev. Lett.*, **74**, 3515 (1995)

[58] L. Blanchet, T. Damour, and B.R. Iyer, *Phys. Rev. D.*, **51**, 5360 (1995)

[59] C. M. Will and A. G. Wiseman, (In preparation)

[60] A. Gopakumar and B. R. Iyer, (1996) (In preparation)

[61] Kanti Jotania, (1996) (In preparation)

[62] S. Chandrasekhar, *Ap. J.*, **158**, 45 (1969)

[63] S. Chandrasekhar and Y. Nutku, *Ap. J.*, **158**, 55 (1969)

[64] S. Chandrasekhar and F. P. Esposito, *Ap. J.*, **159**, 153 (1970)

It is only when we observe the scale of Newton's achievement that comparisons, which have sometimes been made with other men of science, appear altogether inappropriate both with respect to Newton and with respect to others. In fact, only in juxtaposition with Shakespeare and Beethoven is the consideration of Newton appropriate.....

Now, a few remarks concerning the style of *Principia.* Quite unlike his early communications on his optical discoveries, the *Principia* is written in a style of glacial remoteness which makes no concessions to his readers....

It is however, clear that the rigid and lamellated style of the *Principia* is deliberate....

Newton seems to have been remarkably insensitive: impervious to the arts, tactless, and with no real understanding of others....

Newton's most remarkable gift was probably his powers of concentration...

The central paradox of Newton's life is that he deliberately and systematically ignored his supreme mathematical genius and through most of his life neglected the one activity for which he was gifted beyond any man.

— S. CHANDRASEKHAR

DATA ANALYSIS OF GRAVITATIONAL WAVE SIGNALS FROM COALESCING BINARIES

R. Balasubramanian[1]

Inter-University Center for Astronomy & Astrophysics
Ganeshkind
Pune

Abstract

In this lecture I shall be dealing with the techniques of Data analysis which will be used to extract the gravitational wave signal from the noisy output of the gravitational wave detector. The inspiral of binary systems consisting of compact objects such as black-holes and neutron stars can be well modelled. The knowledge of the signal waveform enables us to design detection strategies which are optimised to detect the 'chirp' signal. The basic idea here is to correlate the incoming data with the waveforms that we expect to observe.

[1]rbs@iucaa.ernet.in

Chandrasekhar is one of the most unusual examples of a scientist who has been able to inject his personal style into his work

— A. LIGHTMAN

He has an incomparable style. Good English style is a lost art in physics, but he has it and this wonderful feeling for the essential, and a feeling for beauty.

— V. WEISSKOPF

It's a rewarding aesthetic experience to listen to Chandra's lectures and study the development of theoretical structures at his hands. The pleasure I get is the same as I get when I go to an art gallery and admire paintings.

— LYMAN SPITZER

1 Introduction

Gravitational waves can be thought of as ripples in the curvature of spacetime which travel outward from the source carrying energy and momentum to the observer. These waves can be observed by measuring the proper distances between objects which are held fixed at constant coordinate positions in the observers reference frame.

The linear approximation to Einstein's equation yields the wave equation for the metric perturbation h_{ik} and thus gravitational waves are tensorial in nature. The magnitude of the the perturbation h_{ik} at the observer location can be written in terms of the derivatives of the quadrupole moment tensor

$$h_{ik}(t) = \frac{2G}{rc^4}\ddot{I}_{ik}(t - r/c). \tag{1}$$

This result is valid in the so called quadrupole or the 'Newtonian' approximation. An order of magnitude formula for h_{ik} can be given as

$$h \approx \frac{GE^{\mathrm{kin}}_{\mathrm{non-sph.}}}{c^4 r} \tag{2}$$

Gravitational waves carry a huge amount of energy. For a burst source in our galaxy such as a supernova for which the magnitude of $h_{ik} \approx 10^{-20}$ the flux on Earth would be around 30 Watts.m^{-2}! But the weak interaction of these waves with matter due to the smallness of G makes them hard to detect.

1.1 Sources and their strengths

A variety of sources are expected to be observed by the Earth based detectors which are currently being constructed. These are:

Binary Coalescences
: These are the most promising of all sources. Coalescences can be observed at cosmological distances. The typical magnitude for a NS/NS binary is about $h \approx 10^{-23}$ at a distance of 1000 Mpc. With the planned interferrometric detectors [1, 2] we would be observing the last few minutes before coalescence.

Supernovae
: The strength of these sources depends upon the degree of asymmetry during collapse. These will typically have a magnitude of $h \approx 10^{-18}, 10^{-19}$ for supernovae in our galaxy and will last for a few milliseconds.

Pulsars
: These are continuous wave sources with a typical magnitude of $h \approx 10^{-26}$ at a distance of 10 kpc.

Stochastic background
: Quantitative estimates of the cosmological stochastic background is highly model dependent. Pulsar timing observations have already provided tight constraints in fixing an upper bound on the magnitude. It might be possible to further constrain the strength of this background using the gravitational wave detectors.

1.2 The laser interferometer

When two masses are suspended a distance l apart then the change in the proper distance between them due to a gravitational wave of amplitude h is $\approx hl$. Two types of detectors are currently employed:

- Resonant bar detectors and

- Laser interferometers.

The bar detectors are essentially metallic cylinders (other geometries are being investigated) of large mass. These detectors are sensitive in a narrow frequency region around their natural resonant frequencies.

The laser interferometer illustrated in figure 1, converts path length changes to phase difference between laser light emerging from the two arms.

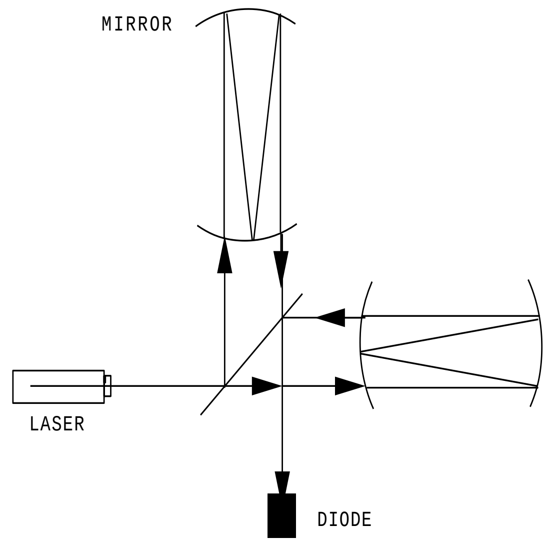


Figure 1: Schematic diagram of the laser interferometer

Such interferometers are being constructed by various groups around the world such as the American LIGO [1], the French-Italian VIRGO [2] and the German-English GEO. The LIGO will comprise of two four kilometre armlength detectors and one two kilometre armlength detector. The VIRGO will consist of a single three kilometre detector. whereas the GEO will have an armlength of 600 meters. A space based interferometric detector is also being planed and the launch is scheduled to take place in the year 2017. The interferometer as opposed to the resonant mass detector takes advantage of the quadrupolar nature of gravitational waves. The fluctuations in the proper distances between the mirrors can be estimated by measuring the shift of the fringe pattern at the photo diode. The laser interferometer is essentially a broadband device, which means that it is sensitive in a wide band of frequencies, though it can be made to operate at the narrowband mode by the use of dual recycling techniques. Ground based detectors will be typically able to measure gravitational waves at frequencies between $1 - 2000$Hz and hence we need to have very long light paths. This can partly be achieved by storing the laser power within a Fabry Perot cavity for a longer period of time. The laser interferometers have the advantage of being scalable, which means that the armlengths can be increased as and when it is feasible.

The performance of the detector will be limited by the various sources of noise. The four main sources of noise are

1. Seismic,
2. Thermal,
3. Photon shot,
4. Quantum noise

The lower cutoff of the detector is fixed by the seismic noise. This cutoff is between $1 - 10$Hz. There is a very steep falloff of this noise beyond the lower cutoff. Observation of gravitational waves at frequencies lower than the seismic cut off will have to be done by space based detectors. The thermal noise is important at intermediate frequencies. As the apparatus will be maintained at room temperature the natural vibrations of the various normal modes of the mirrors and the suspension wires will be substantial. The thermal noise can be reduced by a careful selection of the material used for the construction of the mirrors. Also the effect of the thermal noise can be minimised by increasing the mass of the mirrors. Photon shot noise is important at large frequencies and is caused by the uncertainty in the number of photons at the photodiode. By increasing the power of the laser the effect of this noise can be reduced. The ultimate sensitivity of the detector will be decided by the quantum noise which is relatively less important as compared to the other kinds of noise.

To a large extent the noise in the detector will be Gaussian and hence the power spectrum of the noise describes it completely. The power spectrum $S_h(f)$ is defined by the relation

$$\overline{\tilde{n}(f)\tilde{n}^*(f')} = S_h(f)\delta(f - f'), \tag{3}$$

where $n(f)$ is the Fourier transform of the noise process $n(t)$. The power spectrum is a measure of the noise power per unit frequency and hence its square root measures the equivalent gravitational wave amplitude $h_{\rm eff}(f)$ which has been plotted in figure 2.

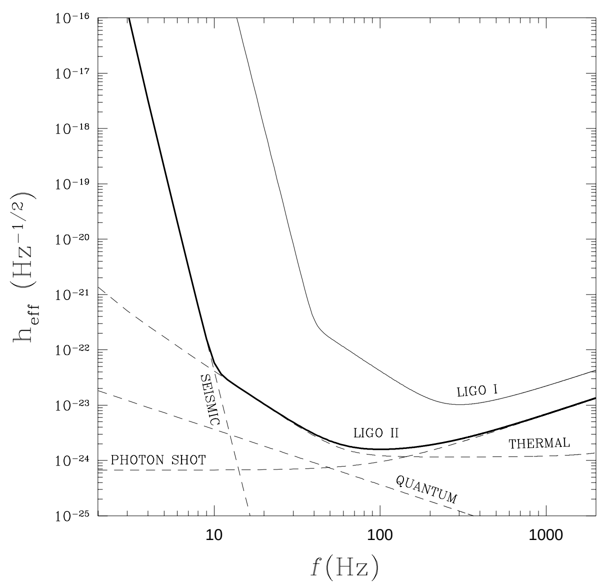


Figure 2: Noise spectrum of the LIGO interferometer

Figure 3 compares the sensitivity of the detector with the strength of the sources. For a signal of amplitude h and a average frequency f_c there is an enhancement of the signal by a factor $\sqrt{n}$ where n is the number of cycles in the neighbourhood of the frequency f_c. This has to be compared with the quantity $h_{\rm rms}$ which is the average strength of the noise over a single cycle of frequency f. For the case of burst sources lasting for a very short period the strength of the burst is to be compared with the quantity $h_{SB} = 11h_{\rm rms}$.

2 Coalescing binaries

Coalescing binaries are promising sources of gravitational waves. The earth based interferometric detectors are ideally suited for measuring such gravitational waves. The key points are summarised

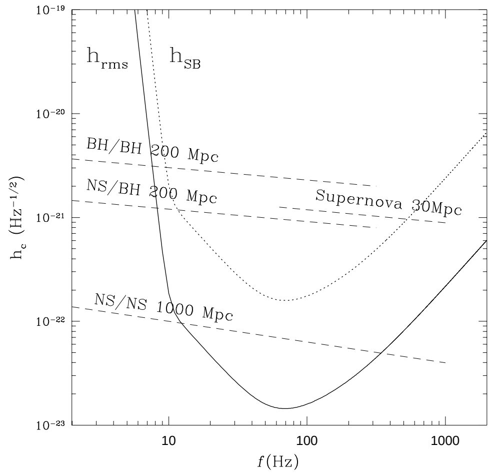


Figure 3: Comparison of the source strengths with the noise

below:

- Systems are well understood and modelled. This means that the waveform can be written down as a function of time and the parameters of the binary system such as the masses of the stars, their spins *etc.*
- The signal lasts for about 15 minutes in the bandwidth of the detector and we can integrate over this interval to enhance the signal to noise ratios.
- The event rates for such coalescences based on binary pulsar statistics have been carried out to a fair degree of reliability. The likely event rates are:
  - $\approx$ 3 NS/NS events per year upto 200Mpc,
  - $\approx$ 100 NS/NS events per year upto 1Gpc.

  These rates have also been extrapolated to BH/BH and BH/NS binaries. With the advanced detector we may be able to observe binaries at cosmological distances   1Gpc.
- The gravitational wave signal carries information of the masses, the spin angular momenta of the two components, the orientation of the binary, its direction, the eccentricity and its distance.
- The parameter information from a collection of such binaries can help determine the cosmologically relevant quantities:
  - the Hubble constant
  - the deceleration parameter
  - the cosmological constant
- Actual coalescence waveforms of BS/BS binaries will yield information about spacetime geometry in the extreme non-linear regime.
- Coalescence of NS/NS binaries will produce waveforms which are sensitive to the equation of state of nuclear matter.
- In the case of BH/NS binaries the tidal disruption of the neutron star can be studied.

2.1 Modelling the chirp waveform

The typical 'chirp' waveform is illustrated in figure 4. Both the amplitude and the frequency increase with time. As the two body problem has not been solved in general relativity post-Newtonian and post-Minkowskian schemes are followed to evaluate the waveform. It turns out that the phase evolution of the waveform is sensitive to the order to which the expansion has been carried out and this plays a major role in the detection strategy. For the purpose of my talk I shall restrict myself to a simplified model of the chirp waveform which is called the restricted post-Newtonian waveform. Here the higher harmonics of the waveform are neglected. Consequently,

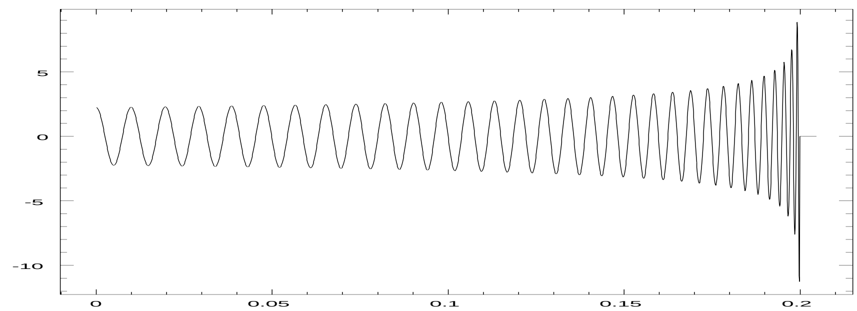


Figure 4: A typical chirp waveform

the restricted post-Newtonian waveforms only contain the dominant frequency equal to twice the orbital frequency of the binary computed up to the relevant order. In the restricted post-Newtonian approximation the gravitational waves from a binary system of stars, modeled as point masses orbiting about each other in a circular orbit, induce a strain $h(t)$ at the detector given by

$$h(t) = A(\pi f(t))^{2/3} \cos\left[\varphi(t)\right], \tag{4}$$

where $f(t)$ is the instantaneous gravitational wave frequency, the constant A involves the distance to the binary, its reduced and total mass, and the antenna pattern of the detector, and the phase of the waveform $\varphi(t)$ contains several pieces corresponding to different post-Newtonian contributions which can be schematically written as

$$\varphi(t) = \varphi_0(t) + \varphi_1(t) + \varphi_{1.5}(t) + \dots. \tag{5}$$

Here $\varphi_0(t)$ is the dominant Newtonian part of the phase and φ_n represents the nth order post-Newtonian correction to it. To further simplify matters we give the formulae only upto the first post-Newtonian order:

$$t - t_a = \tau_0 \left(1 - \left(\frac{f}{f_a}\right)^{-8/3}\right) + \tau_1 \left(1 - \left(\frac{f}{f_a}\right)^{-1}\right)$$

$$\varphi(t) = 2\pi \int_{t_a}^{t} f(t')dt'$$

$$\varphi(t) = \frac{16\pi f_a \tau_0}{5} \left(1 - \left(\frac{f}{f_a}\right)^{-5/3}\right) + 4\pi f_a \tau_1 \left(1 - \left(\frac{f}{f_a}\right)^{-1}\right) + \Phi$$

$$\begin{aligned} \tau_0 &= \frac{5}{256} M^{-1/3} \mu^{-1} \left(\pi f_a\right)^{-8/3} \\ \tau_1 &= \frac{5}{192\mu \left(\pi f_a\right)^2} \left(\frac{743}{336} + \frac{11\mu}{4M}\right) \end{aligned}$$

The quantities τ_0 and τ_1 have the physical significance of representing the Newtonian and post-Newtonian contribution to the time-of-coalescence starting from some fiducial time t_a. Thus in this approximation the time for coalescence is $\tau_0 + \tau_1$. Such a parameterization turns out to be very useful for signal analysis.

3 Detection of the chirp signal

The output of the detector is in general a superposition of a signal and noise. We now discuss how the signal can be extracted from the noisy output of the detector. We first discuss the concept of matched filtering and then develop a geometrical theory of signal analysis. Subsequently we will compare the results of the simulations of the detection process carried out with the theoretically expected results.

3.1 Matched filtering

In order to process the data from the detector the output $g(t)$ of the detector is passed through a linear time-invariant filter. The output of such a detector can be written as

$$c(\tau) = \int_{-\infty}^{\infty} g(t)q(t+\tau)dt = \int_{-\infty}^{\infty} \tilde{g}(f)\tilde{q}(f)e^{2\pi i f \tau} df$$

where $q(t)$ is the impulse response of the detector.

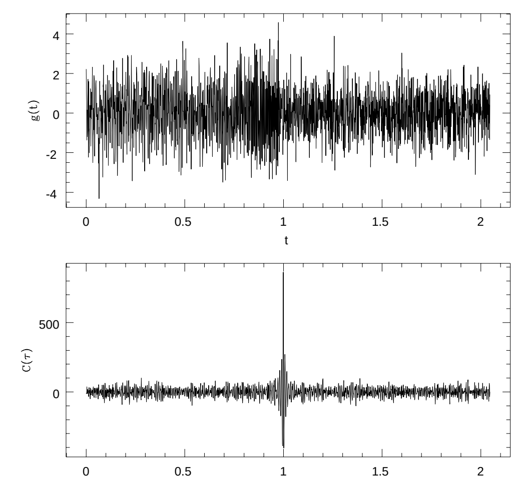


Figure 5: The graph at the top is a typical detector output data train. The noise being large it is difficult to infer the presence of the signal even though it is present. The graph in the bottom is the output of the filter which is matched to the signal present in the data train.

The graph in the top in figure 5 depicts a typical detector output data train which contains a gravitational wave signal from the coalescing binary. It is obvious that the presence of the signal cannot be inferred by inspection. It is necessary to employ special processing to extract the signal from the noise. Such a process is termed as a filter in signal analysis and we shall restrict ourselves to linear transformations of the detector output. The theory of optimal filtering enables us to select an optimal filter suited to the task.

In figure 5, the bottom graph, shows the output of the filter through which the detector output has passed. The key idea here has been to concentrate the signal energy spread over a finite time duration to a small amount of time. The presence of the signal is easily seen above the noise.

3.2 A Geometric approach to Signal analysis

The output of a gravitational wave detector will comprise of data segments, each of duration T seconds, uniformly sampled with a sampling interval of Δ, giving the number of samples in a single data train to be $N = T/\Delta$. Each data train can be considered as a N-tuple $(x^0, x^1, \ldots, x^{N-1})$ x^k being the value of the output of the detector at time $k\Delta$. The set of all such N-tuples constitutes an N-dimensional vector space $\mathcal{V}$ where the addition of two vectors is accomplished by the addition of corresponding time samples. For later convenience we allow each sample to take complex values. A natural basis for this vector space is the *time basis* $\mathbf{e}_m^k = \delta_m^k$ where m and k are the vector and component indices respectively. Another basis which we shall use extensively is the Fourier basis which is related to the time basis by a unitary transformation $\hat{U}$:

$$\tilde{\mathbf{e}}_m = \hat{U}^{mn}\mathbf{e}_n = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\mathbf{e}_n \exp\left[\frac{2\pi i m n}{N}\right], \tag{6}$$

$$\mathbf{e}_m = \hat{U}^{\dagger mn}\tilde{\mathbf{e}}_n = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\tilde{\mathbf{e}}_n \exp\left[-\frac{2\pi i m n}{N}\right]. \tag{7}$$

All vectors in $\mathcal{V}$ are shown in boldface, and the Fourier basis vectors and components of vectors in the Fourier basis are highlighted with a 'tilde'.

A gravitational wave signal from a coalescing binary system can be characterised by a set of parameters $\boldsymbol{\lambda} = (\lambda^0, \lambda^1, \ldots, \lambda^{p-1})$ belonging to some open set of the p-dimensional real space R^p. The set of such signals $\mathbf{s}(t; \boldsymbol{\lambda})$ constitutes a p-dimensional manifold $\mathcal{S}$ which is embedded in the vector space $\mathcal{V}$. The parameters of the binary act as coordinates on the manifold. The basic problem of signal analysis is thus to determine whether the detector output vector $\mathbf{x}$ is the sum of a signal vector and a noise vector, $\mathbf{x} = \mathbf{s} + \mathbf{n}$, or just the noise vector, $\mathbf{x} = \mathbf{n}$, and furthermore to identify which particular signal vector, among all possible. One would also like to estimate the errors in such a measurement. These statements are pictorially represented in figure 6.

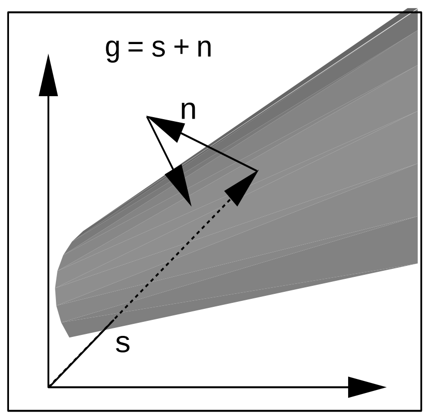


Figure 6: A pictorial representation of the signal manifold.

Essentially the detection process associates the detector output vector with a point on the manifold which is closest to it.

In the absence of the signal the output will contain only noise drawn from a stochastic process which can be described by a probability distribution on the vector space $\mathcal{V}$. The covariance matrix of the noise C^{jk} is defined as,

$$C^{jk} = \overline{n^j n^{*k}}, \tag{8}$$

where an * denotes complex conjugation and an overbar denotes an average over an ensemble. If the noise is assumed to be stationary and ergodic then there exists a noise correlation function

$K(t)$ such that $C_{jk} = K(|j-k|\Delta)$. In the Fourier basis it can be shown that the components of the noise vector are statistically independent [6] and the covariance matrix in the Fourier basis will contain only diagonal terms whose values will be strictly positive: $\tilde{C}_{jj} = \overline{\tilde{n}^j \tilde{n}^{*j}}$. This implies that the covariance matrix has strictly positive eigenvalues. The diagonal elements of this matrix $\tilde{C}_{jj}$ constitute the discrete representation of the power spectrum of the noise $S_n(f)$.

We now discuss how the concept of matched filtering can be used to induce a metric on the signal manifold. The technique of matched filtering involves correlating the detector output with a bank of filters each of which is tuned to detect the gravitational wave from a binary system with a particular set of parameters. The output of the filter, with an impulse response $\mathbf{q}$, is given in the discrete case as

$$c_{(m)} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \tilde{x}^n \tilde{q}^{*n} \exp\left[-2\pi i m n/N\right]. \tag{9}$$

The SNR (ρ) at the output is defined to be the mean of $c_{(m)}$ divided by the square root of its variance:

$$\rho \equiv \frac{\overline{c_{(m)}}}{\left[\overline{\left(c_{(m)} - \overline{c_{(m)}}\right)^2}\right]^{1/2}}. \tag{10}$$

By maximising ρ we can obtain the expression for the optimal filter $\mathbf{q}_{(\mathbf{m})}$ matched to a particular signal $\mathbf{s}(t;\lambda^\mu)$ as

$$\tilde{q}^n_{(m)}(\lambda^\mu) = \frac{\tilde{s}^n(\lambda^\mu)\exp\left[2\pi i m n/N\right]}{\tilde{C}_{nn}}, \tag{11}$$

where ρ has been maximised at the m^{th} data point at the output and where $\mu = 1, 2, \ldots, p$ where p is the number of parameters of the signal. We now introduce a scalar product in $\mathcal{V}$. For any two vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i,j=0}^{N-1} C^{-1}_{ij} x^i y^j = \sum_{n=0}^{N-1} \frac{\tilde{x}^n \tilde{y}^{*n}(\lambda^\mu)}{\tilde{C}_{nn}}. \tag{12}$$

In terms of this scalar product, the output of the optimal filter $\mathbf{q}$, matched to a signal $\mathbf{s}(\lambda^\mu)$, can be written as,

$$c_{(m)}(\lambda^\mu) = \frac{1}{\sqrt{N}} \langle \mathbf{x}, \mathbf{s} \rangle. \tag{13}$$

As $\tilde{C}_{kk}$ is strictly positive the scalar product defined is positive definite. The scalar product defined above on the vector space $\mathcal{V}$ can be used to define a norm on $\mathcal{V}$ which in turn can be used to induce a metric on the manifold. The norm of a vector $\mathbf{x}$ is defined as $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. The norm for the optimal filter can be calculated to give $\rho = \langle \mathbf{s}, \mathbf{s} \rangle^{1/2}$. The norm of the noise vector will be a random variable $\langle \mathbf{n}, \mathbf{n} \rangle^{1/2}$ with a mean value of $\sqrt{N}$ as can be seen by writing the expression for the norm of the noise vector and subsequently taking an ensemble average.

The distance between two points infinitesimally separated on $\mathcal{S}$ can be expressed as a quadratic form in the differences in the values of the parameters at the two points:

$$g_{\mu\nu} d\lambda^\mu d\lambda^\nu \equiv \|\mathbf{s}(\lambda^\mu + d\lambda^\mu) - \mathbf{s}(\lambda^\mu)\|^2 = \left\| \frac{\partial \mathbf{s}}{\partial \lambda^\mu} d\lambda^\mu \right\|^2 \tag{14}$$

$$= \left\langle \frac{\partial \mathbf{s}}{\partial \lambda^\mu}, \frac{\partial \mathbf{s}}{\partial \lambda^\nu} \right\rangle d\lambda^\mu d\lambda^\nu. \tag{15}$$

The components of the metric in the coordinate basis are seen to be the scalar products of the coordinate basis vectors of the manifold.

Since the number of correlations we can perform on-line is finite, we cannot have a filter corresponding to every signal. A single filter though matched to a particular signal will also 'detect' signals in a small neighbourhood of that signal but with a slight loss in the SNR. The drop in the correlation can be related to the metric distance on the manifold between the two infinitesimally separated signal vectors. In statistical theory the matrix $g_{\mu\nu}$ is called the Fisher information matrix.

In section 2.1 we have already introduced the waveform and the four parameters $\lambda^\mu = \{t_a, \Phi, \tau_0, \tau_1\}$. We now introduce an additional parameter for the amplitude and call it $\lambda^0 = \mathcal{A}$. The signal can now be written as $\tilde{s}(f;\boldsymbol{\lambda}) = \mathcal{A}h(f; t_a, \Phi, \tau_0, \tau_1)$, where, $\boldsymbol{\lambda} \equiv \{\mathcal{A}, t_a, \Phi, \tau_0, \tau_1\}$. Numerically the value of the parameter $\mathcal{A}$ will be the same as that of the SNR obtained for the matched filter provided $\mathbf{h}$ has unit norm. We can decompose the signal manifold into a manifold containing normalised chirp waveforms and a one-dimensional manifold corresponding to the parameter $\mathcal{A}$. The normalised chirp manifold can therefore be parameterized by $\{t_a, \Phi, \tau_0, \tau_1\}$. This parameterization is useful as the coordinate basis vector $\frac{\partial}{\partial \mathcal{A}}$ will be orthogonal to all the other basis vectors as will be seen below.

In order to compute the metric, and equivalently the Fisher information matrix, we use the continuum version of the scalar product as given in [11], except that we use the two sided power spectral density. This has the advantage of showing clearly the range of integration in the frequency space though we get the same result using the discrete version of the scalar product. Using the definition of the scalar product we get

$$g_{\mu\nu} = \int_{f_a}^{\infty} \frac{df}{S_n(f)} \frac{\partial \tilde{s}(f;\boldsymbol{\lambda})}{\partial \lambda^\mu} \frac{\partial \tilde{s}^*(f;\boldsymbol{\lambda})}{\partial \lambda^\nu} + \text{c.c.} \tag{16}$$

It turns out that the signal manifold associated with the chirp signal is flat and moreover the particular set of parameters that we have chosen turn out to constitute a cartesian coordinate system on the chirp manifold.

3.3 Choice of filters

It is obvious that as the number of signals is infinite one cannot correlate the output with every possible signal and we must choose a discrete set of templates which will do the job. The most obvious strategy would be to take a finite number of points on the signal manifold, evenly placed on the signal manifold. However it might be much more efficient not to restrict the set of templates to lie on the manifold. Such a strategy has been implemented in [15] and the results obtained are encouraging.

3.4 Estimation of parameters

We have seen that the detection process involves correlating the detector output with a host of templates. The most likely parameters of the binary are same as those of the template which obtains the maximum correlation with the detector output. We would like to compute the errors inherent in such a measurement. We define the error matrix

$$\epsilon^{\mu\nu} = \overline{\left(\lambda^\mu - \overline{\lambda}^\mu\right))\left(\lambda^\nu - \overline{\lambda}^\nu\right)}. \tag{17}$$

It can be shown [6] that there in a lower bound to the errors in such a measurement. This bound is called the Cramer-Rao bound and in the limit of high SNRs the errors converge to the Cramer-Rao bound. The Cramer-Rao bound is valid for unbiased maximum likelihood estimates, conditions which the matched filtering statistic satisfies. It turns out that the lower bound turns out to be nothing but the inverse of the Fisher information matrix or equivalently the inverse of the metric on the signal manifold. Thus,

$$\epsilon^{\mu\nu} \geq g^{-1}_{\mu\nu} = C^{\mu\nu}. \tag{18}$$

In order to test the validity of these bounds we carried out extensive numerical simulations. The simulations were carried out by adding numerous realizations of noise to a given signal and then filtering it through a bank of filters. The simulations were carried out were representative of the actual detection process. The assumptions made are listed below:

1. The initial LIGO power spectrum was used mainly because of computational reasons.
2. We considered BH-NS and BH-BH binaries as their coalescence time is less as compared to low mass binaries. This leads to lesser computational requirements.
3. Only the restricted first post-Newtonian waveform was used. We ignored the spins in the waveform.
4. The inspiral of the binary is abruptly cut of by the plunge phase which occurs when the two stars are approximately at a distance of $6m$ from each other. As the exact onset of the plunge is uncertain we carried out the simulations with the plunge effect both included and excluded.
5. The range of SNRs which we worked with, were between 10 and 20.

3.5 Results of simulations

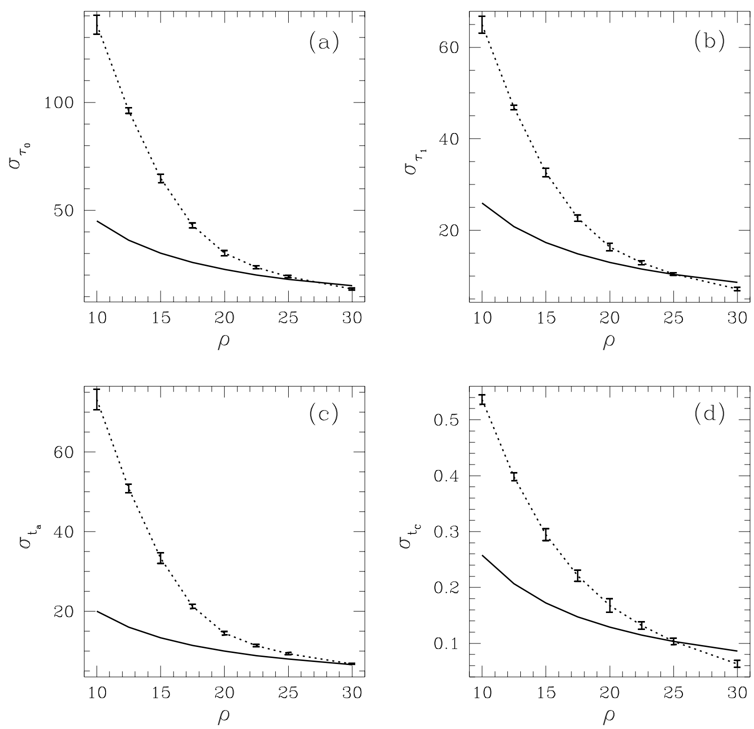


Figure 7: Dependence of the errors in the estimation of parameters of the post-Newtonian waveform *i.e.* $\{\sigma_{\tau_0}, \sigma_{\tau_1}, \sigma_{t_a}, \sigma_{t_C}\}$ as a function of SNR. The solid line represents the analytically computed errors whereas the dotted line represents the errors obtained through Monte Carlo simulations.

The results of the simulations are shown in figure 7. We compare the actual errors obtained with the covariance matrix. The results are shown for the parameters τ_0, τ_1, t_a and $t_C = \tau_0 + \tau_1 + t_a$. The parameter t_C is termed as the instant of coalescence and is very accurately determined.

There is a huge discrepancy in the errors obtained via simulations and the errors obtained by the covariance matrix at SNRs of about 10. Due to the intrinsic weakness of gravitational wave sources, we should not expect to see sources with SNRs higher that 10 very frequently. There is approximately a factor of three difference between the two. This implies that more careful estimation of the errors would have to be carried out. At a SNR of about 30 the errors are in good agreement with the covariance matrix. It is to be noted that the experimental errors are less than the lower bound at high SNRs. This is partly due to numerical noise in the simulations and partly due to the fact that the covariance matrix itself is evaluated using a numerical integration routine. Moreover the accuracy can also be improved by increasing the sampling rate and hence improving the resolution in time.

4 Conclusion

We have discussed here issues in the data analysis of the gravitational wave signals from coalescing binary systems. Though a lot of work has been done in this area a lot of scope remains to improve the detection strategies to be employed. We enumerate below the directions which future work in this area can take.

1. More accurate waveforms are needed so as to improve the chances of detection.
2. To obtain more realistic bounds on the error matrix. As we have seen the Cramer Rao bound works only for very high signal to noise ratios.
3. Further use of differential geometry in choosing templates optimally.
4. Exploration of other data analysis techniques, such as:
    - Periodogram
    - Resampling algorithms
    - non-linear filtering
    - wavelets
5. Optimisation of coincident detections among detectors spread around the globe. This is termed as the network problem.

References

[1] A. Abramovici *et. al.*, Science **256** 325 (1992).

[2] C. Bradaschia *et. al.*, Nucl. Instrum. Methods Phys. Res., Sect A 518 (1990).

[3] B.F. Schutz, in *Gravitational Collapse and Relativity,* Edited by H. Sato and T. Nakamura, (World Scientific, Singapore, 1986), pp. 350-368.

[4] D. Marković, Phys. Rev. D **48**, 4738 (1993).

[5] K.S. Thorne, in *300 Years of Gravitation*, S.W. Hawking and W.Israel (eds.), (Cambridge Univ. Press, 1987).

[6] C.W. Helstrom, *Statistical Theory of Signal Detection,* 2nd. ed, (Pergamon Press, London, 1968).

[7] B.F. Schutz, in *The Detection of Gravitational Radiation,* edited by D. Blair (Cambridge, 1989) pp 406-427.

[8] B.S. Sathyaprakash and S.V. Dhurandhar, Phys. Rev. D **44**, 3819 (1991).

[9] S.V. Dhurandhar and B.S. Sathyaprakash, Phys. Rev. D **49**, 1707 (1994).

[10] C. Cutler and E. Flanagan, Phys.Rev. D **49**, 2658 (1994).

[11] L.S. Finn and D.F. Chernoff, Phys. Rev. D **47**, 2198 (1993).

[12] B.S. Sathyaprakash, Phys. Rev. D **50**, R7111 (1994).

[13] C. Cutler *et al*, Phys. Rev. Lett. **70**, 2984 (1993).

[14] S. V. Dhurandhar and B.F. Schutz, Phys. Rev. D **50**, 2390 (1994)

[15] R. Balasubramanian, S.V. Dhurandhar and B.S. Sathyaprakash, Phys. Rev. D **53**, 3033 (1996).

I have been able to *think* more about stars than I would have, because of the lack of 'paper and pencil.' Paper and pencil necessarily land one in the intricacies of the calculation! I have a number of ideas and I am getting almost impatient to get back to Cambridge to work them out.

— S. CHANDRASEKHAR

GRAVITATIONAL COLLAPSE AND COSMIC CENSORSHIP

T. P. Singh[1]

Theoretical Astrophysics Group
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400005, India.

Abstract

This article gives an elementary overview of the end-state of gravitational collapse according to classical general relativity. The focus of discussion is the formation of black holes and naked singularities in various physically reasonable models of gravitational collapse. Possible implications for the cosmic censorship hypothesis are outlined.

[1] e-mail:tpsingh@tifrvax.tifr.res.in

While an occasion such as this one is personally very gratifying, I must confess to some misgivings as to the appropriateness of selecting for special honor those who have received recognition of a particular kind by their contemporaries. I am perhaps oversensitive to this issue, since I have always remembered what a close friend of earlier years, Professor Edward Arthur Milne, once said. On an occasion, now more than fifty years ago, Milne reminded me that posterity, in time, will give us all our true measure and assign to each of us our due and humble place; and in the end it is the judgement of posterity that really matters. And he further added: he really succeeds who perseveres according to his lights, unaffected by fortune, good or bad. And it is well to remember that there is in general no correlation between the judgment of posterity and the judgment of contemporaries.

I hope you will forgive me if I allow myself a personal reflection. During the seventies, I experienced two major heart episodes. Suppose that one of them had proved fatal, as it well might have. Then there would have been no cause for celebration. But I hope that the judgment by posterity of my efforts in science would not have been diminished on that account. Conversely, I hope that it would not be enhanced on account of a doctor's skills.

— S. CHANDRASEKHAR

1 Introduction

It is expected that a very massive star will not end up either as a white dwarf or as a neutron star, and that it will undergo an intense gravitational collapse towards the end of its life history. The very late stages in the evolution of such a star will necessarily be determined by quantum gravitational effects. Since we do not yet have a quantum theory of gravity, we cannot give a definite description of these very late stages. However, we can at least ask what classical general relativity predicts for the final stages of the evolution. It is possible that the answer given by the classical theory will have some connection with the answer coming from quantum gravity - perhaps the latter will provide a sort of quantum correction to the former.

It is remarkable that seventy years after Einstein proposed the general theory of relativity we do not have a complete understanding of the theory's prediction for the end-state of gravitational collapse. This situation is intimately related with our lack of understanding of the general global properties of the solutions of Einstein's equations. The most significant developments to date in the study of gravitational collapse have been the singularity theorems of Hawking and Penrose [1]. In the context of gravitational collapse, the theorems show that if a trapped surface forms during the collapse of a compact object made out of physically reasonable matter, the spacetime geometry will develop a gravitational singularity. By a gravitational singularity one means that the evolution of geodesics in the spacetime will be incomplete. It is plausible that the formation of a gravitational singularity in a collapsing star will be accompanied by a curvature singularity - one or more curvature scalars will diverge.

The general conditions which will ensure the formation of a trapped surface are not well understood. This is one of the aspects in which our understanding of gravitational collapse is incomplete. In this article, however, we will not be concerned with this particular issue. We will assume that a gravitational singularity does form, either because the conditions of the singularity theorems have been met, or otherwise. All the same it should be mentioned that the astrophysical parameters for very massive collapsing stars are usually such that a trapped surface can be expected to form during gravitational collapse.

It maybe the case that the singularity is not visible to a far-away observer because light is not able to escape the collapsing star. This is essentially what we mean when we say that a black-hole has formed. The singularity is hidden from view by the event horizon, which is the boundary of that spacetime region surrounding the singularity which cannot communicate with the far-away observer (Figure 1).

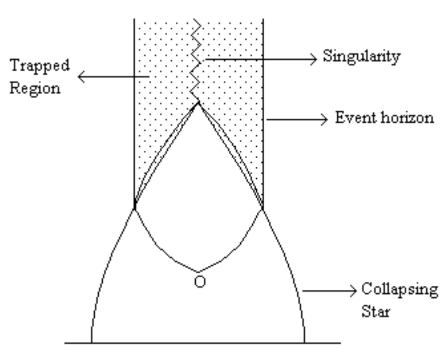


Figure 1. The formation of a black-hole in gravitational collapse. The event horizon starts forming at O and covers the singularity. The boundary of the trapped region inside the star is the apparent horizon.

The singularity theorems of Hawking and Penrose do not imply that the collapsing star which develops a singularity will necessarily become a black-hole. That is, even if a trapped surface and hence a singularity does form during collapse, the trapped surface need not hide the singularity from a far-away observer. This alternate possibility is called a naked singularity (Figure 2).

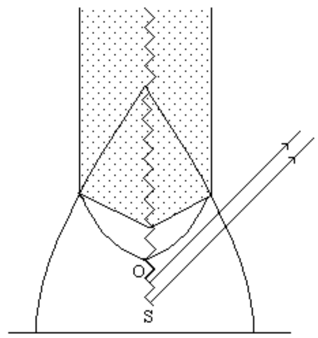


Figure 2. The formation of a naked singularity. The singularity begins to form at S. The event horizon starts forming at O and fails to cover the singularity entirely. Else, the singularity may form without the event horizon forming at all. Light rays are able to escape from the singularity.

In this case, the event-horizon fails to cover the singularity and light from the singularity could escape to infinity. One could form a rough picture of this situation by imagining that a singularity has formed at the center of a collapsing spherical star *before* its boundary has entered the Schwarzschild radius. Such a singularity could be visible to an observer watching the collapse. Also, if a singularity forms without the conditions of the singularity theorems having been met, it could be naked, as the event horizon may not form at all.

Does gravitational collapse end in a black-hole or a naked singularity? Einstein's equations must give a definite answer to this question, but at present we do not know what that answer is. One could well ask, what difference does it make? The difference is very significant. If a singularity is naked, one could prescribe arbitrary data on the singular surface - this would result in total loss of predictability in the future of the singularity. If a singularity is hidden behind an event horizon, predictability would be preserved at least in the spacetime region outside the horizon.

Furthermore, the issue has great importance for black-hole astrophysics and for the theory of black-holes. In terms of their astrophysical properties, naked singularities could be very different from black-holes. While (classical) black-holes are one-way membranes and inflowing matter simply gets sucked into the singularity, quite the opposite may hold for a naked singularity; matter might be thrown out with great intensity from near a naked singularity. The validity of many theorems on black-hole dynamics depends on the assumption of absence of naked singularities. The extensive and successful applications of black-holes in astrophysics, and the detailed studies of their profound and elegant properties does lend strong support to the belief in their existence. Nonetheless, it is an open question as to whether gravitational collapse necessarily ends in a black-hole or could in some cases lead to a naked singularity. If the latter is the case, then one needs to know what kind of stars end as naked singularities.

In the absence of an answer to the above question, the basis for black-hole physics and its applications is the Cosmic Censorship Hypothesis. In somewhat non-rigorous terms, the hypothesis could be stated as

- Gravitational collapse of *physically reasonable matter* starting from *generic initial data* leads to the formation of a black-hole, not a naked singularity.

We will elaborate in a moment on the italics. Such a hypothesis was first considered by Penrose in [2], where he asked as to whether there exists a Cosmic Censor who would always cloth a singularity with an event horizon. (It is interesting, though probably not well-known, that on occasion Penrose has also written in support of naked singularities [3]). The censorship hypothesis has remained unproved despite many serious efforts (for a review of some of the attempts see [4]). Part of the difficulty lies in not having a unique rigorous statement one could try to prove. Studying gravitational collapse using Einstein equations is a formidable task, and evidence for *formation* of a black-hole is as limited as for a naked singularity. This is simply because very few examples of dynamical collapse have been worked out, and not many exact solutions are known. Until recently, the widely held belief has been that the hypothesis must be right; however studies over the last few years have caused at least some relativists to reconsider their earlier view. It is obvious, though not often emphasized, that falsification of the hypothesis does not necessarily rule out black-holes altogether but might allow only a subset of the initial data to result in naked singularities. The former alternative (i.e. no black-holes) is extremely unlikely, given the great richness of black-hole physics and astrophysics.

By 'physically reasonable matter' one usually means that the (classical) matter obeys one or more of the energy conditions (for a discussion see [5]). The weak energy condition for instance requires that the pressures in the collapsing matter not be too negative, and the energy density be positive. By 'generic initial data' one means that the solution of Einstein equations being used to study collapse has as many free functions as are required for the initial data to be arbitrary.

Many people consider naked singularities to be a disaster for general relativity and for physics as such. We have mentioned that naked singularities could result in a total loss of predictability to their future. However, if we make the natural assumption that the ultimate theory of gravity will preserve predictability (in some suitable sense of the concept), the occurrence of naked singularities in general relativity will signal a real need for modification of the theory. Such a definite (and rare) signal could only help improve our understanding of gravitation, instead of being a disaster! To many other people, naked singularities are a positive asset for astrophysics - the possibility of emission of light from high curvature regions close to the singularity might make such singularities extraordinary energy sources. Sometimes a view is expressed that in any case a quantum theory of gravity will avoid singularities altogether - in that case how does it matter whether the classical theory (general relativity) predicts the singularities to be naked or covered? Our discussion above suggests that even if quantum gravity were to avoid singularities, the spacetime regions that are naked according to the classical theory will behave very differently from black-holes in the quantized theory as well. Besides, the relevance of the hypothesis to black-hole astrophysics cannot be overlooked.

Perhaps it is important to mention that the validity of the hypothesis could be discussed at two distinct levels. Firstly, we need to find out if it holds in general relativity. However, even if relativity theory allows naked singularities, it could be that actual stars may not end up as naked singularities. This could happen if the initial conditions necessary for a naked singularity to form are simply not observed in the real world. Thus violation of the hypothesis at a theoretical level might compel us to consider replacing general relativity by a better theory, but may not have observational consequences.

The failure to prove the hypothesis has led, over the last ten years or so, to a change in the approach towards the problem. Attention has shifted to studies of specific examples of gravitational collapse. For instance, people have been studying spherical gravitational collapse, with a particular choice of the matter stress tensor - like dust, perfect fluids, and massless scalar fields. These and other examples often show that collapse of physically reasonable matter can end either in a black-hole or a naked singularity, depending on the choice of initial data. To a degree, such examples go against expectations that the censorship hypothesis is correct. It, however, does remain to be seen as to whether or not generic initial data will lead to naked singularities. These examples can be regarded as good learning exercises - at the very least we learn about the properties of naked singularities which form. Perhaps these very properties might suggest that the singularity, though naked, does not violate the spirit of the hypothesis.

In this article we will attempt an overview of the end state of gravitational collapse according to general relativity, the focus of discussion being the censorship hypothesis. Rather than reviewing the attempts to prove the hypothesis, we will concern ourselves with recent studies of models of collapse. It is relevant to note that these models typically examine the properties of curvature singularities that form during collapse. If a naked curvature singularity does form, the issue of its geodesic incompleteness is to be handled separately - an aspect we will consider briefly towards the end of the review. Also, we will not discuss examples of static naked singularities in general relativity, although many such are known.

In Section II, spherical gravitational collapse for various forms of matter is reviewed. We also discuss some properties of the naked singularities found in these models. Section III is a brief discussion of the limited results on non-spherical collapse. In the last section, a critical comparison of the various existing results and their interpretation is attempted.

2 Spherical Gravitational Collapse

It is only natural that most of the examples studied are for the idealized case of spherical collapse, in asymptotically flat backgrounds. Even here, one does not yet know the conditions for formation of black-holes and naked singularities. Some results are known for specific forms of energy-momentum tensors. Often there are striking similarities amongst results for different kinds of matter, suggesting an underlying pattern. We review results for collapse of dust, perfect fluids with pressure, a radiating star described by the Vaidya metric, massless scalar fields, and spherical collapse for matter with no restriction on the energy-momentum tensor, except the weak energy condition.

2.1 Dust collapse

By dust one means a perfect fluid for which the pressure is negligible and is set to zero. This highly idealized description has the advantage that an exact solution of Einstein equations is known, which describes the collapse of a spherical dust cloud in an asymptotically flat spacetime. This is the Tolman-Bondi solution, given independently by the two authors [6]. The evolution of the cloud is determined once the initial density and velocity distribution of the fluid has been given. Because of spherical symmetry these distributions will be functions only of the radial coordinate r, so that the initial data consists of two arbitrary functions of r. The cloud is assumed to extend upto a finite radius and the interior dust solution is matched to a Schwarzschild exterior. Since we are interested in collapse, the velocity of each fluid element is taken to be towards the center of the cloud. It can be shown that starting from regular initial data, the collapse leads to the formation of a curvature singularity.

A very special case of the Tolman-Bondi solution is a dust cloud whose initial density distribution is homogeneous, and the velocity increases linearly with the physical distance from the center. This of course is the Friedmann solution matched to a Schwarzschild exterior and was used by Oppenheimer and Snyder [7] to provide the first example of dynamical collapse in general relativity. Starting from regular initial data, the cloud develops a curvature singularity at its center which is not visible. This was the first theoretical example of black-hole formation (Figure 3(i)).

For many years, the work of Oppenheimer and Snyder has remained a model for how a black-hole might form in gravitational collapse. The collapsing star will enter its Schwarzschild radius, become trapped and proceed to become singular, and the singularity will be hidden behind the event horizon. The censorship hypothesis was also originally inspired essentially by this work, because not much more was known anyway about properties of gravitational collapse. In hindsight, one might be surprised with the degree of generality attributed to results arising from the study of a model star that is spherically symmetric, homogeneous and made of dust - none of the three properties are obviously true for a real star!

Since the exact solution of Tolman and Bondi was known, it would have been quite natural to extend the work of Oppenheimer and Snyder to *inhomogeneous* dust collapse, described by this

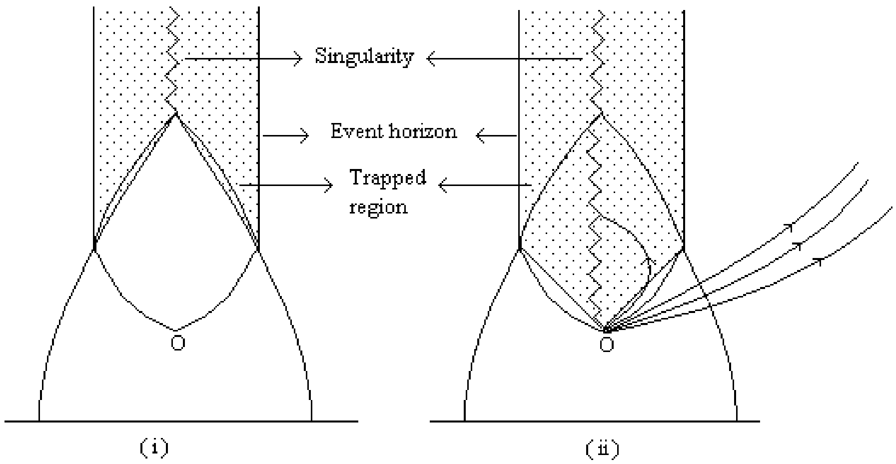


Figure 3. Contrasting the formation of the black-hole in homogeneous dust collapse (Fig. i) with the formation of a globally naked singularity in inhomogeneous dust collapse (Fig. ii). The event horizon begins to form at O. For inhomogeneous collapse leading to a naked singularity, the center is the first point to get trapped, unlike in the homogeneous case. In (ii), a family of rays escapes from the singularity at O.

exact solution. However, for nearly three decades after the paper of Oppenheimer and Snyder was published, very little work appears to have been done on gravitational collapse. Presumably, there were not enough reasons for interest in the subject until the discovery of quasars in the sixties, and the development of singularity theorems. It is also true though that analysis of light propagation in the collapsing inhomogeneous dust star is a difficult task, and tractable methods have been developed only in recent years. The first work to deal with the inhomogeneous Tolman-Bondi model appeared in the seventies [8]. Since then, many investigations of inhomogeneous dust collapse have taken place [9], from the point of view of the censorship hypothesis, and the last word on this particular model has not yet been said.

Yodzis et al. [8] were investigating what are called shell-crossing singularities (or caustics) which form due to the intersection of two collapsing dust-shells at a point other than the center. These are curvature singularities and they are also naked - but are not regarded as violation of censorship since there is evidence [10] that such singularities are gravitationally weak (as discussed later in the article). They are similar to the shell-crossing singularities that occur in Newtonian gravity also and it is believed that spacetime can be extended through such singularities.

Of a more serious nature are the shell-focussing singularities which form at the center of the cloud - they result from the shrinking of collapsing shells to zero radius. It was found by various people that for some of the initial density and velocity distributions, the collapse ends in a naked singularity, whereas for other distributions it ends in a black-hole. Also, both the black-hole and naked singularity solutions result from a non-zero measure set of initial data. In particular, there was found a one-parameter family of solutions (described say by the parameter ξ), such that for $\xi < \xi_c$ the collapse leads to a black-hole, whereas for $\xi > \xi_c$ it leads to a naked singularity.

The space-time diagram for inhomogeneous dust collapse leading to a naked shell-focussing singularity is shown in Figure 3(ii), and should be contrasted with Figure 3(i) for Oppenheimer-Snyder collapse. Of particular importance is the difference in the evolution of trapped surfaces in the two cases, and the fact that for inhomogeneous collapse different shells become singular at different times, unlike in the homogeneous case. For a discussion of trapped surfaces in dust collapse see Jhingan et al. in [9].

At this stage we need to distinguish between a *locally* naked singularity and a *globally* naked singularity. We say the singularity is locally naked if light-rays do emerge from the singularity but fall back to the center without escaping the event-horizon. Such a singularity will be visible to an infalling observer who has entered the Schwarzschild radius, but cannot be seen by an asymptotic observer (Figure 4). A singularity is called globally naked if light-rays emerging from the singularity escape the event-horizon and reach an asymptotic observer. We say the singularity is visible if there are light-rays emerging from the singularity - a visible singularity may be locally or globally naked. We say that a piece of the singularity is covered if it is not even locally naked. In our terminology, a locally naked singularity is also a black-hole. In this article when we call a singularity naked, we mean it is globally naked. The *weak* censorship hypothesis allows for the occurrence of locally naked singularities but not globally naked ones, whereas *strong* censorship does not allow either.

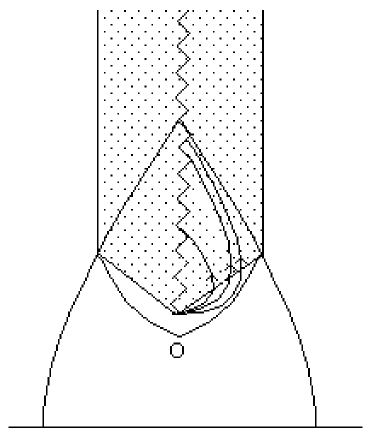


Figure 4. The formation of a locally naked singularity in dust collapse. All rays starting at the singularity fall back into the star.

As an illustration we describe in some detail the gravitational collapse of an inhomogeneous dust cloud, starting from rest. It can be shown that the singularity resulting from the collapse of shells with $r > 0$ is covered. At most, the singularity forming at $r = 0$ (the central singularity) can be naked. The conditions for the central singularity to be naked are given as follows. Let the initial density $\rho(R)$ as a function of the physical radius R be given as a power-series, near the center:

$$\rho(R) = \rho_0 + \frac{1}{2}\rho_2 R^2 + \frac{1}{6}\rho_3 R^3 + ...$$

where ρ_0 is the initial central density, and ρ_2 and ρ_3 are respectively the second and third derivatives at the center. We assume that the density decreases with increasing R, hence the first non-vanishing derivative is negative. It turns out that if $\rho_2 < 0$ the singularity is visible. If $\rho_2 = 0$ and $\rho_3 < 0$ then one defines a parameter $\xi = |\rho_3|/\rho_0^{5/2}$. The singularity is visible for $\xi > 25.47$ and covered for $\xi < 25.47$. If $\rho_2 = \rho_3 = 0$ the singularity is covered and we have the formation of a black-hole. The Oppenheimer-Snyder example is a subset of this case. In the case of a visible singularity, entire families of light-rays emerge from the singularity. Note that $\rho_2 < 0$ is generic and $\rho_2 = 0$ non-generic, hence generic dust collapse leads to a visible singularity.

We find that there is a transition from naked singularity type behaviour to black-hole type behaviour as the density profile is made less inhomogeneous by setting more and more density derivatives to zero. In the cases where the singularity is visible, the initial density distribution through the star then determines whether the singularity is locally or globally naked - examples of both kind occur. When the collapsing dust cloud has an initial velocity, the overall picture regarding the nature of the singularity is essentially the same as described here for the case of a cloud collapsing from rest. A Penrose diagram for the naked singularity is shown in Figure 5. In summary, when one allows for inhomogeneity in the density distribution, the nature of gravitational collapse is quite different from what Oppenheimer and Snyder found for the homogeneous case. Some open issues relating to dust collapse are pointed out later, in the discussion.

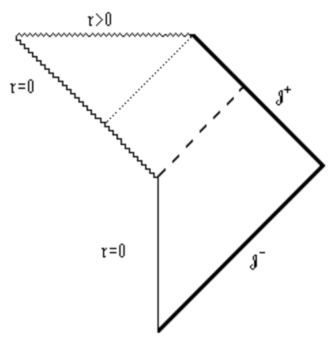


Figure 5. A globally naked central singularity. The dashed line is the Cauchy horizon
the dotted line is the event horizon.

2.2 Including Pressure

The description of collapsing matter as dust is an idealization and a realistic study of collapse must take pressure gradients into account. However, useful exact solutions of Einstein equations describing spherical collapse of matter with pressure are hard to come by, and as a result a clear picture of the kind presented above for dust does not yet exist. It is in fact remarkable that we do not even know how the Tolman-Bondi solution would change under the introduction of "small" pressures. For instance if we are considering an equation of state $p = k\rho$ for a perfect fluid, with k very close to zero, it is not clear whether the solution is a perturbation to the dust solution. In this section we summarize the few known results on collapse of relativistic fluids, vis a vis the censorship hypothesis.

When pressure is included, the energy momentum tensor T_{ik} for matter undergoing spherical collapse is conveniently described in a comoving coordinate system. In these coordinates, T_{ik} is diagonal and its components are the energy density, the radial pressure and the tangential pressure. (The only exception to this general description of the matter is the case of null dust, described by the Vaidya spacetime, and reviewed below). For a perfect fluid, the two pressures are identical. Since perfect fluids are easier to study compared to imperfect ones, they have inevitably received greater attention.

An important early paper on the collapse of perfect fluids is that of Misner and Sharp [11]. They set up the Einstein equations for this system in a useful physical form, bringing out the departures from the Oppenheimer-Volkoff equations of hydrostatic equilibrium. However, they did not consider solutions of these equations. Lifshitz and Khalatnikov, and Podurets [12] worked out the form of the solution near the singularity for the equation of state of radiation, $p = \rho/3$. Their approach has not received much attention, but appears to offer a promising starting point for investigating censorship.

A significant development was the work of Ori and Piran [13], who investigated the self-similar gravitational collapse of a perfect fluid with an equation of state $p = k\rho$. It is readily shown that the collapse leads to the formation of a curvature singularity. The assumption of self-similarity reduces Einstein equations to ordinary differential equations which are solved numerically, along with the equations for radial and non-radial null geodesics. It is then shown that for every value of k (in the range investigated: $0 \leq k \leq 0.4$) there are solutions with a naked singularity, as well as black-hole solutions. Each kind of solution has a non-zero measure in the space of spherical self-similar solutions for this equation of state. The issue as to which initial data (density profile, velocity profile, and value of k) leads to naked singularities has not been fully worked out though, and is an interesting open problem. (We have in mind a comparison with the results for dust, quoted above).

An analytical treatment for this problem was developed by Joshi and Dwivedi [14]. After deriving the Einstein equations for the collapsing self-similar perfect fluid they reduce the geodesic equation, in the neighborhood of the singularity, to an algebraic equation. The free parameters in this algebraic equation are in principle determined by the initial data. The singularity will be naked for those values of the parameters for which this equation admits positive real roots. Since this is an algebraic equation, it will necessarily have positive roots for some of the values of the parameters, and for the initial data corresponding to such values of the parameters the singularity is naked. It is shown that families of non-spacelike geodesics will emerge from the naked singularity. As in the case of Ori and Piran's work, the explicit relation between the initial data and the naked singularity has not yet been worked out. Also it is not clear as to what is the measure of the subset of solutions leading to a naked singularity or a black-hole. This analysis was extended to a self-similar spacetime with a general form for T_{ik} [15].

On physical grounds, imperfect fluids are more realistic than perfect ones; very little is known about their collapse properties though. An interesting paper is the one by Szekeres and Iyer [16], who do not start by assuming an equation of state. Instead they assume the metric components to have a certain power-law form, and also assume that collapse of physically reasonable fluids can be described by such metrics. The singularities resulting in the evolution are analysed, with attention being concentrated on shell-focussing singularities at $r > 0$. They find that although naked singularities can occur, this requires that the radial or tangential pressure must either be negative or equal in magnitude to the density.

Lastly we mention the collapse of null dust (directed radiation), described by the Vaidya spacetime [17]. Since the exact solution is known, the collapse has been thoroughly investigated [18] for the occurrence of naked singularities. One considers an infalling spherical shell of radiation and imagines it as being made of layers of thin shells. A thin shell becomes singular when its radius shrinks to zero. Let the shells be labelled by the advanced time coordinate v, with $v = 0$ for the innermost shell, and let $m(v)$ be the Vaidya mass function. It can be shown that the singularity at $v > 0$ is covered. For $v = 0$ the singularity is naked if $\lambda \equiv 2\; dm(v)/dv|_{v=0}$ is less than or equal to $1/8$, and covered otherwise. These results bear an interesting similarity with those for dust, described in the previous section.

It will be evident from the previous few paragraphs that our understanding of spherical collapse, when pressure gradients are included, is rather incomplete. Ultimately, one would like to develop the kind of clear picture that is available for dust collapse. It is interesting however that the collapse of a self-similar perfect fluid, and of the fluids considered by Szekeres and Iyer, admits both black-hole and naked singularity solutions. This also brings forth an astrophysical issue - what is the relevant equation of state in the final stages of collapse of a star? Could it be that the initial data leading to a naked singularity is not being realised astrophysically?

2.3 Collapse of a scalar field

One could take the viewpoint that the description of matter as a relativistic fluid is phenomenological, and that the censorship hypothesis should be addressed by considering matter as fundamental fields. As a first step, one could study the spherical collapse of a self-gravitating massless scalar field. A good deal of work has been done on this problem over the last few years, and exciting results have been found. Here we can provide only a very brief overview.

In a series of papers, Christodoulou has pioneered analytical studies of scalar collapse [19]. He established the global existence and uniqueness of solutions for the collapsing field, and also gave sufficient conditions for the formation of a trapped surface. For a self-similar scalar collapse he showed that there are initial conditions which result in the formation of naked singularities. He has claimed that such initial conditions are a subset of measure zero and hence that naked singularity formation is an unstable phenomenon.

Christodoulou was also interested in the question of the mass of the black-hole which might form during the collapse of the scalar wave-packet. Given a one parameter family $\mathcal{S}[p]$ of solutions labelled by the parameter p which controls the strength of interaction, it was expected that as p is varied, there would be solutions with $p \rightarrow p_{weak}$ in which the collapsing wave-packet disperses again, and solutions with $p \rightarrow p_{strong}$ which have black-hole formation. For a given family there was expected to be a critical value $p = p_*$ for which the first black-hole appears as p varies from the weak to the strong range. Do the smallest mass black holes have finite or infinitesimal mass [20]? This issue would be of interest for censorship, since an infinitesimal mass would mean one could probe arbitrarily close to the singularity. Of course when one is considering real stars, a finite lower limit on the mass of the collapsing object arises because non-gravitational forces are also involved.

This problem was studied by Choptuik [21] numerically and some remarkable results were found. He confirmed that the family $\mathcal{S}[p]$ has dispersive solutions as well as those forming black-holes, and a transition takes place from one class to the other at a critical $p = p_*$. The smallest black-holes have infinitesimal mass. Near the critical region, the mass M_{bh} of the black-hole scales as $M_{bh} \approx (p - p_*)^\gamma$, where γ is a universal constant (i.e. same for all families) having a value of about 0.37. The near critical evolution can be described by a universal solution of the field equations which also has a periodicity property called echoing, or discrete self-similarity. That is, it remains unchanged under a rescaling $(r, t) \rightarrow (e^{-n\Delta} r, e^{-n\Delta} t)$ of spacetime coordinates. n is an integer, and Δ is about 3.4. Subsequently, these results have been confirmed by others [22]. The occurrence of black-holes with infinitesimal mass goes against the spirit of censorship. The critical solution ($p = p_*$) is a naked singularity. However, since the naked singularity is being realised for one specific solution in the one parameter family, it is a measure zero subset.

Attempts have been made to construct analytical models which will reproduce Choptuik's numerical results [23]. Since it is difficult to make a model with discrete self-similarity, continuous self-similarity is assumed instead. Brady showed that there are solutions which have dispersal, as well as solutions which contain a black-hole or a naked singularity. It would be of interest to relate his results to the naked singularity solutions found by Christodoulou for self-similar collapse. Recently, Gundlach has constructed a solution with discrete self-similarity, which agrees with the critical universal solution found numerically by Choptuik [24].

Similar critical behaviour has also been found in numerical studies of collapse with other forms of matter. Axisymmetric collapse of gravitational waves was shown to have a γ of about 0.36, and $\Delta \simeq 0.6$ [25]. For spherical collapse of radiation (perfect fluid with equation of state $p = \rho/3$) the critical solution has continuous self-similarity, and γ of about 0.36 [26]. However it has become clear now that the critical exponent γ is not independent of the choice of matter. A study of collapse for a perfect fluid with an equation of state $p = k\rho$ shows that γ depends on k [27]. For a given form of matter, there appears to be a unique γ, but the value changes as the form of T_{ik} is changed. Further studies of critical behaviour are reported in [28].

The models described in this section exhibit a naked-singularity like behaviour for near-critical solutions - such solutions are presumably of measure zero on the space of all solutions. In the supercritical region the collapse is said to lead to the formation of a black-hole. This raises a question as to how these results, say the supercritical solutions for radiation fluid collapse, are consistent with those of Ori and Piran, who do find naked singularities. (The Ori-Piran naked singularity lies in the supercritical region). It maybe that when one numerically finds a singularity at the center $r = 0$ one is not easily able to tell whether this is a black-hole or a naked singularity, and this may have to be investigated further.

It is perhaps also relevant to note that collapse of real stars which proceed to become singular is expected to be described by supercritical solutions. Thus the naked singularity observed near the critical region, while of major theoretical interest, may not have astrophysical implications. This further emphasizes the need for investigating whether the singularity being observed numerically in the supercritical region is necessarily covered by the horizon, or could be naked.

It is undoubtedly true that these studies of critical behaviour have opened up an entirely new aspect of gravitational collapse and the related issue of censorship. Obtaining a theoretical understanding of these numerically observed phenomena is an important open problem.

2.4 Spherical collapse with general form of matter

One finds a certain degree of similarity in the collapse behaviour of dust, perfect fluids and scalar fields - in all cases some of the initial distributions lead to black holes, and other distributions lead to naked singularities. This would suggest an underlying pattern which is probably characterized, not by the form of matter, but by some invariants of the gravitational field. The role of matter may simply be that of determining which part of the parameter space these invariants lie in. Hence studies of collapse which put no restriction on T_{ik} apart from an energy condition should prove useful (still maintaining spherical symmetry).

An interesting attempt in this direction has been made by Dwivedi and Joshi in [29], where they generalized their earlier work on dust collapse and self-similar fluids. They assumed a general T_{ik} obeying the weak energy condition, and also that the collapsing matter forms a curvature singularity at $r = 0$ (the central singularity). As we noted earlier, in the comoving coordinate system, matter is described by its energy density and the radial and tangential pressures. Along with these three functions, three functions describing the metric enter a set of five Einstein equations, which are augmented with an equation of state in order to close the system. The geodesic equation for radial null geodesics is written in the limit of approach to the singularity, and it is shown that the occurrence of a visible singularity is equivalent to the occurrence of a positive real root for the geodesic equation, suitably written. Since this equation depends on free initial data, it follows that for a subset of the initial data there will be positive real roots and the singularity will be visible.

This approach needs to be pursued further in order to find out whether or not the naked singularities are generic. Also, it is of interest to work out as to exactly which kind of initial data lead to naked singularities. These are difficult problems, in the absence of known exact solutions. Another interesting attempt at treating spherical collapse without restricting T_{ik} is due to Lake [30]. He concluded that a visible central singularity could form if the mass function in the neighborhood of the singularity satisfies certain conditions. The relation of these conditions with the initial data is not yet apparent.

2.5 Properties of naked singularities

It is evident that energy conditions by themselves do not restrict the occurrence of naked singularities. One would then like to examine in some detail properties of these naked singularities, so as to see if these properties might contain clues for a censorship hypothesis. We review below some of the important features of the naked singularities found in various models.

Curvature strength:

When a collapsing star develops a curvature singularity, the energy density also becomes singular. However, finite physical volumes may or may not be crushed to zero volume as the singularity is approached. This could be used as a criterion for judging the physical seriousness of the singularity, and also for the possible extendibility of spacetime through the singularity [31]. We call a singularity a *strong* curvature singularity if collapsing volume elements do get crushed to zero at the singularity, and a *weak* curvature singularity if they do not. (The terms *weak* and *strong* singularity are sometimes used in the literature with a different meaning. We use them here in Tipler's sense). It is believed that spacetime cannot be extended through a strong singularity, but is possibly extendible through a weak one. A rigorous proof for this is not yet available but is being attempted by some relativists (for detailed studies see [32]). Clarke and Krolak [33] gave necessary and sufficient quantitative criteria for the singularity to be strong, in terms of the rate of growth of curvature along outgoing geodesics, as the singularity is approached.

A strong naked singularity is regarded as a more serious violation of censorship as compared to a weak one. For instance, the shell-crossing type singularities are gravitationally weak [10]. Newman [34] studied the naked central singularity in Tolman-Bondi dust collapse for a wide class of initial data and showed it to be weak. On this basis it was conjectured in [34] that strong naked singularities do not occur in collapse. It was however shown by various people [35] that inclusion of initial data not considered by Newman gives rise also to strong central naked singularities in dust. Interestingly, it has recently become clear that the initial data leading to these strong naked singularities is non-generic, whereas the data leading to a weak naked singularity is generic [36]. In this sense Newman's conjecture does hold for dust collapse! But strong naked singularities have been found in other models - for instance in the naked singularities in the Vaidya spacetime, where they arise from generic initial data. They were also found by Ori and Piran in their study of the self-similar perfect fluid. No results on strength seem to be known for scalar collapse. The general T_{ik} studied by Dwivedi and Joshi would lead to a strong naked singularity for some initial data - however the genericity of such initial data is an open issue. Thus the generality of strong naked singularities remains unclear and it still might be possible to formulate a censorship hypothesis along the lines of Newman's conjecture.

Are naked singularities massless?:

In all known examples of naked singularities, the mass of the collapsing object (well-defined in spherical symmetry, with a vacuum exterior) is found to be zero at the point where the singularity forms. There is evidence that this is a general property of naked singularities in spherical collapse [30]. On the other hand the black hole singularity is always found to be massive. Since a massless singularity might be thought of as having no associated gravitational field, this has led to the suggestion that such singularities do not violate censorship. Note however that even from this *massless* naked singularity entire families of geodesics emerge, and it is not clear whether it is the mass or the outgoing geodesics which are a more important property of the naked singularity!

Redshift:

In known examples of naked singularities for dust and perfect fluids, the redshift along outgoing geodesics emerging from the singularity is found to be infinite (when calculated for observers in the vacuum region). This could be interpreted to mean that no "information" is being transmitted from the naked singularity and could be yet another approach to preserving censorship.

Stability and Genericity:

This of course is the most important issue relating to the naked singularity examples, and a notion of stability is hard to define. The most direct definition of stability (equivalently, genericity) of naked singularities would be simply as follows. If a solution of Einstein equations describing collapse leading to a naked singularity has as many free functions as required for arbitrary initial data, the solution is stable. (One is reminded here of the methods adopted by Belinskii, Lifshitz and Khalatnikov to show that general solutions of Einstein equations contain singularities). Of course progress in such a broad sense is hopelessly difficult, and one talks of stability of a given solution under specific kinds of perturbations. For instance, one would consider stability of the solution against change of initial data, against change of equation of state, and against non-spherical metric perturbations. From these viewpoints, very little is known about the stability of the naked singularity models mentioned in this article. (It is important to note that equally little is known about the stability of the black-holes which form in these models during collapse! Various studies show that the event horizon is stable to small perturbations [37], and hence the singularity is stable, but it could be either naked or covered.)

One very useful way to address stability of naked singularities is to study the blue-shift instability of the Cauchy horizon. One considers ingoing waves starting from null infinity, as they approach the Cauchy horizon. If they develop an infinite blue-shift along the Cauchy horizon, this in some sense is like saying this horizon would be 'destroyed' and the spacetime region beyond, which is exposed to the naked singularity, would no longer be accessible. Hence predictability will be preserved in the observable spacetime region. Interestingly enough, it has been found that the Cauchy horizon in the dust and perfect fluid examples does not have a blue-shift instability.

Quantum effects:

The censorship hypothesis as such is concerned with the nature of singularities in classical general relativity. However, even if naked singularities do occur in the classical theory, one could ask if their formation would be avoided when quantum effects in their vicinity are taken into account. This would be a quantum cosmic censorship. Some investigations have taken place in this direction [38, 39] and this very interesting question deserves to be pursued further. Essentially the idea is to repeat for a naked singularity the kind of calculation Hawking carried out to show that quantum effects cause black-holes to radiate. Since regions of very high curvature are exposed near the naked singularity, intense particle production can be expected. It is typically found in these calculations that as a result of the produced particles, the energy-momentum flux at infinity diverges - in spite of the fact that the naked singularity is massless and the classical outgoing geodesics have infinite redshift! Although back-reaction calculations are hard to carry out, the infinite flux would suggest that the naked singularity formation will be avoided. From the quantum viewpoint, naked singularities appear to be explosive events, and the outgoing flux might be their only possible observational signature. It is worth studying the properties of this flux in detail to understand what observations, if any, can detect naked singularities if they do occur in nature.

Thus we find that properties like curvature strength, masslessness, redshift, blue-shift instability, and quantum effects give a mixed sort of picture regarding the significance of these naked singularity examples. An optimistic assessment of this situation is that there is a good deal of richness in the problem, and much to think about, before we can decide one way or the other.

3 Non-spherical gravitational collapse

As we have seen, there are examples of naked singularities in spherical collapse. By assuming that the evolution can be continued beyond the Cauchy horizon, one can conclude that the collapsing star will eventually shrink below its Schwarzschild radius and the event horizon will form (according to the infalling observer). There is also evidence that the horizon is stable to small perturbations. However, if there are large departures from spherical symmetry, the picture could be different, and the horizon may not form at all. The naked singularity so forming would qualitatively be of a different kind, compared to the ones seen in spherically symmetric spacetimes.

Our knowledge of exact solutions in the non-spherical case is inevitably even more limited than for spherical systems, and one must again resort to introducing some symmetry. An important early study was due to Thorne [40], and was motivated by the work of Lin, Mestel and Shu [41] on the collapse of Newtonian spheroids. Thorne examined the collapse of an infinite cylinder and showed that it develops a curvature singularity without an event horizon forming. Considerations such as these led him to propose what came to be known as the hoop conjecture, which he stated as follows [40]:

- Horizons form when and only when a mass M gets compacted into a region whose circumference in EVERY direction is C $\lesssim 4\pi M$.

According to the conjecture, collapsing objects which become so asymmetric as to attain a circumference which is greater than the bound will not develop horizons; hence if a singularity forms it will be naked. We note that even if the conjecture holds, a naked singularity can form, (as it does sometimes in spherical collapse), but an event horizon will also form. One could say that the naked singularities which form when the hoop conjecture holds are of a less serious nature than those which form when the conjecture does not hold. In the former case an infalling observer cannot communicate with asymptotic observers after crossing the horizon, while in the latter case no such restriction arises.

Important numerical simulations were carried out by Shapiro and Teukolsky [42] to test the hoop conjecture. They studied the gravitational collapse of homogeneous non-rotating oblate and prolate spheroids of collisionless gas, starting from rest. Maximal slicing is used, and the evolution of the matter particles is followed with the help of the Vlasov equation in the self-consistent gravitational field. The development of a singularity is detected by measuring the Riemann invariant at every point on the spatial grid. Since an event horizon can be observed only by tracking null rays indefinitely, they instead search for the formation of an apparent horizon (the boundary of trapped surfaces) - the existence of the apparent horizon can be determined locally. If a spacetime region has an apparent horizon, it will also have an event horizon, to its outside.

They found that oblate spheroids first collapse to thin pancakes, but then the particles overshoot and ultimately the distribution becomes prolate and collapses to a thin spindle. An apparent horizon develops to enclose the spindle, which eventually becomes singular, and a black hole is formed. The minimum exterior circumference in the polar and equatorial directions is consistent with the requirements of the hoop conjecture. The collapse of prolate spheroids leads however to the formation of a spindle singularity, with no evidence for an apparent horizon covering the singularity. The initial dimensions of the spheroid are such that the minimum circumference, at the time of formation of the singularity, exceeds $4\pi M$. The collapse of smaller prolate spheroids leads to spindle singularities that are covered by a horizon, again favouring the conjecture. Shapiro and Teukolsky suggested, noting the absence of an apparent horizon for large prolate configurations, that the resulting spindle singularity is naked, and hence that the hoop conjecture holds.

They were also careful to point out that the absence of an apparent horizon up until the time of singularity formation does not necessarily imply the absence of an event horizon. Hence, strictly one could not conclude that the singularity is necessarily naked. Wald and Iyer [43] showed this mathematically with an example - they demonstrated that Schwarzschild spacetime can be sliced with nonspherical slices, which approach arbitrarily close to the singularity, but do not have any trapped surfaces. Another analytical example showing that absence of apparent horizon does not imply the singularity is naked can be found in [36], where inhomogeneous spherical dust collapse was studied. However, in support of their conclusion Shapiro and Teukolsky pointed out that null rays continue to propagate away from the region of the singularity until when the simulations are terminated, and that the formation of an event horizon is unlikely. It is perhaps fair to conclude that while their numerical simulations are of major importance and their results suggestive, further investigations are necessary to settle the issue. An analytical demonstration analogous to these simulations was worked out in [44].

Another interesting analytical example is the quasi-spherical dust solution due to Szekeres, which also admits naked singularities [45], including those having strong curvature [46].

4 Discussion

We now attempt a critical comparison of the results reviewed here, and discuss their implications for the censorship hypothesis. (For other recent reviews of cosmic censorship see Clarke [4] and Joshi [47]). Let us begin with a quick summary, even though it amounts to repetition.

Very massive stars are expected to end their gravitational collapse in a singularity. There has been around an unproven conjecture that the singularity will be hidden behind an event horizon, and hence such stars will become black-holes. If the conjecture is false, some stars can end up as naked singularities - this will have major implications for black-hole physics and astrophysics, and for classical general relativity. Since a proof for the conjecture has not been forthcoming, relatively modest attempts have been made to study specific examples of gravitational collapse. These studies, which so far have been mostly for spherical collapse, have thrown up some surprises.

The collapse does not always end in a black-hole; for some initial data it ends in a naked singularity, and this is true for various forms of matter.

The spherical gravitational collapse of inhomogeneous dust leads to weak naked singularities for generic initial data; the strong naked singularities which do form for some data are non-generic. The naked singularity, irrespective of whether it is weak or strong, is massless. The outgoing geodesics have an infinite redshift, and the Cauchy horizon does not have a blue-sheet instability. The collapse of a self-similar perfect fluid also exhibits strong curvature naked singularities for some initial data - these are again massless, have infinitely redshifted outgoing geodesics, and the Cauchy horizon is stable. Numerical studies of scalar field collapse suggest that the critical solution is a naked singularity; elsewhere in the data space the collapse ends either in dispersal or a singularity. It seems unclear as to whether this singularity is definitely covered, or could be naked. For collapse of a general form of matter, there is an existence proof that for some of the data the singularity will be naked and strong - its genericity is an open issue. There appear to be no conclusive studies on non-spherical perturbations of these solutions, or on non-spherical collapse as such - the simulations of Shapiro and Teukolsky are however an important progress in this direction.

A broad conclusion is that at this early stage in collapse studies, one does not have enough evidence to take a grand decision about the validity of the censorship hypothesis, one way or the other. Further, it does not appear very useful right now to try and prove a specific version of the hypothesis. Consider, just as an example, the following proposal:

- Gravitational collapse of *physically reasonable matter* starting from *generic initial data* leads to the formation of a black-hole or a naked singularity. The naked singularity is massless and gravitationally weak, the Cauchy horizon does not have a blue-shift instability and the redshift along outgoing geodesics is infinite.

This does not appear to be a very interesting proposal to prove, given the number of properties attached to the naked singularity, nor is it certain that it will survive further studies of collapse models. It is not even clear whether this proposal proves or disproves cosmic censorship!

What is noteworthy however is that naked singularities *do* occur in dynamical collapse, side by side with black-hole solutions, so to say. In order to address the censorship hypothesis, one has to assess the significance of their properties, and also generalise the models studied so far. There are perhaps three important questions: (i) what is the role of the form of matter? (ii) are the naked singularities genuine features of the spacetime geometry? (iii) do they have observational effects? We respond to these questions briefly.

The form of matter in these models is dust, a perfect fluid, a scalar field, or where an existence proof has been given, a general T_{ik} obeying the weak energy condition. There are quite a few views on using dust as a form of matter in these studies, which we try to enumerate. Firstly, since dust collapse can give rise to singularities even in Newtonian gravity or in special relativity, it is said that the (naked) singularities being observed in general relativity have nothing to do with gravitational collapse. This view appears acceptable for shell-crossing and the weak shell-focussing naked singularities. But it is difficult to accept it for the strong naked singularities which crush physical volumes to zero and hence ought to be a genuine general relativistic feature. Secondly, the evolution of collisionless matter is described by the Einstein-Vlasov equations; at any given point in space the particles have a distribution of momenta. Dust is a very special case of these equations, defined by the assumption that all particles have exactly the same momentum. It has been shown [48] that the spherically symmetric Einstein-Vlasov system has global solutions which do not contain singularities, naked or otherwise, and hence censorship is preserved. However, we recall that dust collapse itself has a rich structure, admitting both black-holes and naked singularities, a variety in trapped surface dynamics, and of course includes the classic Oppenheimer-Snyder model of black-hole formation. It would be a little surprising if these dust features turn out to have no connection at all with more realistic collapse models, which do have singularities. Thirdly, it has been suggested, though certainly not universally accepted, that during late stages of collapse matter will effectively behave like dust. In my view, a useful attitude towards the dust collapse results is to treat them as a learning exercise and see if they will survive when more general forms of matter are considered.

Quite naturally, perfect fluids with pressure get more serious consideration than dust. Yet, the naked singularities found in their collapse can be objected to by saying that a fluid description is phenomenological and not fundamental. This objection has been weakened by the results showing naked singularities in scalar field collapse, and also by the existence proofs for naked singularities with a general form of T_{ik}. Also, the assumption of self-similarity made in the explicit examples given by Ori and Piran [13] could be considered as restrictive. The existence proofs by Dwivedi and Joshi relax this assumption. It appears safe to conclude at this stage that the form of matter is not crucial in the examples of naked singularities that have been found so far.

The second question deserves a more serious consideration, and holds the key to the validity of the censorship hypothesis. Are these naked singularities genuine features of the spacetime geometry? What is the relative importance of the properties like masslessness of the naked singularity, curvature strength, redshift along outgoing geodesics, and instability of the Cauchy horizon? Do one or more of these features establish that the naked singularity is not genuine? There is a need to develop what one might call the theory of naked singularities, in order to answer a difficult and important question of this sort. Stability against non-spherical perturbations will also help decide the significance of these examples. The singularity theorems showed that gravitational singularities are not restricted to spherical symmetry; if this is any guide then one would expect both the black-hole as well as naked singularity solutions to persist when sphericity is relaxed. It is indeed difficult to visualize how naked singularities will convert themselves to black-holes when asymmetry is introduced.

The third question, as to whether naked singularities might have observational effects, also deserves attention in view of the examples now available. In spite of limited theoretical evidence for their formation, black-holes have been successfully used to model many observed astrophysical processes. One did not have to wait for the censorship hypothesis to be proved before applications of black-holes could begin. In a similar vein, one ought to give naked singularities a chance, so as to examine if they will or will not have a significant flux emission, due to quantum effects or otherwise.

While there is no clearly defined line of attack for further investigations of the censorship hypothesis, two specific approaches appear to hold promise. Firstly, the methods used by Belinskii, Lifshitz and Khalatnikov to prove the generality of singularities [49] involve construction of solutions of Einstein equations near spacetime singularities. Perhaps one could study propagation of light rays using these solutions, to investigate if the singularities could be naked. Secondly, major advances in numerical relativity have made the subject ripe for studies on the censorship hypothesis. For instance, it should be possible to check numerically if the naked singularities of spherical inhomogeneous dust collapse persist if the collapse is non-spherical. Or, to check for naked singularities in spherical perfect fluid collapse, when the self-similarity assumption is relaxed.

How does a very massive star evolve during the final stages of its collapse? Does it choose to hide behind the event horizon to die a silent death, or does it explode dramatically, exposing the singularity, as if the Big Bang was being reenacted for our benefit? Recent developments in our understanding of classical general relativity leave room for both possibilities. Further studies on gravitational collapse should prove to be exciting, and it remains to be seen whether naked singularities will come to play a role in physics and astrophysics.

Acknowledgements

It is a pleasure to thank Patrick Brady, Fred Cooperstock, I. H. Dwivedi, Sanjay Jhingan, Pankaj Joshi, Kerri Newman, Amos Ori, Ed Seidel, Peter Szekeres, C. S. Unnikrishnan, Cenalo Vaz, Shwetketu Virbhadra and Louis Witten for helpful discussions. I would also like to thank the Institute of Mathematical Sciences, Madras for hospitality.

References

[1] R. Penrose (1965) *Phys. Rev. Lett.*, **14**, 57; S. W. Hawking (1967) *Proc. Roy. Soc. Lond. A*, **300**, 187; S. W. Hawking and Penrose (1970) *Proc. Roy. Soc. Lond. A*, **314**, 529

[2] R. Penrose (1969) *Rivista del Nuovo Cimento*, **1**, 252

[3] R. Penrose (1972) *Nature*, **236**, 377; (1974) *Ann. New York Acad. Sci.*, **224**, 125

[4] C. J. S. Clarke (1993) *Class. Quantum Grav.*, **10**, 1375

[5] S. W. Hawking and G. F. R. Ellis (1973) The large-scale structure of space-time, Section 4.3 (Cambridge University Press)

[6] R. C. Tolman (1934) *Proc. Nat. Acad. Sci., USA*, **20**, 169; H. Bondi (1947) *Mon. Not. Astron. Soc.*, **107**, 410

[7] J. R. Oppenheimer and H. Snyder (1939) *Phys. Rev.*, **56**, 455

[8] P. Yodzis, H.-J Seifert and H. Müller zum Hagen (1973) *Commun. Math. Phys.*, **34**, 135; (1974) *Commun. Math. Phys.*, **37**, 29

[9] D. M. Eardley and L. Smarr (1979) *Phys. Rev. D*, **19**, 2239; D. Christodoulou (1984) *Commun. Math. Phys.*, **93**, 171; R. P. A. C. Newman, (1986) *Class. Quantum Grav.*, **3**, 527; B. Waugh and K. Lake (1988) *Phys. Rev. D*, **38** 1315; V. Gorini, G. Grillo and M. Pelizza, (1989) *Phys. Lett. A*, **135**, 154; G. Grillo (1991) *Class. Quantum Grav.*, **8**, 739; R. N. Henriksen and K. Patel (1991) *Gen. Rel. Gravn.*, **23**, 527; I. H. Dwivedi and S. Dixit (1991) *Prog. Theor. Phys.*, **85**, 433; I. H. Dwivedi and P. S. Joshi (1992) *Class. Quantum Grav.*, **9**, L69; P. S. Joshi and I. H. Dwivedi (1993) *Phys. Rev.*, **D47**, 5357; P. S. Joshi and T. P. Singh (1995) *Phys. Rev.*, **D51** 6778; T. P. Singh and P. S. Joshi (1996) *Class. Quantum Grav.*, **13**, 559; C. S. Unnikrishnan (1996) *Phys. Rev.*, **D53**, R580; H. M. Antia, (1996) *Phys. Rev.*, **D53**, 3472; Sanjay Jhingan, P. S. Joshi and T. P. Singh (1996) gr-qc/9604046; T. P. Singh (1996) The central singularity in spherical dust collapse - a review, in preparation

[10] R. P. A. C. Newman, Ref. 9 above

[11] C. W. Misner and D. H. Sharp (1964) *Phys. Rev.*, **136**, B571

[12] E. M. Lifshitz and I. M. Khalatnikov (1961) *Soviet Physics JETP*, **12**, 108; M. A. Podurets (1966) *Soviet Physics - Doklady*, **11**, 275

[13] A. Ori and T. Piran (1990) *Phys. Rev. D*, **42**, 1068

[14] P. S. Joshi and I. H. Dwivedi (1992) *Commun. Math. Phys.*, **146**, 333

[15] P. S. Joshi and I. H. Dwivedi (1993) *Lett. Math. Phys.*, **27**, 235

[16] P. Szekeres and V. Iyer (1993) *Phys. Rev.*, **D47**, 4362

[17] P. C. Vaidya (1943) *Current Science*, **13**, 183; (1951) *Physical Review*, **83**, 10; (1951) *Proc. Indian Acad. Sci.*, **A33**, 264

[18] W. A. Hiscock, L. G. Williams and D. M. Eardley (1982) *Phys. Rev.*, **D26**, 751; Y. Kuroda (1984) *Prog. Theor. Phys.*, **72**, 63; A. Papapetrou (1985) in A Random Walk in General Relativity, Eds. N. Dadhich, J. K. Rao, J. V. Narlikar and C. V. Vishveshwara (Wiley Eastern, New Delhi); G. P. Hollier (1986) *Class. Quantum Grav.*, **3**, L111; W. Israel (1986) *Can. Jour. Phys.*, **64**, 120; K. Rajagopal and K. Lake (1987) *Phys. Rev.*, **D35**, 1531; I. H Dwivedi and P. S. Joshi (1989) *Class. Quantum Grav.*, **6**, 1599; (1991) *Class. Quantum Grav.*, **8**, 1339; P. S. Joshi and I. H. Dwivedi (1992) *Gen. Rel. Gravn.*, **24**, 129; J. Lemos (1992) *Phys. Rev. Lett.*, **68**, 1447

[19] D. Christodoulou (1986) *Commun. Math. Phys.*, **105**, 337; **106**, 587; (1987) **109**, 591; **109**, 613; (1991) *Commun. Pure Appl. Math.*, **XLIV**, 339; (1993) **XLVI**, 1131; (1994) *Ann. Math.*, **140**, 607

[20] M. W. Choptuik (1994) in Deterministic Chaos in General Relativity, Eds. D. Hobill, A. Burd and A. Coley (Plenum, New York)

[21] M. W. Choptuik (1993) *Phys. Rev. Lett.*, **70**, 9

[22] C. Gundlach, J. Pullin and R. Price (1994) *Phys. Rev.*, **D49**, 890; D. Garfinkle (1995) *Phys. Rev.*, **D51**, 5558; R. S. Hamade and J. M. Stewart (1996) *Class. and Quantum Grav.*, **13**, 497

[23] M. D. Roberts (1989) *Gen. Rel. Gravn.*, **21**, 907; V. Husain, E. Martinez and D. Nunez (1994) *Phys. Rev.*, **D50**, 3783; J. Traschen (1994) *Phys. Rev.*, **D50**, 7144; Y. Oshiro, K. Nakamura and A. Tomimatsu (1994) *Prog. Theor. Phys.*, **91**, 1265; P. R. Brady (1994) *Class. Quantum Grav.*, **11**, 1255; (1995) *Phys. Rev.*, **D51**, 4168

[24] C. Gundlach (1995) *Phys. Rev. Lett.*, **75**, 3214; (1996) gr-qc/9604019

[25] A. M. Abrahams and C. R. Evans (1993) *Phys. Rev. Lett.*, **70**, 2980

[26] C. R. Evans and J. S. Coleman (1994) *Phys. Rev. Lett.*, **72**, 1782

[27] D. Maison (1994) gr-qc/9504008

[28] E. W. Hirschmann and D. M. Eardley (1995) *Phys. Rev.*, **D51**, 4198; (1995) gr-qc/9506078; (1995) gr-qc/9511052; D. M. Eardley, E. W. Hirschmann and J. H. Horne (1995) gr-qc/9505041; R. S. Hamade, J. H. Horne and J. M. Stewart (1995) gr-qc/9511024

[29] I. H. Dwivedi and P. S. Joshi (1994) *Commun. Math. Phys.*, **166**, 117

[30] K. Lake (1992) *Phys. Rev. Lett.*, **68**, 3129

[31] F. J. Tipler (1977) *Phys. Lett.*, **A64**, 8; F. J. Tipler, C. J. S. Clarke and G. F. R. Ellis (1980) in General Relativity and Gravitation (Vol. 2), Ed. A. Held (Plenum, New York)

[32] C. J. S. Clarke (1993) Analysis of spacetime singularities (Cambridge University Press)

[33] C. J. S. Clarke and A. Krolak (1986) *J. Geo. Phys.*, **2**, 127

[34] R. P. A. C. Newman, Ref. [9] above

[35] V. Gorini et al.; G. Grillo; B. Waugh and K. Lake; I. H. Dwivedi and P. S. Joshi; T. P. Singh and P. S. Joshi; Ref. [9] above

[36] Sanjay Jhingan, P. S. Joshi and T. P. Singh, Ref. [9] above

[37] A. G. Doroshkevich, Ya. B. Zeldovich and I. Novikov (1966) *Soviet Physics JETP*, **22**, 122; V. de la Cruz, J. E. Chase and W. Israel (1970) *Phys. Rev. Lett.*, **24**, 423; R. H. Price (1972) *Phys. Rev.*, **D5**, 2419

[38] L. H. Ford and L. Parker (1978) *Phys. Rev.*, **D17**, 1485; W. A. Hiscock et al. Ref. 18 above

[39] C. Vaz and L. Witten (1994) *Phys. Lett.*, **B325**, 27; (1995) *Class. Quantum Grav.*, **12**, 2607; (1995) gr-qc/9511018; (1996) gr-qc/9604064

[40] K. S. Thorne (1972) in Magic Without Magic: *John Archibald Wheeler*, Ed. John R. Klauder, (W. H. Freeman and Co., San Francisco)

[41] C. C. Lin, L. Mestel and F. H. Shu (1965) *Ap. J.*, **142**, 1431

[42] S. L. Shapiro and S. A. Teukolsky (1991) *Phys. Rev. Lett.*, **66**, 994; Phys. Rev. (1992) **D45** 2006.

[43] R. M. Wald and V. Iyer (1991) *Phys. Rev.*, **D44**, 3719

[44] C. Barrabès, W. Israel and P. S. Letelier (1991) *Phys. Lett.*, **A160**, 41

[45] P. Szekeres (1975) *Phys. Rev.*, **D12**, 2941

[46] P. S. Joshi and A. Krolak (1996) gr-qc/9605033

[47] P. S. Joshi (1993) Global Aspects in Gravitation and Cosmology, Clarendon Press, OUP, Oxford, Chapters 6 and 7

[48] A. D. Rendall (1992) *Class. Quantum Grav.*, **9**, L99; (1992) in Approaches to Numerical Relativity Ed. R. d'Inverno (Cambridge University Press); G. Rein and A. D. Rendall (1992) *Commun. Math. Phys.*, **150**, 561; G. Rein, A. D. Rendall and J. Schaeffer (1995) *Commun. Math. Phys.*, **168**, 467

[49] V. A. Belinskii, J. M. Khalatnikov and E. M. Lifshitz, (1970) *Advances in Physics*, **19**, 525

In my entire scientific life, extending over forty-five years, the most shattering experience has been the realization that an exact solution of Einstein's equations of general relativity, discovered by the New Zealand mathematician, Roy Kerr, provides the absolutely exact representation of untold numbers of massive black holes that populate the universe. This "shuddering before the beautiful", this incredible fact that a discovery motivated by a search after the beautiful in mathematics should find an exact replica in Nature, persuades me to say that beauty is that to which the human mind responds at its deepest and most profound. Indeed, everything I have tried to say in this connection has been stated more succintly in the Latin mottos:

The Simple is the seal of the true
Beauty is the splendour of truth

— S. CHANDRASEKHAR

ASPECTS OF ACCRETION PROCESSES ON A ROTATING BLACK HOLE

Sandip K. Chakrabarti[1]

Tata Institute of Fundamental Research
Homi Bhabha Road
Mumbai 400 005

Abstract

We describe the most general nature of accretion and wind flows around a compact object and emphasize on the properties which are special to *black hole* accretion. The angular momentum distribution in the most general solution is far from Keplerian, and the non-Keplerian disks can include standing shock waves. We also present fully time dependent numerical simulation results to show that they agree with these analytical solutions. We describe the spectral properties of these accretion disks and show that the soft and hard states of the black hole candidates could be explained by the change of the accretion rate of the disk. We present fits of the observational data to demonstrate the presence of sub-Keplerian flows around black holes.

[1]e-mail: chakraba@tifrc2.tifr.res.in

The award of Nobel Prize carries with it so much distinction and the number of competing areas and discoveries are so many, that it must of necessity have a sobering effect on an individual who receives the prize. For who will not sobered by the realization that among the past laureates there are some who have achieved a measure of insight into nature that is far beyond the attainment of most? But I am grateful for the award since it is possible that it may provide a measure of encouragement to those, who like myself, have been motivated in their scientific pursuits, principally, for achieving personal perspectives, while wandering, mostly, in the lonely byways of science. When I say personal perspectives, I have in mind the players in Virginia Woolf's 'The Waves'.

There is a square; there is an oblong. The players take the square and place it upon the oblong. They place it very accurately; they make a perfect dwelling place. Very little is left outside. The structure is now visible; what was inchoate is here stated; we are not so various or so mean; we have made oblong and stood them upon squares. This is our triumph; this is our consolation.

— S. CHANDRASEKHAR

1 Introduction

Centers of galaxies are believed to be the seats of massive black holes ($M_{BH} \sim 10^{6-9} M_{\odot}$) and some evolved compact binary systems are believed to be the seats of small mass black holes ($M_{BH} \sim 3 - 10 M_{\odot}$). In the 1970s, the standard accretion disk models around black holes were constructed (Shakura & Sunyaev 1973; Novikov & Thorne 1973) assuming Keplerian angular momentum distribution in the orbiting matter. In the early 1980s, these disks became the favorite candidates for the explanation of the "big blue bump" seen in the UV region of the continuum spectra of active galaxies (Malkan 1982; Sun & Malkan 1989) with some success.

However, from X-ray and γ-ray spectra and line emissions from these objects (see, Chakrabarti, 1996a for a general review), it is becoming clear that the accretion disks cannot be simple Keplerian type nor are as simple as spherically symmetric Bondi flows (Bondi, 1952). To satisfy the inner boundary condition on the horizon, matter must cross the horizon supersonically (Chakrabarti 1990, hereafter C90) and therefore must pass through a sonic point where the radial velocity of the flow is the same as the sound velocity. This means that the angular momentum must deviate from a Keplerian flow. All these considerations require one to solve the most general set of equations which must include effects of rotation, flow pressure, viscosity, advection, heating and cooling processes. Below, we present these equations and discuss the solutions and their implications. For the purpose of the present review we emphasize those properties which are *special* to a black hole accretion.

We consider a perfect flow around a Kerr black hole with the metric (e.g., Novikov & Thorne, 1973; using t, r, θ & z as coordinates and the units which render $G = M_{BH} = c = 1$).

$$ds^2 = -\frac{r^2\Delta}{A}dt^2 + \frac{A}{r^2}(d\phi - \omega dt)^2 + \frac{r^2}{\Delta}dr^2 + dz^2, \tag{1}$$

where,

$$A = r^4 + r^2a^2 + 2ra^2,$$

$$\Delta = r^2 - 2r + a^2,$$

$$\omega = \frac{2ar}{A}.$$

Here, a is the Kerr parameter.

The stress-energy tensor of a perfect fluid with pressure p and mass density $\rho = \rho_0(1+\pi)$, π is the internal energy is give by,

$$T_{\mu\nu} = \rho u_\mu u_\nu + p(g_{\mu\nu} + u_\mu u_\nu). \tag{2}$$

We shall concentrate on the time independent solution of the governing equations (Chakrabarti 1996b, hereafter C96b). The equation for the balance of the radial momentum is obtained from $(u_\mu u_\nu + g_{\mu\nu})T^{\mu\nu}_{;\nu} = 0$. Here the advection term due to significant radial velocity is included. The baryon number conservation equation (continuity equation) is obtained from $(\rho_0 u^\mu)_{;\mu} = 0$. The conservation of angular momentum is obtained from $(\delta^\phi_\mu T^{\mu\nu})_{;\nu} = 0$. Here the angular momentum is allowed to be non-Keplerian. Entropy generation equation is obtained from the first law of thermodynamics along with baryon conservation equation: $(S^\mu)_{;\mu} = \frac{1}{T}[2\eta\sigma_{\mu\nu}\sigma^{\mu\nu}]$, where, T is the temperature of the flow and η is the coefficient of viscosity. $\sigma_{\mu\nu}$ is the shear tensor which is responsible for the viscous transport of angular momentum.

These set of equations are solved simultaneously along with the possibility that the shock waves may form in the flow, where, the following Rankine-Hugoniot conditions must be fulfilled:

$$W_- n^\nu + (W_- + \Sigma_-)(u^\mu_- n_\mu)u^\nu_- = W_+ n^\nu + (W_+ + \Sigma_+)(u^\mu_+ n_\mu)u^\nu_+.$$

Here, n_μ is the four normal vector component across the shock, and W and Σ are vertically integrated pressure and density on the shock surface.

2 Solution Topologies

First note that the above mentioned equations are valid for any compact object whose external spacetime is similar to that of a Kerr black hole. However, on a neutron star surface matter has to stop and corotate with the surface velocity. The inner boundary condition is therefore subsonic. On a black hole, on the other hand, the flow must enter through the horizon with the velocity of light, and therefore must be supersonic. Stationary shock waves may form when matter starts piling up behind the centrifugal barrier (which arises due to centrifugal force $\propto \lambda^2/r^3$). The supersonic flow becomes subsonic at the shock and again becomes supersonic before entering through the horizon. Clearly, the flow has to become supersonic, before forming the shock as well, and therefore pass through another sonic point at a large distance away from the black hole. Thus, as a whole, the flow may deviate from a Keplerian disk and (a) enter through the inner sonic point only, or, (b) enter through the outer sonic point only, or, (c) pass through the outer sonic point, then a shock, and finally through an inner sonic point if the shock conditions are satisfied. If the angular momentum is too small, then the flow has only one sonic point and shocks cannot be formed as in a Bondi flow. For more details, see Chakrabarti (1989, hereafter C89), C90 and C96a.

Three flow parameters govern the topology of the flow: the α parameter (Shakura & Sunyaev, 1973) which determines the viscosity, the location of the inner sonic point r_{in} through which matter must pass through, and the specific angular momentum λ_{in} of the matter at the horizon (or, that at r_{in}). It so happens that these parameters are sufficient to completely determine the solution when a cooling process is provided. To illustrate the flow topologies, we first choose the viscosity of the flow to be negligible (C96b). In the inviscid case, the angular momentum and energy remain constant $hu_\phi = l$ and $hu_t = \mathcal{E}$. Hereafter, we use $\lambda = l/\mathcal{E}$ to be the specific angular momentum. Flow entropy remains constant unless it passes through a shock wave where it goes up within a thin layer.

In Fig. 1 we show *all possible* solutions of weakly viscous accretion flows around a Kerr black hole of rotation parameter $a = 0.5$. (For flows in Schwarzschild geometry, see, C89, C90.) The adiabatic index $\gamma = 4/3$ has been chosen. Vertical equilibrium flow model with the vertical height prescription of NT73 is used. In the central box, we divide the parameter space spanned by $(\lambda, \mathcal{E})$ into nine regions marked by N, O, NSA, SA, SW, NSW, I, O^*, I^*. The horizontal line at $\mathcal{E} = 1$ corresponds to the rest mass of the flow. Surrounding this parameter space, we plot various solution topologies (Mach number $M = v_r/a_s$ vs. logarithmic radial distance where v_r is the radial velocity and a_s is the sound speed) marked with the same notations (except N). Each of these solution topologies has been drawn using flow parameters from the respective region of the central box. Though each contour of each of the boxes represents individual solutions (differing only by specific entropy), the relevant solutions are the ones which are self-crossing, as they are transonic. The crossing points are 'X' type or saddle type sonic points and the contours of circular topology surround 'O' type sonic points. If there are two 'X' type sonic points, the inner one is called the inner sonic point, and the outer one is called the outer sonic point. If there is only one 'X' type sonic point in the entire solution, then the terminology of inner or outer is used according to whether the sonic point is close to or away from the black hole. The solutions from the region 'O' has only the outer sonic point. The solutions from the regions NSA and SA have two 'X' type sonic points with the entropy density S_o at the outer sonic point *less* than the entropy density S_i at the inner sonic point. However, flows from SA pass through a standing shock (See Fig. 2) as the Rankine-Hugoniot condition is satisfied. The entropy generated at the shock is exactly $S_i - S_o$, which is advected towards the hole to enable the flow to pass through the inner sonic point. Rankine-Hugoniot condition is not satisfied for flows from the region NSA. Numerical simulation indicates (Ryu, Chakrabarti & Molteni, 1996) that the solution is very unstable and show periodic changes in emission properties as the flow constantly try to form the shock wave, but fail to do so. The solutions from the region SW and NSW are very similar to those from SA and NSA. However, $S_o \geq S_i$ in these cases. Shocks can form only in winds from the region SW. The shock condition is not satisfied in winds from the region NSW. This may make the NSW flow unstable

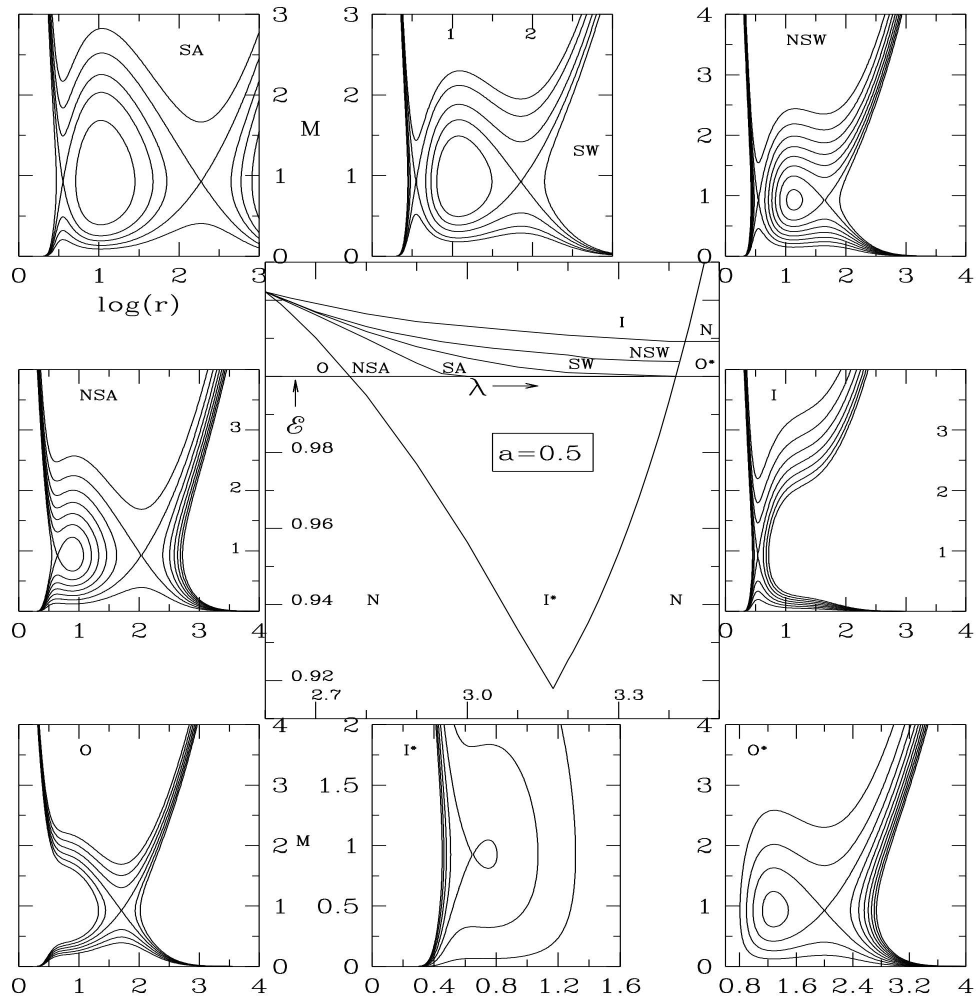


Fig. 1: Classification of the parameter space (central box) in the energy-angular momentum plane in terms of various topology of the black hole accretion. Eight surrounding boxes show the solutions from each of the independent regions of the parameter space. See text for details.

as well. A flow from region I only has the inner sonic point and thus can form shocks (which require the presence of two saddle type sonic points) only if the inflow is already supersonic due to some other physical processes.

The flows from regions I^* and O^* are interesting in the sense that each of them has two sonic points (one 'X' and one 'O') only and neither produces complete and global solution. The region I^* has an inner sonic point but the solution does not extend subsonically to a large distance. The region O^* has an outer sonic point, but the solution does not extend supersonically to the horizon! In both the cases a weakly viscous flow is expected to be unstable. When a significant viscosity is added, the closed topology of I^* is expected to open up (Fig. 3 below; Chakrabarti, 1996c; hereafter C96c) and then the flow can join with a Keplerian disk.

3 Examples of 'Discontinuous' Solutions

In Fig. 1 we presented continuous solutions and mentioned the possibility of shock formation. In Fig. 2(a-b) we give examples of solutions which include shock wave discontinuities. Mach number is plotted along Y-axis and logarithmic radial distance is plotted along X-axis. Fig. 2a is drawn with parameters from the region SA and Fig. 2b is drawn with parameters from the region SW.

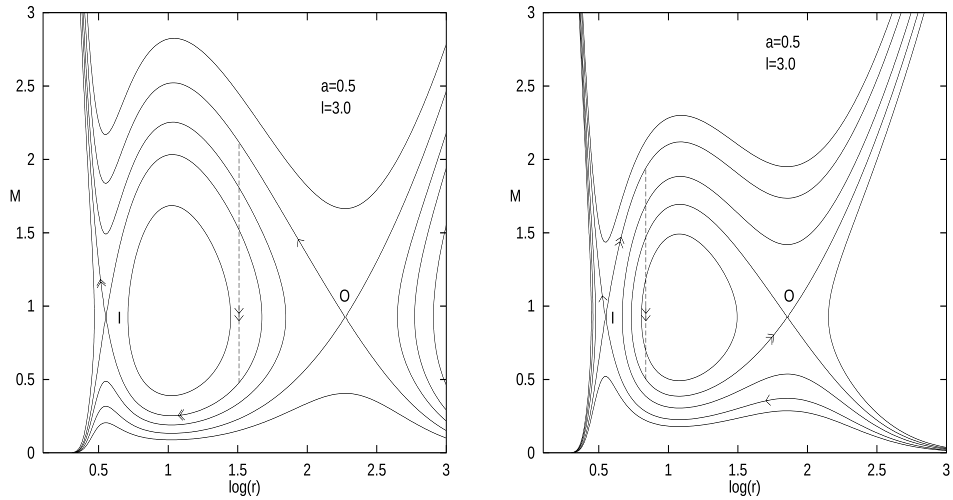


Fig. 2(a-b): Mach number M is plotted against logarithmic radial distance r. Contours are of constant entropy. $a = 0.5$, $\lambda = 3$. In (a), $\mathcal{E} = 1.003$ and in (b), $\mathcal{E} = 1.007$. O and I denote the outer and inner sonic points respectively. Single arrow shows the accretion flow path through O while double arrow traces the path of more stable shocked flow which passes through both I and O.

In Fig. 2a, the single arrowed curve represents a solution coming subsonically from a large distance and becoming supersonic at O, the outer sonic point. Subsequently, the flow jumps onto the subsonic branch (at the place where Rankine-Hugoniot condition is satisfied) along the vertical dashed line and subsequently the flow enters the black hole through the inner sonic point at I (double arrowed curve). The flow chooses to have a shock since the inner sonic point has a higher entropy. The parameters of the flow are $a = 0.5$, $\mathcal{E} = 1.003$ and $\lambda = 3$. In Fig. 2b, on the other hand, the accretion flow can straightaway pass through the inner sonic point (single arrowed curve) and will have no shocks. However, outgoing winds, which may be originated closer to the black hole can have a shock discontinuity (double arrowed curves). Here the flow first passes through I, the shock (vertical dashed line) and the outer sonic point 'O'. The parameters in the case are $a = 0.5$, $\mathcal{E} = 1.007$ and $\lambda = 3$.

Above mentioned discussions are valid for flows around a black hole only. For flows around a neutron star, the inner boundary condition must be subsonic, and therefore, the shock transition (in Fig. 2a, for example) must take the flow to a branch which will remain subsonic thereafter, instead of a branch which is likely to become supersonic again.

4 Properties of Viscous Transonic Flows

The topologies shown above become more complicated once various cooling and viscous heating effects are included (C90, C96c). We present the case of 'isothermal gas' where the flow adjusts heating and cooling in such a way that matter remains at a constant temperature. Figs. 3(a-d) show the 'phase space' of the accretion flow.

We assume Schwarzschild black hole $a = 0$ for simplicity and consider Shakura-Sunyaev (1973) viscosity prescription where the viscous stress at a given location depends on the local thermal pressure p: $t_{r\phi} = -\alpha p$. We note the general change in topology of the flow. First of all, the circular topology of the inviscid flow ('O' type sonic point) is converted into spiral topology as in a damped harmonic oscillator (C90). Secondly, the closed topology has opened up, partially or completely, depending upon flow parameters. In Fig. 3a, the flow parameters are $\alpha = 0.05$, $\lambda_{in} = 1.8$, $r_{in} = 2.8$. The spiral is 'half closed'. This topology is still good for shock formation as shown in the vertical curve (C90). In Fig. 3b, we use, $\alpha = 0.1$, $\lambda_{in} = 1.8$, $r_{in} = 2.8$. Here the

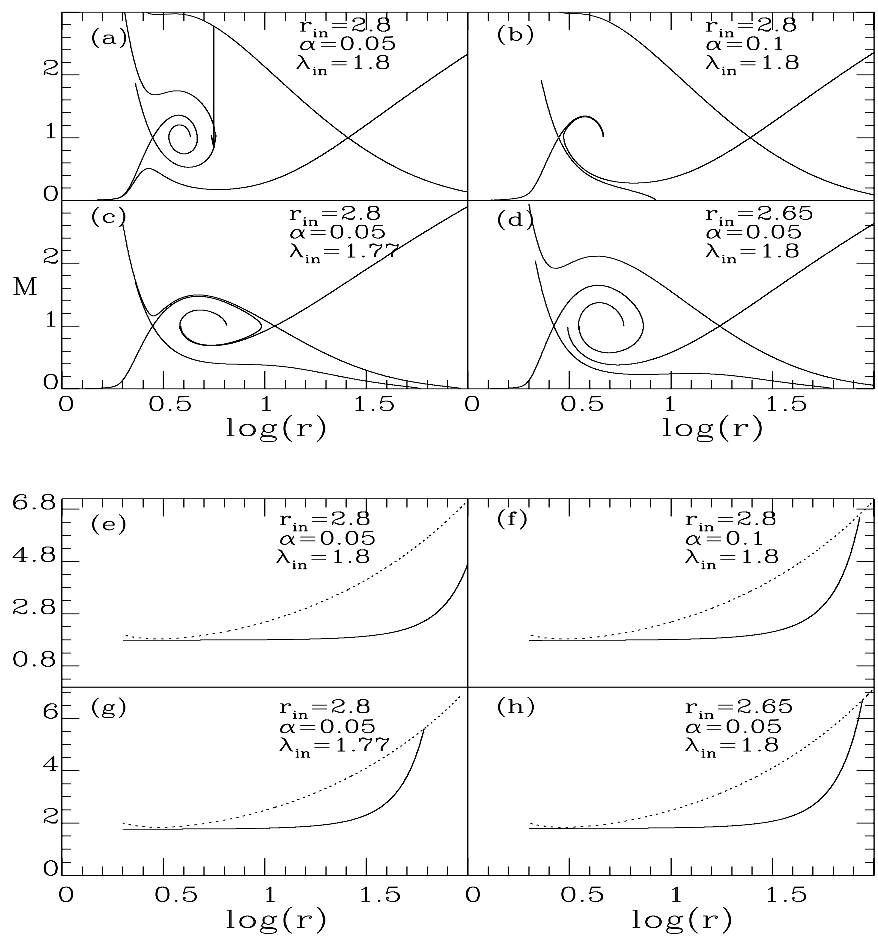


Fig. 3: Mach number variation (a-d) and angular momentum distribution (e-h) of an isothermal viscous transonic flow. Only the topology (a) allows a shock formation. Transition to open (no-shock) topology is initiated by higher viscosity or lower angular momentum or inner sonic point location. In (e-h), flow angular momentum (solid) is compared with Keplerian angular momentum (dotted).

spiral is complete with branches from both sonic points. In Fig. 3c, we use, $\alpha = 0.05$, $\lambda_{in} = 1.77$, $r_{in} = 2.8$ and in Fig. 3d, we use, $\alpha = 0.05$, $\lambda_{in} = 1.8$, $r_{in} = 2.65$. The topology remains similar to Fig. 3b, but whereas Fig. 3b is attained by increasing viscosity, Fig. 3c and Fig. 3d are obtained by decreasing angular momentum and decreasing the inner sonic point respectively. What is common in Fig. 3(b-d), is that both the sonic points allow flows to become Keplerian (where, $M \sim 0$) but since the dissipation in the flow passing through 'I' is higher, we believe that the disk will choose this branch. More importantly, the flow cannot have a standing shock wave with this topology. All these solutions, except where $M \sim 0$, are sub-Keplerian as illustrated in Figs. 3(e-h) where the flow angular momentum (solid curve) is compared with Keplerian angular momentum (dotted curve). The location where the flow joins a Keplerian disk may be somewhat turbulent (description of which is not within the scope of our solution) so as to adjust pressures of the subsonic Keplerian disk and the sub-Keplerian transonic viscous flow. These figures illustrate the existence of critical viscosity parameter α_c, critical angular momentum λ_c, and critical inner sonic point r_c at which the flow topologies are changed (C90).

The above results obtained for isothermal flow can be generalized easily using the most general set of equations with heating and cooling processes (C96c). If Q^+ and Q^- denote the heating and cooling rates, and if, for illustration purpose, one assumes that $f = (Q^+ - Q^-)/Q^- = constant$, then one can easily solve the general equations to find the degree at which the flow deviates from a Keplerian disk. In Fig. 4a we show the ratio $\lambda_{disk}/\lambda_{Kep}$ for various viscosity and cooling parameters α and f. Clearly as the flow starts deviating from a Keplerian disk $\lambda_{disk}/\lambda_{Kep} = 1$, it becomes

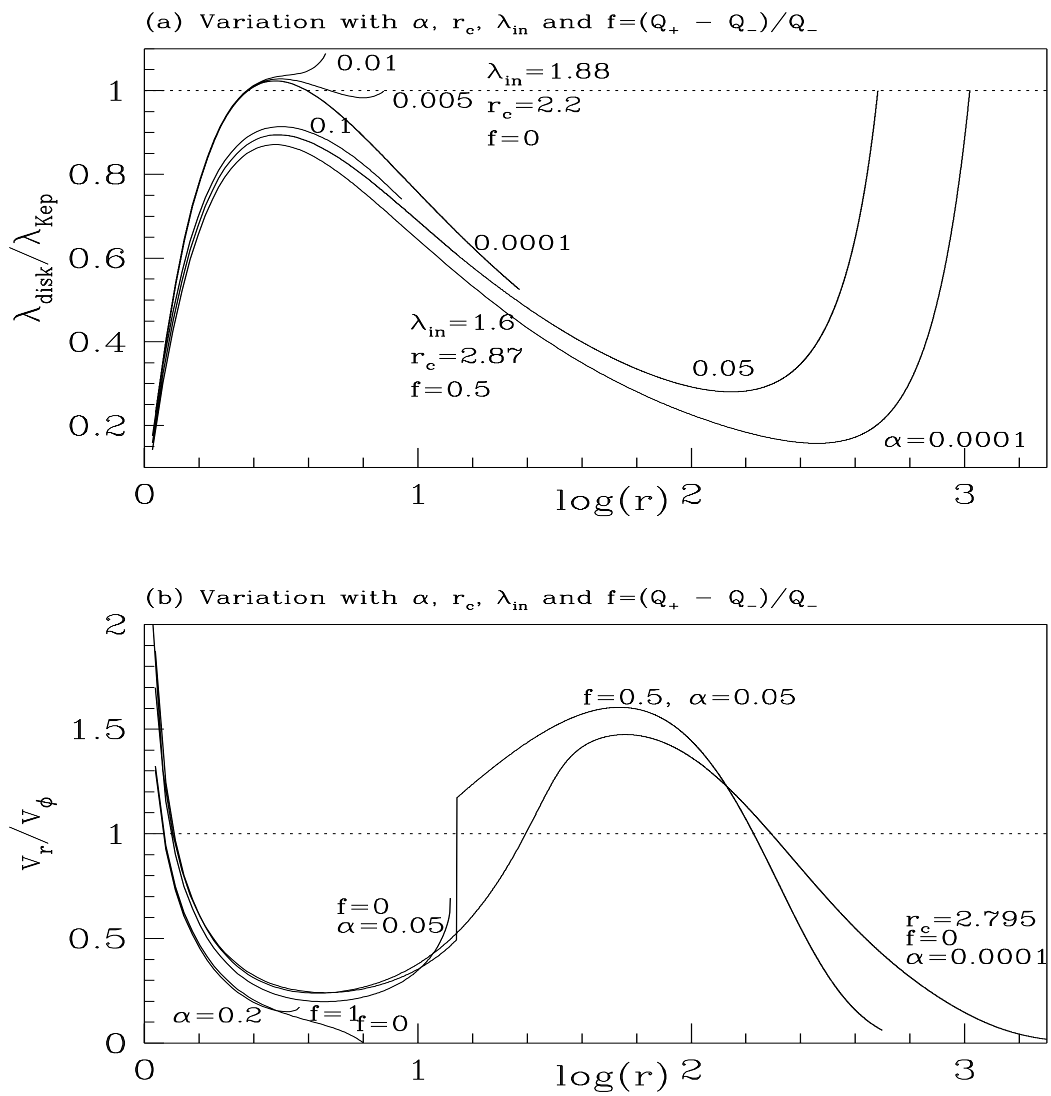


Fig. 4(a-b): Ratio of (a) disk angular momentum to Keplerian angular momentum (λ/λ_{Kep}) and (b) radial to azimuthal velocities (v_r/v_ϕ) are shown for a few solutions. Parameters are marked on the curves. Note that r_{Kep}, where the flow joins a Keplerian disk, depends inversely on the viscosity parameter.

about 20-30% of Keplerian. As it approaches the black hole, it becomes close to Keplerian again (and sometimes super-Keplerian also) before plunging in to the black hole in a very sub-Keplerian manner. In Fig. 4b, we show the ratio v_r/v_ϕ for the same disks. Note that near the outer edge, as the flow deviates from a Keplerian disk, $v_r << v_\phi$, i.e., the flow is rotation dominated. Around $r \sim 100$ the radial velocity becomes dominant, and subsequently, even closer to the black hole, the flow becomes rotation dominated (due to the centrifugal barrier). Near the horizon the radial (advection) velocity dominates once more. The case $f = 0.5$, $\alpha = 0.05$ showing a sudden jump in velocity ratio represents a solution which includes a standing shock. The corresponding angular momentum distribution in the upper panel does not show discontinuity since shear is chosen to be continuous across a shock wave. The region where the disk deviates from a Keplerian disk could be geometrically thick and would produce thick accretion disks. The post-shock region, where the flow suddenly becomes hot and puffed up would also form another thick disk of smaller size.

The solutions presented so far is of most general nature encompassing the entire parameter space. Any other solutions of black hole accretion or winds are special cases of these solutions.

5 Problem of Angular Momentum Transport

A curious property of a black hole accretion flow is that the shear stress $\sigma_{r\phi}$ is not a monotonic function of distance, and it can be negative close to the black hole. In the Newtonian geometry, it has the form $-r\frac{d\Omega}{dr}$ and its magnitude is monotonically increasing inward. For a cold radial flow below the marginally bound orbit this was pointed out by Anderson & Lemos (1988). In Fig. 5(a-b) we show this for the complete and exact global solutions of the accretion flow presented in the earlier Section (C96b). In Fig. 5a, we present the variation of shear and angular velocity

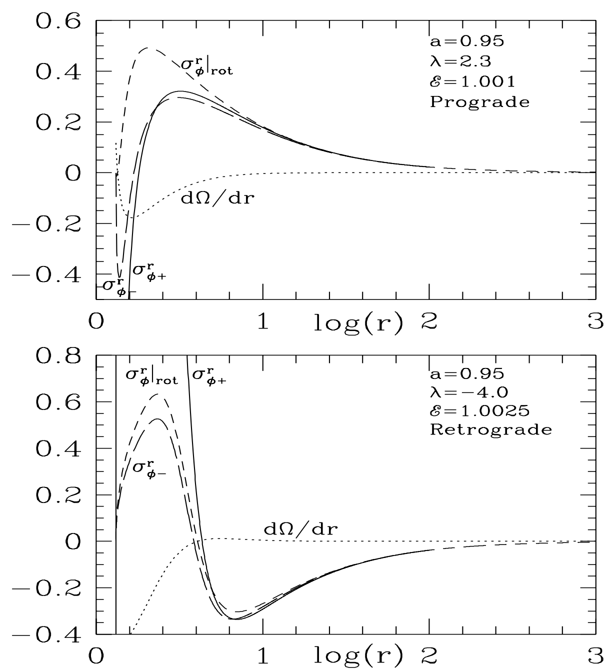


Fig. 5: Comparison of rotational viscous stress $\sigma^r_\phi|_{rot}$ (short dashed curves) with complete viscous stress $\sigma^r_{\phi+}$ (solid) along the supersonic branch (passing through outer sonic point) for (a) a prograde flow (upper panel) and (b) a retrograde flow (lower panel). Also shown is $d\Omega/dr$ (dotted curves). For comparison, results for the subsonic branch is also shown (long dashed). Note the change in sign of the shear near the horizon. $\sigma^r_{\phi+}$ does not vanish on the horizon, but $\sigma^r_\phi|_{rot}$ and $\sigma^r_{\phi-}$ do.

gradient for a typical prograde flow ($a = 0.95$, $\lambda = 2.3$ and $\mathcal{E} = 1.001$) and in Fig. 5b we show the results of a retrograde flow ($a = 0.95$, $\lambda = -4.0$ and $\mathcal{E} = 1.0025$). The solid curves are computed using the most general definition of shear tensor: $\sigma_{\alpha\beta} = 1/2(u_{\alpha;\mu}P^{\mu}_{\beta} + u_{\beta;\mu}P^{\mu}_{\alpha}) - \Theta P_{\alpha\beta}/3$ where $P_{\alpha\beta} = g_{\alpha\beta} + u_\alpha u_\beta$ is the projection tensor and $\Theta = u^{\alpha}_{;\alpha}$ is the expansion. The velocity components have been taken from the exact solution which passes through the outer sonic point. The subscript $+$ sign under σ^r_ϕ refers to the supersonic branch of the solution. We also present the same quantity (long dashed curve) for the subsonic branch ($\sigma^r_{\phi-}$) which passes through the outer sonic point. For comparison, we present the component of shear tensor $\sigma^r_\phi|_{rot}$ (short dashed curve) where we ignore the radial velocity completely $v_r << v_\phi$. We also present the variation of $d\Omega/dr$ (dotted curve) to indicate the relation between the angular velocity gradient and shear.

Some interesting observations emerge from this peculiar shear distribution. First, for prograde flows, the shear reverses its sign and becomes negative just outside the horizon. This shows that very close to the horizon, the angular momentum transport takes place 'towards the black hole' rather than away from it, Second, $\sigma^r_\phi|_{rot}$ and $\sigma^r_{\phi-}$ which have negligible radial velocities, vanish on the horizon whereas $\sigma^r_{\phi+}$ does not. Existence of the negative shear component shows the α viscosity prescription (SS73) is invalid close to a horizon, since pressure cannot be negative. However, the error by choosing a positive (e.g., $-\alpha p$) shear may not be significant, since one requires a (unphysically) large viscosity so as to transport significant angular momentum inwards. In any case, the inward angular momentum transport may change angular momentum distribution near the horizon (from that of an almost constant to something perhaps increasing inwards, similar to the Keplerian distribution below marginally stable orbit). Third, in retrograde flows, the $\sigma^r_{\phi+}$ reverses twice but $\sigma^r_{\phi-}$ and $\sigma^r_\phi|_{rot}$ reverse once since they vanish on the horizon as in the prograde flows.

The physical mechanism underlying the reversal of the shear component is simple. In general relativity, all the energies couple one another. It is well known that the 'pit in the potential' of a black hole is due to coupling between the rotational energy and and the gravitational energy (Chakrabarti, 1993). As matter approaches the black hole, the rotational energy, and therefore 'mass' due to the energy increases which is also attracted by the black hole. This makes gravity much stronger than that of a Newtonian star. When the black hole rotates, there are more coupling terms (such as thise arise out of spin of the black hole and the orbital angular momentum of the matter) which either favour gravity or go against it depending on whether the flow is retrograde or prograde (Chakrabarti & Khanna, 1992) respectively. This is the basic reason why retrograde and prograde flows display different reversal properties.

6 Numerical Simulations of Black Hole Accretion

The steady state solution topologies described in the previous Sections need not be the final solution of a set of time dependent equations. Depending on the stability of the solutions, the flow may or may not settle on the steady state solution. However, it so happens that except for the solutions in regions NSA and NSW, numerical results actually match with analytical results. The accuracy of the matching primarily depends on the ability of the numerical code to conserve angular momentum and energy. In Chakrabarti & Molteni (1993) one dimensional simulations and in Molteni, Lanzafame & Chakrabarti (1994) and Molteni, Ryu & Chakrabarti (1996) two dimensional simulations have been presented using parameters from SA and SW regions. Simulation by diverse codes produced similar results. These suggest that these analytical solutions could be used for rigorous tests of the codes in curved spacetime. The parameters from regions NSA and NSW show unstable behaviours in a multidimensional flow (Chakrabarti et al, 1996, in preparation) though in strictly one dimension they match with the analytical work. Preliminary solutions of cold flows ($\mathcal{E} = 1$) have already been reported in the literature (Ryu, Chakrabarti & Molteni, 1996).

Fig. 6 shows the Mach number and density variations in an one dimensional flow around a Schwarzschild black hole (Molteni, Ryu and Chakrabarti, 1996). The solution is chosen from region SA so that analytically a stable shock is expected in a thin flow of polytropic index $\gamma = 4/3$. The solid curve shows the analytical solution: at the shock the density goes up as matter virtually stops behind the shock and the Mach number goes down from supersonic to subsonic. The long-dashed and the short-dashed curves are the results of simulations using Total Variation Diminishing (TVD) method (see, Harten 1983; Ryu et al, 1995) and the Smoothed Particle Hydrodynamics (SPH) method (see, Monaghan, 1985; Molteni & Sponholz, 1994). The results show that the shocks form in an accretion flow, and they are stable. These simulations also verify the shock free solutions.

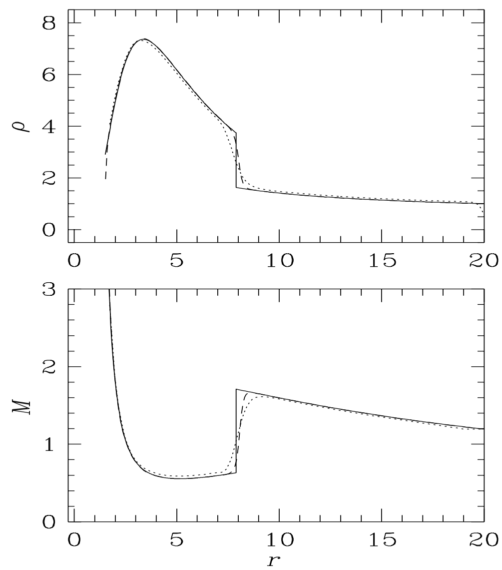


Fig. 6: Comparison of analytical and numerical results in a one-dimensional accretion flow which allows a standing shock. The long and short dashed curves are the results of the TVD and SPH simulations respectively. The solid curve is the analytical result for the same parameters. Upper panel is the mass density in arbitrary units and the lower panel is the Mach number of the flow.

7 Spectral Properties of Generalized Accretion Disks

How does an accretion disk radiate? What are the observational signatures of a black hole? How to distinguish a black hole and a neutron star from observations? In order to answer these questions one clearly needs to solve the hydrodynamic equations in conjunction with radiative transfer. In the first approximation one could construct a realistic accretion disk model based on analytical solutions mentioned above, and then compute the radiations emitted out of this model flow.

In Fig. 3, we showed that above a critical viscosity or below a critical angular momentum or inner sonic point location, the flow does not have a shock wave after deviating from a Keplerian disk. Since with height the sonic point is expected to be closer to a black hole, and a Keplerian disk above a sub-Keplerian flow may be Rayleigh-Taylor unstable, it may be worthwhile to consider the important alternative that the flow viscosity decreases with height. This makes a solution topology of the kind as in Fig. 3b to be on the equatorial plane with a Keplerian disk close to a black hole, while a solution with a shock (Fig. 3a) at a higher elevation.

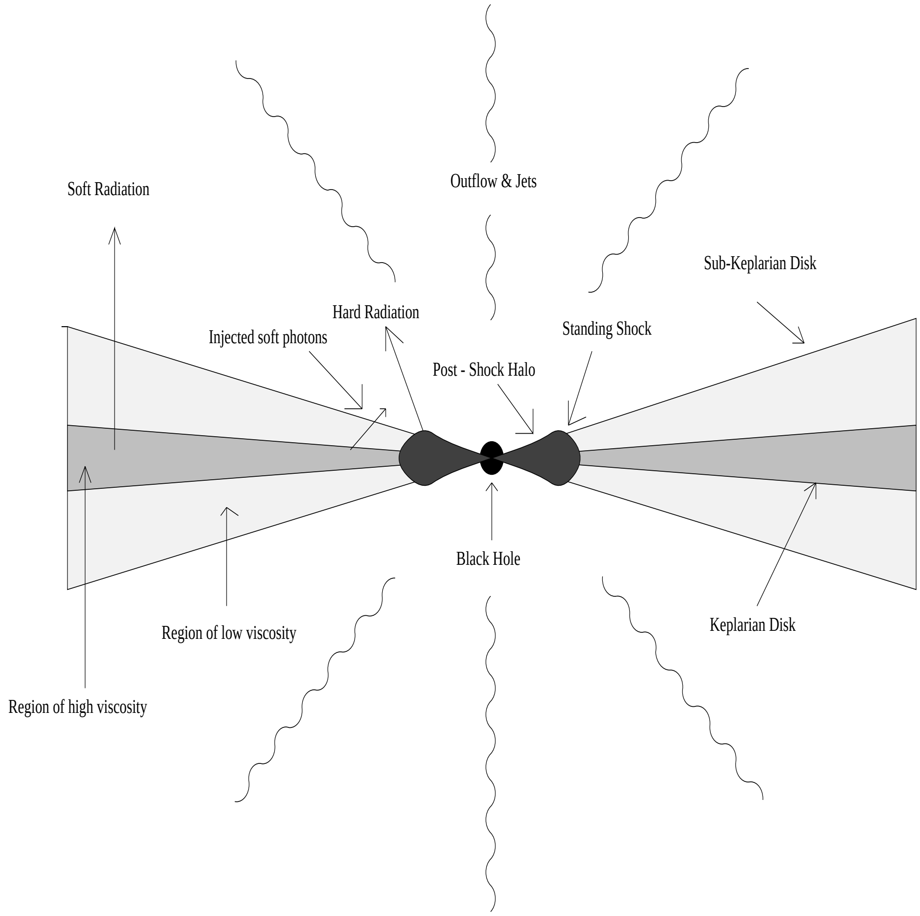


Fig. 7: Schematic diagram of the accretion processes around a black hole. An optically thick, Keplerian disk which produces the soft component is surrounded by an optically thin sub-Keplerian halo which terminates in a standing shock close to the black hole. The postshock flow Comptonizes soft photons from the Keplerian disk and radiates them as the hard component. Iron line features may originate in the rotating winds.

Fig. 7 shows a typical disk model where Keplerian disk is flanked by sub-Keplerian 'halo' which undergoes a shock wave at around $r \sim 10 - 30$ in Schwarzschild geometry and roughly at half as distant in extreme Kerr geometry (Chakrabarti & Titarchuk, 1995). Photons from the Keplerian disks are in the optical-UV range in the case of massive black holes at galactic centers, and in the soft X-ray range in the case of small black holes in X-ray binaries. These photons are intercepted by the post-shock flow and are energized through inverse Compton process, and are eventually re-radiated at higher energies (as soft X-rays in massive black holes, or, in hard X-rays in smaller black holes). First consider a galactic black hole candidate of mass $\sim 5M_\odot$. When accretion rate of the Keplerian component is very low, the soft X-ray from the Keplerian component is weak, but the hard X-ray intensity from the post-shock region is very strong. This is known as the hard state. As the accretion rate increases, electrons in the post-shock region become cooler and the object eventually goes to the soft state where only the soft-X rays from the Keplerian disk dominates. Occasionally, one finds a weaker hard component tail of slope $\alpha_\nu \sim 1.5$ even in the soft state. This is interpreted (Chakrabarti & Titarchuk, 1995) as due to the 'bulk motion Comptonization' where the converging flow energizes soft photons not due to thermal Comptonization (i.e., by thermal motion of the hot electrons) but due to direct transfer of bulk motion momentum of the cool electrons to cooler photons (Blandford & Payne, 1981; Titarchuk, Mastichiadis & Kylafis, 1996).

Figs. 8(a-b) show the variation of intensity L_ν of radiation as a function of frequency ν (Chakrabarti & Titarchuk, 1995).

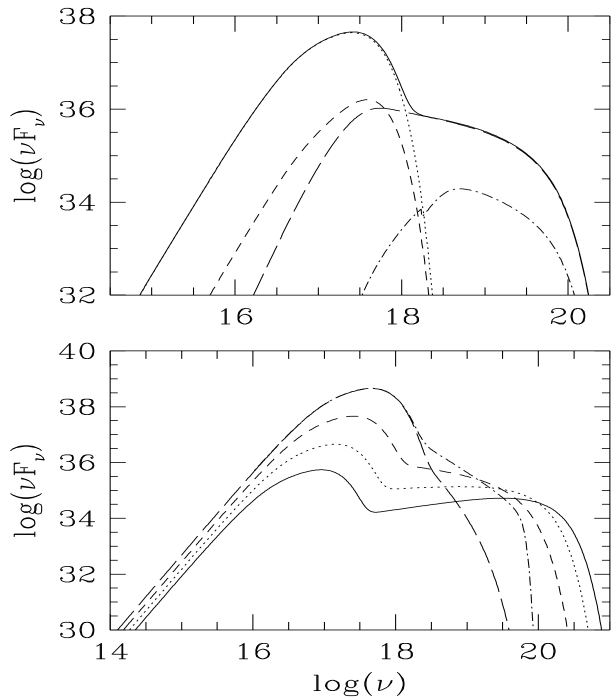


Fig. 8(a-b): Analytical computation of the emitted radiation from a black hole. (a) Contributions of various components to the net spectral shape (solid). Dotted, short dashed, long dashed and dash-dotted curves are the contributions from the Shakura-Sunyaev disk $r > r_s$, the reprocessed hard radiation by the Shakura-Sunyaev disk, reprocessed soft-radiation by the postshock disk $r < r_s$ and the hard radiation reflected from the Shakura-Sunyaev disk along the observer ($\mu = cos\theta = 0.4$). Parameters are $\dot{m}_{disk} = 0.1$, $\dot{m}_h = 1$, and $M = 5\ M_\odot$. (b) Variation of the spectral shape as the accretion rate of the disk is varied. $\dot{m}_{disk} = 0.001$ (solid line), 0.01 (long-dashed line), 0.1 (short-dashed line), and 1 (dotted line). The dash-dotted curve represents the hard component from convergent inflow near the black hole and has the characteristics of the slope ~ 1.5 in soft state.

In Fig. 8a, we show contributions from various components of the disk. The dotted curve represents the black body radiation from the optically thick Keplerian flow. The long dashed curve represents the fraction of these photons which were intercepted by the postshock flow and were energized by hot electrons of the post-shock flow through thermal Comptonization processes. The dash-dotted curve is the reflection of these hot photons from the Keplerian disk. The solid curve gives the overall spectra which has a bump in the soft X-ray and a power law in the 2-50 keV region. The above figure is for $M_{BH} = 5M_\odot$ and dimensionless disk accretion rate of $\dot{m}_{\rm disk} = 0.1\dot{M}_{Edd}$. In Fig. 8b, we show the computed spectra as the accretion rate of the Keplerian component is varied. The solid, long-dashed, short-dashed and the dotted curves are for accretion rates $\dot{m}_{\rm disk} =$ 0.001, 0.01, 0.1, 1 respectively. We also plot the dash-dotted curve which includes the effect of the bulk-motion Comptonization and shows a weak hard tail. The bulk motion Comptonization is the process by which the cool electrons deposit their momentum (while rushing towards the black hole horizon) onto photons and energize them. The spectral index α_ν ($L_\nu \propto \nu^{-\alpha_\nu}$) of the weak hard tail computed from Chakrabarti & Titarchuk (1995) model is around 1.5, and is often observed in black hole candidates in their soft states. The conclusions regarding general spectral shape remain unchanged when computations are done for supermassive black hole.

8 Comparison With Observations and Concluding Remarks

As an example of the success of our understanding of black hole accretion processes, we use the aforementioned theoretical understanding to fit a spectrum of GRS 1009-45 (X-ray Nova in Vela) obtained by MIR-KVANT module experiments (Titarchuk et al. 1996) The source was discovered by WATCH/GRANAT all-sky monitor (Lapshov et al., 1993; IAUC 5864) and verified by BATSE/GRO (Harmon et al., IAUC 5864).

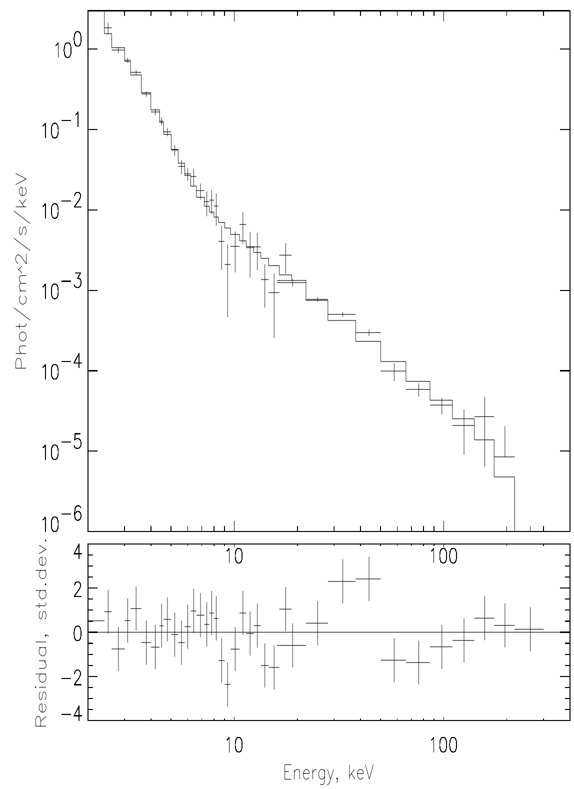


Fig. 9: Spectral fits of broad band X-rays from GRS1009-45 using our composite disk model. Data is obtained from MIR-KVANT experiments as reported in Sunyaev et al. (1994).

Fig. 9 shows the TTM and HEXE data (the photon flux versus energy; Sunyaev et al. 1994) and the best theoretical fit given by the considerations of the converging inflow in the post-shock region. The parameters, namely, mass of the black hole (M), accretion rate in Keplerian disk ($\dot{m}_{disk}$), accretion rate in the sub-Keplerian disk ($\dot{m}_h$), the shock location (R_{sh}), the post-shock temperature (T_{sh}), fraction of the soft photons intercepted by the post-shock bulge (H) and the distance (D) of the object, which are derived from the best-fit of the data seem to be: $M = 3.7554 M_\odot$, $\dot{M}_{disk}/\dot{M}_{Edd} = \dot{m}_{disk} = 1.94$, $\dot{M}_h/\dot{M}_{Edd} = \dot{m}_h = 1.1$, $R_{sh} = 11.74 R_g$, $T_{sh} = 5.59 (KeV)$, $H = 0.015$, $D = 5.0 (Kpc)$ The $\chi^2 = 1.03$ for this fit. The general agreement clearly vindicates the claim that a black hole accretes a significant amount of sub-Keplerian matter. In a neutron star accretion, the weak hard tail is not expected.

Some of the black hole candidates show quasi-periodic oscillations of its spectra in some range of hard X-rays (Dotani, 1992). Moteni, Sponholz, & Chakrabarti (1996) and Ryu, Chakrabarti, Molteni (1996) show that these oscillations could be due to dynamic oscillations of the standing shock waves or even the sub-Keplerian region itself. The frequency and amplitude modulation of the radiation, as well as the time variation of the frequencies match well with observations.

In this review, we showed that understanding accretion processes on a black hole must necessarily include the study of sub-Keplerian flows and how they combine with Keplerian disks farther away. We showed that the so called viscous transonic flows are stable, and constitute the most general form of accretion flows. We showed that the accretion shock could play an important role in energetics of the radiated photons. Indeed we showed that hard and soft states of black hole candidates as well as Quasi-Periodic Oscillations could be explained if shock waves were assumed to be present. Both of these observations along with our model could be used to obtain the mass and the accretion rates of black holes. Success of these solutions clearly depend on more accurate observations of steady state and time varying spectra of the black hole candidates.

References

[1] Anderson, M.R. and Lemos, J.P.S, *MNRAS.*, **233**, 489

[2] Blandford, R.D. and Payne, D. G., 1981 *MNRAS*, **194**, 1033

[3] Bondi, H., 1952 *MNRAS.*, **112**, 195

[4] Chakrabarti, S.K. 1989 *ApJ.*, **347**, 365 (C89)

[5] Chakrabarti, S.K. 1990, Theory of Transonic Astrophysical Flows (Singapore: World Scientific, Singapore, 1990) (C90)

[6] Chakrabarti, S.K. 1996a *Physics Reports*, **266**, No.5-6, 238

[7] Chakrabarti, S. K., 1996b *ApJ.*, (June 20th issue)

[8] Chakrabarti, S. K., 1996c *MNRAS.*, (in press)

[9] Chakrabarti, S. K., 1993 *MNRAS.*, **261**, 625

[10] Chakrabarti, S.K., and Molteni, D. 1993 *ApJ.*, 417, 671

[11] Chakrabarti, S. K. and R. Khanna *MNRAS.*, **256**, 300

[12] Chakrabarti, S. K., and Titarchuk, L. G., 1995 *ApJ.*, 455, 623

[13] Dotani, Y., 1992, in Frontiers in X-ray Astronomy (Tokyo: Universal Academic Press) 152

[14] Harten, A., 1983 *J. Comput. Phys.*, **49**, 357

[15] Harmon, B.A., Zhang, S.N., Fishman, G.J. and Paciesas, W.S., 1993, *IAU. Circ.*, No. 5864

[16] Kaniovsky, A., Borozdin, K., and Sunyaev, R., 1993, *IAU. Circ.*, No.5878

[17] Lapshov, I., Sazonov, S., and Sunyaev, R., 1993, *IAU Circ.*, No.5864

[18] Malkan, M. 1982 *ApJ.*, **254**, 22

[19] Molteni, D., Lanzafame, G., and Chakrabarti, S.K. 1994 *ApJ.*, **425**, 161 (MLC94)

[20] Molteni, D. and Sponholz, H. 1994 *J. Italian Astronomical Soc.*, Vol. 65-N. 4-1994, Ed. G. Bodo and J.C. Miller

[21] Molteni, D., Sponholz, H., and Chakrabarti, S.K. 1996 *ApJ.*, **457**, 805

[22] Molteni, D., Ryu, D., and Chakrabarti, S.K. 1996 *ApJ.*, (Oct. 10th, in press)

[23] Monaghan J.J., 1985 *Comp. Phys. Repts.*, **3**, 71

[24] Novikov, I., and K.S. Thorne. 1973. in: Black Holes, eds. C. DeWitt and B. DeWitt (Gordon and Breach, New York)

[25] Ryu, D., Brown, G., Ostriker, J.P. and Loeb, A., 1995 *ApJ.*, **452**, 364

[26] Ryu, D., Chakrabarti, S.K. and Molteni, D., *ApJ.*, (Submitted)

[27] Shakura, N.I., and Sunyaev, R.A. 1973 *A and A*, 24, 337

[28] Sun, W.H. and Malkan, M. 1989 *ApJ.*, 346, 68

[29] Sunyaev, R. A. et al. 1994 *Astronomy Letters*, **20**, 777

[30] Titarchuk, L., Mastichiadis, A. and Kylafis, N. 1996 *ApJ.*, (submitted)

[31] Titarchuk, L.G., Borozdin, K., Sunyaev, R. and Chakrabarti, S.K., 1996, (in preparation)

During all my discussions with Fermi, I never failed to marvel at the ease and clarity with which he analyzed novel situations in fields in which, one might have supposed, he was not familiar and, indeed, was often not familiar prior to discussion. In the manner in which he reacted to new problems, he always gave me the impression of a musician who, when presented with a new piece of music, at once plays it with a perception and a discernment which one would normally associate only with long practice and study. The fact, of course, was that Fermi was instantly able to bring to bear, on any physical problem with which he was confronted, his profound and deep feeling for physical laws: the result invariably was that the problem was illuminated and clarified. Thus, the motions of interstellar clouds with magnetic lines of force threading through them reminded him of the vibrations of a crystal lattice; and the gravitational instability of a spiral arm of a galaxy suggested to him the instability of a plasma and led him to consider its stabilization by an axial magnetic field.

— S. CHANDRASEKHAR

LARGE SCALE STRUCTURE IN THE UNIVERSE:

THEORY VS OBSERVATIONS

Dipak Munshi[1]

Inter-University Centre for Astronomy & Astrophysics
Post Bag 4
Ganeshkhind
Pune 411 007, India

Abstract

I discuss main observational inputs and analytical methods in formation of large scale structure in the universe with particular reference to approximate methods for understanding gravitational clustering and computational techniques used in simulations. Different statistics which are used to differentiate models of structure formation are also discussed in brief.

[1] e-mail: munshi@iucaa.ernet.in

It's a terrible dilemma, I agree, that for the future of India, the problems of population, poverty, and the misery in which the majority live, have of course to be solved. But does that need absolute stifling of all other things? I do not think so. I have the feeling that, in every country, pure science has a place that is not very different from the cultivation of literature, arts, etc. No one claims that the latter should be eliminated or that the priorities require one to suppress them. Questions arise in the case of science because of the money involved. But there are reasonable choices one can make; there are branches of science which are not expensive. One should not just blindly imitate the rich countries.

— S. CHANDRASEKHAR

1 Introduction

Cosmology although one of the oldest subjects of human enquiry, until the 20th Century lay in the domain of metaphysics, a subject of pure speculation. With little observational data it's practitioners were philosophers and religious leaders rather than scientists. The science of cosmology emerged in the 20th Century when it first became possible to probe great distances through the universe. Observations from ground based observatories and satellites in several frequency bands starting from Gamma rays to radio provided images and data that could begin to discriminate competing theories.

The universe that we observe and seek to understand is astonishingly smooth on larger scales (> 100 Mpc), but is increasingly inhomogeneous on smaller scales (< 30 Mpc). Although evolution of a smooth background universe is well understood in the general frame work of FRW models, any reasonable model of the universe has to explain the origin and evolution of such inhomogeneities. While study of generation of such perturbations is mainly related to study of early universe, here we will be concentrating mainly on observational and theoretical aspects of evolution of these large scale structures.

Observations related to large scale structure in the universe can be mainly divided into three categories. (1) Observations related to distribution of galaxies, (2) Observations related to peculiar velocities i.e. deviations from smooth Hubble flow, (3) The initial condition: along with the positions and velocities of galaxies it is extremely important that we have some idea about the initial condition of the system from which it has evolved this enables us to check that the observed velocities and density inhomogeneities are really consistent with the paradigm of gravitational clustering. One of the major input to observational cosmology from cosmic microwave background studies is to provide information about such initial conditions in terms of initial potential fluctuation.

On the theoretical side the major effort has been devoted to understanding the generation of perturbation, quantifying the statistical nature of density and velocity inhomogeneity and understanding the evolution of inhomogeneities from the initial condition in terms of dynamics of gravitational clustering.

2 Observational Inputs

2.1 Density inhomogeneity and redshift surveys

Over two million galaxies have been mapped in 2D angular surveys although redshift of small fraction of them is actually known at present. The clustering of galaxies are very clear in three dimension surveys such as CfA (Center for Astrophysics) survey and the southern sky redshift survey (SSRS). These two surveys taken together contain over 15,000 galaxies. It is clear from figure-1 that galaxies are not distributed randomly in the universe and they appear to lie preferentially in two dimensional sheets and one dimensional filaments which separate huge voids which have practically no galaxies. The typical length scale characterising this cellular structure is about 30–50 h^{-1}Mpc, h being the Hubble parameter measured in units of 100 km/sec/Mpc.

The most commonly used statistics for describing clustering of galaxies are n-point correlation functions. Lowest order member of this family is 2-point correlation function $\xi(r)$ which estimates the probability in excess of random of finding galaxies at a distance r from a given galaxy.

$$\langle \rho(x)\rho(\vec{x}+\vec{r})\rangle = \bar{\rho}^2(1+\xi(r)). \tag{1}$$

Where $\rho(x)$ is the density at x and averaging is done over different realisation of the density field. On scales $r < 10\ h^{-1}$ Mpc, the two point correlation function can be accurately described by a power law,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-1.8} \tag{2}$$

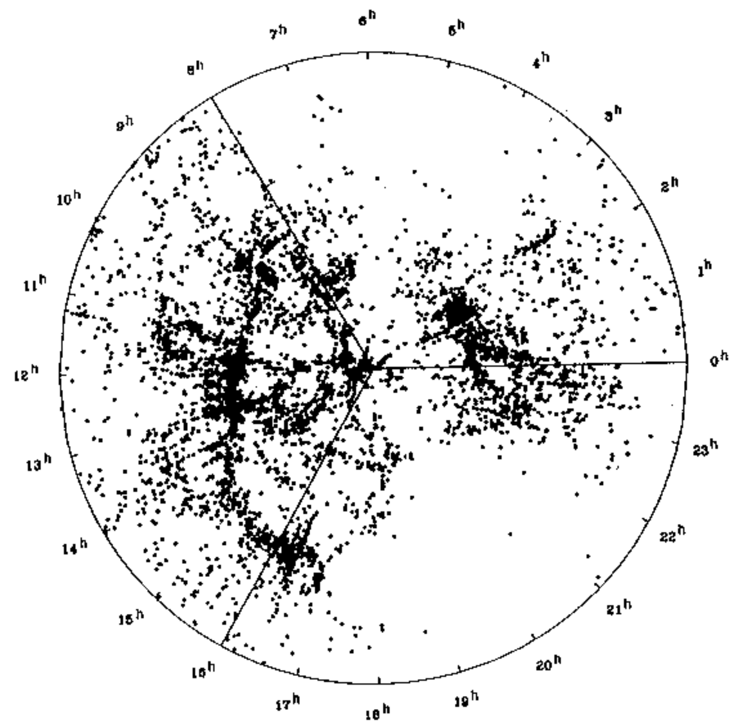


Figure 1: A 360° view that shows a relation between the great wall and the Perseus - Pisces chain. The slice covers the declination region from 20° to 40° and contain all the 6112 galaxies with detected redshift cz < 15,000 km/s. The region that appear devoid of galaxies are obscured by galactic plane. Adopted from Geller, M. J., Huchra, J. P., 1989, Science, 246, 897

where $r_0 \simeq 5$ h^{-1} Mpc. If we decompose δ the density contrast in Fourier modes we can write

$$\delta(x,t) = \frac{(\rho - \bar{\rho})}{\bar{\rho}} = \int d^3k \exp i\vec{k}\vec{x}\delta(k,t) \tag{3}$$

$$\delta(k,t) = |\delta_k(t)| \exp(i\phi_k) \tag{4}$$

For a Gaussian random field the amplitude δ_k has Rayleigh distribution and phases ϕ_k are distributed uniformly in the interval $[0\text{–}2\pi]$. As is well known a Gaussian random field is completely specified by its power spectrum.

$$P(k) = \langle |\delta(k)|^2 \rangle \tag{5}$$

so that the two point correlation function $\xi(r)$ and $P(k)$ are related by the following relations.

$$\xi(r) = \int d^3k \exp(i\vec{k}.\vec{x})P(k) \tag{6}$$

$$P(k) = \frac{1}{(2\pi)^3} \int d^3x \exp(-i\vec{k}\vec{x})\xi(x). \tag{7}$$

We have plotted the power spectrum recovered from different observations in figure - 3. More details about redshift surveys are given in [2]

Lots of important information actually can be gathered by studying the dynamical aspects of groups and rich clusters of galaxies. While groups can typically carry somewhere between ten to

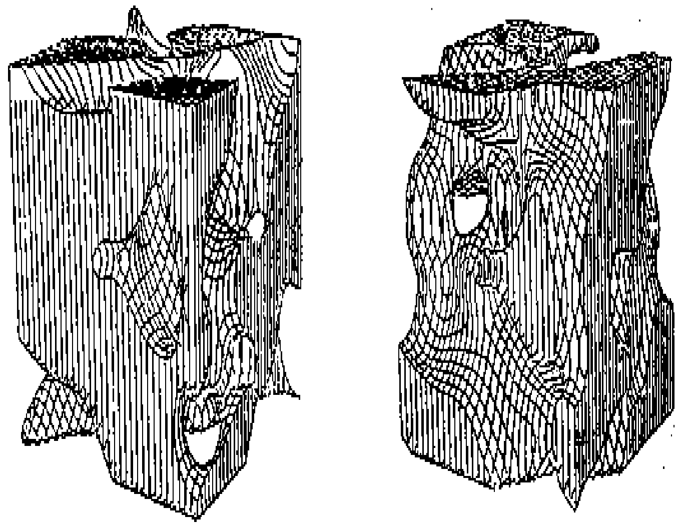


Figure 2: Isodensity contours about the mean value of a Gaussian random field. Left panel: region above the mean density. Right panel: regions below the mean density. Gravitational clustering can modify such initial topology and the signature can be used to differentiate models. Adopted from Gott, J. R., Melott, A. L., Dickinson, M., 1986, ApJ, 306, 341

hundred galaxies depending upon the size, galaxy clusters are generally massive objects containing sometime thousands of galaxies. Dynamical states of these objects and their relative abundance can be used to differentiate between different types of cosmological models. Presence of large density contrast like cluster in the line of sight of CMB observations produces appreciable distortions in the CMB sky.

2.2 Peculiar velocities of galaxies

For a smooth FRW universe relative velocities of two points will be proportional to their separation $\vec{r}$. Presence of inhomogeneities will distort their smooth Hubble flow and we can write the deviation as $\vec{v}_{pec} = \vec{V} - H\vec{r}$. Any deviation from Hubble flow characterises effect of gravitational potential due to density inhomogeneities. Estimates of peculiar velocity depend upon the sample depth, most surveys indicate $v_{pec} \approx 350$ Km/sec, on scales of 50 Mpc. From the dipole anisotropy in CMBR sky one can estimate the peculiar velocity of our own galaxy with respect to CMBR rest frame, which indicate that Milky Way is moving at a velocity $v_{pec} \approx 610$ Km/sec. A detail study of large scale motion of elliptical galaxies made by Lynden-Bell et al. indicate that a sample of galaxies spanning a volume $50\mathrm{h}^{-1}$ Megaparsec cube could be participating in a bulk flow directed towards Centaurus. The large amplitude of the flow led theorists to suggest that a large density fluctuation termed as "the Great Attractor" lying in this direction.

Bulk flows in galaxies are directly related to fluctuation in the gravitational potential. Considerable amount of effort has been paid to different ways to reconstruct density fields from velocity data and compare them with redshift surveys. One of the major advantage in using this method is that peculiar velocity field is sensitive to both luminous and dark mass as compared to galaxy surveys which map only luminous matter in the universe [1].

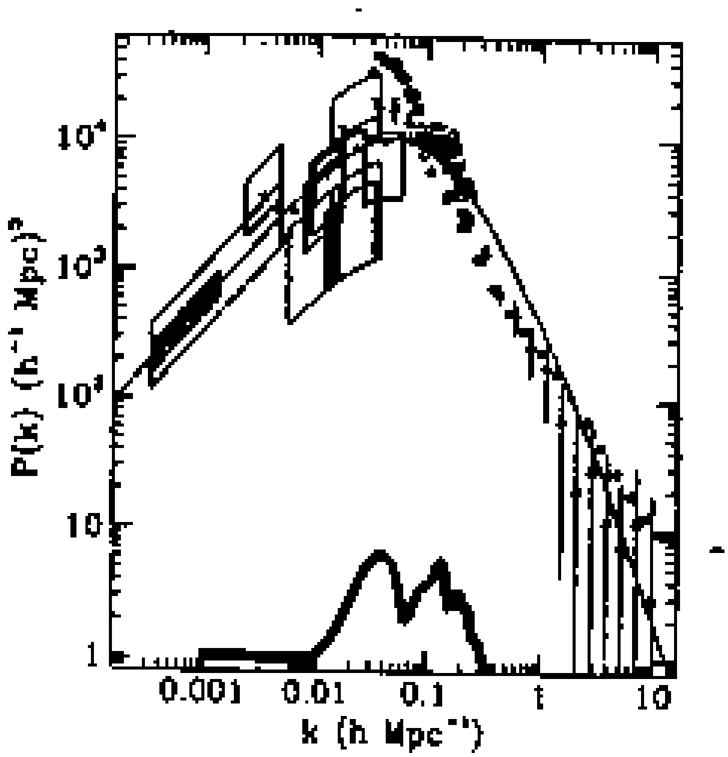


Figure 3: The cold dark matter power spectrum on a range of scales as inferred from cosmic microwave background data and large scale data. The solid line is CDM normalised to reproduce COBE data, stars crosses, squares and triangles are the APM, CfA, IRAS-QDOT and IRAS-1.2Jy surveys respectively. The boxes are $\pm 1\sigma$ values of matter power spectrum inferred from CMB experiments COBE, FIRAS, Tenerife, SP91-13pt, Saskatoon, Python, ARGO, MSAM2, MAX-GUM & MAX-MuP, MSAM3. Adopted from [6]

2.3 Cosmic Microwave Background Anisotropy

Most important discovery in recent years has been the detection of large angle anisotropy ($> 7^o$) in the cosmic microwave background made by COBE satellite in 1992. (These fluctuations are generally believed to have primordial origin unlike the dipole anisotropy discussed earlier. These fluctuations are related to the primordial fluctuation present in matter-energy density at the time when universe started becoming transparent due recombination of protons and electrons and photons started streaming freely. These photons which are reaching the detector now are carrying extremely valuable information about cosmological parameters like Ω, h, Λ, ratio between scalar and gravity wave perturbations, and composition of dark matter particles i.e whether they are made of highly relativistic low mass neutrino (also known as hot dark matter particle or HDM) or more massive and relatively cooler particles which are moving with non-relativistic speed (also known as cold dark matter or CDM). In addition to these they carry information about relative amplitude of perturbation. Other important information like ionization history of the universe is also encoded in the CMB spectra [6].

It is conventional to expand the temperature fluctuation in spherical harmonics

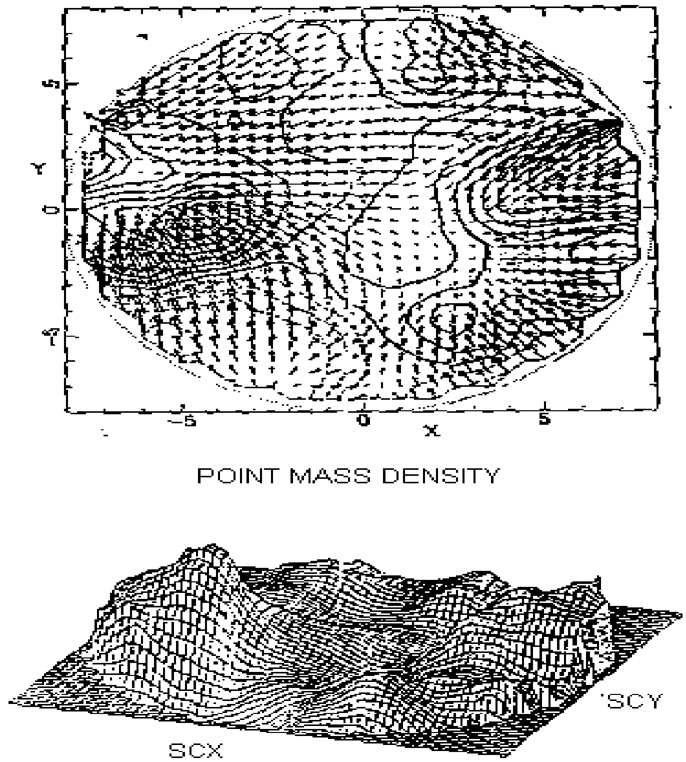


Figure 4: Density contrast δ and peculiar velocity v are plotted for ~ 3000 galaxies. The velocity vectors are projection of the 3-D velocity field in CMB frame. Distances are plotted in unit of 1000 km/s. Contour spacing is 0.2 in δ with heavy contours marking $\delta > 0$ and dashed contours $\delta < 0$. The local group is at the center and great attractor is on the left. Adopted from [1]

$$\frac{\triangle T(\theta,\phi)}{T} = \sum_{lm} a_{lm} Y_{lm}(\theta,\phi) \tag{8}$$

and work in terms of the multiple moments a_{lm}. One can define the sky correlation function as

$$C_{sky}(\theta_{21}) = \left\langle \frac{\triangle T(\hat{n}_1)}{T} \frac{\triangle T(\hat{n}_2)}{T} \right\rangle_{21}. \tag{9}$$

Here n_1, n_2 are unit vectors and the average is taken over the entire sky with separation angle θ_{12} held fixed. Using the properties of Y_{lm}, the correlation function can be written as

$$C(\theta) = \frac{1}{4\pi} \sum_{l} a_l^2 P_l(cos\theta) \tag{10}$$

Typically the conversion from multipole space l to angular scale θ is accomplished by the following approximate formula.

$$\frac{\theta}{1^o} \simeq \frac{60}{l}. \tag{11}$$

A common way of comparing theory and experiments is through a_{lm}. Of course an actual measurement of a temperature difference on the sky involves finite resolution and specific measurements strategies modifying equation (10).

As mentioned earlier COBE observations are made at large angles ($> 7^o$ scales) but after successful measurement of anisotropy lots of other experiments has been set up to study fluctuations at small angular scales. South Pole, Big Plate, PYTHON, MSAM, ARGO, MAX, White Dish, Tenerife and Southpole are experiments which are currently in progress.

While larger angles are free from astro-physical processes because they were outside the horizon during recombination smaller scale carry informations regarding astro-physical processes operating during recombination.

Finally gravitational lensing is another interesting observational input to large scale structure formation. This method can be used to map both luminous and dark matter components.

3 Analytical Methods

3.1 Statistical tools

As mentioned earlier, n-point correlation functions are most commonly used statistics for quantifying clustering of galaxies. Lowest order correlation function, 2-point correlation function and its Fourier transform, power spectrum are most commonly used statistical tools in cosmology. In the standard model of structure formation based on gravitational instability the power spectrum is related to (i) the spectrum of primordial fluctuation generated during inflation and it is also related to (ii) the nature of dark matter in the universe. The power spectrum in model with hot dark matter shows a sharp cutoff on scales greater smaller than $40h^{-1}$ Mpc, which means that super-cluster size objects are the first one to form in this scenario. Galaxies form later due to the fragmentation of these primordial super clusters. This scenario is commonly known as "top-down" scenario for galaxy formation and was proposed by Ya. B. Zeldovich during 1970's and 1980's. The cold dark matter model on the other hand, has power on all scales which is mainly related to the fact that the dark matter particles in this case are moving with relativistic speed. As a result the first objects to collapse in this scenario have low mass and larger objects are formed due to merger of these small scale objects [3], [4].

Determination of higher order correlations are much more difficult than 2-point correlation although they carry additional informations about deviation from Gaussianity. Count in cell statistics and void probability function are commonly used to extract information about higher orders from galaxy catalogues. Other important statistical methods based on fractal analysis has also been studied in this connection which are closely related to count in cell statistics.

Although n-point correlation functions contain valuable informations about statistics of cosmological fields, geometrical aspects like shapes of collapsed objects and topological connectivity of overdense or underdense regions can not be studied using them. One of the first method suggested in this direction was based on percolation theory. Other important statistical indicators based on geometry are shape statistics and minimal spanning tree. Besides geometry of collapsed objects topological properties of overdense and underdense regions have also been studied extensively and they can be used to distinguish different models of structure formation.

All the statistical methods we have described so far are based on description of density field but similar analysis regarding probability distribution function, two point correlation function can also be extended to velocity fields. It has been shown that moments of velocity divergence are more sensitive to cosmological parameters like Ω, Λ etc which can be used to determine these parameters. Statistical properties of velocity field and density field are related to each other if we believe that gravitational clustering is responsible for formation of structure in the universe.

3.2 Dynamical approximations

Analytical treatment of gravitational clustering of collisionless particles is extremely difficult due to complicated non-linear nature of the problem. However tremendous progress have been made in recent years by using approximate methods to model this difficult problem. Most of these approximations rely on simplified dynamics or special initial conditions [5].

Gravitational dynamics in an expanding background is described by the following set of equations.

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla_x[(1+\delta)\vec{v}] = 0. \tag{12}$$

$$\frac{\partial(a\vec{v})}{\partial t} + (\vec{v}.\nabla_x)\vec{v} = -\nabla_x\phi \tag{13}$$

$$\triangle\phi = 4\pi G\rho_b a^2\delta \tag{14}$$

Equation (8) is continuity equation which describe conservation of mass, Equation (9) is related to conservation of momentum also known as Euler equation. The last equation in the series is Poisson equation. All the partial derivatives has been taken with respect to co-moving co-ordinates and the background dynamics of expanding universe enters into the picture through the presence of scale factor $a(t)$. (These set of equation is only valid before trajectory crossing beyond which one can not extend this hydrodynamical description

Roughly speaking one can divide the gravitational clustering in three different regimes. In the quasi-linear regime the density contrast is small and one can linearise these equations by neglecting all higher order non-linear terms. Linear evolution of density growth preserve the Gaussianity, very soon higher order corrections start dominating and one has to take into account corrections to linear calculations. Although cumbersome, these perturbative methods of analysis can give considerable insight into the dynamics of quasi-linear regime. Perturbation theory can be done in Lagrangian co-ordinates where one measures physical quantities in a moving reference frame which is attached to the fluid element. In general Lagrangian perturbation theory perform better than Eulerian Perturbation theory.

In the highly non-linear regime one assumes that structures have already collapsed and they are in a state of statistical equilibrium which have stopped evolving with rest of the universe, which means that Hubble flow is balanced by opposite peculiar velocity field. This assumption also known as stable clustering ansatz helps in simplifying the dynamics in highly non-linear regime. Other efforts in this regime include different methods of closing the system of equation known as BBGKY hierarchy which govern the evolution of n-point correlation functions.

Intermediate regime is most difficult to understand and several dynamical approximations like Zeldovich, Adhesion, Frozen Flow, Frozen Potential, smooth potential and conserving momentum approximations help us a lot to understand one or other aspect of gravitational clustering in this regime.

Other aspects of gravitational clustering like number density of collapsed objects (also known as mass function) has been studied in great detail which also provide us with extremely valuable informations like merger history. Several scaling ansatzs have also been suggested to relate non-linear power spectra with linear spectra which is extremely useful in reconstructing primordial power spectra.

There has also been effort to understand gravitational clustering based on principles of thermodynamics which seems to work well at least for Poisson shot noise initial condition.

It may be clear from the above discussion that some of the aspects of gravitational clustering can be understood by using analytical methods. Nevertheless many situations remain outside the reach of analytical techniques and one is compelled to adopt a numerical approach to solve the equation of motion of particles. In general numerical solutions are also used widely to test the analytical predictions. The simplest way to compute the non-linear evolution of cosmological fluid is to represent it as a discrete set of particles, and then sum the pairwise interaction between particles directly to calculate the Newtonian force on each particle. One then uses this force to update the velocity and position of these particles. Depending upon how force is calculated from distribution of particles, numerical codes can mainly be divided into two categories (1) Particle - Particle or PP code where pairwise force is calculated directly for all the particles. (2) Particle - Mesh PM code where a rectangular grid is used for calculating the density of the particles from their positions, which in turn is used to get the gravitational potential (and hence gravitational force) by using Fast Fourier algorithms which increases the speed of the code by order of magnitude. However PP codes are more accurate in calculating the force. Most of the codes that are commonly used are a hybrid of this two methods, i.e. they calculate the forces due to neighbouring particles by direct summation and for other particles they rely on fast Poisson solver. This method is known as Particle - Particle: Particle - Mesh or P^3M code. With increasing computational power numerical methods can now handle more and more number of particles and also at the same time they are becoming more accurate in force calculations.

Our discussion so far has been limited to gravitational clustering of dark matter particles which being collisionless are easier to model. For comparing simulation data with galaxy catalogues however one has to take baryonic components into account. Including baryonic components means one has to calculate pressure and cooling which makes hydrodynamical codes more complicated in nature.

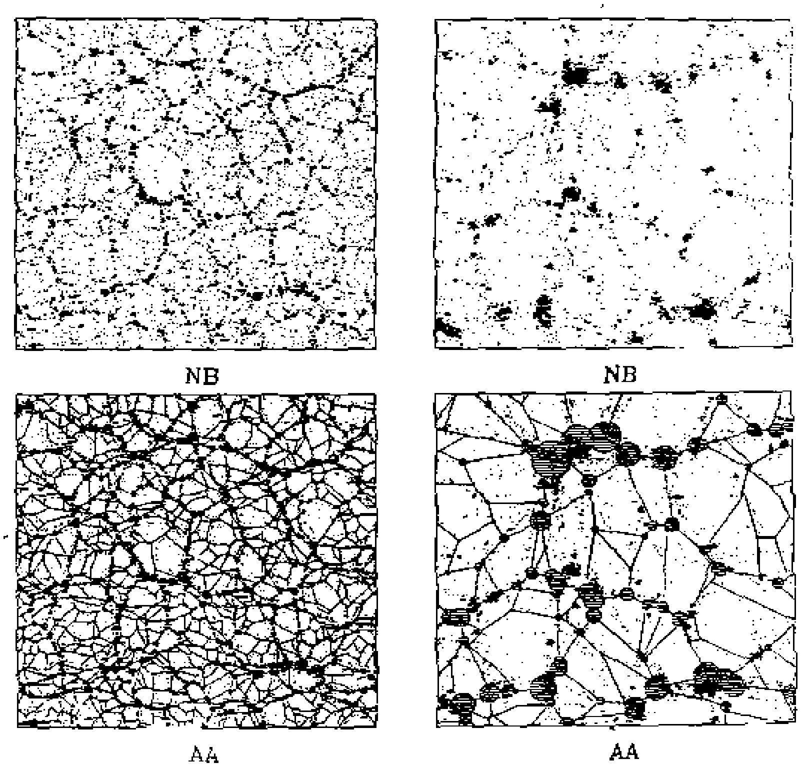


Figure 5: The evolved particle positions from n-body simulations are plotted for initially scale free n = 0 power spectra (Poisson Shot noise) in 2D for two different epochs in the upper two panels. In lower panels we superpose n-body data with predictions from adhesion approximation. Adopted from [5]

For perturbations which are of size comparable to horizon or constituent particles are moving with relativistic speed one has to do actual relativistic calculation and it is not appropriate to use Newtonian picture described above. Calculations regarding cosmic microwave background where one has to calculate path of photons in a perturbed background are important class of problem which fall in this category.

4 Conclusion

The pace of discovery has accelerated markedly in last few decades. In next decade or so, major projects will measure the distribution and velocities of galaxies, dark matter, and radiation at cosmological distances. These results will tightly constrain all present theories of the Universe and may point to fundamentally new paradigm. By its very nature cosmology entails explaining a single series of irreproducible events. Our ability to explore the physical Universe is limited to those regions which are in casual contact. Given these consideration one might wonder whether the basic features of the universe are explainable as a consequence of symmetry and fundamental laws of physics or some key features are largely determined by special initial conditions, extraordinary coincidence and/or physical laws that are untestable locally. One hopes to get some answer in this direction from next generation experiments.

Acknowledgements

It is pleasure for me to acknowledge Varun Sahni, T. Padmanabhan, Shiv Sethi, Tarun Souradeep Ghosh, Somnath Bharadwaj, R. Srianand, Jasjeet Singh Bagla and Sunu Engineer for many fruitful discussions. I thank Prof. Bala Iyer and Prof. G. Date for inviting me to attend 18th IAGRG meeting. This work was financially supported by Council of Scientific and Industrial Research.

References

[1] Dekel, A., 1994 *Annu. Rev. Astron. Astrophy.*, **32**, 371-418

[2] Efstathiou, G., 1994 Observations of Large-Scale Structure in the universe, Ed. Schaeffer, R. Les Houches Lecture Notes (To be published in Elsevier Science Publishers)

[3] Padmanabhan, T., 1993 Structure Formation in the Universe (Cambridge University Press)

[4] Peebles, P. J. E., 1993; Principles of Physical Cosmology, (Princeton University Press)

[5] Sahni, V., and Coles P., 1995, *Physics Report*, **262**, 1

[6] White, M., Douglas, S., Silk, J., 1994 *Annu. Rev. Astron. Astrophy.*, **32**, 319-370

The pursuit of science has often been compared to the scaling of mountains, high and not so high. But who amongst us can hope, even in imagination, to scale the Everest and reach its summit when the sky is blue and the air is still, and in the stillness of the air survey the entire Himalayan range in the dazzling white of the snow stretching to infinity? None of us can hope for a comparable vision of nature and of the universe around us. But there is nothing mean or lowly in standing in the valley below and awaiting the sun to rise over Kinchinjunga.

— S. CHANDRASEKHAR

SOME NON-NINEAR ASPECTS OF COSMOLOGICAL STRUCTURE FORMATION

Somnath Bharadwaj[1]

Joint Astronomy Program
Department of Physics
Indian Institute of Science
Bangalore 560 012

and

Raman Research Institute
C.V.Raman Avenue
Sadashivanagar
Bangalore 560 080

[1]e-mail: somnath@mri.ernet.in

I don't see how we can *demand* of somebody else anything. I mean, when we are not experiencing the same thing. We should be glad if more people show courage, but that is not to say that we must *require* of them such qualities. Besides, you ask yourself, how many scientists, colleagues, and friends of yourself here stand up for what is right?... Well, really, does respect for someone's scientific achievements allow one to smother one's conscience to that extent?

— S. CHANDRASEKHAR

In the first part of the talk we consider two analytic models which can be used to follow the evolution of cosmological structures, namely the Zel'dovich Approximation and the Spherical Collapse Model. These methods can be used well beyond the epoch when the linear perturbation theory breaks down and the deviation from the background universe are quite large. Details of this part are not presented here and the reader is referred to any of the following reference (Padmanabhan 1993).

In the second part of this talk we shall present some results from our study of the perturbative evolution of cosmological correlation. This approach is limited in that it can be used to study the evolution only a little beyond the linear epoch i.e. when the deviation from the background universe are mildly nonlinear. The main points are presented below.

1. The universe is modeled as a system of particles interacting only through the Newtonian gravitational force. This is appropriate on scales much smaller than the horizon scale but large enough that gravity is the dominant force. We consider an ensemble of such systems and we set up the equations of the BBGKY hierarchy in the fluid limit to study the evolution of some of the statistical properties of such an ensemble. A convenient parameter is used for the evolution instead of the cosmic time.

2. The initial conditions are chosen such that the deviation of any of the systems from the uniform state can be characterized by a small parameter ϵ. The initial conditions are also such that all the members of the ensemble have a single streamed flow. These initial conditions allow us to associate powers of ϵ with various statistical quantities for the ensemble. We consider the second and the third equations of the BBGKY hierarchy. By taking velocity moments of these equations we obtain equations for perturbatively evolving the two point and the three point correlation functions. At the lowest order of perturbation these equations give us the linear evolution of the initial two point and three point correlation functions.

3. We consider a situation where the initial disturbances are such that the density fluctuation is a random Gaussian field in a universe with the critical density. For these initial conditions the initial three point correlation function is zero. We calculate the nonlinearity-induced three point correlation function at the lowest order of perturbation for which it is non-zero. We obtain a general expression for this in terms of the linear two point correlation function and its average over a sphere. This investigation brings out the limitations of the commonly used hierarchical form.

4. In general the evolution of the two point correlation is influenced by the three point correlation function. The BBGKY hierarchy equations are used to calculate perturbatively, the lowest order nonlinear correction to the two point correlation and the pair velocity for Gaussian initial conditions. Our formalism is valid even if the flow becomes multi-streamed as the evolution proceeds. We compare our results with the results obtained using the hydrodynamic equations which neglect pressure and other effects of multi-streaming. We find that the two match, indicating that there are no effects of multi-streaming at the lowest order of nonlinearity.

5. We study the two point correlation induced at large scales for the case when it is initially zero there. Based on an analytic study confirmed by numerical results we conclude that this has a universal x^{-6} behaviour.

6. We numerically study a class of initial conditions where the power spectrum at small k has the form k^n with $0 < n \leq 3$ and we calculate the nonlinear correction to the two point correlation, its average over a sphere and the pair velocity over a large dynamical range. We find that at small separations the effect of the nonlinear term is to enhance the clustering whereas at intermediate scales it can act to either increase or decrease the clustering. We also find that the small scales significantly influence the evolution at large scales and this may lead to a possible early breakdown of linear theory at large scales due to spatial nonlocality. We

obtain a simple fitting formula for the nonlinear corrections at large scales and we interpret this in terms of a diffusion process. We also investigate the case with $n = 0$ and we find that it differs from the other cases.

7. We use the perturbative calculations described above to numerically investigate a widely discussed universal relation between the pair velocity and the average of the two point correlation. We find that in the weakly nonlinear regime there is no universal relation between these two quantities.

8. The Zel'dovich approximation (ZA) is used to study some of the issues that have been studied perturbatively for the full gravitational dynamics (GD) in the previous chapters. We investigate whether it is possible to study perturbatively the transition between a single streamed flow and a multi-streamed flow. We do this by calculating the evolution of the two point correlation function using two methods: a.) Distribution functions b.) Hydrodynamic equations without pressure and vorticity. The latter method breaks down once multi-streaming occurs whereas the former does not. We find that the two methods give the same results to all orders in a perturbative expansion. We thus conclude that we cannot study the transition from a single stream flow to a multi-stream flow in a perturbative expansion. We expect this conclusion to hold even if we use the full GD instead of ZA, as already checked at the lowest order of nonlinearity.

9. We calculate nonperturbative expressions for the evolution of the two point correlation function, the pair velocity and its dispersion in the Zel'dovich approximation. We numerically investigate these formulae at various scales.

10. We use ZA to look analytically at the evolution of the two point correlation function at large spatial separations and we find that until the onset of multi-streaming the evolution can be described by a diffusion process where the linear evolution at large scales gets modified by the rearrangement of matter on small scales. We compare these results with the lowest order nonlinear results from GD. We find that the difference is only in the numerical value of the diffusion coefficient and we interpret this physically.

11. We also use ZA to study the induced three point correlation function. At the lowest order of nonlinearity we find that, as in the case of GD, the three point correlation does not necessarily have the hierarchical form. We also find that at large separations the effect of the higher order terms for the three point correlation function is very similar to that for the two point correlation and it can be described in terms of a diffusion process.

References

[1] Padmanabhan T. 1993, Structure Formation in the Universe, (Cambridge, Cambridge University Press)

[2] Somnath B. 1994 *ApJ.*, **428**, 419

[3] Somnath B. 1996 *ApJ.*, **460**, 50

[4] Somnath B. 1996 *ApJ*, (To appear in 10 Nov issue)

RADIATIVE CORRECTIONS TO GRAVITATIONAL COUPLING OF NEUTRINOS AND NEUTRINO OSCILLATIONS

Subhendra Mohanty[1]

Physical Research Laboratory
Ahmedabad 380 009 *[1cm]

Abstract

It is well known that if the gravitational coupling of neutrino's is flavour dependent then this can give rise to neutrino oscillations which can explain the solar and atmospheric neutrino anomalies. Once the universality of the gravitational coupling (κ) i.e., the gravitational gauge invariance is violated at the tree level then there is no way of protecting κ from radiative corrections. We calculate the corrections to κ arising from leading order weak interaction diagrams and find that for a given neutrino flavour $\kappa_i \simeq \kappa(1+G_F\, m_{l_i}^2)$ (where $l_i = e, \mu, \tau$ is the corresponding charged fermion). For the neutrino oscillations in the sun with non-degenerate neutrino masses and these gravitational couplings we find that a resonant conversion takes place for neutrinos with masses in the eV scale for both solar and atmospheric neutrinos.

[1] e-mail: mohanty@prl.ernet.in

A journal is what the authors write. The editor doesn't solicit articles; the articles come to him. If the editor has in some way encouraged publication of good papers, promptly, efficiently, and fairly, he has done a little service, but the credit for the quality is not the editor's. It belongs to the astronomical community.

— S. CHANDRASEKHAR

1 Introduction

We consider applications in the regime of linearised gravity in the Minkowski background. The metric $g^{\mu\nu}$ is written as

$$g^{\mu\nu} = \eta^{\mu\nu} + \kappa \ h^{\mu\nu} \tag{1}$$

where $\eta^{\mu\nu}$ is the Minkowski space metric and $h^{\mu\nu}$ is regarded as the graviton field, with coupling $\kappa = \sqrt{8\pi G}$. The coupling of matter to gravitons is determined by expanding the general covariant Lagrangian $\mathcal{L}$ of matter in curved background in the Minkowski background and retaining the linear terms in κ. The universal matter-graviton interaction term thus obtained is given by

$$\delta\mathcal{L} = (\frac{\delta\mathcal{L}}{\delta g^{\mu\nu}})|_{g^{\mu\nu}=h^{\mu\nu}} \ \ (\kappa h^{\mu\nu}) = \kappa T_{\mu\nu} \ h^{\mu\nu} \tag{2}$$

where $T^{\mu\nu}$ is the matter field stress tensor in the Minkowski background. The general covariance of Einstein's gravity reduces in the linearised case to the transformations

$$\begin{aligned} x^\mu &\rightarrow x'^\mu = x^\mu + \epsilon^\mu \\ h^{\mu\nu} &\rightarrow h'^{\mu\nu} = h^{\mu\nu} - \partial^{(\mu}\epsilon^{\nu)} \end{aligned} \tag{3}$$

The graviton matter coupling are of the universal form **??** and are invariant under the "gauge transformation" 3 and Lorentz transformations. If one starts with a general theory of spin 2 massless prticles interacting with matter then in general one could have different couplings κ_i for different matter species. This theory would then violate the equivalence principle. It was shown by Weinberg [**?**] that if one imposes the invariance under gauge transformation 3 of a general n-point scattering amplitude with some gravitons in the external legs then the universality of κ for all matter (including self coupling with gravitons) follows. In other words gravitational gauge invariance implies the equivalence principle.
We would like to explore the phenomenological consequences of dropping the gravitational gauge invariance or the equivalence principle. In a gauge theory, gauge invariance in the form of the Ward identity $q^\mu \Pi_{\mu\nu}(q^2) = 0$ on the vacuum polarisation amplitude of the gauge bosons ensures that $\Pi_{\mu\nu}(q^2)$ is transverse i.e., $\Pi_{\mu\nu}(q^2) = (g_{\mu\nu}q^2 - q_\mu q_\nu)$ and the gauge boson does not acquire a mass due to radiative corrections to the propagator. The Ward identity for the gravitational two point function is given by [11]

$$q^\mu \Pi_{\mu\nu\rho\sigma} + (\delta_{\mu\rho}q_\sigma + \delta_{\mu\sigma}q_\rho - \delta_{\rho\sigma})\Pi^{\alpha}_{\mu\nu\,\alpha} = 0 \tag{4}$$

Due to the presence of the second term in 4 gravitational gauge invariance does not ensure that the vacuum polarisation of the graviton be transverse and the graviton is not protected from acquiring a mass by radiative corrections. Since gauge invariance in gravity does not play the same role as in QED one can in principle explore the consequences of dropping it without running into any fundamental problem.

2 Graviton-neutrino vertex corrections by weak interactions

We compute diagrams of the form shown in Fig 1. which give rise to weak corrections to the tree level graviton-photon interaction

$$\kappa h^{\mu\nu}(\frac{1}{4})\bar{\nu}(p_1)\gamma_{(\mu}(p_1 + p_2)_{\nu)}\nu(p_2) \tag{5}$$

The graviton charged lepton vertex in Fig 1. is of the same form as 5. The graviton-vector boson vertex is given by

$$\kappa h^{\mu\nu}(F^{\alpha}_{\mu}F_{\alpha\nu} + \frac{1}{4}\eta_{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}) \tag{6}$$

We chose the harmonic gauge where the coupling of $h^{\mu\nu}$ is to $\tilde{T}_{\mu\nu} = (T^{\mu\nu} - \frac{1}{2}g^{\mu\nu}T^{\alpha}_{\alpha})$ and therefore the mass dependent terms are traced out from $\tilde{T}_{\mu\nu}$ in 6. The diagram Fig.1a has the amplitude given by

$$\begin{aligned} i\mathcal{L} &= \frac{1}{4}(\frac{ig}{\sqrt{2}})^2 \int \frac{d^4p}{(2\pi)^4}\ \bar{\nu}_i(k_2)\ \gamma_{\alpha'}\ P_L\ i(\not{k}_2 - \not{p} + m_i)\ \gamma_{(\mu}(k_1+k_2)_{\nu)} i(\not{k}_1 - \not{p} + m_i) \\ & \gamma_\alpha\ P_L\ (-ig^{\alpha\alpha'})\ I(k_1,k_2,p)\ \nu_i(k_1) \end{aligned}$$

where

$$I(k_1,k_2,p) = \int dxdy \frac{2x}{[(p^2 - M_w^2)xy + ((k_1-p)^2 - m_i^2)x(1-y) + ((k_2-p)^2 - m_i^2)(1-x)y]^3} \quad (7)$$

Here ν_i with $(i = e, \mu, \tau)$ are the flavour eigenstates of the neutrino and m_i are the corresponding charged lepton masses. In the limit of zero momentum transfer $(k_1 - k_2) \to 0$ the leading order term in the amplitude 7 is of the form

$$i\mathcal{L} = \frac{1}{4}(\frac{g}{4\pi})^2[(\frac{m_i}{M_w})^2(1 + ln(\frac{{M_w}^2 - {m_i}^2}{{M_w}^2})(\kappa\ h^{\mu\nu})\ \bar{\nu}_i(k_2)\ \gamma_{(\mu}(k_1+k_2)_{\nu)}\nu_i(k_1) \quad (8)$$

The contribution of diagram shown in Fig 1b. is suppressed by a factor $({m_\nu}^2/{M_w}^2$ with respect to the contribution of diagram 1a.
Comparing the coefficients of the vertex operators given by 5 and 8 we see that the extra contribution to the diagrams Fig 1 is a flavour dependent term given by

$$\kappa_i = \kappa\ (1 + \Delta f_i) \quad (9)$$

with

$$\Delta f_i = (\frac{g}{4\pi})^2[(\frac{m_i}{M_w})^2(1 + ln(\frac{{M_w}^2 - {m_i}^2}{{M_w}^2}) \quad (10)$$

In the flavour basis the gravitational couplings of the neutrinos are diagonal due to the fact that weak interactions conserve the lepton number in the absence of neutrino masses. With non-zero neutrino masses there are off-diagonal terms in the κ_{ij} matrix but the off diagonal terms are suppressed w.r.t the diagonal ones by a factor $\sim {m_\nu}^2/{M_w}^2$.

2.1 Neutrino oscillations by VEP and mass terms

When violation of equivalence principle (VEP) occurs by weak radiative corrections then the flavour weak interaction couplings and the gravitational couplings are both diagonal in the same basis. For neutrino oscillations to occur during propagation there has to be a mass matrix with nonzero off-diagonal elements.
For simplicity, we shall assume there are only two neutrino flavors. In the flavour basis the evolution of neutrino flavor in a medium is described by [3, 4]

$$\begin{aligned} i\frac{d}{dt}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} &= \Big\{ \frac{\Delta m^2}{4E} U_M \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} U_M^\dagger \\ &- 2E|\phi(r)|\Delta f\ \ U_G \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} U_G^\dagger \\ &+ \frac{\sqrt{2}}{2} G_F N_e \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \Big\} \begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} \end{aligned} \quad (11)$$

Here $\Delta m^2 \equiv m_2^2 - m_1^2$ denotes the difference in neutrino vacuum masses, N_e is the electron density of the medium, and G_F is Fermi's constant. The first term describes the contribution from vacuum masses, the second term describes the contribution from equivalence principle breaking, and the third term describes the contribution from a background of normal matter. There are two unitary matrices which parametrize the mixing, U_M and U_G, and they are generally completely unrelated. If VEP occurs via weak interaction loops then U_G is diagonal in the flavour basis and Δf_i is given by 10. The condition of resonant oscillations is when the trace of r.h.s becomes zero

$$\frac{\Delta m_\nu{}^2}{2E} Cos(2\theta) - 2E\Phi\Delta f = 0. \tag{12}$$

Where θ is the mixing angle (in vacuum) between the mass and the weak eigenstates and Φ is the Newtonian potential. For the case of solar neutrinos $\Phi \sim 10^{-6}$ and $E \sim 10MeV$ and the resonant conversion occurs for neutrino mass differnce of the order $\Delta m_\nu \sim eV$. Interestingly for the atmospheric neutrinos, $\Phi \sim 10^{-9}$ and $E \sim 1GeV$ and again resonant conversion occurs for neutrino mass difference of the order $\Delta m_\nu \sim eV$.

Acknowledgement

This talk is based on work in progress in collaboration with Anjan Joshipura.

References

[1] S. Weinberg, *Phys. Rev.*, **135**, B1049 (1964)

[2] Oxford symposium on Quantum Gravity, Oxfor University Press, (1975)

[3] M. Gasperini, *Phys. Rev. D*, **38**, 2635 (1988); *Phys. Rev. D*, **39**, 3606 (1989)

[4] A. Halprin and C. N. Leung, *Phys. Rev. Lett.*, **67**, 1833 (1991); *Nucl. Phys. B*, (Proc. Suppl.) **28A**, 139 (1992)

In assessing an artist, one often distinguishes an early, a middle, and a late period; and the distinction is generally one of growing maturity and depth. But this is not the way a scientist is assessed: he (or she) is assessed by the significance of one or more of the discoveries that he (or she) may have made in the realm of ideas or in the realm of facts. And, it is often the case that the most "important" discovery of a scientist is his first. In contrast, the deepest creation of an artist is equally often his last. I continue to be puzzled by this dichotomy.

— S. CHANDRASEKHAR

TOPOLOGICAL DEFECTS IN COSMOLOGY

Pijushpani Bhattacharjee[1]

Indian Institute of Astrophysics
Koramangala
Bangalore 560 034, India

Abstract

Topological Defects like magnetic monopoles, cosmic strings, domain walls, etc., could have formed during symmetry breaking phase transitions in the early universe. We give an overview of the implications of topological defects for cosmology, with particular emphasis on their possible role as sources of the extremely high energy cosmic rays detected in several recent cosmic ray experiments; these experiments may indeed provide the signature of existence of topological defects in the Universe.

[1]e-mail: pijush@iiap.ernet.in

…still I cannot close without stating the simple fact that I and everyone of my colleagues who share the title of astrophysicist is personally in your debt. It is a very happy event to see a portion of that debt recognized in this fashion.

— A.A.PENZIAS.

Please accept my warmest congratulations. It is hard to imagine a Nobel Award that would be as widely and enthusiastically approved by physicists and astronomers.

— E.M.PURCELL.

1 Introduction

Extended particle-like non-dissipative solutions of classical non-linear field theories, variously called *solitons*, *lumps*, etc., have been known and studied for a long time. For a comprehensive review of the subject, see for example, Refs. [1, 2]. The importance of these solutions came into sharp focus with the discovery in the early seventies that spontaneously broken gauge theories (SBGTs) possessed these finite-energy, non-singular, soliton-like solutions, namely, the *vortex* solution of Nielsen and Olesen simplest of spontaneously broken gauge theories, and the *magnetic monopole* solution of 't Hooft [4] and Polyakov [5] in the simplest spontaneously broken *non-abelian* gauge theory.

A fundamental notion in a SBGT is the existence of more than one degenerate vacua (indeed a whole continuum of them, in general), which allows one to choose non-trivial boundary conditions for the fields and thereby construct topologically non-trivial solutions. Typically, these solutions represent extended objects with a 'core', at the centre of which the symmetry under consideration is *unbroken*, while outside the core the symmetry is spontaneously broken. It is in this sense that these solutions are generally called topological *defects* (TDs)— they represent regions of space within which the underlying fields are constrained, due to the special topological properties of these solutions, to remain in the 'false' vacuum of unbroken symmetry as opposed to the 'true' vacua representing the broken symmetry outside the regions. The 'center' of a defect could be a point, a line, or a 2-dimensional surface, corresponding respectively to point-like, line-like and surface-like defects; in the cosmological context, these are usually referred to as monopole, cosmic strings, and domain walls, respectively.

Topological Defects as physical objects can be formed in a system that exhibits the phenomenon of spontaneous symmetry breaking (SSB). Often, the phenomenon of SSB in a physical system is manifested in the form of a *phase transition* as the system is cooled through a critical temperature. Laboratory examples of TDs include vortex filaments associated with superfluid phase transition in liquid ^{4}He, magnetic flux tubes associated with superconducting phase transition in Type-II superconductors, disclinations and dislocations associated with isotropic-to-nematic phase transition in liquid crystals, and so on. Similarly, in the cosmological context, TDs could have been formed during SSB phase transitions in the early Universe as the Universe expanded and cooled through certain critical temperatures.

The link between temperature and symmetries is well known. One well-known example is the Ferromagnet: A piece of iron cooled below the Curie temperature becomes a ferromagnet; the full rotation symmetry of the piece of iron (in which the spins are randomly aligned) is broken below the Curie temperature — the spins get aligned along a particular direction which defines the magnetization vector. So, a higher temperature in general corresponds to more symmetry, and symmetry can be 'broken' as temperature is reduced. Conversely, a symmetry can also be restored at high temperatures— the ferromagnet loses its magnetization when heated above the Curie temperature, and the full rotation symmetry is restored. When a symmetry is broken, it need not always be *completely* broken, however; in the above case of ferromagnet, there is a residual symmetry left unbroken, namely the invariance under rotation about the direction of magnetization. Perhaps we should also mention here that there are counterexamples, at least theoretical models[6] in which a symmetry broken at a high temperature can be *restored* at a lower temperature. These models are, however, rather special.

In the example of ferromagnet above, the relevant symmetry is the rotation symmetry in the physical space— it is a *space-time* symmetry associated with a transformation of the space-time coordinates. It was shown in the mid-seventies [7, 8, 9] that gauge theories involving *internal* symmetries (i.e., symmetries associated with transformations of the fields themselves) may also possess similar behaviour— the internal symmetries broken at low temperatures can be restored at higher temperatures. Thus within the context of the Big-Bang model of the Universe, according to which the Universe was hotter (and denser) in the past than it is now, it is possible that the spontaneously broken (internal) symmetries incorporated in unified gauge theories (of which the electroweak theory is the most successful example) were actually unbroken at some sufficiently early time. In particular, it is possible that the full symmetry of the so-called Grand Unified Theories (GUTs) that describe unified strong and electroweak interactions was unbroken at some epoch in the early Universe, and as the Universe expanded and cooled, a series of phase transitions took place by way of which the full GUT symmetry was successively broken to the presently observed fundamental unbroken symmetry of the Universe, namely, the symmetry under the group $\mathrm{SU}(3)_{\mathrm{c}} \times \mathrm{U}(1)_{\mathrm{em}}$, where $\mathrm{SU}(3)_{\mathrm{c}}$ denotes the 'color' symmetry group that describes strong interaction of quarks and gluons, and $\mathrm{U}(1)_{\mathrm{em}}$ describes the electromagnetic interaction. Various kinds of TDs, namely, magnetic monopoles, cosmic strings, domain walls, textures, etc. [10] would have formed at some of these symmetry breaking phase transitions. As we shall see below, the typical mass (energy) scale of the TDs are roughly of same order as the energy (or temperature) scale of the symmetry breaking phase transition at which the particular kind of TD is formed. Thus, TDs formed at a GUT scale phase transition (at an energy or temperature scale of $\sim 10^{16}\,\mathrm{GeV}$) would be very massive and so they would have tremendous implications for cosmology mainly through their gravitational interactions with matter. For a nice summary of the subject of TDs, their formation, evolution and their (mainly gravitational) implications for cosmology in general, see the recent monograph by Vilenkin and Shellard[11]. In this talk I will not deal much with those well-known gravitational implications of TDs, but will rather concentrate on some *non-gravitational* implications of TDs; in particular, I shall discuss the possibility that TDs, under certain circumstances, may be sources of extremely energetic particles in the Universe, and that some of the currently operating and planned future detectors of extremely high energy cosmic rays (with energy $\gtrsim 10^{20}\,\mathrm{eV}$) may provide direct signatures of TDs in the Universe.

In section 2 we briefly discuss the nature of TDs in theories that exhibit SSB. For illustration we explicitly discuss three different kinds of TDs, namely, domain walls, cosmic strings and monopoles, by considering the simplest models of SSB in which these objects are possible, and also discuss the general topological classification of TDs. In section 3 we discuss the mechanism of cosmological formation of TDs during phase transitions in the early Universe, and briefly discuss the cosmological constraints on domain wall and magnetic monopole defects. Finally, in section 4, we discuss the possibility that TDs, under certain circumstances, may be the sources of the observed highest-energy cosmic rays.

Unless otherwise stated, I use natural units with $\hbar = c = k_B = 1$, where k_B denotes the Boltzmann constant. I should emphasize that this talk is not intended to be an exhaustive review of the subject of TDs—-for this purpose, see Ref.[11].

2 Topological defects and their classification

2.1 Spontaneous Symmetry Breaking (SSB) and Topological Defects (TDs)

2.1.1 Domain Wall

Consider the theory of a single scalar field ϕ described by the Lagrangian density

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - V(\phi)\,, \tag{1}$$

where the "potential" $V(\phi)$ is chosen as

$$V(\phi) = \frac{1}{4}\lambda\left(\phi^2 - \eta^2\right)^2\,, \tag{2}$$

where η is a constant. The shape of this potential is shown schematically in Fig. 1.

This Lagrangian with the above "double-well" potential is invariant under the symmetry $\phi \to -\phi$. For the case of a static field ($\dot{\phi} = 0$) the equilibrium or the lowest-energy state of the field corresponds to the minimum of the potential $V(\phi)$. The above potential has *two* minima, and so there are two lowest-energy or ground states of the theory corresponding to $\phi \equiv \phi_+ = \eta$ and

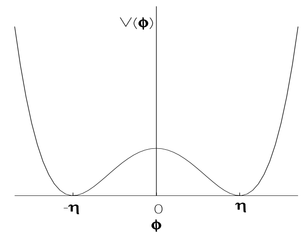


Figure 1: The double-well potential

$\phi \equiv \phi_- = -\eta$. The symmetry $\phi \to -\phi$ is *spontaneously broken* because although the Lagrangian is invariant under the above symmetry, the ground states themselves are not ($\phi_+ \leftrightarrow \phi_-$).

Now suppose in one region of the universe we find the field to be in the state $\phi = \phi_+$ and in a neighbouring region $\phi = \phi_-$. (We shall discuss later why we may expect this to happen in general in a cosmological situation.) Then the spatial continuity of the field ϕ implies that along any line joining the two regions the field must pass through the value zero at least once. Indeed, in 3 spatial dimensions, the set of points where the field is zero forms a 2-dimensional surface — a "domain wall" that separates a region where $\phi = \phi_+$ from a neighbouring region where $\phi = \phi_-$. The domain wall region (in which $\phi \approx 0$) has a higher energy density than the background (in which $\phi = \pm\eta$) because of the higher potential energy density $V(\phi)$ there and also because of the contribution from the non-zero gradient term, $(\nabla\phi)^2$, across the wall.

The equation of motion obtained from the Lagrangian (1) admits a closed form analytical solution that represents a domain wall. Thus, e.g., one can check that there exists a solution, namely,

$$\phi_{\rm wall}(x) = \eta \tanh(x/\Delta)\,, \tag{3}$$

which represents an infinite, static domain wall in the yz-plane. (An arbitrarily moving domain wall solution can be obtained by applying an appropriate Lorentz transformation to the above solution). The 'core' of the wall is at $x = 0$, and the wall has a "thickness" characterized by $\Delta = (\frac{2}{\lambda})^{1/2}\eta^{-1}$. The above solution, shown schematically in Fig. 2, interpolates between $\phi = -\eta$ at $x = -\infty$ and $\phi = +\eta$ at $x = +\infty$.

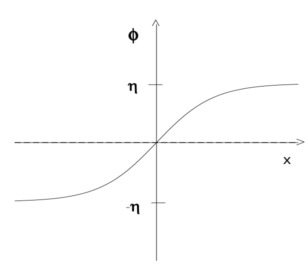


Figure 2: The domain wall solution

The topological stability of the solution is ensured by the above non-trivial boundary conditions.

(A trivial boundary condition would be one in which ϕ takes the same value, either $+\eta$ or $-\eta$, at both $x = +\infty$ and $x = -\infty$.) Thus no continuous rearrangement of the field configuration can get rid of the "kink" (see Fig. 2) and the associated energy density of the solution. Thus a domain wall once formed cannot disappear by itself unless it annihilates with an "anti-wall" (characterized by the solution with the opposite boundary conditions, i.e., $\phi = +\eta$ at $x = -\infty$ and $\phi = -\eta$ at $x = +\infty$).

The characteristic thickness of the wall, Δ, is determined by the need to minimize the surface energy density (Σ) associated with the wall. There are two contributions to Σ: (a) The gradient term, $\sim \Delta \times (\nabla\phi)^2 \sim \eta^2/\Delta$, which tends to make Δ as large as possible, and (b) the potential energy term, $\sim \Delta \times V(\phi) \sim \Delta\lambda\eta^4$, which tends to make Δ as small as possible. The balance between these two competing terms gives $\Delta \sim \lambda^{-1/2}\eta^{-1}$. Substituting this back in the gradient term and the potential energy term, one gets $\Sigma \sim \lambda^{1/2}\eta^3$. A more rigorous calculation in terms of the full energy-momentum tensor obtained from the Lagrangian (1– 2) gives $\Sigma = \frac{2\sqrt{2}}{3}\lambda^{1/2}\eta^3$.

We have discussed the domain wall solution in some details because it illustrates several general features of all defect solutions (except for textures). These general features are the following: (a) The mass/energy density scale associated with the defect is essentially fixed by the magnitude of the "vacuum expectation value" (VEV), $\langle\phi\rangle$ of the scalar field ϕ (i.e., its value in the symmetry-breaking vacuum, which for the potential (2) is η). (b) The asymptotic configuration of the symmetry breaking scalar field determines the existence of the defect. Note that existence of multiple degenerate vacua (allowed in theories with SSB) is an essential pre-requisite, for it allows us to choose the non-trivial asymptotic field boundary conditions required for the defect solutions to exist. (c) The 'location' of the defect is the point of space where the scalar field is forced to remain in the symmetric state.

These general features will be more apparent when we consider other kinds of possible topological defects below.

2.1.2 Vortices and Cosmic Strings

Domain walls as topological defects arise in theories with spontaneously broken *discrete* symmetries. To illustrate other kinds of defects, let us now consider a theory with a spontaneously broken *continuous* symmetry. More specifically, consider now the theory of a *complex* scalar field ϕ described by the Lagrangian density

$$\mathcal{L} = (\partial_\mu\phi^*)(\partial^\mu\phi) - V(\phi)\,, \quad V(\phi) = \frac{1}{2}\lambda\left(|\phi|^2 - \frac{1}{2}\eta^2\right)^2 . \tag{4}$$

The potential $V(\phi)$ in (4) is the "Maxican-hat" potential, the generalization of the "double-well" potential of Eq. (2) to the case of a complex-, i.e., two real-component scalar field. The Lagrangian (4) has a global U(1) symmetry, under transformations $\phi \to \phi\exp(i\alpha)$, with α a real constant. A ground state, or vacuum solution, corresponding to the minimum of the potential, is $\phi_0 = (\eta/\sqrt{2})\exp(i\alpha_0)$, where α_0 is an arbitrary real constant. This is not the only ground state, however. Indeed, there is now an infinite number of degenerate ground states, obtained by simply applying U(1) transformations $\exp(i\alpha)$ with different values of α to the ground state ϕ_0. Thus any chosen ground state is not invariant under U(1) transformations, although the Lagrangian (4) is. In other words, the global U(1) symmetry of the Lagrangian (4) is spontaneously broken by the vacuum. This theory has a scalar particle with mass $m_s = \sqrt{\lambda}\,\eta$, and also a massless scalar particle, the so-called Nambu-Goldstone boson, associated with the spontaneously broken global symmetry.

The existence of the whole infinity of ground states again allows us to construct a defect solution by choosing the appropriate asymptotic boundary conditions on the complex field ϕ. To illustrate the "vortex" defect, let us go over to cylindrical polar coordinates $\{\rho, \theta, z\}$, and consider a closed circular loop C of large radius in the xy-plane. Let us assume that on the loop C the complex field ϕ has the following cylindrically symmetric configuration:

$$\phi(\rho \to \infty, \theta) = \frac{\eta}{\sqrt{2}}\exp(in\theta)\,, \tag{5}$$

where n is an integer. The field configuration (5) thus defines a mapping of a circle in the xy-plane into the manifold of vacua which in the present case is also a circle. Let us take the field configuration (5) as defining our boundary conditions for the field ϕ. It is then easy to see that if the field ϕ is spatially continuous, and if it is to satisfy the boundary conditions (5), then there must be at least one point on the disk bounded by the closed curve C at which $\phi = 0$. This point in the xy-plane defines the location of a topologically stable defect, the *vortex*, where the potential energy density is maximum and the symmetry is unbroken. In fact one can show that the Euler-Lagrange equations of motion derived from the Lagrangian (4) admit static solutions (with non-zero energy density) representing vortex *lines* in 3-dimensional space. For example, for an infinitely long straight static vortex line along the z-axis at origin, the general solution can be written as

$$\phi(\rho,\theta) = \frac{\eta}{\sqrt{2}}\, f(\xi)\, \exp(in\theta)\,, \tag{6}$$

where $\xi = m_s\rho$. Unfortunately, unlike the case of domain wall solution, it is not possible to write the function $f(\xi)$ in terms of any known function; however, it can be easily obtained numerically. The boundary condition on f is that $f \to 1$ for $\xi \to \infty$, and the continuity of ϕ demands that $f \to 0$ as $\xi \to 0$. Thus at all points on a vortex line the field ϕ is zero, and far away from this 'nodal line' the field takes on one of the vacuum values given by Eq. (5). Moreover, the phase of the complex field ϕ turns by an integral multiple of 2π as we make one complete circuit along any closed curve encircling any point on the vortex line. The above vortex defects are usually referred to as "global" vortices or "global" strings, because they arise in theories with a spontaneously broken *global* U(1) symmetry. The global strings are closely related to the vortex filaments in superfluid ^{4}He where the complex scalar field represents the wave function of the condensed ^{4}He atoms. The core of a global string has a thickness $\sim m_s^{-1}$ within which most of the energy density of the defect is located. However, it turns out that a global string has a formally infinite energy per unit length; within a cylinder of radius ρ around a straight vortex line the energy per unit length is proportional to $\eta^2 \ln(m_s\rho)$, which diverges logarithmically with ρ. This divergence is traced to the existence of the massless Nambu-Goldstone scalar particle in the theory, which gives rise to long range interactions in the problem. However, in a finite system or in the cosmological context (where the cosmological horizon or the distance to the nearest string provides a cutoff radius), the global strings have effectively finite energy per unit length.

Let us now consider a theory with a spontaneously broken *local* U(1) symmetry. The simplest theory of this type is the so-called Abelian Higgs model, which is the theory of a complex scalar field ϕ interacting with a vector field A_μ. It is described by the Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) - V(\phi)\,, \tag{7}$$

where $D_\mu \equiv \partial_\mu + ieA_\mu$ is the covariant derivative (e being the gauge coupling constant), $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$, and the potential $V(\phi)$ is the same as in Eq. (4). The Lagrangian (7) is invariant under the *local* (i.e., space-time dependent) U(1) transformations

$$\phi \to \phi\exp\{i\alpha(x)\}\,, \quad A_\mu \to A_\mu - \frac{1}{e}\partial_\mu\alpha(x)\,, \tag{8}$$

where α is a real single-valued, but otherwise arbitrary, function of space-time. The form of the potential $V(\phi)$ again implies that the local U(1) symmetry is spontaneously broken by the vacuum states of the field ϕ. So, like the global string defects described above, we can obtain now a "local" or "gauge" string defect solution which has the same asymptotic configuration for the field ϕ as implied by Eq. (5). However, the presence of the gauge field A_μ now makes the properties of the local string solution very different from the global string. First note that the particle spectrum of the theory again contains a massive scalar field (the so-called "Higgs boson") with a mass $m_H = \sqrt{\lambda}\eta$, but in contrast to the global case, the massless Nambu-Goldstone boson which, by virtue of its long-range interactions, was responsible for making the energy per unit length of global strings formally infinite, is now absent. Instead, there is now a massive gauge field A_μ with a mass

$m_v = e\eta$. The asymptotic configuration of the gauge field in the local vortex string solution can be chosen such that, by its coupling with the ϕ field through the covariant derivative $D_\mu\phi$, it cancels the divergent part of $\int (\vec{\nabla}\phi)^2 \mathrm{d}^2\mathrm{x}$ which in the global string case had led to an infinite energy per unit length. In other words, the energy per unit length of the local string, denoted by μ hereafter, is finite. The asymptotic configuration of the gauge field for a straight static local string along the z-axis can be written, with an appropriate 'gauge choice', as

$$A_0 = 0\,, \quad \vec{A}(\rho \to \infty) = \frac{n}{e\rho}\hat{\theta}\,, \tag{9}$$

where $\hat{\theta}$ is the unit vector in the azimuthal direction, and the integer n is the same integer that describes the winding number of the ϕ as in Eq. (5). All components of the gauge field, except the azimuthal component, are zero everywhere. Also, ϕ is zero at the origin, i.e., the local U(1) symmetry is unbroken at the centre of the vortex. The "magnetic" field $\vec{B} = \vec{\nabla} \times \vec{A}$ is along the $\hat{z}$ direction. The full solutions for ϕ and A_θ can be obtained numerically (see [11] for details). The solutions are parametrized by the quantity $\beta = \lambda/e^2 = (m_H/m_v)^2$. An important point to note is that the total magnetic flux is in fact quantized: By Stokes theorem,

$$\int \mathrm{d}^2\mathrm{x}\vec{B}\cdot\hat{z} = \int_0^{2\pi} \rho\mathrm{d}\theta\frac{n}{e\rho} = n\frac{2\pi}{e}\,. \tag{10}$$

The local string thus contains a tube of "magnetic" flux quantized in integer units of a basic flux quantum of magnitude $\frac{2\pi}{e}$. The magnetic flux also makes a local string 'orientable'— if the solution characterized by a positive value of n is defined as the "string" ($\vec{B}$ along $+\hat{z}$), then that with $-n$ can be called "antistring" ($\vec{B}$ along $-\hat{z}$). It can be shown that for $\beta > 1$, a local string with $|n| > 1$ is unstable with respect to dissociation into n number of strings each with $n = 1$. If, however, $\beta < 1$, then a local string is stable for any value of n. For $\beta = 1$, the n-string is neutrally stable. All this can be understood in terms of the ranges of the attractive (for the scalar field) and repulsive (for the vector field) interactions in the theory. Most importantly, however, an $|n| = 1$ string is always 'topologically' stable; it can never relax to the 'trivial' configuration with $n = 0$ because the winding number n is a topologically conserved quantity. Thus a string once formed cannot by itself disappear unless it annihilates with an antistring. (We shall see later that the latter eventuality of annihilation is indeed possible under certain physical circumstances, and it has important implications.)

Finally, as already mentioned, the energy per unit length of the local string is finite, $\mu \sim \eta^2$, and is determined primarily (up to a numerical factor of $O(1)$) by the vacuum expectation value of the scalar field ϕ. Actually μ also depends on the parameter β, but that dependence is rather mild (see [11]). The size of the core, or the thickness of the string is $\sim \eta^{-1}$.

The local strings are analogous to quantized magnetic flux tubes that appear inside a Type II superconductor placed in an external magnetic field of strength greater than some critical value. This is not surprising since the Abelian Higgs model Lagrangian (7) is, in fact, just the relativistic version of the phenomenological Ginzburg-Landau Lagrangian that in condensed matter physics describes the superconducting phase transition in certain conductors at low temperatures. The (local) string solution of the abelian higgs model was first obtained by Nielsen and Olesen [3].

In the cosmological context, the vortex strings resulting from U(1) symmetry breaking phase transitions in the Universe are called *cosmic strings*.

2.1.3 Magnetic monopoles

Domain walls and vortex strings are examples of 2-dimensional (surface-like) and 1-dimensional (line-like) defects, respectively, in the 3-dimensional space. It is also possible to have *point-like* defects, called *magnetic monopoles* (for reasons that will become clear shortly). These appear in some spontaneously broken *non-abelian* gauge theories. The simplest theory allowing magnetic monopole solution [4, 5] incorporates the SSB scheme SO(3)→ U(1). Here SO(3) refers to the rotation group in the 3-dimensional "internal" space generated by a triplet of scalar fields, ϕ^a, with $a = 1, 2, 3$. (Thus the scalar field ϕ in the present case can be regarded as a "vector", $\vec{\phi}$, in the 3-dimensional internal space.) The potential $V(\phi)$ in the Lagrangian that incorporates this symmetry breaking can be chosen as

$$V(\phi) = \frac{1}{4}\lambda(\phi^a\phi^a - \eta^2)^2 \,. \tag{11}$$

Clearly, this potential is invariant under transformations belonging to the group SO(3); these transformations preserve the 'length' $|\phi| = \sqrt{\phi^a\phi^a}$ of the scalar field. One can write down the full Lagrangian invariant under local SO(3) transformations in the standard way by introducing the required triplet of gauge fields A_μ^a ($a = 1, 2, 3$), and the associated covariant derivatives. In the vacuum state, $|\phi| = \eta$. However, the 'direction' of $\vec{\phi}$ in the internal space is not fixed. Thus, there is again a whole manifold of degenerate vacua. The vacuum manifold is just S^2, a 2-sphere of 'radius' η. The SO(3) symmetry is, therefore, spontaneously broken. However, it is not completely broken; all rotations about a given ϕ^a leave it invariant. Thus a subgroup U(1) of SO(3) remains unbroken. 't Hooft [4] and Polyakov [5] showed that this theory admits a spherically symmetric static defect solution which has the following asymptotic boundary conditions:

$$\phi^a(r \to \infty) = \eta\hat{\mathbf{r}}^a \,, \tag{12}$$

$$A_i^a(r \to \infty) = \epsilon^{iaj}\frac{\hat{\mathbf{r}}^j}{er} \,, \quad A_0^a = 0 \,. \tag{13}$$

Here r is the spatial radial coordinate, and $\hat{\mathbf{r}}$ the radial unit vector in the spherical polar coordinate system. The 'hedgehog' configuration of the scalar field triplet, implied by Eq.(13), guarantees the topological stability of this defect solution. Like in the domain wall and cosmic string case, the continuity again implies that the scalar field must vanish as $r \to 0$. Thus the field $\vec{\phi}$ retains the full SO(3) symmetry at $r = 0$, while far away from this point only a U(1) symmetry remains. The remarkable property of this defect solution, as shown by 't Hooft[4] and Polyakov[5], is that the asymptotic configuration of the gauge field corresponds to a radially outward magnetic field corresponding to the *unbroken* symmetry U(1). Now, in the context of realistic unified theories of fundamental interactions in nature, the only known unbroken U(1) symmetry is the one that corresponds to $\mathrm{U(1)_{em}}$, the gauge group of electromagnetism. One can, therefore, identify the above unbroken U(1) symmetry with $\mathrm{U(1)_{em}}$. Therefore, the "magnetic" field of the defect at infinity is actually the magnetic field we are familiar with in electromagnetism. Moreover, the radial nature of the magnetic field implies that the defect is actually a magnetic *monopole*; it can be further shown that the magnetic charge carried by this solution is $g = -4\pi/e$. The energy associated with the monopole is $\sim \frac{4\pi}{e}\eta$, which can be regarded as the mass of the monopole. Since any model of GUT must necessarily incorporate in it the unbroken $\mathrm{U(1)_{em}}$ symmetry, it follows that the monopole solution is allowed in all realistic GUT models. They would be very massive, with a mass of $\sim 10^{16}$ GeV, for a GUT symmetry breaking scale given by $\eta \sim 10^{14}$ GeV. This has important cosmological consequences. Note that, like domain walls and cosmic strings, the topological stability implies that a monopole once formed cannot by itself disappear; the only way it can do so is by annihilating with an antimonopole whose asymptotic configuration is given by Eq.(13) with a negative sign on the right hand side.

There is a vast amount of literature on magnetic monopoles because of their fascinating and often intriguing properties, for which the reader can consult, e.g., Refs. [11, 12, 13].

2.1.4 Other defects

There is a variety of other possible topological defects discussed in literature. These include, global texture, superconducting cosmic strings and domain walls, hybrid or composite defects (such as a monopole-antimonopole pair connected by a cosmic string in between, domain walls bounded by cosmic strings, and so on), and the so-called embedded or semi-local defects. We will not discuss these here; the reader may consult Ref.[11] for details.

2.2 Topological classification of Defects

In general, in a theory with a spontaneous symmetry breaking from a symmetry group G to a symmetry group H (which is a subgroup of G), $G \to H$, the degenerate vacua of the scalar field (that implements this symmetry breaking through its potential energy density) form a manifold $\mathcal{M}$. The existence and the kind of possible TDs are determined by the topology of the vacuum manifold $\mathcal{M}$, in particular, by its homotopy groups. We will not go into the details here, but will simply mention that domain walls, cosmic strings, magnetic monopoles, and texture solutions are allowed if $\pi_0(\mathcal{M})$, $\pi_1(\mathcal{M})$, $\pi_2(\mathcal{M})$ and $\pi_3(\mathcal{M})$ are, respectively, non-trivial. Here $\pi_n(\mathcal{M})$ is the n-th homotopy group of the manifold $\mathcal{M}$. A non-trivial $\pi_n(\mathcal{M})$ (for $n \geq 1$) means that not all n-spheres (S^n) in $\mathcal{M}$ can be shrunk to a point. Note that $\pi_0(\mathcal{M})$, although it is often referred to as the "zeroth homotopy group" of $\mathcal{M}$, is not in general a group; it simply counts the disconnected components of $\mathcal{M}$. An important point is that the vacuum manifold $\mathcal{M}$ may, in general, be identified with the coset space G/H; $\mathcal{M} \equiv \mathcal{G}/\mathcal{H}$. This allows one to calculate the relevant homotopy groups just by knowing the topological structures of the groups G and H. For more details on the homotopy classification of TDs, see, e.g., Ref. [11].

3 Cosmological formation of topological defects

3.1 The Kibble mechanism

We have already indicated in the Introduction that several spontaneously symmetry breaking phase transitions may actually have occurred in the early Universe as it expanded and cooled. In the non-zero (actually, high)- temperature environment of the early Universe the potential $V(\phi)$ used in our discussions above is actually a temperature dependent object. In fact, it is the relevant free-energy density of the field, or, more specifically, the so-called finite temperature *effective potential*, $V_T(\phi)$. Consider, for example, the double-well potential of Fig. 1. It can be shown that above a certain critical temperature $T_c \sim \eta$ (where η is the vacuum expectation value (VEV) of ϕ in the broken symmetry phase), the shape of the potential changes to a simple harmonic oscillator potential with a single minimum at $\phi = 0$, but below T_c the potential develops the double-well shape with two degenerate minima at non-zero values of ϕ. Thus initially at $T > T_c$, the Universe can be in an unbroken symmetry phase ($\phi = 0$), but as it cools below T_c, the symmetry is broken, a phase transition takes place, and the field settles down to a minimum of the potential, i.e., ϕ develops a non-zero VEV, $\langle\phi\rangle \neq 0$. However, there is no special reason for all points in the Universe to settle down to the same vacuum. Thus, e.g., for the case of the double-well potential of Eq. (1), as the Universe cools below T_c, the $\langle\phi\rangle$ at different points of the Universe would be randomly chosen to be either $+\eta$ or $-\eta$. Moreover, as first pointed out by Kibble [10], the value of $\langle\phi\rangle$ chosen at two different points would in general be uncorrelated if the two points are separated by a distance greater than the correlation length, ξ, of the field ϕ at the phase transition temperature T_c. But causality necessarily implies that $\xi(t) < d_H(t) \sim t$, where $d_H(t)$ is the causal particle horizon distance in the expanding Universe at time t. Hence, after the phase transition, the field ϕ would lie at different points of the vacuum manifold $\mathcal{M}$ at different spatial points. It will then happen, just by chance, that the field configuration on some asymptotic surfaces (with respect to some points, lines, or 2-surfaces) will be topologically non-trivial like in the case of domain wall-, cosmic string-, and monopole solutions discussed earlier, and so a defect will be present somewhere inside the asymptotic surface. This is the so-called "Kibble mechanism" of formation of TDs in the Universe. Recent laboratory experiments studying defect formation in phase transitions in liquid crystal [14] and in liquid ^{4}He [15] give qualitative support to this general idea of defect formation in cosmology; for more details on this see, e.g., Zurek [16] and references therein.

One simple consequence of the Kibble mechanism is that it gives a rough *lower limit* on the density of defects formed at a phase transition: At least one defect should be formed per horizon-size volume of the Universe at the time of the phase transition.

Defects could also be formed by quantum mechanical nucleation of a defect-antidefect pair during an inflationary phase of the Universe [17]. However, since this is a quantum mechanical process, the rate and number density of formation of defects by this process is small.

3.2 Cosmological constraints on defects

3.2.1 Domain Walls

A topological defect once formed in the early Universe cannot by itself disappear (unless it somehow collapses or annihilates with the corresponding anti-defect), and so it should be present in the Universe today. There are, however, rather strong cosmological constraints on certain kinds of TDs. For example, domain walls formed at a GUT scale phase transition with $\eta \sim 10^{15}$ GeV would have a mass per unit area of the wall of order $\sim \eta^3 \sim 4.6 \times 10^{48}\,\mathrm{g\,cm^{-2}}$. Even a single domain wall of such huge mass density stretched across one Hubble volume (of radius $\sim H_0^{-1} \simeq 10^{28}$cm) of our Universe today would make the Universe very far from the almost homogeneous and isotropic Universe we know today. It would also create unacceptably large fluctuation in the cosmic microwave background radiation (CMBR). We can, therefore, conclude that theories which predict GUT scale domain walls are certainly ruled out; see e.g., Ref. [18]. 'Lighter' cosmological domain walls, those formed in a phase transition at a temperatures below ~ 10 MeV, are, however, not apriori ruled out.

3.2.2 Monopoles

There are similar constraints on monopoles. A monopole formed at a GUT scale phase transition would have a mass $\sim 10^{16}$ GeV. It is not very hard to show that for such massive monopoles, even the lower limit on the monopole number density implied by the Kibble mechanism would make the total mass density of the Universe far exceed the critical mass density for closure of the Universe [19, 12] soon after the formation of the monopoles. In other words, GUT-scale monopoles, if their number densities are not drastically reduced by some mechanism subsequent to their formation, are incompatible with the expanding Universe today. The only way to reduce the number density of the monopoles is by annihilation with antimonopoles, and this has been shown [19] to be too slow a process to reduce their number density significantly. This leads to the famous *monopole problem*: There is an apparent conflict of GUTs with the standard cosmology, because essentially all GUT models necessarily predict the existence of monopoles. Solution of this monopole problem was one of the motivating factors that, in fact, led to the development of the idea of *inflation*[20] (see e.g., Ref. [12] for details). Inflation basically solves the monopole problem by diluting the number density of monopoles to very low levels during the exponentially expanding phase of the Universe envisaged in the theory of inflation. There are also particle physics solution of the monopole problem; see, e.g., Ref. [6].

Apart from the cosmological constraints discussed above, there is also a variety of *astrophysical* constraints on monopoles; see Ref. [12] for a review.

It should be mentioned that monopoles could also be produced by thermal monopole- anti-monopole pair production process in the early Universe, and such a process could leave behind a small but interesting relic abundance of monopoles in the Universe today.

3.2.3 Cosmic strings and Textures

In contrast to domain walls and monopoles which lead to "problems", the other two major kinds of TDs, namely, cosmic strings and textures are generally found to be cosmologically 'useful'. In particular, they form the basis of some of the most successful theories of formation of galaxies and large scale structure in the Universe. I will not discuss this aspect of the usefulness of cosmic strings and textures in this talk; see, Ref.[11] for a contemporary review of this topic. It should be mentioned, however, that cosmological and astrophysical constraints also exist for these defects. For example, oscillating closed loops of cosmic strings emit gravitational waves creating a stochastic gravitational wave background in the Universe today. Cosmic strings as well as textures can also give rise to distinctive patterns of distortion in the CMBR. Cosmic strings, because of their distinctive gravitational fields, can also act as gravitational lenses. At the present moment, it can be said that no current observations rule out the existence of GUT scale cosmic strings and textures. For detailed discussions of these topics, see Ref.[11].

4 Topological defects as sources of extremely high energy cosmic rays

The discussions of constraints on cosmological TDs in the previous section were concerned with mainly their gravitational effects. This is due to the fact that GUT-scale TDs are expected to be very massive, and hence their gravity provides the dominant mode of their interactions with other matter in the Universe. However, one should remember that, at the microphysical level, TDs represent highly non-trivial classical configurations of the underlying quantum fields in the theory, and as such have important (non-gravitational) particle physics implications too. Indeed, under certain circumstances, TDs can be sources of extremely energetic particles. In this section we discuss the possibility that such energetic particles created by TDs can provide the signature of TDs in the Universe. We also discuss the possibility that TDs may even be the *sources* of extremely high energy (EHE) (energy $\gtrsim 10^{20}$ eV) cosmic rays. (For a review of cosmic rays at the highest energies, see, for example, Refs. [21, 22, 23, 24].)

TDs can be thought of as 'made' of massive quanta of gauge- and higgs bosons (hereafter referred to as the "X" particles) of the underlying spontaneously broken gauge theory. The X particles are 'trapped' in topologically non-trivial configurations which under normal dynamical evolution cannot 'unwind'. The X particles are thus not free and cannot decay as long as they are trapped inside TDs. Thus, owing to topological stability, the TDs formed in the early Universe, and along with them the massive X particles trapped inside the TDs, can survive down to the present epoch. The TDs can nevertheless be occasionally, and in certain circumstances, frequently, destroyed in physical processes like collapse or annihilation. This may happen, for example, if monopole annihilates with an antimonopole [25, 26], or if a closed loop of cosmic string collapses [27, 28], or in the so-called "cusp evaporation" process for cosmic strings[29]. When a TD is physically destroyed, the energy stored in the TD is released in the form of massive X particles that 'constitute' the TD. The released X particles would then decay, essentially instantaneously, typically into fundamental particles like quarks and leptons. Hadronization of the quarks would then produce jets of hadrons containing mainly light mesons (pions) together with a small fraction of nucleons (protons and neutrons). The gamma rays and neutrinos from the decay of the neutral and charged pions, respectively, would thus be the most abundant particles in the final decay products of the massive X particles. If the TDs under consideration were originally formed at a GUT energy scale, the mass m_X of the X particles released from the TDs can be $\sim 10^{25}$eV. The decay of the X particles released from the TDs can thus give rise to protons, neutrons, gamma rays, electron-positrons and neutrinos with energies up to $\sim m_X$. The relevant experiments that may be able to detect these extremely energetic particles are the large extensive air shower experiments set up to detect the extremely high energy cosmic rays (see, for example, Sokolsky[21] for a review of the relevant cosmic ray experiments).

Recent cosmic ray experiments have detected several cosmic ray events with energies above 10^{20} eV [30]. Several analyses[31, 32, 33] have shown that it is indeed very difficult, if not impossible, to explain the origin of these energetic events in terms of the standard diffusive shock acceleration mechanism (see, e.g., Drury[34]) of producing high energy cosmic rays in known astrophysical environments such as in Active Galactic Nuclei (AGN), in the lobes of powerful radio galaxies, and so on. As mentioned above, and as detailed studies [35, 36, 26, 31, 33] have shown, TDs can indeed provide a natural and fundamentally different mechanism of production of extremely high energy cosmic rays.

The release of X particles from TDs may occur continually at all cosmological epochs after the formation of the TDs under consideration in the early Universe. However, only the X particles released in relatively recent epochs are likely, if at all, to contribute to the present-day flux of protons and gamma rays in the ultrahigh energy (UHE) energy region (i.e., energy $\gtrsim 10^{18}\,$eV), because those released in the earlier epochs would have to traverse greater distances through the cosmic background radiation fields in order to reach us and would, therefore, lose significant fractions of their energy in collision with the photons of the background radiation. However, neutrinos can come from relatively earlier cosmological epochs because they suffer little energy loss as a consequence of their small cross section of interaction with the relevant particles of the background medium.

In the TD model, the shapes of the energy spectra of protons, gamma rays and neutrinos at production (i.e., the injection spectra) at any time are determined primarily by the physics of fragmentation of quarks into hadrons and not by any astrophysical parameters. At any given time, the injection spectra are, therefore, also independent of the specific kind of TDs responsible for release of X particles. Different kinds of TDs, however, evolve in different ways. Therefore, the absolute magnitude and the rate of production of the various particles, and so also the final evolved spectra, will be different for different kinds of TDs. However, in the highest energy regions, the shapes of the proton as well as gamma ray spectra become insensitive to the kind of TDs producing them, because to survive at these energies, the protons and photons must originate at relatively close (i.e., non-cosmological) distances for which the cosmological evolution is immaterial. The shape of the neutrino spectrum will, however, remain sensitive to the kind of TDs producing them and their cosmological evolution because neutrinos can propagate over large (cosmological) distance scales without much attenuation.

The injection spectra of nucleons, gamma rays, and neutrinos in the TD model have been calculated[27, 35, 36] by extrapolating the available models of hadronization of quarks as described by the theory of Quantum Chromo-Dynamics (QCD) to extremely high energies. This gives approximately power-law differential injection spectra for nucleons, gamma rays and neutrinos all with a power-law index $\alpha \sim 1.3$. It is to be emphasized, however, that there is a great deal of uncertainty involved in extrapolating the low-energy QCD based models of hadronization of quarks to extremely high energies involved in the present situation. It is also possible that the value of α for nucleons may be somewhat different from that for, say, gamma rays. The main point, however, is that the injection spectra of cosmic ray particles in the TD model can in principle be considerably *flatter* than in the standard shock acceleration scenarios which, by and large, produce differential injection spectra with $\alpha \geq 2$.

The typical shapes of the final proton, gamma ray and neutrino spectra, including the effects due to propagation through the extragalactic medium, are shown in Figure 3.

Fig. 3 was actually obtained for a specific TD model, namely, that involving annihilation of magnetic monopole-antimonopole pairs[26]; however, as explained above, the spectra have more or less a universal *shape* independent of the specific kind of TD-process one considers, especially at the highest energies, and hence the spectra shown in Fig. 3 are representative of the particle spectra expected in TD models in general. One major uncertainty in this scenario is the absolute magnitude of the cosmic ray flux produced by TDs, i.e., the 'normalization' of the predicted flux. This clearly depends on the specific process of particle production involving specific kind of TDs. The normalization of the particle fluxes in Fig. 3 implies a monopole abundance that is well below the stringent astrophysical upper limit on the monopole abundance — the so-called "Parker limit" (see, Ref.[12]) — and is, therefore, quite plausible.

From Fig.3 we see that TD models are probably not relevant for cosmic rays below about $5 \times 10^{19}\,$eV; it is only at energies above this energy that TD models become a viable option. It is to be mentioned here that the shape of the proton spectra in the acceleration models (not shown in Fig. 3), which typically yield power-law injection spectra with index $\alpha \geq 2$, would correspond to a sharply falling curve at energies beyond $\sim 10^{20}\,$eV, and so would be unable to explain the two highest energy events indicated in the figure. The 'dip' of the proton curve of Fig. 3 beginning at $\sim 10^{20}\,$eV and its subsequent 'recovery' at $\sim 10^{21}\,$eV are characteristic features induced by the propagation effects. The 'recovery' of the proton spectrum after the 'dip' is, however, a feature that is not shared by the proton spectra one gets in standard acceleration models. The recovery

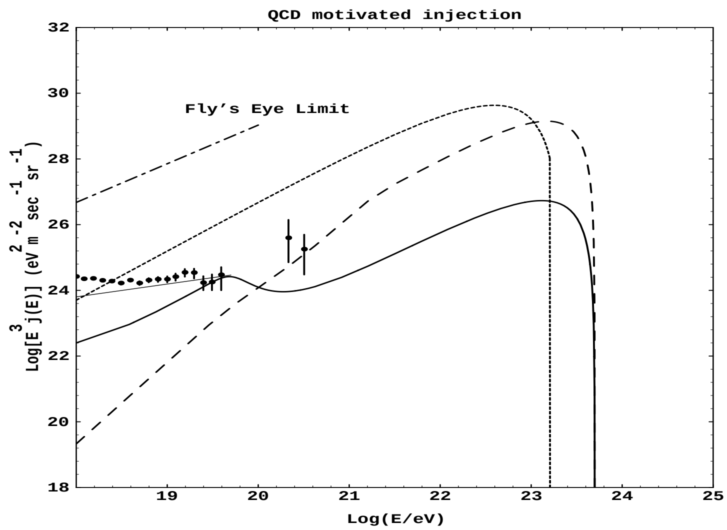


Figure 3: The proton (solid line), gamma ray (long-dashed line) and the neutrino (short-dashed line) spectra in the TD scenario including the effect of propagation through the cosmic background medium. The X particle mass is taken to be 10^{15} GeV and an extrapolation of QCD-based hadronization spectrum to the relevant high energies has been used to obtain the injection spectra. The combined proton and gamma ray flux has been normalized at $10^{19.7}$ eV to the "extragalactic flux component" (thin solid line; see Bird et al in Ref. [30]) fitted to the data from the Fly's Eye experiment (filled circles with error bars; Bird et al in Ref. [30]) . Also shown (dash-dotted line) is an approximate limit on the neutrino flux determined from non-detection of deeply penetrating particles by the Fly's Eye detector [21].

in the present case is essentially due to flatter nature of the proton injection spectrum in the TD model as compared to that in the acceleration models. Note also from Fig. 3 that the two highest-energy events are naturally explained in the TD model if these are gamma ray events, and not protons. Experimentally, it is hard to determine the composition of the events at these energies with full certainty. Although, the traditionally favoured composition for these events are protons, a gamma-ray composition is certainly not ruled out (see, Bird et al in Ref.[30]).

Besides the characteristic shapes of the UHE cosmic ray (UHECR) spectra, the TD models of UHECR origin have two definitive predictions. Firstly, the UHECR should consist of only 'fundamental' particles like protons, neutrons, gamma rays, neutrinos, electrons, positrons, and so on (and perhaps their antiparticles too), but definitely *no nuclei* such as ^{4}He or Fe. (There is no way hadronization of quarks would directly give rise to nuclei!). Secondly, the UHECR should be highly rich in gamma-rays and neutrinos. These predictions can be used as crucial tests of the TD model in future UHECR experiments with large-area detector coverage, such as the proposed Pierre Auger project [37], which could thus provide the *signatures* of existence of TDs in the Universe.

References

[1] S. Coleman, *Aspects of Symmetry* (Cambridge Univ. Press, Cambridge, 1989).

[2] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).

[3] H.B. Nielsen and P. Olesen, *Nucl. Phys.* **B61**, 45 (1973).

[4] G. 't Hooft, *Nucl. Phys.* **B79**, 276 (1974).

[5] A.M. Polyakov, *JETP Lett.* **20**, 194 (1974).

[6] P. Langacker and S.-Y. Pi, *Phys. Rev. Lett.* **45**, 1 (1980).

[7] D.A. Kirzhnits and A.D. Linde, *Sov. Phys. JETP* **40**, 628 (1974); *Ann. Phys.* **101**, 195 (1976).

[8] L. Dolan and R. Jackiw, *Phys. Rev.* **D9**, 3320 (1974).

[9] S. Weinberg, *Phys. Rev.* **D9**, 3357 (1974).

[10] T.W.B. Kibble, *J. Phys.* **A9**, 1387 (1976).

[11] A. Vilenkin and E.P.S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge Univ. Press, Cambridge, 1994).

[12] E.W. Kolb and M.S. Turner, *The Early Universe* (Addison-Wesley, Redwood City, California, 1990).

[13] P. Goddard and D.I. Olive, *Rep. Prog. Phys.* **41**, 1357 (1978).

[14] I. Chuang, R. Durrer, N. Turok and B. Yurke, *Science* **251**, 1336 (1990); M.J. Bowick, L. Chander, E.A. Schiff and A.M. Srivastava, *Science* **263**, 943 (1994).

[15] P.C. Hendry, N.S. Lawson, R.A.M. Lee, P.V.E. McClintock and C.H.D. Williams, *Nature* **368**, 315 (1994).

[16] W.H. Zurek, *Nature* **317**, 505 (1985); in *Formation and Interactions of Topological Defects* (plenum Press, New York, 1995), (eds.: A.-C. Davis and R. Brandenberger), p. 349.

[17] R. Basu, A.H. Guth and A. Vilenkin, *Phys. Rev.* **D44**, 340 (1991).

[18] Ya.B. Zel'dovich, I.Yu. Kobzarev and L.B. Okun, *Sov. Phys. JETP* **40**, 1 (1975).

[19] J. Preskill, *Phys. Rev. Lett.* **43**, 1365 (1979).

[20] A.H. Guth, *Phys. Rev.* **D23**, 347 (1981).

[21] P. Sokolsky, *Introduction to Ultrahigh Energy Cosmic Ray Physics* (Addison-Wesley, Redwood City, California, 1988).

[22] *Astrophysical Aspects of the Most Energetic Cosmic Rays*, eds.: M. Nagano and F. Takahara (World Scientific, Singapore, 1991).

[23] *Proc. of Tokyo Workshop on Techniques for the Study of Extremely High Energy Cosmic Rays*, (ICRR, Univ. of Tokyo, 1993).

[24] *Cosmic Rays above* 10^{19} eV — 1992, eds. M. Boratov et al., *Nucl. Phys. B.* (Proc. Suppl) **28B** (1992).

[25] C. T. Hill, *Nucl. Phys.* **B224**, 469 (1983).

[26] P. Bhattacharjee and G. Sigl, *Phys. Rev.* **D51**, 4079 (1995).

[27] P. Bhattacharjee, in Ref. [22], p. 382.

[28] P. Bhattacharjee and N.C. Rana, *Phys. Lett.* **B246**, 365 (1990).

[29] P. Bhattacharjee, *Phys. Rev.* **D40**, 3968 (1989).

[30] D. J. Bird *et al., Phys. Rev. Lett.* **71**, 3401 (1993); *Ap. J.* **424**, 491 (1994); *Ap. J.* **441**, 144 (1995); N. Hayashida et al., *Phys. Rev. Lett.* **73**, 3491 (1994); S. Yoshida et al., *Astropart. Phys.* **3**, 105 (1995); N.N. Efimov et al., in Ref.[22], p. 20; T.A. Egorov, in Ref. [23], p. 35; A.A. Watson, in Ref.[22], p. 2.

[31] G. Sigl, D.N. Schramm and P. Bhattacharjee, *Astropart. Phys.* **2**, 401 (1994).

[32] J.W. Elbert and P. Sommers, *Ap. J.* **441** 151 (1995).

[33] G. Sigl, S. Lee, D.N. Schramm and P. Bhattacharjee, *Science*, **270**, 1977 (1995).

[34] L.O'C. Drury, *Contemp. Phys.* **35**, 231 (1994).

[35] P. Bhattacharjee, C.T. Hill and D.N. Schramm, *Phys. Rev. Lett.* **69**, 567 (1992).

[36] F.A. Aharonian, P. Bhattacharjee and D.N. Schramm, *Phys. Rev.* **D46**, 4188 (1992).

[37] See, Ref.[24].

I cannot allow myself to congratulate you, but allow me to tell you that when I heard the news, I was turned upside down. For some reason, I feel so happy and proud that your long work has found such a world recognition. Not that you need–if I may dare to express my humble feelings–any wordly gratification, for you and my father belong to the same kind, but still, I cannot help feeling immensely happy and proud. After all, you are a little bit my father. Someone said: the rich and the poor are two locked caskets, of which each contains the key to the other (it was Karen Blixen, A Danish writer). How wonderful!

— KAREN CHALLONGE

GENERALISED RAYCHAUDHURI EQUATIONS FOR STRINGS AND MEMBRANES

Sayan Kar[1 2]

Institute of Physics
Sachivalaya Marg
Bhubaneswar 751 005

Abstract

A recent generalisation of the Raychaudhuri equations for timelike geodesic congruences to families of D dimensional extremal, timelike, Nambu–Goto surfaces embedded in an N dimensional Lorentzian background is reviewed. Specialising to $D = 2$ (i.e the case of string worldsheets) we reduce the equation for the generalised expansion $\theta_a, (a = \sigma, \tau)$ to a second order, linear, hyperbolic partial differential equation which resembles a variable–mass wave equation in $1 + 1$ dimensions. Consequences, such as a generalisation of geodesic focussing to families of worldsheets as well as exactly solvable cases are explored and analysed in some detail. Several possible directions of future research are also pointed out.

[1]e-mail: sayan@iopb.ernet.in

[2]Address after September 1, 1996 : Inter-University Centre for Astronomy & Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007

I first constructed proofs for myself. Then I compared my proofs with those of Newton. The experience was a sobering one. Each time, I was left in sheer wonder at the elegance, the careful arrangement, the imperial style, the incredible originality, and above all the astonishing lightness of Newton's proofs; and each time I felt like a schoolboy admonished by his master.

— S. CHANDRASEKHAR

1 Introduction

It is a well established fact today that the proof of the existence of spacetime singularities in the general theory of relativity (GR) largely relies on the consequences obtained from the Raychaudhuri equations for null/timelike geodesic congruences [1,2,3]. Even though the applications of the Raychaudhuri equations are mostly confined to the domain of GR, it is important to note that these equations contain some basic statements about the nature of geodesics in a Riemannian/pseudo–Riemannian geometry. GR comes into the picture when one assumes the Einstein field equation and thereby reduces one crucial term containing information about geometry into an object related to matter stress–energy. Subsequently, if one imposes an Energy condition (such as the Weak Energy condition which implies that the energy density of matter is always positive in all frames of reference) it is possible to derive the fact that geodesic congruences necessarily converge within a finite value of the affine parameter. This is known as the focussing theorem, which, along with other assumptions about causality, essentially imply the existence of spacetime singularities.

In string and membrane theories, the notion of the point particle (and its associated world–line) which is basic to GR as well as other relativistic field theories, gets replaced by the string/membrane (and its corresponding world–surface). This is a radically new concept and has paid rich dividends in recent times. For instance, it is claimed that quantum gravity as well as a unification of forces comes out naturally from quantizing string theories.

If one accepts the string/membrane viewpoint then it should, in principle, be possible to derive the corresponding generalized Raychaudhuri equations for timelike/null worldsheet congruences and arrive at similar focussing and singularity theorems in Classical String theory. Very recently, Capovilla and Guven [4] have written down the generalized Raychaudhuri equations for timelike worldsheet congruences. In this talk, we shall first give a brief review of these equations. Thereafter, we construct explicit examples of these rather complicated set of equations by specializing to certain simple extremal families of surfaces. Our principal aim is to extract some information regarding *focussing of families of surfaces* in a way similar to the results for geodesic congruences in GR.

Finally, we shall summarize some recent work in progress and point out several future directions of research.

2 Review

2.1 What are Raychaudhuri Equations ?

An useful way of visualising the content of the Raychaudhuri equations is to look at an analogy with fluid flow. The flow lines of a fluid (which are the integral curves of the velocity field) form a congruence of curves. If we focus our attention on the cross–sectional area enclosing a certain number of these curves we find that it is different at different points. The gradient of the velocity field contains crucial information about the behaviour of this area. What can happen to this area as we move along the family of curves? The obvious answer is–it can *expand* (i.e. a smaller circle may become a larger circle which is concentric to the former), it can become *sheared* (i.e. a circle can become an ellipse) or it can twist or *rotate*. The information about each of these objects (i.e. the expansion, shear and the rotation) is encoded in the gradient of the velocity field). Recall that this gradient is a second rank tensor. Such a quantity can be split into its symmetric traceless, antisymmetric and trace parts – these are respectively the shear, rotation and expansion of the congruence /flow lines. For geodesic congruences in Riemannian/pseudo–Riemannian geometry we have to replace the flow lines with geodesics and the velocity field by the tangent vectors to the geodesic curves.

Therefore, expressed in mathematical language we have :

$$v_{\mu;\nu} = \sigma_{\mu\nu} + \omega_{\mu\nu} + \frac{1}{3}\theta h_{\mu\nu} \tag{1}$$

The *evolution* equation for each of these quantities – expansion, shear, rotation along the geodesic congruence are what are known as the Raychaudhuri equations.
For example, the equation for the expansion θ for a timelike geodesic congruence turns out to be :

$$\frac{d\theta}{d\lambda} + \frac{1}{3}\theta^2 + \sigma^2 - \omega^2 = -R_{\mu\nu}\xi^\mu\xi^\nu \tag{2}$$

An analysis of the nature of the solutions of this equation leads us to the concept of geodesic focussing. First, let us convert this equation into a second order linear equation by redefining $\theta = 3\frac{F'}{F}$. This results in the following :

$$\frac{d^2F}{d\lambda^2} + \frac{1}{3}\left(\sigma^2 - \omega^2 + R_{\mu\nu}\xi^\mu\xi^\nu\right)F = 0 \tag{3}$$

Assuming zero rotation, one can prove that if θ is negative somewhere it has to go to $-\infty$ within a finite value of the affine parameter if the coefficient of the second term in the above equation is greater than or equal to zero. This largely follows from the theorems on the existence of zeros of a general class of such second order, linear equations proofs of which can be found in the paper by Tipler [6].

What happens when $\theta \to -\infty$? If A_1 is the area at λ_1 and A_2 at λ_2 we can write θ as $\theta = \frac{A_2 - A_1}{A_2}$. Therefore θ can go to $-\infty$ if $A_2 \to 0$. Thus, the family of geodesics must focus at a point (i.e. at $\lambda = \lambda_2$).
A word about the constraints on the geometry which are necessary for focussing. Notice that if the rotation is zero we must have $R_{\mu\nu}\xi^\mu\xi^\nu \geq 0$. Using the Einstein field equations one can write this as $T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T \geq 0$. This is what is known as the Strong Energy Condition. An Energy Condition, in general, defines a certain class of matter which is physically possible. For instance, the Weak Energy Condition states that $T_{\mu\nu}\xi^\mu\xi^\nu \geq 0$ which physically implies the positivity of energy density in all frames of reference.

We now ask the following questions:

Suppose we replace extremal curves (geodesics) by extremal D dimensional timelike surfaces embedded in an N dimensional background.

- Are there generalisations of the Raychaudhuri equations?
- Is there an analog of geodesic focussing?

The remaining part of this article is devoted to answering these two questions.

2.2 Geometry of Embedded Surfaces

We begin with a brief account of the differential geometry of embedded surfaces.

A D dimensional surface in an N dimensional background is defined through the embedding $x^\mu = X^\mu(\xi^a)$ where ξ^a are the coordinates on the surface and x^μ are the ones in the background. Furthermore, we construct an orthonormal basis (E^μ_a, n^μ_i) consisting of D tangents and $N-D$ normals at each point on the surface. E^μ_a, n^μ_i satisfy the following properties.

$$g(E_a, E_b) = \eta_{ab} \quad ; \quad g(E_a, n_i) = 0 \quad ; \quad g(n_i, n_j) = \delta_{ij} \tag{4}$$

We can write down the Gauss–Weingarten equations using the usual definitions of extrinsic curvature, twist potential and the worldsheet Ricci rotation coefficients.

$$D_a E_b = \gamma^c_{ab}E_c - K^i_{ab}n_i \tag{5}$$

$$D_a n_i = K^i_{ab}E^b + \omega^{ij}_a n_j \tag{6}$$

where $D_a \equiv E^\mu_a D_\mu$ (D_μ being the usual spacetime covariant derivative). The quantities K^i_{ab} (extrinsic curvature), ω^{ij}_a and γ^c_{ab} are defined as :

$$K^i_{ab} = -g(D_a E_b, n^i) = K^i_{ba} \tag{7}$$
$$\omega^{ij}_a = g(D_a n^i, n^j) \tag{8}$$
$$\gamma_{abc} = g(D_a E_b, E_c) = -\gamma_{acb} \tag{9}$$

In order to analyse deformations normal to the worldsheet we need to consider the normal gradients of the spacetime basis set. The corresponding analogs of the Gauss–Weingarten equations are:

$$D_i E_a = J_{aij} n^j + S_{abi} E^b \tag{10}$$
$$D_i n_j = -J_{aij} E^a + \gamma^k_{ij} n_k \tag{11}$$

where $D_i \equiv n^\mu_i D_\mu$. The quantities J^{ij}_a, S_{abi} and γ^k_{ij} are defined as :

$$S^i_{ab} = g(D^i E_a, E_b) = -S^i_{ba} \tag{12}$$
$$\gamma_{ijk} = g(D_i n_j, n_k) = -\gamma_{ikj} \tag{13}$$
$$J^{ij}_a = g(D^i E_a, n^j) \tag{14}$$

2.3 Sketch of Derivation

The full set of equations governing the evolution of deformations can now be obtained by taking the worldsheet gradient of J^{ij}_a. This turns out to be (for details see Appendix of Ref. [4]).

$$\tilde{\nabla}_b J^{ij}_a = -\tilde{\nabla}^i K^j_{ab} - J^i_{bk} J^{kj}_a - K^i_{bc} K^{cj}_a - g(R(E_b, n^i)E_a, n^j) \tag{15}$$

where the extrinsic curvature tensor components are $K^i_{ab} = -g_{\mu\nu} E^\alpha_a (D_\alpha E^\mu_b) n^{\nu i}$

On tracing over worldsheet indices we get

$$\tilde{\nabla}_a J^{aij} = -J^i_{ak} J^{akj} - K^i_{ac} K^{acj} - g(R(E_a, n^i)E^a, n^j) \tag{16}$$

where we have used the equation for extremal membranes (i.e. $K^i = 0$)

The antisymmetric part of (12) is given as:

$$\tilde{\nabla}_b J^{ij}_a - \tilde{\nabla}_a J^{ij}_b = G^{ij}_{ab} \tag{17}$$

where $g(R(Y_1, Y_2)Y_3, Y_4) = R_{\alpha\beta\mu\nu} Y^\alpha_1 Y^\beta_2 Y^\mu_3 Y^\nu_4$ and

$$G^{ij}_{ab} = -J^i_{bk} J^{kj}_a - K^i_{bc} K^{cj}_a - g(R(E_b, n^i)E_a, n^j) - (a \to b) \tag{18}$$

One can further split J_{aij} into its symmetric traceless, trace and antisymmetric parts ($J^{ij}_a = \Sigma^{ij}_a + \Lambda^{ij}_a + \frac{1}{N-D}\delta^{ij}\theta_a$) and obtain the evolution equations for each of these quantities. The one we shall be concerned with mostly is given as

$$\Delta\gamma + \frac{1}{2}\partial_a\gamma\partial^a\gamma + (M^2)^i_i = 0 \tag{19}$$

with the quantity $(M^2)^{ij}$ given as :

$$(M^2)^{ij} = K^i_{ab} K^{abj} + R_{\mu\nu\rho\sigma} E^\mu_a n^{\nu i} E^{\rho a} n^{\sigma j} \tag{20}$$

∇_a is the worldsheet covariant derivative ($\Delta = \nabla^a \nabla_a$) and $\partial_a \gamma = \theta_a$. Notice that we have set Σ^{ij}_a and Λ^{ij}_a equal to zero. This is possible only if the symmetric traceless part of $(M^2)^{ij}$ is zero. One can check this by looking at the full set of generalized Raychaudhuri equations involving Σ^{ij}_a, Λ^{ij}_a and θ_a [4]. For geodesic curves the usual Raychaudhuri equations can be obtained by noting that $K^i_{00} = 0$, the J_{aij} are related to their spacetime counterparts $J_{\mu\nu a}$ through the relation $J_{\mu\nu a} = n^i_\mu n^j_\nu J_{aij}$, and the θ is defined by contracting with the projection tensor $h_{\mu\nu}$.

The θ_a or γ basically tell us how the spacetime basis vectors change along the normal directions as we move along the surface. If θ_a diverges somewhere, it induces a divergence in J_{aij}, which, in turn means that the gradients of the spacetime basis along the normals have a discontinuity. Thus the family of worldsheets meet along a curve and a cusp/kink is formed. This, we claim, is a focussing effect for extremal surfaces analogous to geodesic focussing in GR where families of geodesics focus at a point if certain specific conditions on the matter stress energy are obeyed.

2.4 Is the evolution of J_{aij} constrained?

For each pair (ij) there are D quantities J^{ij}_a. To analyse whether the evolution of J^{ij}_a is constraint–free or not we split the antisymmetric equations into two sets.

$$\tilde{\nabla}_0 J^{ij}_A - \tilde{\nabla}_A J^{ij}_0 = G^{ij}_{ab} \tag{21}$$
$$\tilde{\nabla}_B J^{ij}_A - \tilde{\nabla}_A J^{ij}_B = G^{ij}_{ab} \tag{22}$$

Thus first of these contains a total of D equations for each ij. Eqn (12) on the other hand contains one equation. Thus the total number of equations contained in these is D which is the number necessary to specify the evolution of the quantity J^{ij}_a. Therefore the second set above is actually a set of constraints on the evolution.

However, note that the number of equations in this second set is equal to $\frac{1}{2}(D-1)(D-2)$. Therefore for $D = 1$ (curves) and $D = 2$ (string worldsheets) these equations are vacuous and the evolution of J^{ij}_a is entirely constraint–free! Even more surprising is the fact that for any general $D > 2$ also the second set of equations are *identities* (for a proof see [4]). Therefore, if the constraints and equations of motion are satisfied at any initial time they continue to be so for all future values.

3 Focussing of string worldsheets [5]

Two dimensional timelike surfaces embedded in a four dimensional background are the objects of discussion in this section. We begin by writing down the generalised Raychaudhuri equation for the case in which Σ^{ij}_a and Λ^{ij}_a are set to zero (i.e implicitly assuming that $(M^2)^{ij}$ does not have a nonzero symmetric traceless part.Thus we have

$$-\frac{\partial^2 F}{\partial \tau^2} + \frac{\partial^2 F}{\partial \sigma^2} + \Omega^2(\sigma,\tau)(M^2)^i_i(\sigma,\tau)F = 0 \tag{23}$$

where Ω^2 is the conformal factor of the induced metric written in isothermal coordinates. Notice that the above equation is a second–order, linear, hyperbolic partial differential equation. On the contrary, the Raychaudhuri equation for curves is an linear, second order, ordinary differential equation. The easiest way to analyse the solutions of this equation is to assume separability of the quantity $\Omega^2(M^2)^i_i$. Then, we have

$$\Omega^2(M^2)^i_i = M_1^2(\tau) + M_2^2(\sigma) \tag{24}$$
$$F(\tau,\sigma) = F_1(\tau) \times F_2(\sigma) \tag{25}$$

With these we can now split the partial differential equation into two ordinary differential equations given by

$$\frac{d^2F_1}{d\tau^2} + (\omega^2 - M_1^2(\tau))F_1 = 0 \tag{26}$$

$$\frac{d^2F_2}{d\sigma^2} + (\omega^2 + M_2^2(\sigma))F_2 = 0 \tag{27}$$

Since the expansions along the τ and σ directions can be written as $\theta_\tau = \frac{\dot{F}_1}{F_1}$ and $\theta_\sigma = \frac{F_2'}{F_2}$ we can analyse focussing effects by locating the zeros of F_1 and F_2 much in the same way as for geodesic curves [6]. The well–known theorems on the existence of zeros of ordinary differential equations as discussed in [6] make our job much simpler. The theorems essentially state that the solutions of equations of the type $\frac{d^2A}{dx^2} + H(x)A = 0$ will have at least one zero iff $H(x)$ is positive definite. Thus for our case here, we can conclude that, focussing along the τ and σ directions will take place only if

$$\omega^2 \geq max[M_1^2(\tau)] \qquad ; \qquad \omega^2 \geq max[-M_2^2(\sigma)] \tag{28}$$

For stationary strings, one notes that $(M^2)^{ij}$ will not have any dependence on τ. Thus we can set M_1^2 equal to zero. Thus, focussing will entirely depend on the sign of the quantity M_2^2. We can write M_2^2 alternatively as follows. Consider the Gauss–Codazzi integrability condition:

$$R_{\mu\nu\alpha\beta}E^\mu_a E^\nu_b E^\alpha_c E^\beta_d = R_{abcd} - K_{aci}K^{aci} + K_{bdi}K^{bdi} \tag{29}$$

Trace the above expression on both sides with $\eta^{ac}\eta^{bd}$ and rearrange terms to obtain :

$$K_{abi}K^{abi} = -{}^2R + K^iK_i + R_{\mu\nu\alpha\beta}E^\mu_a E^{\alpha a}E^\nu_b E^{\beta b} \tag{30}$$

Thereafter, use this expression and the fact that $n^{i\mu}n^\nu_i = g^{\mu\nu} - E^{\mu a}E^\nu_a$ in the original expression for $(M^2)^i_i$ (see Eqn.(11)) to get

$$M_2^2 = -{}^2R + R_{\mu\nu}E^{\mu a}E^\nu_a \tag{31}$$

One can notice the following features from the above expression:

(i) If the background spacetime is a vacuum solution of the Einstein equations then the positivity of M_2^2 is guaranteed iff ${}^2R \leq 0$. Thus all string configurations in vacuum spacetimes which have negative Ricci curvature everywhere will necessarily imply focussing. This includes the well known string solutions in Schwarzschild and Kerr backgrounds.

(ii) If the background spacetime is a solution of the Einstein equations then we can replace the second and third terms in the expressions for M_2^2 by the corresponding ones involving the Energy Momentum tensor $T_{\mu\nu}$ and its trace. Thus we have

$$M_2^2 = \left(-\frac{1}{2}g_{\mu\nu}{}^2R + T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}\right)E^{\mu a}E^\nu_a \tag{32}$$

Notice that if we split the quantity $E^\mu_a E^{\nu a}$ into two terms such as $E^\mu_\tau E^{\nu\tau}$ and $E^\mu_\sigma E^{\nu\sigma}$ then we have :

$$M_2^2 = -{}^2R + \left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}\right)E^{\mu\tau}E^\nu_\tau + \left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}\right)E^{\mu\sigma}E^\nu_\sigma \tag{33}$$

The second term in the above equation is the L. H. S. of the Strong Energy Condition (SEC). Apart from this we have two other terms which are entirely dependent on the fact that we are dealing with extended objects. The positivity of the whole quantity can therefore be thought of as an *Energy Condition* for the case of strings. Thus even if the background spacetime satisfies the SEC, focussing of string world–sheets is not guaranteed–worldsheet curvature and the projection of the combination $T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T$ along the σ direction have an important role to play in deciding focussing/defocussing.

Let us now try to understand the consequences of the above equations for certain specific flat and curved backgrounds for which the string solutions are known.

3.1 Rindler Spacetime

The metric for four dimensional Rindler spacetime is given as

$$ds^2 = -a^2x^2dt^2 + dx^2 + dy^2 + dz^2 \tag{34}$$

We recall from [7] the a string solution in a Rindler spacetime:

$$\begin{aligned} t = \tau \quad ; \quad x = ba\cosh a\sigma_c \quad ; \\ y = ba^2\sigma_c \quad ; \quad z = z_0 \quad (constant) \end{aligned} \tag{35}$$

where $d\sigma_c = \frac{d\sigma}{a^2x^2}$ and b is an integration constant. The orthonormal set of tangents and normals to the worldsheet can be chosen to be as follows:

$$E^\mu_\tau \equiv \left(\frac{1}{ax}, 0, 0, 0\right) \quad ; \quad E^\mu_\sigma \equiv (0, \tanh a\sigma_c, secha\sigma_c, 0) \tag{36}$$

$$n^\mu_1 \equiv (0, 0, 0, 1) \quad ; \quad n^\nu_2 \equiv (0, secha\sigma_c, -\tanh a\sigma_c, 0) \tag{37}$$

In the worldsheet coordinates τ, σ_c the induced metric is flat and the components of the extrinsic curvature tensor turn out to be

$$\begin{aligned} K^1_{ab} = 0 \quad ; \quad K^2_{\tau\tau} = -K^2_{\sigma_c\sigma_c} = \frac{1}{ba\cosh^2 a\sigma_c} \quad ; \\ K^2_{\sigma\tau} = 0 \end{aligned} \tag{38}$$

The quantity $(M^2)^i_i$ which is dependent only on the extrinsic curvature of the worldsheet (the background spacetime being flat) turns out to be

$$(M^2)^i_i = \frac{2}{b^2a^2\cosh^4 a\sigma_c} \tag{39}$$

Therefore the generalized Raychaudhuri equation turns out to be

$$-\frac{\partial^2 F}{\partial\tau^2} + \frac{\partial^2 F}{\partial\sigma_c^2} + \frac{2a^2}{\cosh^2 a\sigma_c}F = 0 \tag{40}$$

Separating variables ($F = T(\tau)\Sigma(\sigma)$) one gets the harmonic oscillator equation for T and the Poschl Teller equation for positive eigenvalues for Σ which is given as:

$$\frac{d^2\Sigma}{d\sigma^2} + \left(\omega^2 + \frac{2a^2}{\cosh^2\sigma}\right)\Sigma = 0 \tag{41}$$

From the results of Tipler [6] on the zeros of differential equations one can conclude that focussing will occur ($H(\sigma) > 0$ always).

Several other examples can be found in [5] and [9].

4 Focussing of hypersurfaces [5]

We now move on to the special case of timelike hypersurfaces. Here we have D quantities J_a but only one normal defined at each point on the surface. The Eqn. (8) turns out to be:

$$\partial_b J_a - \partial_a J_b = 0 \tag{42}$$

Therefore one can write $J_a = \partial_a\gamma$ and the traced equation (7) becomes,

$$\Delta\gamma + (\partial_a\gamma)(\partial^a\gamma) + M^2 = 0 \tag{43}$$

with

$$M^2 = K_{ab}K^{ab} + R_{\nu\sigma}n^\nu n^\sigma$$
$$= -{}^2R + R_{\mu\nu}E^{\mu a}E^\nu_a = -{}^2R + {}^3R - R_{\mu\nu}n^\mu n^\nu \tag{44}$$

where we have used $n^\mu n^\nu = g^{\mu\nu} - E^\mu_a E^{\nu a}$ and the Gauss–Codazzi integrability condition. If we assume that the background spacetime satisfies the Einstein equation then we have:

$$M^2 = -\left({}^2Rg_{\mu\nu} + T_{\mu\nu} + Tg_{\mu\nu}\right)n^\mu n^\nu \tag{45}$$

Thus, for stationary two dimensional hypersurfaces (strings in 3D backgrounds) we have the same conclusions as obtained in the previous section. For a two–dimensional hypersurface in three–dimensional flat background the task is even simpler. M^2 can be shown to be equal to the negative of the Ricci scalar of the membrane's induced metric and ${}^2R \leq 0$ guarantees focussing.

Let us now turn to a specific case where the equations are exactly solvable.

4.1 Hypersurfaces in a $2+1$ Curved Background

Our backgound spacetime here is curved, Lorentzian background and $2+1$ dimensional. The metric we choose is that of a Lorentzian wormhole in $2+1$ dimensions given as :

$$ds^2 = -dt^2 + dl^2 + \left(b_0^2 + l^2\right)d\theta^2 \tag{46}$$

A string configuration in this background can be easily found by solving the geodesic equations in the $2D$ spacelike hypersurface [8]. This turns out to be

$$t = \tau \quad ; \quad l = \sigma \quad ; \quad \theta = \theta_0 \tag{47}$$

The tangents and normal vectors are simple enough:

$$E^\mu_\tau \equiv (1,0,0) \quad ; \quad E^\mu_\sigma \equiv (0,1,0) \quad ; \quad n^\mu = \left(0, 0, \frac{1}{b_0^2 + l^2}\right) \tag{48}$$

The extrinsic curvature tensor components are all zero as the induced metric is flat. Using the Riemann tensor components (which can be evaluated simply using the standard formula) we can write down the generalised Raychaudhuri equation. This turns out to be :

$$-\frac{\partial^2 F}{\partial\tau^2} + \frac{\partial^2 F}{\partial\sigma^2} + \left(-\frac{b_0^2}{(b_0^2+\sigma^2)^2}\right)F = 0 \tag{49}$$

A separation of variables $F = T(\tau)\Sigma(\sigma)$ will result in two equations–one of which is the usual Harmonic Oscillator and the other given by:

$$\frac{d^2\Sigma}{d\sigma^2} + \left(\omega^2 - \frac{b_0^2}{(b_0^2+\sigma^2)^2}\right)\Sigma = 0 \tag{50}$$

The above equation can be recast into the one for Radial Oblate Spheroidal Functions by a simple change of variables – $\Sigma' = \sqrt{b_0^2 + \sigma^2}\Sigma$.

$$(1+\xi^2)\frac{d^2\Sigma'}{d\xi^2} + 2\xi\frac{d\Sigma'}{d\xi} + \left(\omega^2 b_0^2(1+\xi^2)\right)\Sigma' = 0 \tag{51}$$

where $\xi = \frac{\sigma}{b_0}$.

The general equation for Radial Oblate Spheroidal Functions is given as :

$$(1+\xi^2)\frac{d^2V_{mn}}{d\xi^2}+2\xi\frac{dV_{mn}}{d\xi}+\left(-\lambda_{mn}+k^2\xi^2-\frac{m^2}{1+\xi^2}\right)V_{mn}=0 \tag{52}$$

Assuming $m=0$ and $\lambda_{0n}=-k^2=-\omega^2 b_0^2$ we get the equation for our case. The solutions are finite at infinity and behave like simple sine/cosine waves in the variable σ. Consulting the tables in [10] we conclude that only for $n=0,1$ we can have λ_{0n} to be negative. In general, the scattering problem for the Schroedinger–like equation has been analysed numerically in [11].

As regards focussing, one can say from the differential equations and the theorems stated in [5] that the function Σ' will always have zeros if $\omega^2 \geq \frac{1}{b_0^2}$. Even from the series representations (see [9]) of the Radial Oblate Spheroidal Functions we can exactly locate the zeros and obtain explicitly the focal curves. However, we shall not attempt such a task here.

5 Work in progress and future directions

A systematic study of the generalised Raychaudhuri equations has only begun. A large number of open problems therefore exist in this subfield. Here we report briefly on some recently obtained results and list a few of the outstanding issues.

(i) Although we have been able to derive an analog of geodesic focussing for extremal timelike membranes by looking at some special cases a general treatment of the problem is still lacking. For example, recall that we made the crucial assumption of separability in the equation for the case of strings and hypersurfaces. In order to avoid this assumption, a way out could be to define an initial value problem with the expansions in different directions taking on a negative value at specific locations on the surface. The initial value as well as the partial differential equation can be recast into one Volterra integral equation of the second kind. Solving this would then give the necessary condition for focussing. Note that in this approach one does not assume an Energy Condition at the outset because one does not have a choice to do so. Some progress along these directions are to be reported in [12].

(ii) What does one need to have an energy condition in this case? Recall that in the Raychaudhuri equation for geodesic congruences the appearance of the term $R_{\mu\nu}\xi^\mu\xi^\nu$ and the Einstein equation relating geometry to matter were the deciding factors. One could translate the *purely geometric* term into a term containing properties of matter stress energy. Focussing was a consequence of assuming certain physically relevant features of matter. To have such a situation in the case of strings we need to have an *Einstein equation* in string theory. This may sound outrageous but we will see why it need not be so.

General relativity as a theory has a unique feature in comparison to all other theories that we know of. *The motion of test particles can be derived from the Einstein field equations.* We must remember that this is a fact which is not true in all other theories. For example, in electrodynamics the Lorentz force law *cannot* in any way be derived from the Maxwell equations. The basic question therefore reduces to the following :

What is the field equation from which one can arrive at the equation of motion for test strings ?

An answer to this question will help us in analysing worldsheet focussing by assuming an Energy condition much in the same way as one does it for geodesic congruences.

(iii) It is important to note that we have restricted ourselves exclusively to extremal Nambu–Goto type membranes while deriving the generalised Raychaudhuri equations and exploring its consequences. What are the corresponding equations for actions other than Nambu–Goto (rigidity corrections, presence of antisymmetric tensor fields, supersymmetric generalisations etc.)? It has been found [13] that in the presence of antisymmetric tensor fields one has several nontrivialities appearing. Firstly, one cannot set the $\Sigma_{aij}, \Lambda_{aij}$ equal to zero and work with only the equation of the expansions. If one sets Σ_{aij} equal to zero then one has to identify the components of Λ_{aij} with the projections of $H_{\mu\nu\rho}$. Therefore, we can now attribute a physical meaning to Λ_{aij} by associating

it with the background antisymmetric tensor field's projections. Further work is clearly needed to understand the consequences of actions other than Nambu–Goto and is in progress.
(iv) Finally, we indicate a possible application in a totally different area–the theory of biological (amphiphilic) membranes. Deformations of these membranes ($D = 2$ hypersurfaces in a $D = 3$ Euclidean background) can be analysed using the same formalism as presented here and may turn out to be useful in understanding the fluctuations of these objects. The simplest case of the catenoidal membrane is discussed in detail in the appendix to [5].

Acknowledgement

I thank B. R. Iyer and G. Date for inviting me to this meeting and for the excellent hospitality provided at the Institute of Mathematical Sciences, Madras.

References

[1] A. K. Raychaudhuri, *Phys. Rev.*, **98**, 1123 (1955)

[2] S. W. Hawking and G. F. R. Ellis, The Large Scale Structure of Spacetime, (Cambridge University Press, Cambridge, 1972)

[3] R. M. Wald, General Relativity, (University of Chicago Press, 1985)

[4] R. Capovilla and J. Guven, *Phys. Rev.*, **D 52**, 1072 (1995)

[5] S. Kar, *Phys. Rev.*, **D 53**, 2071 (1996)

[6] F. Tipler, *Ann. Phys.*, **108**, 1 (1977)

[7] A. L. Larsen and V. P. Frolov, *Nucl. Phys.*, **B412**, (1994)

[8] S. Flugge, Practical Quantum Mechanics, (Springer Verlag, Berlin)

[9] S. Kar, *Phys. Rev.*, **D 52**, 2036 (1995)

[10] M. A. Abramowitz and I. R. Stegun, Handbook of Mathematical Functions, Tables and Formulae, (Dover Publications, New York, 1965)

[11] S. Kar, S. N. Minwalla, D. Mishra, D. Sahdev **D51**, 1632 (1995)

[12] S. Kar, (In Preparation)

[13] S. Kar, *Generalised Raychaudhuri Equations for strings in the presence of an antisymmetric tensor field* IP–BBSR–96/21

So far, my remarks have been confined to what we may all concede as great ideas conceived by great minds. It does not, however, follow that beauty is experienced only in the context of great ideas and by great minds. This is no more true than that the joys of creativity are restricted to a fortunate few. They are, indeed accessible to each one of us provided we are attuned to the perception of strangeness in the proportion and the conformity of the parts to one another and to the whole. And there is satisfaction also to be gained from harmoniously organizing a domain of science with order, pattern, and coherence.

— S. CHANDRASEKHAR

AN OVERVIEW OF EXACT SOLUTIONS OF EINSTEIN'S EQUATIONS

D.C.Srivastava[1]

Department of Physics
University of Gorakhpur
Gorakhpur 273 009, India

Abstract

In this note an overview of reviews on exact solutions for stationary axisymmetric fields and anisotropic cosmological models has been presented. Senovilla class of singularity free space time is reviewed.

[1]e-mail: dcs@gkpu.ernet.in

A final observation bearing on the attitude to physical problems that we have maintained in our studies of the mathematical theories of black holes and of colliding waves: in the theory of spacetimes with two Killing fields one timelike and one spacelike or both spacelike- the basic governing equations are one or more Laplace's equations, the Ernst equations, and the X- and Y-equations. The simplest solutions of the Laplace and the Ernst equations have provided all the fundamental solutions describing black holes and colliding waves. But the simplest solution of the X- and Y- equations have been left out. At long last it has found its place: it provides the first nontrivial binary black hole solution with supporting strings.

— S. CHANDRASEKHAR.

1 Introduction

Einstein's field equations

$$G^{ij} = -8\pi T^{ij} \tag{1.1}$$

where T^{ij} represents the energy momentum tensor of the matter and field, are second order non linear partial differential equations. There are basically three approaches to study them : exact solutions, approximation schemes and numerical computation. These approaches are listed in order of decreasing aesthetic appeal. The exact solutions as explained below, are necessarily very special and constitute a measure of zero in the space of all solutions. However, exact solutions play an important role in generating tests for numerical codes [Centrella (1986)] and also for providing checks on the validity of the approximation schemes. They can also provide models for important, though oversimplified, physical situations e.g. the Schwarzchild, and Kerr black holes and the Friedmann cosmological model on which almost all of the relativistic astrophysics and astrophysical cosmology is based. Besides, they can also provide counter-examples to conjectures [Misner (1963)]; for a discussion of these aspects one may refer to [MacCallum (1984, 1985 a,b)].

Due to the nonlinearity of the Einstein's equations it is impossible to solve these equations in their full generality. The usual practice is to assume, at the outset, some symmetry e.g. Cylindrical, Planar, Axial or the ones described in terms of Killing vectors. In addition to this, one has to resort to some specific energy momentum tensor. The most commonly used energy momentum tensors are that characterising dust, perfect fluid, electromagnetic field, imperfect fluid and radiation field. At this point it is to be noted that the solutions where energy momentum tensor is generalized by way of introducing, say null radiation, electromagnetic field, heat conducting fluid etc., and which do not leave any of its trace in the metric coefficients are not important unless a physical interpretation of decomposing the energy momentum tensor and an explanation of how ambiguities in the decomposition are to be removed is provided. [MacCallum (1987), Krasinski (1993)].

Exact solutions is one of the topics which attract many scientists working in General Relativity. About 10% of the research papers at the International conferences on the subject and 40% on the average at the Indian conferences are devoted to this topic. Hence, it is very difficult to present an overview of the subject in such a short duration unless some selection rules and procedures are adopted. Fortunately, in the literature, good review articles are available from time to time on this topic and it would be easier to paint a 'broad-brush' picture of the state of affairs by looking at these reviews. I would present an overview of these reviews with the year 1974 as the cut off. Besides, we have to narrow down our interests to (i) stationary axi-symmetric fields and (ii) inhomogeneous cosmologies. This restriction, is guided mainly by the fact that the physically relevant solutions fall into these two subtopics but reflects my personal interests and prejudices as well.

2 Stationary Axisymmetric Fields

Kinnersley (1974) describes briefly all known vacuum solutions including electrovacuum and their inter-relationships. He discusses space times admitting two commuting 2-forming Killing vectors containing (i) Tomimatsu-Sato solution (ii) stationary axially symmetric solutions and various transformation theorems for generating solutions from other known solutions. Here, the algebraically special solutions with non-diverging rays and with non rotating rays have also been discussed.

Starting from the early seventies and for about a decade there has been active interest and considerable progress in obtaining new axisymmetric stationary solutions of Einstein's field equations. The main motivation was the close resemblance of Einstein's equations, for this case, with the non-linear differential equations e.g. Sine-Gordon or Korteweg-de Vries (KdV) equations arising in other branches of Physics. Consequently, extensive use of Backlund transformations and Soliton techniques has been made. The various topics e.g. Hoensaelers, Kinnersley, Xanthopolous (H K X) transformations, Kinnersley-Chitre(K C) transformation, Geroch group have been discussed thoroughly and every aspect has been considered in detail at the Seminar with the title "Solutions of Einstein's Equation : Techniques and Results" edited by Hoensealers and Dietz (1994). For the benefit of the reader list of topics covered there are included as Appendix I. For an earlier review of the topic one may refer to Kramer and Stephani (1980).

2.1 Anisotropic Cosmology

The well known Friedmann-Lemaitre Robertson-Walker (FLRW) model, represents Universe in which all space points and all directions at any space point are equivalent. If one assumes the matter to be represented by dust the Universe turns out to be homogeneous. The isotropy on a large scale is confirmed by experiments but the issue whether the Universe was homogeneous and isotropic already at its early stages and has it still these properties in very distant regions is so far unresolved. Besides, there are mathematical arguments in favour of inhomogeneous models [Tavakole and Ellis (1988)]. It is to be pointed that even at the very early stage of Relativistic Cosmology several physicists were able to see the importance and necessity of investigating inhomogeneous cosmological models [Dingle (1933), Tolman (1934)]. For a detailed discussion of these aspects one may refer to a recent review by Krasinski (1993) where one may find topics e.g. memorable statements about the cosmological principle, why one should consider inhomogeneous models of the universe. In view of the above there has always been active interest in inhomogeneous cosmological models and in the literature several good review articles e.g. MacCallum (1984), Jantzen (1987), Kramer and stephani (1980) are available for a recent review one may refer to Krasinski [1993, 1994].

2.1.1 Senovilla Class of Solution

Senovilla (1990) has discovered an important solution representing a perfect fluid distribution with cylindrical symmetry and obeying an equation of state $\rho = 3p$. This solution represents the distribution for $-\infty < t < \infty$ having the curvature as well as matter invariants regular and smooth everywhere. The fluid flow lines have inhomogeneous expansion, nonvanishing shear and anisotropic acceleration. Until now the conventional and prevalent view of Cosmology was that of FLRW which represents spherically symmetric isotropic and homogeneous perfect fluid distribution with vanishing shear and homogeneous acceleration. The anisotropic and inhomogeneous models obtained by earlier workers viz. Wainwright and Goode (1980), Feinstein and Senovilla (1980), Davidson (1991), possess space - like Big Bang singularity. Consequently, it was considered that this would represent general singularity structure in other models as well. This experience was strongly aided by the singularity theorems, according to which if one adheres to Einstein's theory of General Relativity and assumes physically reasonable conditions of positivity of energy, causality and regularity etc, the initial singularity ($t = 0$) is inescapable. Chinea, Fernandez-Jambrina and Senovilla (1992) have carried out the analysis of geodesics and have found that the Senovilla spacetime is geodesically complete. Senovilla and his coworkers have made a thorough scrutiny to unravel how this solution could avoid the powerful singularity theorem. Senovilla argues that the general reason presented by Hawking and Ellis (1973) in establishing the singularity theorem makes use of the assumption of geodesic motion of the cosmological matter which is clearly not supported by any theoretical reasoning. He further finds that the singularity free nature of his solution is in accordance with Raychaudhuri's work [1955] according to which the presence of acceleration or rotation may prevent the existence of a universal singularity in our past [Senovilla (1995)]. The other reason is attributed to the shear of the fluid flow- lines which is nonvanishing and to the nonexistence of compact surfaces. This is interpreted to mean that nowhere in the spacetime gravity becomes strong enough to focus geodesics in a small compact region for the trapping of the all particles including photons to take place [Dadhich, Tikekar and Patel (1993)]. Recently Ruiz and Senovilla (1992) have obtained a general class of inhomogeneous perfect fluid solution which contains singular solutions due to Wainwright and Goode (1980), Feinstein and Senovilla (1989) and the solution being discussed presently. It is easier to understand Senovilla's solution if we express FLRW solution in cylindrical coordinates (τ, r, ϕ, z) as

$$ds^2 = T^2[-d\tau^2 + dr^2 + (1 + M\sum^2)d\phi^2 + \sum^2 dz^2] \tag{2.1}$$

where T is an arbitrary function of time and M is an arbitrary constant ; its sign is related the usual curvature index in FLRW model. $\sum$ is a function of r given by

$$\sum^{\prime 2} = 1 + M\sum^2 \tag{2.2}$$

where a prime indicates derivative with respect to r. The density ρ and the pressure p of the fluid are given by

$$\rho = \frac{3}{T^2}\left(\frac{\dot{T}^2}{T^2} - M\right) \quad , \quad p = -\frac{1}{T^2}\left(\frac{2\dot{T}}{T} - \frac{\dot{T}^2}{T^2} - M\right) \tag{2.3}$$

where an overhead dot represents derivative with respect to time. The Senovilla solution is given by

$$\begin{aligned} ds^2 &= T^{2(1+n)}\sum^{2n(n-1)}\left[-d\tau^2 + dr^2\right] \\ &\quad + T^{2(1+n)}\sum^{2n}\sum^{\prime 2} d\phi^2 \\ &\quad + T^{2(1-n)}\sum^2(1-n)dz^2 \end{aligned} \tag{2.4}$$

where $\sum$ is a function of r satisfying

$$\sum^{\prime 2} = M\sum^2 + 1 - nK\sum^{2(1-2n)} \tag{2.5}$$

and M, n and k are arbitrary constants. There arises two cases :

$$(i) \quad n = 0 \quad \text{and} \quad T \quad \text{arbitrary} \tag{2.6}$$

$$(ii) \quad n = 0 \quad \text{and} \quad T^2 = \left\{ \begin{array}{lll} A\cos h(2\sqrt{M}C) + B\sin h(2\sqrt{\tau}) & , & M > 0 \\ A\tau + B & , & M = 0 \\ A\cos h(2-\sqrt{-M}\tau) + B\sin h(2\sqrt{-M}\tau) & , & M < 0 \end{array} \right\} \tag{2.7}$$

The expansion θ, acceleration a_i and shear σ_{ij} have their nonvanishing components as given by

$$\left.\begin{aligned} \theta &= (n+3)\sum^{n(1-n)}\frac{\dot{T}}{T^{n+2}} \\ a_1 &= n(n-1)\sum^{n(1-n)}\frac{\sum}{T^{n+1}\sum} \\ \sigma_{11} &= \sigma_{22} = \sigma_{33} = \frac{2n}{3}\sum^{n(1-n)}\frac{\dot{T}}{T^{n+2}} \end{aligned}\right\} \tag{2.8}$$

The anisotropy of the fluid motion, the scale factor $R(\tau)$ and deceleration parameter q are given by

$$\left.\begin{aligned} \frac{\sigma}{\theta} &= \frac{2}{\sqrt{3}}\left[1 - \frac{3}{n+3}\right] \\ R(\tau) &= T^{(n+3)/3} \\ q &= 2 - \frac{n}{n+3}\left(1 - \frac{T\dot{T}}{T^2}\right) \end{aligned}\right\} \tag{2.9}$$

The solution reduces to FLRW solution for $n = 0$. The physical scenario described by this solution is as follows. The distribution if ; we assume it initially to contract, collapses to $r = 0$ and then reverses its motion and expands for ever. This model may be compatible with observational fact that the present Universe is to be homogeneous and expanding provided one prefers to have the dynamics of the distribution as given by

$$T(\tau) \sim \begin{cases} T_{SF} \quad \text{around } \tau = 0 \text{ with } \quad \text{any } n \\ \\ T_{FLRW} \quad \text{for } C = \bar{\tau} \quad \text{with} \quad n = 0 \end{cases} \tag{2.10}$$

where T_{SF} represents the time scale for singularity free Senovilla solution valid up to $\bar{\tau}$, the instant of bounce and T_{FLRW} is the time scale as given by FLRW solution. However, there are many conceptual issues to be resolved. For a discussion of these one may refer to Senovilla (1995). Some recent works attempting to establish the uniqueness of Senovilla class of cylindrically symmetric solution among separable, irrotational space times are going on and encouraging results are reported [Dadhich, Patel, Govinder and Leach (1994), Dadhich, and Patel (1995).]

2.1.2 Spherically Symmetric Non-vanishing Shear Models

I have remarked above that the existence of the shear of the four velocity is attributed to be a reason for resulting in a singularity free solution. For spherically symmetric perfect fluid distribution with vanishing shear there is no dearth of exact solutions. [Srivastava (1987,1992)] But the investigations with non-vanishing shear are rare and may be counted. The solutions for a stiff matter [Wesson (1978), Vaidya (1968)], and for equation of state $\rho = \rho(p)$, under the assumption of separability of metric coefficientists [Vandenberg and Wils (1985)] are available. The additional assumptions to yield a solution are (i) existence of self similarity [Collins and Lang (1987)], (ii) The four velocity begin perpendicular to surface = constant [Vaidya 1968)]. Biech and Das (1990) have shown null coordinates to be useful for obtaining solutions of this class. This investigation also presents reviews of the earlier works. Recently an interesting result has emerged out of the earlier works of Mc Vittie and Wiltshire (1977) ; the solution is found to have shear. [Bonnor and Knutsen (1993), and Knutsen (1995)]. This investigation employs noncomoving coordinates. It is interesting to remark that investigations for solutions with shear has been a topic of interest in the very early days. [Narlikar (1936), Narlikar and Moghe (1935)].

2.1.3 Shearfree Models

In contrast to the above there is sufficient reason to study the perfect fluid distributions with vanishing shear. As has been discussed by Collins (1986) shear free solutions would retain the feature of isotropy of local motions but redshifts need not be isotropic. It is shown following Ellis (1971) that relative recessional motion of the neighbouring galaxies is isotropic iff $\sigma_{ij} = 0$ whereas for the relative red shift to be isotropic one would need in additon vanishing of $\dot{U}_i$. Further the isotropy of the transverse motion of the neighbouring galaxies would require $\sigma_{ij} = w_{ij} = 0$. Thus $\sigma_{ij} = 0$ is a common feature of all these aspects. For a review of shear free perfect fluid solutions one may refer to Barnes (1983). Besides, there is another strong reason to study shear free models. It is because of following conjecture due to Collins [1986].

$$\sigma_{ij} \; = \; 0 \quad \Rightarrow \quad w\theta \; = \; 0 \tag{2.11}$$

This conjecture is based on the observation that for shear free motions;it must have either $w = 0$ or $\theta = 0$. The conjecture has been proved for (i) and parallel to each other [White and Collins (1984)] and (ii) the vanishing magnetic part of the Weyl tensor [White (1986)]. The importance of the study is in obtaining a counter-example to the conjecture because in Newtonian theory it is fairly readily shown that for because in Newtonian theory it is fairly readily shown that for self gravitating shear free fluids neither vorticity nor expansion vanish. The intuitive picture in general relativity would therefore, be illusive if conjecture holds. At this point I would like to quote Chandrasekher.

"On the relativistic theory, the frequencies of oscillation of the non-radial models (as we have shown) depend only on the distribution of the energy density and the pressure in the static configuration and the equation of state only to the extent of its adiabatic exponent. If this is a true representation of the physical situation then it must be valid in the Newtonian theory as well : the true nature of an object cannot change with modes and manner of one's perception".

— S. CHANDRASEKHAR (1991).

3 Commentary

Kinnersely (1974) while presenting his talk on "Recent progress in exact solutions" at the GR7 International Conference has remarked "the study of exact solutions has acquired a rather low reputation in the past, for which there are several explanations. Most of known exact solutions describe situations which are frankly unphysical and these do have a tendency to distract attention from useful ones. But the situation is also partially the fault of us who work in the field. We toss in null currents, macroscopic neutrino fields and tachyons for the sake of greater "generality"; We seem to take delight at the invention of confusing anti-intuitive notion ; and when all is done we leave our newborn wobbling on its vierbien without any visible means of interpretation".

More or less similar views have been expressed by later workers. There is another aspect of the problem. We have too many solutions rather than too few solutions. Simple minded attempts to derive a new solution from natural assumptions are likely to result in yet another discovery. As a side effect, some simple results are published again and again while important fine developments remain unnoticed for a long time [Krasinski (1993)]. For example, the investigation of Kustaanheimo and Kvist (1948) pertaining to shear free perfect fluids could only be noticed after the publication of book by Kramer et al. (1980). To avoid such repetition and also for our own sake we should not stop just after obtaining a solution but should look for invariant characterisation. At present the literature is full of materials where these aspects are clearly discussed e.g. MacCallum (1987), Krasinski (1994), Kramer et al. (1980).

In fact even in the early seventies it was being felt that number of exact solutions was growing; consequently to any serious minded research worker in the field it was necessary to have a stock - taking of known solutions in order to avoid duplication. The notable investigations among these are Kramer, Neugebauer and Stephani (1972) and Kinnersley (1974). Kramer et. al started in 1975 an exhaustive survey of the literature on exact solutions of Einstein's field equations and have published their results in the form of a book [Kramer et. al (1980)]. It is devoted to the classification and construction of exact solutions. They have discussed the classification schemes for exact solutions viz. (i) algebraic classification of conformal curvature tensor (Petrov types), (ii) algebraic classification of the Ricci tensor (Plebanski types) and the physical characterisation of the energy momentum tensor, (iii) the existence and structure of the preferred vector fields and (iv) the group of symmetry admitted by the metric. However, they have based the classification of the known solutions primarily on (i) symmetry group and (ii) Petrov types and subdivided these to include other main classification schemes. Besides, these studies have remarkably brought to our notice many old works to name a few, Wyman (1946), Narlikar (1947), Kustaanheimo and Qvist (1948) which were otherwise almost out of sight but are more rigorous and deep as compared to the later works which rederive their results.

3.1 What should we be doing?

An uptodate account of exact solutions has been presented by MacCallum at ICGC 87 and the material covered there under the titles what solutions do we know? Can we find physically interesting solutions?, Can we put realistic physics into our solution? Can we interpret the solutions we find? Do we find interesting mathematics? is relevant even today and the reader may refer to it

[MacCallum (1987)]. However, I would conclude my comments, which are of course, not completely independent.

- Tackle whole class of solutions at once.
- Analyse the symmetry structure of field equations viz. Lie, Painle've analysis. A good recent reference to Painle've analysis of which I am aware may be the lecture notes (unpublished) of Leach, Govinder (1994).
- Explore new solutions for the well identified field e.g. solution having two Killing vector fields, source for Kerrmetric, anisotropic inhomogeneous cosmologies, radiating star, colliding plane waves, etc.
- Apply more efforts towards physical interpretation of the known solutions rather than to increase the number of solutions.
- Have more physical motivation in obtaining exact solution than the mathematical skill. I would explain this point by an example from one of my work [Srinivastava (1987)] in appendix II.
- Obtain the invariant characterisation of the solution e.g. kinematical parameters, classification of Weyl tensor, and of principal null eigen vectors after one has obtained it.

Acknowledgements

I am thankful to Professors R.M.Misra and N.Dadhich for various useful discussions.

APPENDIX – I

Some topics in the Proceedings, Seminar on Solutions of Einstein's field Equations: Techniques and results

Sl.No.	Title	Page	Authors
1.	Backlund transformations in general relativity	1-25	Kramer, D. and Neugehauer, G.
2.	Vector Backlund transformation and associated superposition principle	55-67	Chinea, F.J.
3.	H K X transformations : An introduction	68-84	Hoensaelers, C.
4.	The Geroch group is a Banch Lie group	113-127	Schmodt, B.G.
5.	On the homogeneous Hilbert problem for effecting Kinnersley-Chitre transformations	128-175	Hauser, I.
6.	Non-iterative methods for constructing exact solutions of Einstein equations	186-198	Guo, D.S.
7.	Inverse scattering, differential geometry, Einstein Maxwell solutions and one soliton Backlund transformations	199-234	Gurses, M.
8.	N Kerr particles	311-320	Yamazaki, M.
9.	Algebraically special, shearfree diverging and twisting vacuum and Einstein-Maxwell fields	321-333	Stephani, H.
10.	Exact solutions in Cosmology	334-366	MacCallum, M.A.H.

APPENDIX – II

The purpose of this appendix is to show, with the help of an example, how physical motivation renders a solution not to be new which otherwise had appeared as new. The solution refers to spherically symmetric perfect fluid distribution executing shearfree motion [Srivastava (1987a)].

We start with the metric ansatz

$$dS^2 \;=\; \left(\frac{1-\Phi}{1+\Phi}\right)^2 dt^2 - \left[\frac{(1+\Phi)^2 S}{V}\right]^2 \left(dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right) \qquad (II.1)$$

where $V = V(r,t), S = S(t)$ and $\Phi = \Phi(r,t)$ are arbitrary functions of their arguments. An analysis of field equations reveals that

$$\Phi \;=\; \frac{1}{C}\left[\int \left(\frac{C}{S}\right) S dt + U(r)\right] \qquad (II.2)$$

$$C \;=\; bS[1-2as]^{-1/2} \qquad (II.3)$$

$$U'' - \frac{U'}{r} + \frac{2U'V'}{V} = \begin{cases} -\frac{1}{2}V\tilde{F} & , \quad a = 0 \\ \frac{a}{2}UV\tilde{F} & , \quad a \neq 0 \end{cases} \qquad (II.4)$$

$$\frac{V''}{V} - \frac{V'}{rV} = \frac{a}{2}V\tilde{F} \qquad (II.5)$$

where a and b are arbitrary constants and $U(r)$ and $\tilde{F}(r)$ are arbitrary functions of r. It is to be noted for later reference that $\tilde{F}$ is related to Weyl conformal curvature invariant as

$$\psi_2 = -\frac{1}{3}\tilde{F}\left[\frac{S}{V}(1+\Phi)\right]^{-3} \qquad (II.6)$$

Hereafter the analysis gets divided into two classes.

$$CaseI \ : \ a = 0 \ , \quad CaseII \ : \ a \neq 0 \qquad (II.7)$$

In the first case the functions U, V and $\tilde{F}$ are determined explicitly and the solution becomes the one obtained earlier by Glass and Mashhoon (1976). Whereas in the second case one has to solve equations :

$$U'V^2 = q^4(a^2U^2 - b^2) \ , q = const. \qquad (II.8)$$

$$\frac{4a^2b^2U'^2}{(a^2U^2-b^2)^3} - \frac{2U'''}{U'} + \left(\frac{U'}{U}\right)^2 = \frac{1}{r^2} \qquad (II.9)$$

These equations can be solved in principle but it had not been possible to find $U(r)$ explicitly. At this stage I thought that the two cases ought to be new precisely because of the way they appear in (II.4). However, before the solution for the case (ii) could be claimed to be new its invariant classification is needed. For this, one way is to evaluate $\tilde{F}(r)$ for the two cases. Fortunately, following the analysis made by Stephani (1983) I could evaluate the function $\tilde{F}(r)$ for the second case and met happy surprise to see that (II.9) is a perfect third order differential equation.

$$\left(\frac{a^2U^2-b^2}{U_x}\right)_{xxx} = 0 \, , (\ \ )_x = \frac{d}{dx}(\ \ ) \ , x = r^2 \qquad (II.10)$$

The function $\tilde{F}(r)$ for the case (ii) has the same functional form as in the case (i) hence the two cases should be the same. This means that by a suitable parameterisation equations (II.4)-(II.5) should reduce to corresponding ones with $a = 0$. The required parameterisation is found to be

$$\left.\begin{array}{lcl} U \rightarrow u & = & \frac{aU-b}{aU+b} \\ V \rightarrow v & = & \frac{-2ab^2V}{(aU+b)^2} \end{array}\right\} \qquad (II.11)$$

which in no way is an obvious parameterisation. This illustrates the point.

References

[1] Barnes, A. (1983) Shear free flows of a perfect fluid in classical general relativity, Proc. conf. on classical (non quantum) general relativity, eds. W.B. Bonnor, J.N. Islam and M.A.H. MacCallum, City University, London, Cambridge University Press, 15

[2] Biech, T. and Das, A. (1990) On exact shearing perfect fluid solutions of the spherically symmetric Einstein field equations, *Canad. J. Phys.*, **68**, 121

[3] Bonnor, W.B. and Knutsen, H. (1993) *Int. J. Theor. Phys.*, **32**, 1061

[4] Chandrashekhar, S. (1991) *Proc. Roy. Soc. London*, **A432**, 247-279

[5] Centrella, J.M., Shapiro, S.L., Evans, C.R., Hawley, J.F. and Teukolosky, S.A. (1986) Test-bed calculations in numerical relativity; in Dynamical Space-times and Numerical relativity, ed. J.M. Centrella, p.326-344, Cambridge University Press

[6] Chinea, F.J., Fernandez, Jambrina. L and Senovilla, J.M.M. (1992) Singularity-free space times, *Phys. Rev.*, **45D**, 2, 481

[7] Collins, C.B. (1986) Shear free fluids in general relativity, *Canad. J. Phys.*, **64**, 2, 191-199

[8] Collins, C.B. and Lang, J.M. (1987) *Class. Quant. Grav.*, **4**, 61

[9] Dadhich, N. and Patel, L.K., Govinder, K. and Leach, P.G.L. (1994) Characterisation of orthogonal perfect fluid cosmological space-times, IUCAA, Preprint 27/94

[10] Dadhich, N.K. and Patel, L.K. (1995a) On the uniqueness of nonsingular inhomogeneous cosmological models, IUCAA preprint; and Dadhich, N. in; (1995) "Inhomogeneous Cosmological Models" eds. A. Molina and J.M.M. Senovilla, p. 63, World Sci.; invited talk at Spanish Relativity Society Conference, Sept. 12-14, 1994

[11] Dadhich, N. and Patel, L.K. (1995b) The role of shear in the expanding G orthogonal perfect fluid models, IUCAA preprint presented at ICGC-95, IUCAA, Pune, India, Dec. 13-19, 1995, To appear (1996) it Gen. Rel. Grav.

[12] Dadhich, N.K., Patel, L.K. and Tikekar, R. 'On singularity free spacetimes', (1995) *Pramana, Journal of Physics*, **44**, 4, 303-316

[13] Dadhich, N., Tikekar, R. and Patel, L.K. (1993) On singularity free cosmological models, *Curr. Sci.*, **65**, 9, 694

[14] Davidson, W. (1991) *J. Math. Phys.*, **32**, 1560

[15] Dingle, H. (1933) *Mon. Not. Roy. Astron. Soc.*, **94**, 134

[16] Ellis, G.F.R. (1971) Relativistic cosmology, in General Relativity and Cosmology, Course XLVII, Varenna, Italy (1969) ed. Sach, R.K. Academic Press, Inc. N.Y. (1971) 104

[17] Fenstein, A. and Senovilla, J.M.M. (1989) *Class. Quant. Grav.*, **6**, L89

[18] Glass, E.N., and Mashhoon, B. (1976) *Astrophys. J*, **205**, 570

[19] Hoensealers, C. and Dietz, W. eds. (1984) Solutions of Einstein's Field Equations: Techniques and Results, Retzbach, Germany, (1983) 205, Berlin, Springer Verlag

[20] Hawking, S.W. and Ellis, G.F.R. (1973) The Large Scale Structure of the Space Times, Cambridge Univ. Press, Cambridge, 350

[21] Jantzen, R.T. (1987) Exact cosmological solutions of gravitational theories, *Phys. Lett.*, **186**, 290-296

[22] Kinnersley, W. (1974) Recent progress in exact solution; Proc. 7th International conference on general relativity and gravitation, GR7 June 23-28, Israel, John Willy, N.Y.

[23] Knutsen, H. (1995) On a class of spherically symmetric perfect fluid distributions in noncomoving coordinates, *Class. Quant. Grav.*, **12**, 2817

[24] Kramer, D., Neugebauer, G. and Stephani, H. (1972) Konstruktion and charakterisierung von gravitationsfelderen, *Fortschritt. der. Phys.*, **20**, 1

[25] Kramer, D. and Stephani, H., (1980) Exact Solutions of Einsteins Field equations; in Proc. 9th International Conf. on General Relativity and Gravitation, Jena., ed. E. Schmutzer, Cambridge University Press, 75-92

[26] Kramer, D., Stephani, H., Mac Callum, M.A.H. and Herlt. E., (1980) Exact Solutions of Einstein's field equations, Cambridge University Press

[27] Krasinski, A. (1993) Physics in an inhomogeneous Universe preprint 1993/10, University of Cape Town, South Africa; published as (1996) "Inhomogeneous Cosmological Models", Cambridge Univ. Press

[28] Krasinski, A., (1994) Bibliography on inhomogeneous cosmological models, *Acta cosmologica*, **20**, 67

[29] Kustaanheimo, P. and Qvist, B., (1948) A note on some general solutions of the Einstein field equations in a spherically symmetric world, Soc. Sci., *Fenn. Comment. Phys- Math XIII*, **16**, 1-12

[30] Leach, P. and Govinder, K. (1994) Group and singularity analysis of differential equations, lecture at "Discussion meeting on Group and Singularity Analysis of differential equations, (unpublished), IUCAA, Pune, India

[31] MacCallum, M.A.H. (1984) Exact Solutions in Cosmology; in Solutions of Einstein equations : Techniques and results ed. C. Hoenselaers and W. Dietz, *Lecture Notes in Physics*, **205**, 334-336, springer verlag

[32] Mac Callum, M.A.H., (1985a) relativistic cosmological models in; Observational and Theoretical Aspects of Relativistic Astrophysics and Cosmology, Proc. Santander School (1984) ed. J. Sanz and L.I. Goicoechera, Singapore, World Scientific

[33] MacCallum, M.A.H. (1985b) Understanding the solutions of Einstein's equations, *Quart. J. Roy. Astro. Society* **26**, 127-136

[34] MacCallum, M.A.H. (1987) An overview of Exact solutions of Einstein's equations and their classification, Proc. ICGC-87, Dec. 14-19, Goa, India

[35] Mcvittie, G.C. and Wiltshire, R.J. (1977) *Int. J. Theor. Phys.*, **16**, 12, 1

[36] Misner, C.W. (1963) Taub-Nut space as counterexample to almost anything; in Relativity Theory and Astrophysics, 1, Relativity and Cosmology, ed. J. Ehlers, *Lectures in Applied Mathematics*, **8**, 160-169; Providence, R.I.; American Mathematical Society

[37] Narlikar, V.V. and Moghe, D.N. (1935) *Phil. Mag.*, **20**, 1104

[38] Narlikar, V.V. (1936) *Phil, Mag.*, **22**, 767

[39] Narlikar, V.V., (1947) Some new results regarding spherically symmetric fields in relativity, Curr., Sci. letter to the editor, (India) 113-114

[40] Raychaudhuri, A.K. (1955) *Phys. Rev.*, **90**, 1123

[41] Ruiz, E. and Senovilla, J.M.M. (1992) General class of inhomogeneous perfect fluids, *Phys. Rev.*, **45D**, 6, 1995

[42] Senovilla, J.M.M. (1990) New class of inhomogeneous cosmological perfect fluid solutions without Big Bang singularity, *Phy. Rev. lett.*, **64**, 19, 2219

[43] Senovilla, J.M.M. (1995) Towards realistic singularity free cosmological models, presented at ICGC-95, IUCAA, Pune, India, Dec., 13-19, 95

[44] Srivastava, D.C. (1987a) A class of interior solutions of Einstein field equations for spherically symmetric perfect fluid in shearfree motion, Abstract, Contributed papers at 11th International Conference on General Relativity and Gravitation, Sweden, July 6-12, 1986, 1, 280

[45] Srivastava, D.C. (1987b) Exact solutions for shear free motion of spherically symmetric perfect fluid distributions in general relativity, *Class. Quant. Grav.*, **4**, 1093-1117

[46] Srivastava, D.C. (1992) Exact solutions for shear free motion of spherically symmetric charged perfect fluid distributions in general relativity Fortschr. der. Physik, 40, 31-72

[47] Stephani, H. (1983) A new interior solution of Einstein's field equations for a spherically symmetric perfect fluid distribution in shear free motion, *J. Phys. A. Math. Gen.*, **16**, 3529

[48] Tavakol, R.K. and Ellis, G.F.R. (1988) *Phys. Lett.*, **A130**, 217

[49] Tolman, R.C. (1934) *Proc. Nat. Acad. Sci.*, (USA) **20**, 169

[50] Vaidya, P.C. (1968) *Phys. Rev.*, **174**, 1615

[51] Vandenbergh, N. and Wils, P. (1985) *Gen. Rel. Grav.*, **17**, 223

[52] Wainwright, J. and Good, S.W. (1980) *Phy. Rev.*, **D22**, 1906

[53] Wesson, P.S. (1978) *J. Math. Phys.*, **19**, 2213

[54] White, A.J. and Collins, C.B. (1984) *J. Math.*, **25(2)**, 332

[55] White, A.J. (1986) *J. Math. Phys.*, **25**, (4) 995

[56] Wyman, M. (1946) Equation of state for radially spherically symmetric distribution of matter, *Phys. Rev.*, **70**, 396-400

QUANTUM GRAVITY AND STRING THEORY

Jnanadeva Maharana [1]

Institute of Physics,
Bhubaneswar 751005, INDIA

Abstract

The recent developments in string theory are discussed and it is argued that string theory holds the most promising prospect of solving some of the important problems of quantum gravity in a unified manner.

[1]e-mail: maharana@iopb.ernet.in

The fact that two standard theorems of the subject, the non-existence of solutions describing assemblages of black holes except those of the extreme Reissner Nordstrom type and the rule $|Q| \leq |M|$ for isolated charged black holes can be transgressed with a minimal violation of the smoothness requirements on the spacetime manifold by allowing only simple conical singularities suggests the fragility of the theorems of general relativity to the strict enforcement of smoothness conditions....Could one conclude from the very natural way in which strings emerge in the binary black-hole solution that *strings* are indeed *predicted* by *general theory of relativity*.

—S. CHANDRASEKHAR

1 Introduction

The most outstanding unsolved problem in this century is to synthesise the two magnificent theories of our time: quantum mechanics and general theory of relativity; to construct a theory which will describe gravitational effects and microscopic distances which is the order of the Planck scale. Some of the most brilliant minds have tried to quantize gravity eversince the birth of quantum mechanics and the problem has eluded generations of physicists over a period of half a century. It is accepted that the conventional method of renormalisation, so successful in describing quantum electrodynamics and in the construction of the so called 'Standard Model', when applied to Einstein's theory fails to remove the infinities in the perturbation theory. The Newton's constant appearing in the Einstein- Hilbert action carries a dimension and the power counting theorem tells us that the resulting theory is not renormalisable.

The gravitational interaction is very feeble in its strength, however our present knowledge of fundamental interactions and the facts about cosmology force us to conclude that we cannot ignore the phenomena associated with quantum gravity. Let us recall that the experimental observation of the 3^0K microwave background radiation and the asymmetry in its distribution support the Big Bang hypothesis and it is accepted that our Universe is 15 to 18 billion year old (could be a little less if the new experimental data remain unaltered) and it is still expanding. Thus if we go back in time, our Universe was very small in size and therefore, in the Planck epoch of 10^{-43} sec the laws of evolution of the Universe cannot be described by classical mechanics. We are naturally led to ask what is the quantum mechanical formulation of gravitation and what is the quantum mechanical wave equation associated with the quantum time evolution of the Universe ? Furthermore, the Universe today is large and the motion of the celestial bodies are described by classical physics in an appropriate frame work. Therefore, it is pertinent to seek answer to the question how did the quantum Universe evolve and at what stage one goes over to the classical description from the quantum description.

The next issue concerns the problem of vanishing cosmological constant. It it probably the physical parameter measured closest to zero. When the cosmological constant is viewed from the macro physics perspective, it is found to be small because the Universe is large and flat. However, it can be interpreted from another view point in the microscopic frame work. The cosmological constant will appear in the action like any other set of parameters that appear in physics such as masses and coupling constants when we write down an action. It is natural to expect that such parameters are controlled by the ultraviolet behavior of the theory and the short distance scale, in this case the Planck mass scale, will determine parameter. Thus we notice that the cosmological constant problem is related to the physics at the Planck scale and the effects of quantum gravity have an important role to play if we are to resolve this fundamental problem. It is amusing to note from the above discussions that the cosmological constant plays a dual role and it is probably the only parameter in physics which provides us a relation between the very large distance scale, dealing with the size of our Universe, and very short distance scale like the Planck length.

It is well known that black holes appear as macroscopic objects at final stages of stellar evolution. Classically, we expect that the matter falls into a black hole and disappears behind the event horizon and we have no means of communication between regions outside the event horizon and inside. Hawking has argued that a black hole radiates due to quantum mechanical effects. The distribution is that of a black body and the temperature is inversely proportional to the mass of the black hole. When the black hole is quite massive the calculations for the emission of the radiation can be carried out in the semiclassical frame work, as was done by Hawking in his seminal paper, where the metric is treated as a fixed background. It is justified for very massive 'black holes' since the energy loss at a given instant is very small and the metric does not get affected significantly due to a very small change in the mass of the black hole. However, at a later stage, when black hole has lost most of its energy such that the mass is quite small and the Schwarzschild radius is indeed very much reduced, say order of the Planck length, then a small change in mass due to Hawking radiation will affect the metric configuration significantly. Thus we must take into account the change in metric which is called the back reaction effect. Thus at this scale the semiclassical approximation is no longer valid and the problem deserves attention in the frame work of quantum gravity. It is also well known that when we look at the matter falling into a black hole we can prepare these incoming objects in quantum mechanically pure states. However, the Hawking radiation has a thermal distribution of a black body and thus the out going state is a mixed one. Consequently, we are not in a position to describe this phenomena by a unitary S-matrix. Thus we might have to re-examine foundations of quantum mechanics itself, if we are to describe the phenomena associated with a quantum black hole.

In view of the preceding observations, it is obvious that it is essential to construct a quantum theory of gravity if we want to consistently formulate laws of fundamental processes. It is accepted that the string theory has provided a basis to understand phenomena associated with quantum gravity in a unified frame work.

Let us recapitulate essential aspects of string theory [1]. Since it is a one dimensional object, as it evolves the string traces out a surface, the world sheet, just as when a particle evolves it traces the world line. The theory should be invariant under deformation of the world sheet and therefore, the conformal invariance of the string theory at the quantum level imposes stringent constraints on the theory. For example, the consistency of the theory demands that the bosonic strings live in twenty-six spacetime dimensions whereas such critical dimensions for the superstrings is required to be ten. Furthermore, if we consider the evolution of a closed string in the background of its massless excitations, then the conformal invariance restricts the configurations of these massless fields. We can envisage, in the first quantised approach, the evolution of the string in these backgrounds as a σ-model and these backgrounds find a natural interpretation as coupling constants of the theory. Now, conformal invariance of the quantum theory demands that the trace of the two-dimensional stress energy momentum tensor associated with the two dimensional world sheet metric must be traceless. This constraint translates to the condition that the β-functions associated associated with these coupling constants must vanish. Thus, the constraints on backgrounds are expressed as differential equations known as ‘equations of motion’. Since the β-functions are computed perturbatively, these equations of motions can be obtained order by order in the perturbation theory in the σ-model sense. These equations motion can be utilised to construct the string effective action in the following way: we look for an action in the target spacetime dimensions (26 for bosonic string in critical dimensions and 10 for superstrings) involving these background fields so that the variation of this action with respect to these backgrounds exactly reproduce the β-function equations alluded to above. The string effective action plays an important role in understanding of several aspects of string theory.

It is well known that the string theory possesses a very rich symmetry structure and these symmetries have a central role in unravelling dynamics of string theory. We have mentioned the symmetry of the string on the worldsheet which was so essential in deriving the constraints on the backgrounds. These are several symmetries in the target space. Of particular interest, are local symmetries such as general coordinate transformation associated with the presence of graviton in the theory, Abelian gauge transformation related to the antisymmetric tensor field and nonabelian gauge transformations present due to the massless nonabelian gauge fields in the compactified string theories. Indeed, the string effective action is invariant under these symmetry transformations.

There are another class of symmetries in the target space which involve global transformations. One such symmetry is T-duality. In its simplest form it tells us that if we have a string theory with one of its spatial dimension compactified on a circle of radius R, it is equivalent to a string theory compactified on a circle of radius $\frac{1}{R}$. The T-duality and its generalisation $O(d,d)$ symmetry can bested perturbatively and have several applications. Another important symmetry, which has attracted considerable attention in recent times, is the S-duality. Under this transformation, the strong and weak coupling phases of string theory are related. Under certain circumstances, S-duality transformation relates two different string theories and naturally it has played a very important role in our understanding of string dynamics in diverse dimensions. Naturally, S-duality symmetry cannot be tested in the perturbative frame work and it gives us a deep understanding of the string theory in its nonperturbative regime.

I would like to discuss now the vacuum configurations of string theory. When we seek solutions to the equations of motion we arrive at some background field configurations. For example, if we consider a string only in its graviton background, then the manifold of the target space has to be Ricci-flat. Each of these solutions to the equation of motion can be interpreted as the ground state of the associated string theory and thus called the vacuum of the theory. Thus if we get a cosmological solution, a black hole solution or a wormhole solution, each of these solutions corresponds to an acceptable vacuum configuration of the theory.

The transformations associated with T-duality and its generalized version, the noncompact global group $O(d,d)$, transform one set of vacuum background field configurations to another set of gauge inequivalent background fields. Furthermore, it is possible to cast the effective action, after it is dimensionally reduced to lower dimensions with d-compact coordinates, in a manifestly $O(d,d)$ invariant form.

It is natural to expect that string theory should be able to address fundamental questions pertaining to quantum gravity in a unified way and provide answers. Indeed, considerable attention has been focussed to understand the issues mentioned above and there has been definite progress in several directions. As an illustration, I shall present below an application of the string theory to black hole physics which exhibits how T-duality and S-duality play important roles in the context of black holes. This work has been done in collaboration with Amit Ghosh at the Saha Institute of Nuclear Physics.

Recently, Hawking and Ross [2] have implemented S-duality transformation to derive interesting results for the electrically and magnetically charged black holes. It is well known that the partition function for electrically charged black holes is analogous to that of the grand canonical ensemble with a chemical potential; whereas the partition function for the magnetically charged black holes has the interpretation of canonical partition function. Naturally, the two partition functions differ from one another. In order to investigate the implications of S-duality the partition function for electrically charged black hole is projected to one with definite charge and then it is argued that the partition functions for the electrically and magnetically charged black holes are identical.

It is recognized that the consequences of S-duality are interesting and surprising [3]. In the recent past, considerable attention has been focussed to investigate the implications of S-duality in string theory [4] and supersymmetric field theories [5]. This transformation, in its simplest form, when applied to electrodynamics, interchanges the role of electric and magnetic fields. However, in theories with larger number of field contents, the transformation rules for the fields are to be appropriately defined so that the equations of motion remain invariant under S-duality transformations. Note however, that the action is not necessarily invariant under these transformations. If we toroidally compactify ten dimensional heterotic string effective action [6] to four dimensions, the reduced effective action is endowed with N=4 supersymmetry. One of the attractive features of the four dimensional theory is that the quantum corrections are restricted due to the existence of non-renormalization theorems. It is expected that this theory respects strong-weak coupling duality symmetry and might provide deeper insight into the non-perturbative aspects of string theory. Furthermore, the four dimensional reduced action can be cast in a manifestly O(6,22) invariant form. Consequently, one can generate new background field configurations by implementing suitable O(6,22) and S-duality transformations on a known solution of the effective action. It is well known that string effective action admits black hole solutions[7]. Sen [8] has obtained interesting solutions such that they preserve N=4 supersymmetry, saturating Bogomol'nyi bound in the extremal limit.

Our aim here is to investigate the consequences of S-duality transformations for the black hole solutions of the four dimensional heterotic string effective action and investigate the thermodynamic properties of these black holes. First, we compute the partition function of an electrically charged black hole which resembles a grand partition function with a chemical potential. Then we adopt the techniques due to Coleman Preskill and Wilczek [9] to project to a partition function with definite charges. The free energy, entropy and the thermodynamical potential are derived. The entropy is shown to be zero in the extremal limit. Our result is of importance due to the fact that these extremal black hole solutions saturate Bogomol'nyi bound and the partition functions are expected not to get quantum corrections due to N=4 supersymmetry [14]. Our next result provide an important consequence of S-duality. We envisage electrically charged black hole solutions with zero axion field. Subsequently we implemented the S-duality transformation such that the electrically charged solutions go over to the magnetic ones. Furthermore, it is argued on the grounds of S-duality that the charged projected partition function for electrically charged black hole is identical to that of the magnetically charged black hole and hence the entropy is also vanishing for the magnetically charged black hole. While calculating the partition functions we use the Euclidean formulation throughout the paper.

The massless bosonic sector of four dimensional heterotic string effective action consists of dilaton, graviton, axion (dual to the antisymmetric tensor field) and 28 abelian gauge fields denoted respectively by ϕ, $G_{\mu\nu}$, ψ and $A_\mu^{(a)}$, $a = 1 \cdots 28$. The effective action, written in Einstein metric, $g_{\mu\nu} = e^{-\phi} G_{\mu\nu}$ takes the following form,

$$\begin{aligned} S &= \int d^4x \sqrt{-g} \Big\{ R + \frac{1}{2\lambda_2^2} \partial_\mu \bar{\lambda} \partial^\mu \lambda + \frac{1}{8} \mathrm{tr}(\partial_\mu ML \partial^\mu ML) \\ &- \lambda_2 F_{\mu\nu}^{(a)} (LML)_{ab} F^{(b)\mu\nu} + \lambda_1 F_{\mu\nu}^{(a)} L_{ab} \tilde{F}^{(b)\mu\nu} \Big\} + \text{ Boundary terms} \end{aligned} \tag{1}$$

where,

$$F_{\mu\nu}^{(a)} = \partial_\mu A_\nu^{(a)} - \partial_\nu A_\mu^{(a)}. \tag{2}$$

M and L are 28×28 symmetric matrices; whereas, M parametrizes the coset $\frac{O(6,22)}{O(6)\otimes O(22)}$, L has 22 eigenvalues -1 and 6 eigenvalues $+1$ and has the following form

$$L = \begin{pmatrix} -I_{22} & 0 \\ 0 & I_6 \end{pmatrix}. \tag{3}$$

M satisfies the property, $MLM^T = L$. λ is a complex scalar field defined in terms of the axion and the dilaton $\lambda = \psi + ie^{-\phi}$. The equations of motion associated with (1) remain invariant under the S-duality transformations

$$\begin{aligned} \lambda &\to \lambda' = \frac{a\lambda + b}{c\lambda + d}, \qquad ad - bc = 1 \\ F_{\mu\nu}^{(a)} &\to (c\lambda_1 + d) F_{\mu\nu}^{(a)} + c\lambda_2 (ML)_{ab} \tilde{F}_{\mu\nu}^{(a)}. \end{aligned} \tag{4}$$

where a,b,c and d are real. The backgrounds $g_{\mu\nu}$ and M remain invariant under the transformation.

The invariance property of the action under the non-compact global transformations has been utilized by Sen to generate charged rotating black hole solutions [8] starting from an uncharged rotating black hole solution with $\phi = 0$, $B_{\mu\nu} = 0$, $A_\mu^{(a)} = 0$, $M = I_{28}$. A special case corresponding to non-rotating charged black hole is given by the following background configurations,

$$ds^2 = -\Delta^{-1/2}(r^2 - 2mr)dt^2 + \Delta^{1/2}(r^2 - 2mr)^{-1} dr^2 + \Delta^{1/2} d\Omega_{II}^2 \tag{5}$$

where, $\Delta = r^4 + 2mr^3(\cosh\alpha \cosh\gamma - 1) + m^2 r^2 (\cosh\alpha - \cosh\gamma)^2$. The dilaton is given by,

$$e^\phi = r^2 / \Delta^{1/2} \tag{6}$$

The gauge fields are,

$$\begin{aligned} A_t^{(a)} &= -\frac{n^{(a)}}{\sqrt{2}} \frac{mr}{\Delta} \sinh\alpha [r^2 \cosh\alpha + mr(\cosh\alpha - \cosh\gamma)] \qquad 1 \le a \le 22 \\ &= -\frac{p^{(a-22)}}{\sqrt{2}} \frac{mr}{\Delta} \sinh\gamma [r^2 \cosh\alpha + mr(\cosh\gamma - \cosh\alpha)] \qquad 23 \le a \le 28 \end{aligned} \tag{7}$$

The M-matrix, satisfying the equations of motion is,

$$M = I_{28} + \begin{pmatrix} Pnn^T & Qnp^T \\ Ppn^T & Ppp^T \end{pmatrix} \tag{8}$$

where, $P = 2\frac{m^2r^2}{\Delta}\sinh^2\alpha\sinh^2\gamma$ and $Q = 2\frac{mr}{\Delta}\sinh\alpha\sinh\gamma[r^2 + mr(\cosh\alpha\cosh\gamma - 1)]$. The mass of this black hole is

$$M = \frac{m}{2}(1 + \cosh\alpha\cosh\gamma) \tag{9}$$

The 28 charges are given by

$$\begin{aligned} Q^{(a)} &= \frac{n^{(a)}}{\sqrt{2}} m \sinh\alpha\cosh\gamma \qquad 1 \le a \le 22 \\ &= \frac{p^{(a-22)}}{\sqrt{2}} m \sinh\gamma\cosh\alpha \qquad 23 \le a \le 28 \end{aligned} \tag{10}$$

Note that the "boost" angles α, γ are the parameters [8,10] that appear in global non-compact transformation to obtain charged solution from the uncharged one.

The horizon of this black hole is at $r = 2m$, there is a curvature singularity at $r = 0$, the area of the event horizon is $A_H = 8\pi m^2(\cosh\alpha + \cosh\gamma)$ and the Hawking-temperature is given by $T_H = \frac{1}{4\pi m(\cosh\alpha + \cosh\gamma)}$.

There are two extremal limits of this black hole which preserve supersymmetry and consequently saturate the Bogomol'nyi bound for the mass:

(I) $m \to 0$, $\gamma \to \infty$, while keeping $m\cosh\gamma = m_0$ finite and α is finite but arbitrary. For this case $\Delta = r^2(r^2 + 2m_0 r\cosh\alpha + m_0^2)$, the mass is $M = \frac{m_0}{2}\cosh\alpha$ and the charges are

$$Q_L^{(a)} = \frac{n^{(a)}}{\sqrt{2}} m_0 \sinh\alpha \quad 1 \le a \le 22, \qquad Q_R^{(a)} = \frac{p^{(a-22)}}{\sqrt{2}} m_0 \cosh\alpha \quad 23 \le a \le 28 \tag{11}$$

The horizon of this black hole is at $r = 0$, hence $A_H = 0$ and temperature , $T_H = \frac{1}{8\pi M}\cosh\alpha$.

The expression for the gauge fields become,

$$A_{tL}^{(a)} = -\frac{n^{(a)}}{\sqrt{2}}\frac{m_0 r^3}{\Delta}\sinh\alpha \quad 1 \le a \le 22, \quad A_{tR}^{(a)} = -\frac{p^{(a-22)}}{\sqrt{2}}\frac{m_0 r^3}{\Delta}\cosh\alpha \quad 23 \le a \le 28 \tag{12}$$

The Bogomol'nyi bound is saturated and $M^2 = \frac{1}{2}\vec{Q}_R^2$, where R stands for the right hand sector and the index a runs from 23 to 28. Also $Q_L^{(a)} = \sqrt{2}Mn^{(a)}\tanh\alpha$.

(II) The other black hole corresponds to limits: $m \to 0$, $\alpha = \gamma \to \infty$ such that $m\cosh^2\alpha = m_0$ is finite. The parameters are $\Delta = r^2(r^2 + 2m_0 r)$, $M = \frac{m_0}{2}$,

the charges are,

$$Q_L^{(a)} = \frac{n^{(a)}}{\sqrt{2}} m_0 \quad 1 \le a \le 22, \qquad Q_R^{(a)} = \frac{p^{(a-22)}}{\sqrt{2}} m_0 \quad 23 \le a \le 28. \tag{13}$$

The horizon is at $r = 0$, so $A_H = 0$ and $T_H = \infty$. The gauge fields are given by,

$$\begin{aligned} A_t^{(a)} &= -\frac{n^{(a)}}{\sqrt{2}}\frac{m_0 r^3}{\Delta} \qquad 1 \le a \le 22 \\ &= -\frac{p^{(a-22)}}{\sqrt{2}}\frac{m_0 r^3}{\Delta} \qquad 23 \le a \le 28 \end{aligned} \tag{14}$$

The mass saturates the Bogomol'nyi bound $M^2 = \frac{1}{2}\vec{Q}_R^2 = \frac{1}{2}\vec{Q}_L^2$.

We mention in passing that the thermodynamic properties follow naturally from the non-extremal solution. Define $\Phi^{(a)} = A_t^{(a)}|_{r=2m}$ to be the electrostatic potential at the horizon. Then the following relations hold.

$$\begin{aligned} T_H d\frac{A_H}{4} &= dM - \sum_a \Phi^{(a)} dQ^{(a)} \\ T_H \frac{A_H}{2} &= M - \sum_a \Phi^{(a)} Q^{(a)} \end{aligned} \tag{15}$$

The explicit form of $\Phi^{(a)}$ is given by,

$$\begin{aligned} \Phi_L^{(a)} &= \frac{n^{(a)}}{\sqrt{2}} \frac{\sinh\alpha}{\cosh\alpha + \cosh\gamma} \qquad 1 \le a \le 22 \\ \Phi_R^{(a)} &= \frac{p^{(a-22)}}{\sqrt{2}} \frac{\sinh\gamma}{\cosh\alpha + \cosh\gamma} \qquad 23 \le a \le 28 \end{aligned} \tag{16}$$

For case (I),

$$\Phi_L^{(a)} = 0 \quad 1 \le a \le 22, \qquad \Phi_R^{(a)} = \frac{p^{(a-22)}}{\sqrt{2}} \quad 23 \le a \le 28 \tag{17}$$

and for case (II),

$$\Phi_L^{(a)} = \frac{n^{(a)}}{2\sqrt{2}} \quad 1 \le a \le 22, \qquad \Phi_R^{(a)} = \frac{p^{(a-22)}}{2\sqrt{2}} \quad 23 \le a \le 28 \tag{18}$$

Now we turn to equations of motion,

$$\begin{aligned} R_{\mu\nu} = \frac{1}{2\lambda_2^2}(\partial_\mu\bar\lambda\partial_\nu\lambda + \partial_\nu\bar\lambda\partial_\mu\lambda) &+ \frac{1}{8}Tr(\partial_\mu ML\partial_\nu ML) - g_{\mu\nu}\frac{1}{16}Tr(\partial_\mu ML\partial^\mu ML) \\ &+ 2\lambda_2 F_{\mu\rho}^{(a)}(LML)_{ab}F_\nu^{(b)\rho} - \frac{1}{2}\lambda_2 g_{\mu\nu}F_{\rho\sigma}^{(a)}(LML)_{ab}F^{(b)\rho\sigma} \\ D_\mu(-\lambda_2(ML)_{ab}F^{(b)\mu\nu} &+ \lambda_1\tilde F^{(a)\mu\nu}) = 0 \end{aligned} \tag{19}$$

where $\tilde F^{(a)\mu\nu} = \frac{1}{2\sqrt{-g}}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}^{(a)}$ is the usual dual tensor satisfying Bianchi identities.

The first equation in (19) corresponds to the Einstein equation, where R has been eliminated in favour of the matter energy-momentum stress tensor and the second one corresponds to the gauge field equation.

In order to compute the partition function, we need to determine the action on shell. One of the efficient ways to obtain this action is to take the trace of $R_{\mu\nu}$ in (19) to obtain the scalar curvature, R, and use the expression in (1). A straightforward calculation shows that the contributions come from the well known gravitational boundary term as well as from the gauge field surface term

$$\frac{1}{8\pi}\int_{\Sigma^\infty} \sqrt{-g}d^3x \; n_r[-\lambda_2 A_t^{(a)}(LML)_{ab}F^{(b)tr} + \lambda_1 A_t^{(a)}L_{ab}F^{(b)tr}] \tag{20}$$

Now we proceed to compute the partition function for the electrically charged black hole. It is necessary to specify the time integral of the fourth component of the vector potential on the boundary. We follow the procedure of [9] to carry on this computation for the problem at hand where the effect of 28 gauge bosons are to be taken into account. The gauge potential and the electric field strength have the following form for asymptotically large r

$$\begin{aligned} \vec{A}_{tL/R} &= \frac{\vec{\omega}_{L/R}}{\beta e}(1 - m\frac{\lambda_{L/R}}{r}) \\ \vec{F}_{trL/R} &= \frac{m\lambda_{L/R}\ \vec{\omega}_{L/R}}{\beta e}\frac{1}{r^2} \end{aligned} \tag{21}$$

A few comments are in order at this point: the subscript L(R) refers to the gauge fields arising from the compactification of the left(right) moving string coordinates. We recall that 22 gauge fields arise from compactification of left hand sector and other 6 of them come from the right hand sector. Here $\vec{\omega}_{L/R}$ are the generalized version of the parameter introduced in [9]. Here β is the inverse temperature and we have introduced the parameter e to keep track of some power countings; however, note that e will not appear in our expressions for partition function and entropy. The equations (21) are used in (20) to compute the boundary term contributions. The constants $\lambda_{L/R}$ have the following form

$$\lambda_L = \cosh\gamma(\cosh\alpha + \cosh\gamma) \qquad \lambda_R = \cosh\alpha(\cosh\alpha + \cosh\gamma) \tag{22}$$

The geometry is asymptotically flat and the boundary is taken to be at $r = \infty$. As we have to subtract the flat space contribution according to the prescription of Gibbons and Hawking [11,9], we shall first take the radius to be large and finite and eventually take the limit $r \to \infty$ after the subtraction.

$$\begin{aligned} S_{boundary} - S^{flat}_{boundary} &= -\frac{1}{2}\beta(g_{rr})^{-1/2}\partial_r[\Delta^{1/2}\{(g_{rr}^{-1/2} - 1\}] \\ &= \frac{1}{2}\beta M \end{aligned} \tag{23}$$

where M is the mass of the black hole given by (9). The contribution of (20) together with (23) can be written as

$$Z(\beta, \vec{\omega}) = \exp[-\frac{1}{2}\beta M - \frac{1}{2e^2}\frac{\lambda_L\vec{\omega}_L^2 + \lambda_R\vec{\omega}_R^2}{4\pi(\cosh\alpha + \cosh\gamma)}] \tag{24}$$

This equation can be interpreted as a grand canonical partition function for electrically charged black holes where $\vec{\omega}$, collectively representing $\vec{\omega}_{L,R}$, play the role of chemical potential. In order to derive the partition function with specific charge configurations, we have to introduce a generalization of the projection technique of [9].

$$\begin{aligned} Z(\beta, \vec{Q}) &= e^{-\beta F(\beta, \vec{Q})} \\ &= \int_{-\infty}^{\infty} \Pi_a \frac{d\omega^{(a)}}{2\pi} \exp[-\frac{i}{e}(\vec{\omega}_L \cdot \vec{Q}_L + \vec{\omega}_R \cdot \vec{Q}_R)]\ Z(\beta, \vec{\omega}) \end{aligned} \tag{25}$$

where $\vec{Q}$ stands for both $\vec{Q}_L$ and $\vec{Q}_R$. This integral can be evaluated by the saddle point approximation after inserting the expression for $Z(\beta, \vec{\omega})$ given by (24). Thus

$$\begin{aligned} F(\beta, \vec{Q}) &= \frac{1}{2}M + 2\pi(\cosh\alpha + \cosh\gamma)\left(\frac{\vec{Q}_L^2}{\beta\lambda_L} + \frac{\vec{Q}_R^2}{\beta\lambda_R}\right) \\ &= \frac{1}{2}M + \frac{1}{2}(\vec{\Phi}_L \cdot \vec{Q}_L + \vec{\Phi}_R \cdot \vec{Q}_R) \end{aligned} \tag{26}$$

where $\Phi_{L/R}$ are the twenty two/ six components of the potentials given in (16). The thermodynamic potential is given by

$$\begin{aligned}\Omega(\beta,\vec{\Phi}) &= M - T\mathcal{S} - \vec{\Phi}\cdot\vec{Q} \\ &= F(\beta,\vec{Q}) - \vec{\Phi}\cdot\vec{Q}\end{aligned} \tag{27}$$

Thus the entropy, $\mathcal{S}$ is given by

$$\begin{aligned}\mathcal{S} &= \beta(M-F) \\ &= \frac{A_H}{4}\end{aligned} \tag{28}$$

We have used the relation (27) in arriving at (28) . Notice that in both the extremal limits; case (I) and case (II), mentioned earlier, *the entropy vanishes identically.* It is important to note that the quantum corrections to this entropy are vanishing due to the non-renormalization theorems.

Now we proceed to estimate the partition function for magnetically charged black holes which are obtained from the electrically charged ones by S-duality transformation. In general axion and dilaton are transformed to new configurations along with the gauge field strengths as given by (4). The electrically charged black hole corresponds to $\lambda_1 = 0$ and ϕ and $A_t^{(a)}$ given by (6) and (7). Our purpose is to implement a special class of duality transformation under which $\lambda_1' = 0$ and

$$\begin{aligned}\phi &\rightarrow -\phi \\ F_{\mu\nu}^{(a)} &\rightarrow -\lambda_2(ML)_{ab}F_{\mu\nu}^{(b)}.\end{aligned} \tag{29}$$

This is accomplished by a simple choice of parameters $a = d = 0$ and $b = -c = -1$. Note that under this transformation $(\vec{Q}_L,\vec{Q}_R) \rightarrow (-\vec{Q}_L,\vec{Q}_R)$. For the magnetically charged black holes the string coupling blows up at the horizon and strictly speaking the semiclassical approximation breaks down. Hence it becomes technically impossible to calculate the partition function and entropy for such black holes. Now since by $N = 4$ supersymmetry $Z(\beta,\vec{Q})$, or equivalently the free energy $F(\beta,\vec{Q})$ is an exact result, on the basis of S-duality on expects that this answer should remain the same in the strong-coupling region, namely for the magnetically charged solutions under consideration. One evidence is that the magnetic charges $(\vec{Q}_L,\vec{Q}_R)$ are to be regarded as the electric charges $(-\vec{Q}_L,\vec{Q}_R)$ in the dual theory and the projected partition function being a quadratic function of charges is indeed invariant under the S-duality transformations (29). Then we can equate the free energy obtained with $M - T\mathcal{S}$ to obtain the expression for the entropy; $\mathcal{S} = \frac{A_H}{4}$, which is the same for the electric case[2].

We see that our results derived from the string effective action imply that both the extremal black holes have vanishing entropy which is conformative of the works [12]. In the past it has been proposed that extremal black holes might be identified with elementary particles [13]/ massive states of the string theory [14]. Computation of partition functions and entropy for the extremal case are protected from quantum corrections due to presence of N=4 supersymmetry [15]. S-duality is believed to be exact symmetry of string theory [16], which has far reaching consequences. Indeed our work indicates another interesting implication of S-duality.

I would like to conclude this talk with a few remarks. Although string theory holds the promise of unifying all fundamental interactions, there are five different string theories: type IIA, type IIB, typeI with gauge group $SO(32)$, heterotic string theories, one with gauge group $SO(32)$ and another with $E_8 \times E_8$. Therefore, when we say that the string theory unifies fundamental interactions which string theory we should choose. There has been rapid progress during the last few months

[2]One should be careful about the statement regarding the use of supersymmetry, because it appears to be a symmetry only in the extremal limits. To really see that the argument goes through one has to ensure that the free energy has well defined limits for both the solutions I and II. In both the cases $F = M$ and as $\mathcal{S} = \beta(M-F)$, entropy vanishes automatically.

in our understanding of the string dynamics in the nonperturbative regime and we have made progress in revealing intimate relations among various string theories. Some of these results are nonperturbative in nature and thus are beyond the reach of perturbative calculations. One of the important breakthrough came with the realization of string/string duality in six dimensions. We know that in four dimensions, both magnetic and electrically charged particles can be accommodated. The point charged particle couples to the vector potential whereas the magnetic monopole appears as a soliton of the theory. The string couples to a two form potential and the corresponding field strength is a three form.Its dual is also a three form tensor in six dimensions. Thus if we have an 'electrically charged' string in six dimensions its dual is a 'magnetic string' and the Dirac quantization condition in this case tells us that the strong coupling phase of one theory is the weak coupling phase of the other and vice versa. This is an example of an S-duality transformation and we can test these results from string theories. Furthermore, it has been shown by Witten that the strong coupling limit of the type IIA theory takes it to $D = 11$, $N = 1$ supergravity theory. There has been intensive activities to construct string theories in diverse dimensions using various methods of compactifications and test the duality conjectures. We find all the different string theories are connected to one another in various interesting ways and there is now a strong belief that one has a hint for the existence of one fundamental theory. All the string theories we construct might be different phases of that unique theory. We have seen facets of this theory in so many ways, but the consistent quantum theory, the **M** theory, is yet to be discovered. This is the most challenging problem for us.

References

[1] *Superstrings*, M. B. Green, J. H. Schwarz and E. Witten, Cambridge University Press

[2] S. W. Hawking and S. F. Ross, *Duality between electric and magnetic black holes*, DAMTP/R-95/8

[3] C. Montonen and D. Olive, *Phys. Lett.*, **B72**, 1977 (1977); P. Goddard, J. Nyuts and D. Olive, *Nucl. Phys.*, **B125**, 1 (1977); H. Osborn, *Phys. Lett.*, **B83**, 321 (1979)

[4] A. Font, L. Ibanez, D. Lust and F. Quevedo, *Phys. Lett.*, **B249**, 35 (1990); S. J. Rey, *Phys. Rev.*, **D43**, 526 (1991); S. Kalara and D. Nanopoulos, *Phys. Lett.*, **267**, 343 (1991); A. Shapere, S. Trivedi and F. Wilczek, *Mod. Phys. Lett.*, **A6**, 2677 (1991); A. Sen, *Nucl. Phys.*, **B404**, 109 (1993), *Phys. Lett.*, **B303**, 22 (1993), *Mod. Phys. Lett.*, **A8**, 2023 (1993); J. Schwarz and A. Sen, **B312**, 105 (1993), *Nucl. Phys.*, **B411**, 35 (1994); P. Binetruy, *Phys. Lett.*, **B315**, 80 (1993); A. Sen, *Int. J. Mod. Phys.*, **A9**, 3707 (1994), *Phys. Lett.*, **B329**, 217 (1994); J. Gauntlett and J. Harvey, preprint EFI-94-36; L. Girardello, A. Giveon, M. Porrati and A. Zaffaroni, Preprint NYU-TH-94/06/02, preprint NYU-TH-94/12/01; M. Duff and R. Khuri, *Nucl. Phys.*, **B411**, 473 (1994); J. Schwarz, preprint CALT-68-1965; J. Harvey, G. Moore and A. Strominger, preprint EFI-95-01; A. Giveon, M. Porrati and E. Ravinivici, *Phys. Rep.*, **244C**, 77 (1994)

[5] C. Vafa and E. Witten, preprint HUTP-94-A017; N. Seiberg and E. Witten, *Nucl. Phys.*, **B426**, 19 (1994); *Nucl. Phys.*, **B431**, 484 (1994); A. Ceresole, R. D'Auria and S. Ferrara, *Phys. Lett.*, **B339**, 71 (1994); M. Bershadsky, A. Johansen, V. Sadov and C. Vafa, preprint HUTP-95-A004

[6] J. Maharana and J. Schwarz, *Nucl. Phys.*, **B390**, 4 (1993); S. Hassan and A. Sen, *Nucl. Phys.*, **B375**, 103 (1993)

[7] J. Horne and G. Horowitz, *Phys. Rev.*, **D46**, 1340 (1992); G. Horowitz, *Trieste 1992 Proceeding*s hep-th/9210119A, for a review and A. Sen, *Phys. Rev. Lett.*, **69**, 1006 (992)

[8] A. Sen, preprint TIFR-TH-94-47, hep-th/941187

[9] S. Coleman, J. Preskill and F. Wilczek, *Nucl. Phys.*, **B378**, 175 (1992)

[10] M. Gasperini, J. Maharana and G. Veneziano, *Phys. Lett.*, **B272**, 167 (1991); *Phys. Lett.*, **B296**, 51 (1992); A. Sen, *Phys. Lett.*, **B271**, 2951 (1991); *Phys. Lett.*, **B274**, 34 (1991)

[11] G. Gibbons and S. W. Hawking, *Phys. Rev.*, **D15**, 2752 (1977)

[12] R. Kallosh, A. Linde, T. Ortin, A. Peet and A. Van Proyen, *Phys. Rev.*, **D46**, 5278 (1992); G. W. Gibbons and R. E. Kallosh, *Phys. Rev.*, **D51**, 2839 (1995); S. W. Hawking, G. Horowitz and S. F. Ross, gr-qc/9409013

[13] S. W. Hawking, *Monthly Notices of Roy. Astron. Soc.*, **152**, 75 (1971); A. Salam, *Quantum Gravity; An Oxford Symposium*, Eds. C. Isham, R. Penrose and D. Sciama, Oxford Univ. Press (1975); G. 't Hooft, *Nucl. Phys.*, **B335**, 138 (1990); R. Kallosh and T. Ortin, *Phys. Rev.*, **D48**, 742 (1993)

[14] J. G. Russo and L. Susskind, preprint UTTG-9-94, hep-th/9405117; M. J. Duff and J. Rahmfeld, CTP-TAMU-25-94, hep-th/9406105; M. J. Duff, R. R. Khuri, R. Minusian and J. Rahmfeld, *Nucl. Phys.*, **B418**, 195 (1994)

[15] R. Kallosh, *Phys. Lett.*, **B282**, 80 (1992); R. Kallosh, A. Linde, T. Ortin, A. Peet and A. Van Proyen, *Phys. Rev.*, **D46**, 5278 (1992)

[16] J. H. Schwarz, *Proceedings of Workshop on Physics at the Planck Scale*, Puri December 1994; E. Witten, preprint IASSNS-HEP-95-18, hep-th/9503124

EIKONAL APPROACH TO PLANCK SCALE PHYSICS

Saurya Das[1]

The Institute of Mathematical Sciences
CIT Campus
Madras - 600 113, India

Abstract

We consider gravitational scattering of point particles with Planckian centre-of-mass energy and fixed low momentum transfers in the framework of general relativity and dilaton gravity. The geometry around the particles are modelled by arbitrary black hole metrics of general relativity to calculate the scattering amplitudes. However, for dilaton gravity, this modelling can be done *only* by extremal black hole metrics. This is consistent with the conjecture that extremal black holes are elementary particles.

[1]e-mail: saurya@imsc.ernet.in

After the early preparatory years, my scientific work has followed a certain pattern motivated, principally, by a quest after perspectives. In practice, this quest has consisted in my choosing (after some trials and tribulations) a certain area which appears amenable to cultivation and compatible with my taste, abilities, and temperament. And when after some years of study, I feel that I have accumulated a sufficient body of knowledge and achieved a view of my own, I have the urge to present my point of view *ab initio*, in a coherent account with order, form, and structure.

— S. CHANDRASEKHAR

Actually, completion of the theory was never achieved. Chandra wanted— and he tried very hard— to decouple and/or separate the perturbations of the Kerr-Newman black hole. It is well known in the scientific community that he did try and did fail. Since Chandra failed, no one seems willing to give this problem a serious try, and the perturbations of Kerr-Newman solution have remained an unsolved problem for the dozen years since Chandra gave up. Perhaps, for the sake of science, he should have kept his failure secret. On the other hand, it is very likely that the Kerr- Newman perturbations cannot be separated, and his documented failure will have saved many scientist-years of fruitless effort.

— BASILIS XANTHOPOULOS

1 Introduction

It is well known that the quantum effects of gravity come into play at the Planck energy scale (which is about $M_{pl} \sim 10^{19}\ GeV$). The space-time curvatures become too large for classical general relativity to hold good at this scale. There has been various attempts to understand these effects and to formulate a quantum theory of gravitation. Notable among them are string theory and the Ashtekar formalism. However, till date, there is no fully satisfactory renormalisable theory of quantum gravity. We address the issue of Planck scale physics in the context of particle scattering via gravitational interaction at very high energies. The kinematics of particle scattering can be expressed by the two independent Mandelstam variables s, and t, which are Lorentz scalars. They are respectively the squares of the centre-of-mass energy and the momentum transfer in the scattering process. Newton's constant G being a dimensional constant and equal to M_{pl}^{-2}, the Planck scale can arise in two ways, either when $Gs \sim 1$ or when $Gt \sim 1$. Thus, the most general quantum gravitational scenario involves both s and t approaching the Planck scale.

The eikonal approximation, on the other hand, is characterised by scattering at high s and fixed low t. Physically, this signifies scattering of particles at very high velocities (and kinetic energies) and at large impact parameters, such that the interaction is weak and the particles deviate slightly form their initial trajectories. In other words, they scatter almost in the forward direction. We would restrict our analyses to this approximation and try to extract whatever information is available about the Planck scale effects as reflected in the scattering amplitudes. The motivation to study this kinematical regime is, as we shall see, that the scattering amplitudes in this approximation can be exactly calculated and expressed in a closed form. These of course become significant only at Planckian centre-of-mass energies and we can obtain some quantitative results about quantum gravity in this kinematical domain.

Without loss of generality, two-particle scattering processes are considered. An inertial frame is chosen, in which one of the particles move at almost the speed of light and the other is relatively slow. Then one of these point particles is modelled as the source of an appropriate metric of general relativity. For example, neutral particles are modelled by Schwarzschild metric and electrically or magnetically charged particles are modelled by Reissner-Nordström metric. Finally, the quantum mechanical wave-function of the other particle in the background of this space-time is analysed to deduce the corresponding scattering amplitude. The effect of electromagnetism is also studied when the particles also carry electric and/or magnetic charges.

Next, we replace these black hole metrics by their counterparts in dilaton gravity, i.e. those that arise in low energy string theory. Here, we see that the above mentioned modelling cannot be done by generic black holes as was the case for general relativity. If we consider charged particles, for example, the modelling can be successfully done only by extremal black holes to be able to calculate the scattering amplitudes. This supports the conjecture that extremal black holes can indeed be *identified* with elementary particles.

2 Eikonal Scattering in General Relativity

We begin by considering the scattering of neutral point particles. The space time around any such particle, when it is static, is obtained by solving Einstein's equations and given by the well known Schwarzschild metric:

$$ds^2 = \left(1 - \frac{2GM}{r}\right) dt^2 - \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 - r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right) . \qquad (1)$$

To obtain the corresponding space-time when the particle is moving at a very high velocity, say along the positive z-axis, we perform a Lorentz transformation on the metric tensor components with the velocity parameter β according to:

$$t' = \gamma (t - \beta z) ,$$

$$z' = \gamma(z - \beta t) ,$$

where $\gamma = 1/\sqrt{1-\beta^2}$. Simultaneously, the mass M is parametrised as

$$M = \frac{P_0}{\gamma} .$$

This is to ensure that the energy of the boosted particle remains finite and equals P_0. The metric tensor components transform as a symmetric second rank tensor. Dropping the primes and taking the limit $\beta \to 1$, the final form of the metric is:

$$ds^2 = dx^- \left[dx^+ - \frac{2GP_0}{|x^-|} dx^- \right] - dx_\perp^2 , \tag{2}$$

where $x^\pm \equiv t \pm z$, the light cone coordinates. The above geometry was first obtained in [1] and then in [2]. Defining the new coordinate differentials $d\tilde{x}^\mu$ as

$$d\tilde{x}^+ = dx^+ - \frac{2GP_0}{|x^-|} dx^- , d\tilde{x}^- = dx^- , d\tilde{x}_\perp = dx_\perp ,$$

Eq.(2) can be re-written as,

$$ds^2 = d\tilde{x}^- d\tilde{x}^+ - d\tilde{x}_\perp^2 . \tag{3}$$

The above form of the infinitesimal line element seems to indicate that we have simply arrived at flat space-time by a coordinate re-definition. However, writing the finite forms of these re-definitions, we get:

$$\tilde{x}^+ = x^+ - 2GP_0 \theta\left(x^-\right) \ln x_\perp , \tag{4}$$
$$\tilde{x}^- = x^- , \tag{5}$$
$$\tilde{x}_\perp = x_\perp . \tag{6}$$

Note that the transformations are continuous everywhere except at $x^- = 0$, which is the trajectory of the boosted particle, where there is a step function discontinuity in the coordinate x^+. Calculation of the Riemann-Christoffel curvature tensor reveals that they are Dirac-delta functions (derivatives of the θ functions), which are non-vanishing only at $x^- = 0$. Thus, all space-time curvatures are localised on the two-dimensional transverse plane, perpendicular to the trajectory of the boosted particle and travelling along with it. We call this infinite plane the *shock-wave*. The space-time in front of and behind the shock-wave is Minkowskian. It is analogous to the case of a boosted charged particle, where the electromagnetic fields tend to become concentrated along the direction perpendicular to the particle trajectory, and in the limit $\beta \to 1$, they are completely localised on the plane fronted (electromagnetic) shock-wave. Note that the coordinate x^- is however, continuous at all points and serves as a bonafide affine parameter for any test particle in the above background. The classical geometry is depicted in Fig(1), which can be thought of as two Minkowski spaces, pertaining to $x^- < 0$ and $x^- > 0$ respectively and glued along the null plane $x^- = 0$, which is the trajectory of the boosted particle and the shock-wave. Having found the geometry of the luminal particle, now we concentrate on the other particle, assumed to be relatively slow. It serves as the test-particle in the above background. Before the shock wave comes and hits this particle, it is free from any interactions, and its wave-function is given by

$$\psi_< = e^{ip\cdot x} = e^{i\left[p_+ x^+ + p_- x^- + \vec{p}_\perp \cdot \vec{x}_\perp\right]} , \qquad x^- < 0 . \tag{7}$$

The moment when it is hit by the shock-wave, the x^+ coordinate undergoes a discrete shift given by Eq.(4) and the wave function picks up a space-time dependent phase factor. Simplifying, we get the final wave function to be:

$$\psi_> = e^{-iGs \ln x_\perp^2} \psi_< . \tag{8}$$

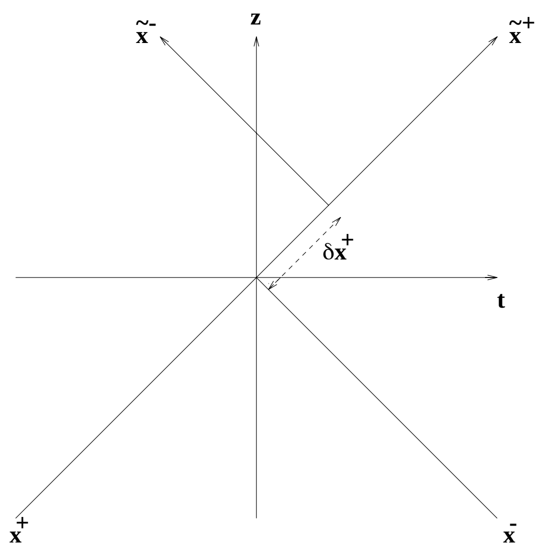


Fig.1: Shock-wave geometry of boosted Schwarzschild metric.

Here we have used the identity $2p_-p_0 = s$. To calculate the scattering amplitude from this wave function, we expand it in terms of a complete set of momentum eigenstates and perform an inverse Fourier transform to obtain the expansion coefficients. The latter can be identified with the scattering amplitude modulo kinematical factors. The detailed calculation is done in ref. [3], and the final result is:

$$f(s,t) = \frac{Gs}{t}\frac{\Gamma(1-iGs)}{\Gamma(1+iGs)}\left(\frac{-1}{t}\right)^{-iGs} . \tag{9}$$

Note that the above expression is simply the Rutherford Scattering amplitude with the gravitational coupling constant $-Gs$ replacing its electromagnetic counterpart α. As advertised, it captures the gravitational interactions between point particles at the Planck scale and is insignificant for sub-Planckian energies, when $Gs \ll 1$.

Without going into the details, which the reader will find in [5], we summarise in brief the situation when electromagnetic interactions are included in the scattering process. That is, the scattering particles also carry electric or magnetic charges. If they carry electric charges e and e', then the scattering amplitude is modified to:

$$f(s,t) = \frac{Gs-ee'}{t}\frac{\Gamma\left(1-i(Gs-ee')\right)}{\Gamma\left(1+i(Gs-ee')\right)}\left(\frac{-1}{t}\right)^{-i(Gs-ee')} . \tag{10}$$

The remarkable fact about this expression is that it can be obtained from the purely gravitational result in Eq.(9) simply by making the replacement $Gs \to Gs - ee'$. This means that the gravitational and electromagnetic coupling constants simply add up to give the effective coupling and there is no interference between them. This is quite unique and holds only in the eikonal approximation, because as we know, the two forces do affect each other in a non-trivial manner in generic cases. This decoupling is reminiscent of the Newtonian limit, where gravitation and electromagnetic interactions can be assumed not to affect each other. However, as far as the velocities and the energies of the particles are concerned, we are far removed from the Newtonian (low velocity) regime.

If, on the other hand, one of the particles carry an electric charge e, and the other a magnetic charge g, then the scattering amplitude is [6]:

$$f(s,t) \;=\; \left(\frac{n}{2}-iGs\right)\frac{\Gamma\left(\frac{n}{2}-iGs\right)}{\Gamma\left(\frac{n}{2}+iGs\right)}\left(\frac{-1}{t}\right)^{1-iGs} \tag{11}$$

Here, the two couplings do not add up in a simple manner as in the previous case, but the same in not expected intuitively, because with magnetic monopoles, the interaction is no longer central in nature like gravitation or electromagnetism involving charges only. Comparing Eqs. (10) and (11), we find that the electromagnetic contribution to the former is insignificant (with ee' typically of the order of the fine structure constant 1/137), while for the latter, it is comparable to the gravitational contribution (both couplings being of the order of unity). In short, gravitation dominates overwhelmingly over electromagnetism at Planckian energies in the absence of magnetic monopoles. Introduction of the latter entails drastic changes in the result.

3 Shock Waves in Dilaton Gravity

As noted earlier, the results in the previous section were derived in the framework of general relativity, and all the black hole metrics used to model the particles satisfy the Einstein's equations. Now, string theory, in the low energy limit provides us an alternative theory of gravitation, known as *dilaton gravity*, where in addition to the metric tensor, a scalar field called the **dilaton** is also an independent degree of freedom. We will not go into the details as to how this theory emerges as a low energy *effective* theory from string theory. Instead, we will analyse the scattering situation envisaged in section 2 using dilaton gravity. We would like to investigate whether the scattering amplitudes are modified in this framework, and whether the electromagnetic and gravitational decoupling still hold good. We will see that modelling the scattering particles by dilatonic black holes poses some generic pathologies, which are removable only under certain specific conditions and when these are satisfied the decoupling exists just as in the case of general relativity. The counterpart of the Reissner-Nordström metric in dilaton gravity is given by the following expression (in the so-called 'string metric') [7]:

$$ds^2 \;=\; \left(1-\frac{\alpha}{Mr}\right)^{-1}\left[\left(1-\frac{2GM}{r}\right)dt^2-\left(1-\frac{2GM}{r}\right)^{-1}dr^2\right]-r^2 d\Omega^2 \;, \tag{12}$$

where $\alpha = Q^2 e^{2\phi_0}$, Q being the electric charge and ϕ_0 the asymptotic value of the dilaton field. Note that setting $Q = 0$ reproduces the Schwarzschild metric, which means that the dilaton field has non-trivial effects only when we consider charged solutions. Like the Reissner-Nordström solution, the above metric has two horizons r_+ and r_-, given by:

$$r_+ \;=\; 2GM \;,$$
$$r_- \;=\; \frac{\alpha}{M} \;.$$

However, here a crucial difference is that the inner horizon r_- is a space-time singularity, where the curvature tensor blows up. For black holes of large masses, this singularity is hidden behind the event horizon r_+ and there is no naked singularity, which however develops as the mass decreases. We will see that this singularity plays a crucial role in eikonal scattering. Analogous to the Schwarzschild or the Reissner-Nordström metric, we apply a Lorentz boost to the dilaton gravity metric along the positive z-axis and take the limit $\beta \to 1$. The resultant metric is of the form [7]:

$$ds^2 \;=\; dx^-\left[\frac{1-\alpha/2P_0|x^-|}{1-\alpha/P_0|x^-|}\right]\left[dx^+ - \frac{4GP_0/|x^-|}{1-\alpha/P_0|x^-|}\;dx^-\right]-dx_\perp^2 \;. \tag{13}$$

As before, to express this in a form representing Minkowski space, we define the new coordinates

$$d\tilde{x}^- \;=\; dx^-\left[\frac{1-\alpha/2P_0|x^-|}{1-\alpha/P_0|x^-|}\right] \tag{14}$$

$$d\tilde{x}^+ \;=\; dx^+ - \frac{4GP_0/|x^-|}{1-\alpha/P_0|x^-|}\;dx^- \;\;. \tag{15}$$

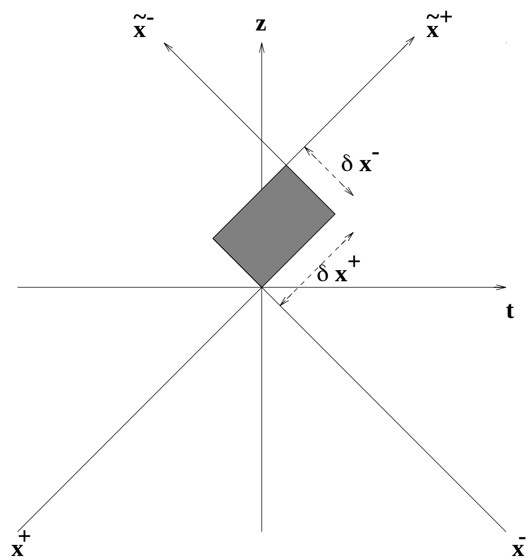


Fig.2: Shock-wave geometry of boosted non-extremal dilaton gravity metric.

Note that here both the coordinates x^- and x^+ are discontinuous for non-vanishing α. This is rather disturbing, because as we saw in the previous section, x^- served as the continuous affine parameter for the test particle, and a discontinuity in it signals a breakdown in the description of the evolution of the particle in terms of geodesics. This is indeed confirmed by writing the classical geodesic equations of the test particle in the background of the boosted dilaton metric and trying to solve it perturbatively in powers of the mass M. The failure of the latter indicates that the null geodesics are incomplete in this case and the curvature singularity at $r^- = \alpha/M$ shows up as an extended naked singularity in the boosted limit and renders eikonal scattering impossible. The classical geometry in this case is shown in Fig.(2), where the rectangular shaded region denotes the finite region of singularity that originated from the singular horizon r_-. Thus, we see that modelling the particles by dilaton gravity metric gives rise to singularities which makes the subsequent calculation of scattering amplitudes impossible.

To circumvent this difficulty, we can try various possibilities. In particular, we can examine the case by imposing the *extremal limit* $r_+ = r_-$. Then, the metric assumes the form

$$ds^2 \;=\; dt^2 - \frac{dr^2}{\left(1 - \frac{2GM}{r}\right)^2} - r^2 d\Omega^2 \;, \tag{16}$$

which is perfectly regular everywhere. The curvature singularity has simply disappeared when the two horizons coincide ! This is precisely the motivation behind considering this limit. Note that the metric above is entirely distinct from the Schwarzschild metric. However, on performing the usual Lorentz boost on it along with the limit $\beta \to 1$, it can be easily verified that the boosted version coincides with the metric (2). This is a pleasant surprise, because now the test particle will 'see' an identical background as in the case of Schwarzschild, and the corresponding scattering amplitude is simply the eikonal result (9).

Thus it is clear that the curvature singularity shows up as an extended singular region and makes it impossible to calculate scattering amplitude. The extremal limit, on the other hand, plays a special role whose imposition reproduces the elegant eikonal results. In the next section, we will probe into the details of the role of this singularity by invoking a different formalism.

4 Scattering in Static Dilaton Gravity Background

In this section, we try to establish rigorously, the role of the naked singularity in dilaton gravity metric in eikonal scattering. Here we consider the scattering of the fast particle in the background of the other static particle. Note that the definitions of 'source' and 'test' has been reversed. The Klein-Gordon equation for the wave function ϕ of the fast particle is given by:

$$g_{\mu\nu} \ D^\mu D^\nu \Phi \ = \ 0 \ , \tag{17}$$

where D^μ denotes the covariant derivative. The above equation can be simplified to

$$\frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g} g^{\mu\nu} \partial_\nu \Phi \right) \ = \ 0 \ . \tag{18}$$

We assume a solution of the form

$$\Phi \ = \ f(r) \ Y_{lm}(\theta, \phi) \ e^{iEt} \ , \tag{19}$$

where we have split the wave function into a radial part, the spherical harmonics and the exponential of asymptotically measured energy E. This decomposition results from the spherically symmetric and static nature of the dilaton gravity metric (12). With this ansatz, the following radial equation is obtained from Eq.(18) [9]:

$$r^2 \Lambda \frac{d^2 f}{dr^2} + \frac{d\left(r^2 \Lambda\right)}{dr} \ \frac{df}{dr} - \left[\frac{l(l+1)}{\Delta} - \frac{E^2 r^2}{\Lambda} \right] \ f \ = \ 0 \ , \tag{20}$$

where we have defined the quantities $\Lambda \equiv 1 - 2GM/r$ and $\Delta \equiv 1 - \alpha/Mr$. As expected, we recover the radial equation for Schwarzschild background for $\alpha = 0, \Delta = 1$, and the subsequent scattering amplitude (9) [10]. However, for generic values of Δ, it can be shown that the above radial equation does not admit of a solution at all. This follows from an elementary theorem in ordinary differential equations since the coefficient of f in eq.(20) suffers an infinite discontinuity at $\Delta = 0$. Thus, as in the previous section, we conclude that the dilaton gravity metric (12) cannot be used to successfully model the high energy particles. Moreover, we are now in a position to understand the physical origin of the pathology. The curvature singularity at $r_- = \alpha/M$ grows indefinitely large as we take the limit $M \to 0$ and eventually fills all space around the dilatonic particle. Thus, the other particle, at arbitrary impact parameter, is forced to hit this singularity and get trapped, signalling the breakdown of the scattering process. This was reflected in the non-existence of solutions of the classical geodesic equations in section 3. Imposing the extremal limit, on the other hand, eliminated this singularity, and thus the scattering amplitude became calculable, which yielded the eikonal result. This can be verified using the Klein-Gordon radial equation also. In the extremal limit, Eq.(20) assumes the form [10]:

$$\frac{d^2 f}{dr^2} + \frac{1}{r^2 \Lambda} \frac{d\left(r^2 \lambda\right)}{dr} \frac{df}{dr} - \frac{1}{\Lambda^2} \left[\frac{l(l+1)}{r^2} - E^2 \right] \ f \ = \ 0 \ . \tag{21}$$

The above radial equation can be expanded in powers of GM/r and this recovers the Schwarzschild radial equation to the lowest order. The scattering amplitude (9) follows immediately. This re-emphasises the importance of the extremal limit for Planckian scattering via dilaton gravity.

5 Perturbative approach

In the previous two sections, we have explicitly used a classical solution of the low energy string effective action to arrive at the scattering amplitudes (in the extremal limit). However, the problem can be approached without the help of such explicit solutions, at the level of the action itself. Historically, the eikonal scattering amplitude was derived in the context of quantum electrodynamics by summing an infinite subset of Feynman diagrams, known as *ladder diagrams*, with certain kinematical restrictions on the matter propagators. It was shown that this infinite sum can be expressed in a neat closed form [12]. A crucial assumption required to arrive at the eikonal result is the assumption that the scattering particles have well defined *classical* trajectories, which differ slightly from the free particle trajectories. The sum of ladder diagrams were seen to converge for gravitational scattering as well in [13] which reproduced the amplitude (9) exactly. For dilaton gravity, however, the assumption regarding classical trajectories is invalidated because, as we saw in the previous sections, an incoming particle is swallowed up by the expanding curvature singularity and there are no well defined scattering solutions. Thus, a priori, it seems impossible to construct and calculate ladder diagrams from dilaton gravity action given by:

$$S \;=\; \int d^4x\sqrt{-g}e^{-2\phi}\left[-\frac{R}{G}+F_{\mu\nu}F^{\mu\nu}+2\partial_\mu\phi\partial^\mu\phi\right] \;. \tag{22}$$

First, we simplify the action by linearising the metric as well as the dilaton field,

$$g_{\mu\nu} \;=\; \eta_{\mu\nu}+h_{\mu\nu}\;, \tag{23}$$

$$\phi \;=\; \phi_0+f\;, \tag{24}$$

where $\eta_{\mu\nu}$ is the flat Minkowskian metric and ϕ_0 is some constant. Retaining terms to leading order in these quantum fluctuations, the action (22) reduces to:

$$\begin{aligned} S \;&=\; \frac{e^{-2\phi_0}}{G}\int d^4x(1-2f)\frac{1}{8}h_{\mu\nu}\left[\eta^{\mu\lambda}\eta^{\nu\sigma}+\eta^{\mu\sigma}\eta^{\nu\lambda}-\eta^{\mu\nu}\eta^{\lambda\sigma}\right]\Box h_{\lambda\sigma} \\ &-\; e^{-2\phi_0}\int d^4x\left(1+\frac{1}{2}h_\alpha{}^\alpha\right)(1-2f)\left[-4\partial_\mu f\partial^\mu f+F^2+\partial_\mu\chi\partial^\mu\chi\right] \;. \end{aligned} \tag{25}$$

Here we have also included the action for the massless matter field χ representing the scattering particles. In addition to the graviton and matter propagators and the matter-graviton interaction vertex already calculated in [13], now we have a dilaton propagator and a matter-dilaton interaction vertex. The factors associated with them can be read of fom the linearised action, and turns out to be $-i/(p^2+m^2-i\epsilon)$ and $-2p\cdot p'$ respectively, where p and p' are the momenta associated with the external matter lines. With these, the new infinite set of ladder diagrams with dilaton exchanges can be computed in a straightforward manner. The details of the calculation is done in [9], and the final result is:

$$i\mathcal{M} \;=\; \frac{ip_1^2p_2^2}{-t}\frac{\Gamma(1-ip_1^2p_2^2/Ep)}{\Gamma(1+ip_1^2p_2^2/Ep)}\left(\frac{4}{-t}\right)^{-i\frac{p_1^2p_2^2}{Ep}} \;. \tag{26}$$

Now, if we make the momenta p_1 and p_2 on-shell and replace them by m^2 and eventually take the massless limit $m\to 0$, then we see that the dilaton amplitude vanishes identically and we are simply left with the gravitational result (9)!

At this point, we investigate the circumstances under which the action can be linearised, because without the latter, the eikonal sum can never be attempted. The metric $g_{\mu\nu}$ is linearised under the assumption that there are small graviton fluctuations over a Minkowskian background. As for the dilaton field, we can try to estimate its quantum fluctuation f by looking at the classical solution for ϕ obtained by minimising the action (22):

$$e^{2\phi} \;=\; e^{2\phi_0}\left(1-\frac{\alpha}{Mr}\right)\;, \tag{27}$$

which implies that the fluctuations over ϕ_0 are of the order of

$$f\sim|\phi-\phi_0|\sim|\ln(-\alpha/Mr)|\;. \tag{28}$$

Smallness of this requires $|\alpha/Mr|\to 0$, or in other words *alpha* should scale at least as M^2 as we take $M\to 0$. But this is equivalent to the extremal limit, when $\alpha=2GM^2$! Thus, even in a solution-independent approach, where we look at the action and impose certain restrictions on it, the extremal limit seems to emerge in a natural way, if we want to obtain well defined scattering amplitudes.

6 Conclusion

Our analysis of Planckian scattering in the light of dilaton gravity unambiguously point to the fact that there are important constraints to be satisfied while trying to model the scattering particles by a suitable metric. Namely, the extremality condition should be necessarily imposed on the parameters to obtain physically meaningful scattering amplitudes. Otherwise, the curvature singularity at a finite radius inevitably shows up as pathologies in the calculations. In the shock-wave approach, where we boosted the dilaton gravity metric, this was seen to give rise to the a discontinuous affine parameter. Next, while trying to solve the Klein-Gordon equation in the background of the above metric, we saw that there were no scattering solutions to the corresponding radial equation. Finally, in the approach of perturbation theory, we showed that the eikonal scattering amplitude can be reproduced if and only if the dilaton field was linearised and the lowest order terms in its quantum fluctuation is retained in the action. Once again, this linearisation is consistent with the extremality condition. There is yet another solution-independent way to arrive at identical conclusions, starting from the dilaton gravity action, where one uses the so called 'Verlinde-scaling' to incorporate the eikonal kinematics. The interested reader may refer to [9] for a detailed discussion of this approach. The important point to note is that no such restrictions were ever necessary in the general relativistic framework to calculate scattering amplitudes. Thus, it is perhaps correct to say that the theory of gravity that emerges from string theory incorporates certain problematic features, at least in the context of Planckian scattering. But the same theory contains the cure to this problem also, namely in the form of the extremal limit! The latter constraint, once imposed, removes the pathologies altogether and reproduces the finite amplitudes of general relativity. It is also curious to note the consistency of these results with the well known conjecture that extremal black holes are actually elementary particles [15]. Here, we are considering scattering of point particles, which can also be regarded as 'elementary'. Thus, it seems logical in the spirit of the conjecture, to model them as extremal black holes.

References

[1] P. Aichelburg and R. Sexl, *Gen. Rel. Grav.*, **2**, 303 (1971)

[2] T. Dray and G. 't Hooft, *Nucl. Phys.*, **B 253**, 173 (1985)

[3] G. 't Hooft, *Phys. Lett.*, **B198**, (1987) 61; *Nucl. Phys.*, **B304**, 867 (1988)

[4] R. Jackiw, D. Kabat and M. Ortiz, *Phys. Lett.*, **B277**, 148 (1992)

[5] S. Das and P. Majumdar, *Phys. Rev.*, **D51**, 5664 (1995)

[6] S. Das and P. Majumdar, *Phys. Rev. Lett.*, **72**, 2524 (1994)

[7] H. Garfinkle, G. Horowitz and A. Strominger, *Phys. Rev.*, **D43**, 3140 (1991); *err. Phys. Rev.*, **D45**, 3888 (1992)

[8] S. Das and P. Majumdar, *Phys. Lett.*, **B348**, 349 (1995)

[9] S. Das and P. Majumdar, IMSc preprint 95/25, hep-th/9512209 (1995)

[10] S. Das and P. Majumdar, IMSc preprint 95/8, hep-th/9504060 (1995)

[11] E. L. Ince, Ordinary Differential Equations (Dover, New York, 1956), p.73

[12] H. D. I. Abarbanel and C. Itzykson, *Phys. Rev. Lett.*, **23**, 53 (1969); M. Lévy and J. Sucher, *Phys. Rev.*, **186**, 1656 (1969)

[13] D. Kabat and M. Ortiz, Nucl. Phys. **B388**, 570 (1992)

[14] H. Verlinde and E. Verlinde, Nucl. Phys. **B371**, 246 (1992)

[15] A. Sen, *Mod. Phys. Lett.*, **A10**, 2081 (1995) and references therein

BLACK HOLE ENTROPY

P. Mitra[1]

Saha Institute of Nuclear Physics
Block AF, Bidhannagar
Calcutta 700 064, INDIA

Abstract

Major developments in the history of the subject are critically reviewed in this talk.

[1]e-mail mitra@tnp.saha.ernet.in

The parable entitled "Not lost but gone before", is about larvae of dragonflies deposited at the bottom of the pond. A constant source of mystery for these larvae was what happens to them, when on reaching the stage of chrysalis they pass through the surface of the pond never to return. And each larva, as it approaches the chrysalis stage and feels compelled to rise to the surface of the pond, promises to return and tell those that remain behind what really happens, and confirm or deny a rumour attributed to a frog that when a larva emerges on the other side of their world it becomes a marvellous creature with long slender body and iridescent wings. But on emerging from the surface of the pond as a fully-formed dragonfly, it is unable to penetrate the surface no matter how it tries and how long it hovers. And the history books of larvae do not record any instance of one of them returning to tell them what happens to it when it crosses the dome of their world. And the parable ends with the cry

...Will none of you in pity,
To those you left behind, disclose the secret?

— S. CHANDRASEKHAR

1 Introduction

A black hole is classically thought of as a region of intense gravitational field from which no form of energy – not even light – can leak out. The best known example is the Schwarzschild black hole solution of Einstein's equation. It is described by the metric

$$ds^2 = -(1-\frac{2M}{r})dt^2 + (1-\frac{2M}{r})^{-1}dr^2 + r^2 d\Omega^2 \tag{1}$$

It has a *horizon* at $r = 2M$, which is a singularity of this coordinate system, but the curvature is not singular there, and regular *Kruskal* coordinates may be chosen. There is, however, a curvature singularity at $r = 0$, which is to be regarded as the location of a point source of mass M.

Another example is the Reissner - Nordström solution of the Einstein - Maxwell equations. The metric is given by

$$ds^2 = -(1-\frac{2M}{r}+\frac{Q^2}{r^2})dt^2 + (1-\frac{2M}{r}+\frac{Q^2}{r^2})^{-1}dr^2 + r^2 d\Omega^2 \tag{2}$$

and the electric field by

$$F_{tr} = \frac{Q}{r^2}, \tag{3}$$

with M and Q denoting the mass and the charge respectively. There are apparent singularities at

$$r_\pm = M \pm \sqrt{M^2 - Q^2} \tag{4}$$

provided $M \geq Q$. This inequality must hold if a naked singularity is to be avoided and then there is a horizon at r_+. The limiting case when $Q = M$ and $r_+ = r_-$ is referred to as the extremal case. There is again a curvature singularity at $r = 0$, which is to be regarded as the location of a point source with mass M and charge Q.

Classically, a black hole is stable and does not radiate. So a black hole may, at that level, be assigned zero temperature and correspondingly zero entropy. But the situation changes when quantum field theory is brought in to describe the interaction of matter with a black hole background. A careful definition of the vacuum using regular (Kruskal) coordinates shows that there is a net flow of particles away from the black hole, with a *thermal spectrum* corresponding to a temperature determined by the black hole parameters like mass and charge.

The strong gravitational field may be imagined to produce a particle - antiparticle pair by polarizing the vacuum. If this occurs outside the horizon, one of the pair may fall in, with the other moving away to infinity, thus contributing to an outward flow.

If the process continues, it actually becomes faster. This is because a black hole gets *hotter* when it loses mass. The black hole is thus expected to evaporate completely! A regular spacetime should be left behind, with all the matter moving away.

This process of evaporation leads to a puzzle. A black hole can be imagined to start off in a pure state. When it evaporates, there is only thermal radiation, which is in a mixed state. Such a transition would appear to be non-unitary and to involve loss of information. Can it really occur?

Three different answers have been proposed.

1. Yes. Quantum gravity is non-unitary, and indeed the laws of quantum mechanics as known to us should be modified to accommodate this kind of process.

2. No. The thermal nature of the radiation is only an approximation. There are correlations between different times and the state is really pure.

3. No. The process of evaporation stops at some point to be determined by an yet unknown theory of quantum gravity. This means that the black hole leaves a *remnant.*

It has to be emphasized that each of these proposals involves some new physics, *i.e.,* there is no way of understanding black hole evaporation in terms of known principles.

It is to be hoped that some day a clear answer will emerge. We shall not discuss this question any further, but shall pass on to a closely related topic: black hole entropy. Entropy is, of course, a measure of *lack of information.*

2 Black hole entropy in the seventies

A precursor of the idea of entropy in the context of black holes was the so-called area theorem [1]. According to this theorem, the area of the horizon of a system of black holes always increases in a class of spacetimes. The asymmetry in time is built into the definition of this class: these spacetimes are predictable from partial Cauchy hypersurfaces. This result is certainly reminiscent of thermodynamical entropy.

Some other observations made around that time were collected together into a set of *laws of black hole mechanics* analogous to the laws of thermodynamics [2].

- The zeroeth law states that the surface gravity κ remains constant on the horizon of a black hole.
- The first law states that
$$\frac{\kappa dA}{8\pi} = dM - \phi dQ, \tag{5}$$
where A represents the area of the horizon and ϕ the potential at the horizon. For the Reissner - Nordström black hole,
$$\kappa = \frac{r_+ - r_-}{2r_+^2},\ \phi = Q/r_+,\ A = 4\pi r_+^2. \tag{6}$$
- The second law is just the area theorem already stated.

.

When these observations were made, there was no obvious connection with thermodynamics, it was only a matter of analogy. But it was soon realized [3] that the existence of a horizon imposes a limitation on the amount of information available and hence may lead to an entropy, which should then be measured by the geometric quantity associated with the horizon, namely its area. Thus, upto a factor, A should represent the entropy and $\frac{\kappa}{8\pi}$ the temperature.

Not everyone accepted this interpretation of the laws of black hole mechanics, and in any case the undetermined factor left a question mark. Fortunately, the problem was solved very soon. It was discovered that quantum theory causes dramatic changes in the behaviour of black hole spacetimes. A scalar field theory in the background of a Schwarzschild black hole indicates the occurrence of radiation of particles [4] at a temperature

$$T = \frac{\hbar}{8\pi M} = \frac{\hbar\kappa}{2\pi}. \tag{7}$$

This demonstrated the connection of the laws of black hole mechanics with thermodynamics and fixed the scale factor. It involves Planck's constant and is a quantum effect.

For a Schwarzschild black hole, the first law of *thermodynamics* can be written as

$$TdS = dM \tag{8}$$

and can be integrated, because of (7), to yield

$$S = \frac{4\pi M^2}{\hbar} = \frac{A}{4\hbar}. \tag{9}$$

Although the expression for T given above is specific to the case of Schwarzschild black holes, the relation between the temperature and the surface gravity given in (7) is more generally valid in the case of black holes having $g_{tt} \sim (1 - \frac{r_h}{r})$. The first law of black hole mechanics then becomes

$$Td\frac{A}{4\hbar} = dM - \phi dQ. \tag{10}$$

Comparison with the first law of thermodynamics

$$TdS = dM - \tilde{\phi} dQ \tag{11}$$

is not straightforward because the chemical potential $\tilde{\phi}$ is not clearly known. However, one way of satisfying these two equations involves the identification

$$S = \frac{A}{4\hbar}, \qquad \tilde{\phi} = \phi. \tag{12}$$

In another approach, the grand partition function is used. For charged black holes [5] it can be related to the classical action by

$$Z_{\text{grand}} = e^{-\frac{M - TS - \tilde{\phi} Q}{T}} \approx e^{-I/\hbar}, \tag{13}$$

where the functional integral over all configurations is semiclassically approximated by the weight factor with the classical action I. The action is given by a quarter of the area of the horizon when the Euclidean time goes over one period, *i.e.,* from zero to $1/T$. Consequently,

$$M = T(S + \frac{A}{4\hbar}) + \tilde{\phi} Q. \tag{14}$$

Now there is a standard formula named after Smarr [6],

$$M = \frac{\kappa A}{4\pi} + \phi Q, \tag{15}$$

which can be rewritten as

$$M = T\frac{A}{2\hbar} + \phi Q. \tag{16}$$

Comparison with (14) suggests once again the relations (12). Although the result is the same, it should be noted that there is a new input: the functional integral. There is a hope that corrections to the above formulas may be obtained by improving the approximation used in the calculation of the functional integral.

Before closing this section, let us briefly point out that the entropy can be different from $A/(4\hbar)$ [7]. The difference between the first laws of thermodynamics and black hole mechanics can be written as

$$TdF = (\phi - \tilde{\phi})dQ, \tag{17}$$

where $F \equiv S - A/(4\hbar)$. Hence,

$$\frac{\partial F}{\partial M} = 0, \quad \frac{\partial F}{\partial Q} = \frac{\phi - \tilde{\phi}}{T}. \tag{18}$$

The first equation can be satisfied by an arbitrary function of Q, and the second equation serves to fix $\tilde{\phi}$ rather than to put any constraint on F.

What happens in the functional integral approach? The difference between (14) and (16) is

$$TF = (\phi - \tilde{\phi})Q. \tag{19}$$

By itself this does not impose any restriction but it can be used to restrict F in conjunction with (18). We observe that

$$F = Q\frac{\partial F}{\partial Q}, \tag{20}$$

which means that F is a homogeneous function of Q of degree 1. This is true even if there are several charges (Q has several components).

3 Matter around black hole

The investigation of field theory in the background of a black hole [4] had given a physical meaning to the temperature of a black hole but the entropy remained mysterious. An attempt was then made [8] to study the entropy of matter in such a background. It is convenient to employ what is called the brick wall boundary condition. Then the wave function is cut off just outside the horizon at r_h. Mathematically,

$$\varphi(x) = 0 \qquad \text{at } r = r_h + \epsilon \tag{21}$$

where ϵ is a small, positive, quantity and signifies an ultraviolet cut-off. There is also an infrared cut-off

$$\varphi(x) = 0 \qquad \text{at } r = L \tag{22}$$

with the box size $L >> r_h$.

The wave equation for a scalar is

$$\frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}g^{\mu\nu}\partial_\nu\varphi) - m^2\varphi = 0. \tag{23}$$

A solution of the form

$$\varphi = e^{-iEt} f_{El} Y_{lm_l} \tag{24}$$

satisfies the radial equation

$$-g^{tt}E^2 f_{El} \quad + \quad \frac{1}{\sqrt{-g}}\frac{\partial}{\partial r}[\sqrt{-g}g^{rr}\frac{\partial f_{El}}{\partial r}] - [l(l+1)g^{\theta\theta} + m^2]f_{El} = 0. \tag{25}$$

An r- dependent radial wave number can be introduced from this equation by

$$\begin{aligned} k^2(r,l,E)f_{El} &= -g_{rr}\frac{1}{\sqrt{-g}}\frac{\partial}{\partial r}[\sqrt{-g}g^{rr}\frac{\partial f_{El}}{\partial r}] \\ &= g_{rr}[-g^{tt}E^2 - l(l+1)g^{\theta\theta} - m^2]f_{El}. \end{aligned} \tag{26}$$

Only such values of E are to be considered here as make the above expression nonnegative. The values are further restricted by the semiclassical quantization condition

$$n_r\pi = \int_{r_h+\epsilon}^{L} dr \; k(r,l,E), \tag{27}$$

where n_r has to be a nonnegative integer.

To find the free energy F at inverse temperature β one has to sum over states with all possible single- particle combinations:

$$\begin{aligned} \beta F &= \sum_{n_r,l,m_l} \log(1-e^{-\beta E}) \\ &\approx \int dl \; (2l+1) \int dn_r \log(1-e^{-\beta E}) \\ &= -\int dl \; (2l+1) \int d(\beta E) \; (e^{\beta E}-1)^{-1} n_r \\ &= -\frac{\beta}{\pi}\int dl \; (2l+1)\int dE \; (e^{\beta E}-1)^{-1}\int_{r_h+\epsilon}^{L} dr \; \sqrt{g_{rr}} \\ &\quad \sqrt{-g^{tt}E^2 - l(l+1)g^{\theta\theta} - m^2} \\ &= -\frac{2\beta}{3\pi}\int_{r_h+\epsilon}^{L} dr \; \sqrt{g_{rr}}g^{\theta\theta} \\ &\quad \int dE \; (e^{\beta E}-1)^{-1}[-g^{tt}E^2 - m^2]^{3/2}. \end{aligned} \tag{28}$$

Here the limits of integration for l, E are such that the arguments of the square roots are nonnegative. The l integration is straightforward and has been explicitly carried out. The E integral can be evaluated only approximately.

The contribution to the r integral from large values of r yields the expression for the free energy valid in flat spacetime because of the asymptotic flatness:

$$F_0 = -\frac{2}{9\pi} L^3 \int_m^\infty dE \frac{(E^2 - m^2)^{3/2}}{e^{\beta E} - 1}. \tag{29}$$

This part has to be subtracted out [8]. The contribution of the black hole is singular in the limit $\epsilon \to 0$. For ordinary black holes the leading singularity is linear:

$$F \approx -\frac{2\pi^3}{45\epsilon} (\frac{r_h}{\beta})^4, \tag{30}$$

where the lower limit of the E integral has been approximately set equal to zero. (There are corrections involving $m^2\beta^2$ which will be ignored here.) β is now replaced by the reciprocal of the black hole temperature and the cutoff ϵ in r replaced by one in the "proper" radial variable defined by

$$d\tilde{r} = \sqrt{g_{rr}} dr, \tag{31}$$

whence $\tilde{\epsilon} \propto \sqrt{r_h \epsilon}$. The contribution to the entropy due to the presence of the black hole is obtained from the formula

$$S = \beta^2 \frac{\partial F}{\partial \beta} \tag{32}$$

to be

$$S = \frac{A}{360\pi\tilde{\epsilon}^2}. \tag{33}$$

The appearance of the area led to a lot of interest. The divergence in the limit of vanishing cutoff ϵ is clearly due to the concentration of the matter states near the horizon. The question then was how the finite result $A/(4[G]\hbar)$ is to be obtained from this expression. It was suggested that different species of matter, which have to be summed over, might renormalize the Newton constant in the denominator of the entropy and might even produce it! However, an alternative point of view is that the calculation indicated above refers to the entropy of the matter rather than that of the black hole. Whereas the calculation of the temperature of matter can tell us about the temperature of the black hole, the entropy of one need not have any connection with that of the other. In support of this view one can cite the case of *extremal* dilatonic black holes, which carry mass and magnetic charge, but the charge has the maximum value that can exist for a given mass without giving rise to a naked singularity. For these black holes, the entropy as calculated by the above procedure is nonzero, though the area vanishes [9].

4 Counting of states: string based black holes

While the previous section reviewed an attempt to calculate the entropy by counting states, it was an entropy in the background of a black hole rather than the entropy of a black hole itself, which remained unexplained. Recently, it has been possible to make some progress in this direction, though in the context of special black holes arising from string theory [10]. In four dimensions the massless bosonic fields of the heterotic string obtained by toroidal compactification lead to an effective action with an unbroken $U(1)^{28}$ gauge symmetry:

$$\begin{aligned} S &= \frac{1}{32\pi} \int d^4x \sqrt{-G} e^{-\Phi} [R_G + G^{\mu\nu} \nabla_\mu \Phi \nabla_\nu \Phi + \frac{1}{8} G^{\mu\nu} Tr(\partial_\mu \mathcal{M} \mathcal{L} \partial_\nu \mathcal{M} \mathcal{L}) \\ &\quad - G^{\mu\mu'} G^{\nu\nu'} F^{(a)}_{\mu\nu} (\mathcal{L}\mathcal{M}\mathcal{L})_{ab} F^{(b)}_{\mu'\nu'} - \frac{1}{12} G^{\mu\mu'} G^{\nu\nu'} G^{\rho\rho'} H_{\mu\nu\rho} H_{\mu'\nu'\rho'}]. \end{aligned} \tag{34}$$

Here, $\mathcal{L} = \begin{pmatrix} -I_{22} & \\ & I_6 \end{pmatrix}$, with I representing an identity matrix, $\mathcal{M}$ a symmetric 28 dimensional matrix of scalar fields satisfying $\mathcal{MLM} = \mathcal{L}$, and there are 28 gauge field tensors $F^{(a)}_{\mu\nu} = \partial_\mu A^{(a)}_\nu - \partial_\mu A^{(a)}_\mu$, $\quad a = 1, ...28$ as well as a third rank tensor $H_{\mu\nu\rho} = \partial_\mu B_{\nu\rho} + 2A^{(a)}_\mu \mathcal{L}_{ab} F^{(b)}_{\nu\rho} +$ cyclic permutations of μ, ν, ρ corresponding to an antisymmetric tensor field B. The canonical metric defined by $g_{\mu\nu} = e^{-\Phi} G_{\mu\nu}$ possesses black hole solutions.

The dilaton field is nontrivial, though H still vanishes in the solutions to be considered. The metric $g_{\mu\nu}$ and the dilaton Φ are given by

$$\begin{aligned} ds^2 &\equiv g_{\mu\nu} dx^\mu dx^\nu \\ &= -\frac{r^2 - 2mr}{\Delta^{1/2}} dt^2 + \frac{\Delta^{1/2}}{r^2 - 2mr} dr^2 + \Delta^{1/2} d\Omega^2_{II} \end{aligned} \tag{35}$$

with

$$\begin{aligned} \Delta &= r^2 \big[r^2 + 2mr(\cosh\alpha \cosh\gamma - 1) + m^2 (\cosh\alpha - \cosh\gamma)^2 \big], \\ \text{and} \qquad e^\Phi &= \frac{g^2 r^2}{\Delta^{1/2}}. \end{aligned} \tag{36}$$

Here α, γ are real parameters and g is a constant. The time components of the gauge fields are given by

$$\vec{A}_t = \begin{cases} -\frac{g\vec{n}_L}{\sqrt{2}} \frac{mr \sinh\alpha}{\Delta} [r^2 \cosh\gamma + mr(\cosh\alpha - \cosh\gamma)] & L = 1, ...22 \\ -\frac{g\vec{n}_R}{\sqrt{2}} \frac{mr \sinh\gamma}{\Delta} [r^2 \cosh\alpha + mr(\cosh\alpha - \cosh\gamma)] & R = 23, ...28 \end{cases} \tag{37}$$

with $\vec{n}_L, \vec{n}_R$ denoting respectively 22-component and 6-component unit vectors and

$$\mathcal{M} = I_{28} + \begin{pmatrix} P n_L n_L^T & Q n_L n_R^T \\ Q n_R n_L^T & P n_R n_R^T \end{pmatrix}, \tag{38}$$

where

$$\begin{aligned} P &= \frac{2m^2 r^2}{\Delta} \sinh^2\alpha \sinh^2\gamma \\ Q &= -\frac{2mr}{\Delta} \sinh\alpha \sinh\gamma [r^2 + mr(\cosh\alpha \cosh\gamma - 1)]. \end{aligned} \tag{39}$$

All other backgrounds vanish for this solution.

The ADM mass of the black hole and its charges are given by

$$\begin{aligned} M &= \frac{1}{4} m (1 + \cosh\alpha \cosh\gamma) \\ \vec{Q} &= \begin{cases} \frac{g\vec{n}_L}{\sqrt{2}} m \sinh\alpha \cosh\gamma & L = 1, ...22 \\ \frac{g\vec{n}_R}{\sqrt{2}} m \sinh\gamma \cosh\alpha & R = 23, ...28 \end{cases} \end{aligned} \tag{40}$$

The area of the horizon, which is at $r = 2m$, is

$$A_H = 8\pi m^2 (\cosh\alpha + \cosh\gamma), \tag{41}$$

and the inverse temperature (as defined in terms of the surface gravity) is given by

$$\beta_H = 4\pi m (\cosh\alpha + \cosh\gamma). \tag{42}$$

One has to consider the special case

$$m \to 0, \quad \gamma \to \infty, \quad \text{with} \quad m \cosh\gamma = m_0, \quad \alpha = \text{finite}. \tag{43}$$

Then

$$A_H = 0, \quad T_H = \frac{1}{4\pi m_0}, \tag{44}$$

and

$$M = \frac{m_0}{4}\cosh\alpha, \quad \vec{Q}_L = \frac{g m_0}{\sqrt{2}}\sinh\alpha\,\vec{n}_L, \quad \vec{Q}_R = \frac{g m_0}{\sqrt{2}}\cosh\alpha\,\vec{n}_R. \tag{45}$$

Consequently,

$$M^2 = \frac{1}{8g^2}\vec{Q}_R^2. \tag{46}$$

The black hole is Bogomol'nyi saturated.

This black hole can be identified with a class of massive string states [11], and this is what allows its entropy to be determined by direct counting. The density of states in heterotic string theory is given for a large number N of oscillators by [12] as

$$\rho \approx \text{const.} N^{-23/2} e^{2a\sqrt{N}}, \tag{47}$$

where $a_L = 2\pi, a_R = \sqrt{2}\pi$. The numbers of oscillators in the left and right sectors are related to the mass and charges of the corresponding states by the usual formula

$$M^2 = \frac{g^2}{8}\left(\frac{\vec{Q}_L^2}{g^4} + 2N_L - 2\right) = \frac{g^2}{8}\left(\frac{\vec{Q}_R^2}{g^4} + 2N_R - 1\right). \tag{48}$$

To find the level density in terms of the ADM mass of a black hole, one has to combine this formula with the relation between the mass and the charges as applicable for the solution describing that black hole. Here, $N_R = \frac{1}{2}$ and the entropy arises from large values of N_L.

$$S = \log\rho \approx 4\pi\sqrt{N_L} \approx \frac{8\pi}{g}\sqrt{M^2 - \frac{Q_L^2}{8g^2}} = \frac{8\pi}{g\cosh\alpha}M = \frac{2\pi}{g}m_0. \tag{49}$$

While this result is nonzero, it must be remembered that in this limiting case the horizon has a vanishing area. So the area formula is *not* supported. If one desires to express the entropy as the area of something, one can try to construct a surface using various prescriptions [10].

However, these limiting cases describe special black holes where the mass and the charges are related with one another, so that these are *extremal cases*. It is known [13] that extremal and nonextremal cases in the euclidean version are topologically different, so that continuity need not hold. If the derivation of an expression for the thermodynamic entropy [5] is attempted afresh for the extremal case, with due attention paid to the fact that *the mass and charges are no longer independent as in the usual cases,* one obtains a form proportional to the mass of the black hole with an undetermined scale [14], as we now show.

Let $\vec{\Phi}$ represent the chemical potential corresponding to the charge $\vec{Q}$. We can make use of the $O(22) \times O(6)$ symmetry to write

$$\vec{\Phi} = \begin{cases} \frac{\sqrt{2} f_L \vec{n}_L}{4g} & L = 1, \ldots 22 \\ \frac{\sqrt{2} f_R \vec{n}_R}{4g} & R = 23, \ldots 28 \end{cases}, \tag{50}$$

where f_L, f_R are unknown functions of m_0 and α. There are standard expressions for the chemical potential in nonextremal cases, but we cannot use them for two reasons: first, extremal black holes may not be continuously connected to nonextremal black holes [13], and secondly, the standard expressions are calculated by differentiating the mass with respect to charges at constant *area* in the anticipation that constant area and constant entropy are synonymous, but this is an assumption we would not like to make. Only such thermodynamic processes are considered here which leave the black hole in the class being considered, *i.e.,* all variations are in the parameters m_0, α and the unit vectors $\vec{n}_L, \vec{n}_R$. Other processes too can occur but are not needed for this discussion.

Once again, in the leading semiclassical approximation, the partition function can be taken to be the exponential of the negative classical action, which vanishes in this case as the area vanishes. Hence the thermodynamic potential vanishes too and we have, as in (14),

$$TS = M - \vec{\Phi} \cdot \vec{Q} = \frac{(\cosh\alpha - f_L \sinh\alpha - f_R \cosh\alpha) m_0}{4}. \tag{51}$$

Using (44), we then have

$$S = \pi m_0^2 (\cosh\alpha - f_L \sinh\alpha - f_R \cosh\alpha). \tag{52}$$

Further, the first law of thermodynamics,

$$TdS = dM - \vec{\Phi} \cdot d\vec{Q}, \tag{53}$$

takes the form

$$T\frac{\partial S}{\partial m_0} = \frac{\partial M}{\partial m_0} - \vec{\Phi} \cdot \frac{\partial \vec{Q}}{\partial m_0} = \frac{(\cosh\alpha - f_L \sinh\alpha - f_R \cosh\alpha)}{4}. \tag{54}$$

This can be written in view of (52) as

$$\frac{\partial S}{\partial m_0} = \frac{S}{m_0}, \tag{55}$$

whence

$$S = k(\alpha) m_0, \tag{56}$$

with $k(\alpha)$ now an undetermined function of α. This function cannot be fixed by considering the analogue of (54) where the m_0-derivatives are replaced by α-derivatives; what happens is that f_L, f_R get expressed in terms of k. The string answer (49) for the entropy is indeed of the form (56), with $k(\alpha)$ actually taking the constant value $\frac{2\pi}{g}$.

5 Conclusion

We have come a long way from the seventies, when what seemed like analogues of the laws of thermodynamics were discovered for black hole physics. First it became clear that the surface gravity, which was the analogue of temperature entering those laws, is indeed proportional to the temperature, the proportionality factor involving Planck's constant. Then it became apparent that the area of the horizon, which was the analogue of entropy, naturally enters the expression for the entropy of matter in the background of a simple black hole. The next step should be to derive the area expression for the entropy of the black hole itself from a counting of states. But this would involve a quantum theory of gravity. What has been achieved in this direction is the embedding of some special black holes in string theory, leading to a microscopic calculation of the entropy. This has *not* produced the area as the answer, but that is no longer an occasion for surprise: for these special black holes the thermodynamical approach also leads to a form different from the area but proportional to the mass instead. It is to be hoped that such calculations will soon be extended to more familiar black holes.

PS: For a new approach not discussed in this talk, see [15].

References

[1] S. Hawking, *Phys. Rev. Letters*, **26**, 1344 (1971)

[2] J. Bardeen, B. Carter and S. Hawking, *Comm. Math. Phys.*, **31**, 161 (1973)

[3] J. Bekenstein, *Phys. Rev.*, **D7**, 2333 (1973); *Phys. Rev.*, **D9**, 3292 (1974)

[4] S. Hawking, *Comm. Math. Phys.*, **43**, 199 (1975)

[5] G. Gibbons and S. Hawking, *Phys. Rev.*, **D15**, 2752 (1977)

[6] L. Smarr, *Phys. Rev. Letters*, **30**, 71 (1973)

[7] A. Ghosh and P. Mitra, gr-qc/9408040; P. Mitra, gr-qc/9503042, to appear in Proceedings of the Puri workshop on *Physics at the Planck scale*, 1994, *eds.* A. Kumar and J. Maharana, World Scientific, Singapore

[8] G. 't Hooft, *Nucl. Phys.*, **B256**, 727 (1985)

[9] A. Ghosh and P. Mitra, *Phys. Rev. Letters*, **73**, 2521 (1994)

[10] A. Sen, *Mod. Phys. Lett.*, **A10**, 2081 (1995)

[11] M. Duff, R. R. Khuri, R. Minasian and J. Rahmfeld, *Nucl. Phys.*, **B418**, 195 (1994); M. Duff and J. Rahmfeld, *Phys. Letters*, **B345**, 441 (1995)

[12] J. G. Russo and L. Susskind, *Nuc. Phys.*, **B437**, 611 (1995)

[13] S. Hawking, G. Horowitz and S. Ross, *Phys. Rev.*, **D51**, 4302 (1995)

[14] A. Ghosh and P. Mitra, hep-th/9509090; see also *Phys. Letters*, **B357**, 295 (1995); hep-th/9602057

[15] J. Maldacena and A. Strominger, hep-th/9603060

‘The Mathematical Theory of Black Holes’ has probably surpassed Hawking’s ‘Brief History of Time’ in the ratio of copies sold to copies actually read!

— MARTIN REES

ASHTEKAR APPROACH TO QUANTUM GRAVITY

G. Date[1]

Institute of Mathematical Sciences,
Madras 600 113, INDIA

Abstract

Some of the recent developments in the Ashtekar program of quantization of gravity are discussed in this talk.

[1]e-mail: shyam@imsc.ernet.in

However, it is well to remember that a theory that is derived from simple and elegant basic ideas need not be a "correct" one. In the case of general relativity, the 'simple and the elegant' ideas gave rise to a mathematical structure so consistent and rich in content that one's confidence in the theory results from these latter facts. Perhaps I could explain my meaning as follows: One commonly says non-linear equations are difficult to solve. Indeed, Einstein remarks in one of his papers, "If only it were not so damnably difficult to find rigorous solutions". This remark is consistent with his earlier reaction to Schwarzschild's discovery of the exact solution for the space-time about a point mass, that he was "astonished that the problem could be solved exactly". The experience during the past twenty-five years has been exactly in the opposite: so many problems of physical significance have been solved exactly that one is almost tempted to say that the test whether a question one asks of general relativity is physically significant or not is the solvability of the problem when it is properly formulated.

— S. CHANDRASEKHAR

There is no doubt that Chandra was enchanted by general relativity. He worked in this area for over thirty years, nearly half of his career. He found in it the 'strangeness in proportion' which to him was an essential feature of beauty. He had heard from Henry Moore that 'great sculptures should be viewed from all distances since new aspects of beauty will be revealed at every scale'. And he found that general relativity, and especially the mathematical theory of black holes, did just that.

— ABHAY ASHTEKAR

It has been a long standing occupation of theoretical physicists to construct *a non-perturbative* quantum theory of gravity. There are several different approaches and efforts of which canonical quantization is one.

Within canonical quantization approach itself there are at least two main versions:

a) Geometro-dynamics á la ADM et al ;

b) Connection-dynamics á la Ashtekar et al.

In this talk, I will try to discuss some of the relatively recent developments, in the approach b). Even this area is quite vast. The classical aspects of the Ashtekar variables, while interesting in their own right, are by now fairly well discussed and understood [1]. Therefore I will *not* discuss these. Their stronger appeal is when one considers canonical quantization of Einstein gravity. While quantization in the connection formulation is "easier" than in the geometro-dynamical formulation, it is still mathematically quite sophisticated and probably not very familiar. My aim in this talk is to give a concise sketch of a generalization of the quantization prescription and some of the recent developments in carrying out these steps in the Ashtekar formulation. I will indicate how some physical operators are defined and finally discuss how a semiclassical description is obtained. For further details I will refer you to the literature cited at the end.

The talk is divided in five parts along the following lines:

1. A minimal introduction to the canonical formulation(s) ;

2. Sketch of the steps in the (algebraic) quantization program;

3. Implementation of some of the steps and the current status;

4. Some Operators, "weaves" and semiclassical picture;

5. Summary.

1 Canonical Formulation(s):

That the Einstein equations can be interpreted in a Hamiltonian framework is well known eg. the ADM formulation. This is "geometro-dynamics" and is summarised below [2].

Geometro-dynamics (variables) Σ : Orientable, compact 3-manifold without boundary q_{ij} : Riemannian metric on Σ $q \equiv \det(q_{ij})$ π^{ij} : symmetric tensor density of weight 1 on Σ $\pi \equiv q_{ij}\pi^{ij}$ $\bar\nabla$: metric connection defined by q_{ij} Γ : locally coordinatised by $(q_{ij}, \pi^{ij})_{p \in \Sigma}$ $C_i \equiv -2q_{ij}\bar\nabla_k \pi^{jk}$ $C \equiv -\sqrt{q}\bar{R}(q_{ij}) \pm \frac{1}{\sqrt{q}}(\pi^{ij}\pi_{ij} - \frac{1}{2}\pi^2)$

Geometro-dynamics Constraints $C_i \approx 0$ vector constraint (diffeomorphisms of Σ) $C \approx 0$ Scalar constraint ("evolution in $\Sigma \times$ R ") $C_{\vec N} \equiv \int_\Sigma N^i C_i$ "smeared" vector constraint, $\vec N$ arbitrary $C_N \equiv \int_\Sigma NC$ "smeared" scalar constraint, N arbitrary $H_{(T)} \equiv C_{\vec N} + C_N$ The total Hamiltonian

Note: If Σ is noncompact (asymptotically flat) then there is an extra term in H_T but for our purposes it is not necessary.

The Constraint algebra

$$\begin{array}{lcllcl}
\{C_{\vec N}, C_{\vec M}\} & = & -C_{\vec K} & , & K^i & \equiv & (\vec N \cdot \vec\nabla \vec M^i - \vec M \cdot \vec\nabla \vec N^i) \\
\{C_{\vec N}, C_M\} & = & -C_K & , & K & \equiv & (\vec N \cdot \vec\nabla M) \\
\{C_N, C_M\} & = & \pm C_{\vec L} & , & L^i & \equiv & q^{ij}(N\partial_j M - M\partial_j N)
\end{array}$$

$\pm$ refer to Minkowskian/Euclidean theory in the four dimensional sense.

Provided that q_{ij}, π^{ij} 's satisfy the constraints at t = 0, $q_{ij}(t), \pi^{ij}(t)$ can be shown to construct a four dimensional space-time which solves the Einstein equations.

There is an alternative canonical formulation first discovered by Ashtekar [3] and elaborated by Ashtekar and others [1]. For convenience of presentation, I will deviate from the historical path and present instead a line of argument given by Bengtsson [4].

With the hindsight of Ashtekar's discovery, one could ask:

Can one define a set of "constraints" on the phase space of a non-abelian gauge theory such that the constraints satisfy the same algebra as above? Further, Can one show that this structure can also be derived via a four dimensional action which is equivalent to the Einstein-Hilbert action as far as the equations of motion are concerned?

The answer to both these questions is YES!

<u>Connection-dynamics</u>

Σ	:	as before
A_i^a	:	1-form on Σ taking values in a Lie algebra $\underline{G}$
E_a^i	:	vector density on Σ taking values in the dual of $\underline{G}$
		i is a spatial index while a is an algebra index
F_{ij}^a	$\equiv$	$\partial_i A_j^a - \partial_j A_i^a + f_{bc}^a A_i^b A_j^c$
Γ	:	locally coordinatised by A_i^a, E_a^i i.e. ,
		$\{A_i^a, E_b^j\}_{P.B.} \equiv \delta_i^j \delta_b^a \delta^3(x - x')$
$\mathcal{G}^a$	$\equiv$	$\partial_i E^{ia} + f_c^{ab} A_i^c E_b^i$, ("Gauss law")
$\mathcal{V}_i$	$\equiv$	$E^{ja} F_{aij}$, ("Vector constraint")
$\mathcal{S}$	$\equiv$	$\frac{1}{2} f_c^{ab} E_a^i E_b^j F_{ij}^c$, ("Scalar constraint")

Impose these constraints on Γ. As before one defines the "smeared" form of these constraints. Then, it was shown by Bengtsson that the constraint algebra is the same (modulo the Gauss law constraint) as that of geometro-dynamics *provided*

One chooses the Lie algebra to be that of SO(3) for the Euclidean case and that of complexified SO(3) for the Minkowskian case.

Note:

(a) The above constraint structure can be derived from the 4-dimensional Palatini form of the action, $\int e \wedge e \wedge R(\omega)$, with the spin connection restricted to (anti)self-dual part of the SO(1,3) connection. (This also requires complexification.)[1]

(b) Complexification requires the A's and E's to be complex valued to begin with. One therefore deals with complex gravity.

To recover real gravity one must demand that certain combinations of A's and E's must be real. These are the "Reality Conditions" . "Reality condition" look like further constraints. How to incorporate these has been a major difficulty for the quantization program.

This is the formulation that we will concentrate on.

Note that in either formulations one is dealing with an infinite dimensional constrained system.

Following Dirac, one has two alternative routes to follow. Either use additional "gauge fixing" constraints to get the reduced phase space Γ^*, use Dirac brackets and proceed with quantization OR quantise with original phase space Γ and impose the constraints as operator equations on the

Hilbert space to get to the "physical" Hilbert space.

Either are easily stated than implemented! Herein lies all the technical sophistication needed. Ashtekar approach takes the second route.

Let me list a few technical problem that can and do arise.

a. Being an infinite dimensional system, one has to regulate various formal operators suggested by the quantization prescription. This involves both the short distance singularities and the factor ordering problems. When the constraints are non-polynomial, as in the case of geometro-dynamics, these problems are too severe to handle. In the connection-dynamics formulation however the constraints are low order polynomials (barring the "reality condition") and there is chance that regularisation issues can be handled.

b. Even after regularisation issues are tackled and one has successfully defined the operator constraints, the equation $\hat{C} \mid phys\rangle = 0$ may admit NO solution in the original Hilbert space (i.e. solution may be non-normalisable).

If this is so then one needs to modify the Dirac procedure to be able to construct a non-trivial decent physical Hilbert space. It is here the so called algebraic quantization [5] procedure is adopted. As it may not be very familiar let me sketch briefly the steps in the procedure.

2 Algebraic Quantization Scheme:

In the following Γ denotes the phase space and C_i's denote the first class constraints generically.

step 1. Choose a subset, S, of the set of all complex valued smooth function on Γ such that (i) S is large enough to allow any sufficiently regular function on Γ to be expressed as a sum of products of its elements ; (ii) S is closed under the Poisson brackets and also closed under complex conjugation; (iii) elements of S are to have unambiguous quantum analogues.

In short, elements of S should play the role played by the q's and p's when $\Gamma \sim R^{2n}$.

step 2. Associate to S, an abstract, free algebra, B_{aux} such that

$$\begin{aligned} F &\rightarrow \tilde{F} \qquad\qquad \forall\ F \in S \ \Rightarrow \\ [\tilde{F}, \tilde{G}] &= i\hbar \widetilde{\{F, G\}}_{PB} \end{aligned}$$

Equip B_{aux} with a complex conjugation operation such that if $G = F^*$ in S then $\tilde{G} = (\tilde{F})^*$. Denote B_{aux} now as B^*_{aux} .

step 3. Let $\mathcal{H}_{aux}$ denote the Hilbert space on which B^*_{aux} is $\star$-represented i.e. if $\tilde{F}$ is represented by operator $\hat{F}$ on $\mathcal{H}_{aux}$, then

$$\widehat{(\widetilde{F^*})} = \hat{F}^{\dagger}$$

$\mathcal{H}_{aux}$ is like the Hilbert space one would construct from the Γ. These intermediate steps are needed when Γ is topologically complicated making it impossible to have global coordinates for Γ. $\mathcal{H}_{phy}$ is to be constructed from $\mathcal{H}_{aux}$ (usually a subspace).

step 4. Let C_i denote the self adjoint operators representing the first class constraints (OR their integrated versions). Since these may not admit solutions in $\mathcal{H}_{aux}$ one needs to generalise the procedure of constructing $\mathcal{H}_{phy}$. One aims at seeking solutions of the operator constraints in terms of "generalised eigenvectors". This is done as follows.

a. Let Φ denote a suitable dense subspace of $\mathcal{H}_{aux}$. Let Φ' denote its topological dual (i.e. the vector space of continuous complex valued linear functions on Φ with a suitable, different topology on $\mathcal{H}_{aux}$). Typically, $\Phi \subset \mathcal{H}_{aux} \Rightarrow \Phi' \supset \mathcal{H}^{dual}_{aux} \sim \mathcal{H}_{aux}$. Thus Φ should be "small" enough to have a "large" enough dual. Further $\hat{C}_i$ should act invariantly on Φ.

b. Apart from physical states one also wants to define physical operators. Having chosen a suitable Φ one defines a set of operators, $B^{\star}_{phy}$, which consists of those $\hat{F}$ which leave Φ invariant, commute with the constraint operators and so do the adjoints, $(\hat{F})^{\dagger}$.

step 5. Now one attempts to define an anti-linear map $\eta : \Phi \to \Phi'$ such that $\forall \chi_1, \chi_2 \in \Phi$,

$$
\begin{array}{llcccc}
(i) & [\hat{C}_i \eta(\chi_1)]\chi_2 & \equiv & (\eta\chi_1)(\hat{C}_i\chi_2) & = & 0 \qquad ; \\
(ii) & (\eta\chi_1)(\chi_2) & = & [(\eta\chi_2)(\chi_1)]^{\star} & , & (\eta\chi_i)(\chi_i) \geq 0 \qquad ; \\
(iii) & (\hat{A}\tilde{\psi})(\chi) & \equiv & \tilde{\psi}(\hat{A}\chi) & \forall & \tilde{\psi} \in \Phi' \quad , \quad \forall \ \chi \in \Phi
\end{array}
$$

step 6. The final step is to define physical Hilbert space and operators.

Let $\mathcal{V}_{phy}$ be the span of $\eta\Phi \subset \Phi'$. Define a new inner product on $\mathcal{V}_{phy}$ as,

$$
\langle \eta\chi_1, \eta\chi_2 \rangle_{phy} \equiv (\eta\chi_2)(\chi_1) \quad \forall \quad \chi_i \in \Phi
$$

and define $\mathcal{H}_{phy}$ to be the completion of $\mathcal{V}_{phy}$.

Operators defined via,

$$
\hat{A}(\eta\Phi) \equiv \eta(\hat{A}\Phi),
$$

are actually densely defined $\forall A \in B^{\star}_{phy}$ and as such define the physical operators, $\hat{A}_{phy}$.

This then is the quantization procedure. Note that $\mathcal{H}_{phy}$ is a subspace of the topological dual Φ' and NOT of $\mathcal{H}_{aux}$.

There are several inputs needed. One has to *choose* S, Φ, and also restrict one to $B^{\star}_{phy}$ to get physical operators. The general scheme permits this freedom but does not single out any natural choices. These are to be made in each physical context separately.

The connection-dynamical formulation actually allows one to carry out all these steps with the exception of the scalar constraint. I will now sketch how these steps are carried out in the connection-dynamical formulation (excluding the scalar constraint).

3 Implementation of Quantization Scheme and current status [6]:

The phase space: A priori the configuration space is the space $\mathcal{A}$ of all gauge potentials A^a_i 's. The Gauss law constraint expresses gauge invariance (excluding "large" gauge transformations, one could however promote the gauge group to include these "large" gauge transformations.) One can "solve" the Gauss law constraint classically i.e. deal with configuration space, $\mathcal{A}/\mathcal{G}$, the space of equivalence classes of the gauge connections. This is done and then the phase space is defined to be:

$$\Gamma \equiv \text{cotangent bundle of } \mathcal{A}/\mathcal{G}$$

<u>The set S:</u> For every closed curve γ in Σ define,

$$U_\gamma(s,t) \equiv \mathcal{P} \exp \int_{\gamma(s)}^{\gamma(t)} \mathrm{A} \;, \qquad \text{for t = s + 1 the curve retraces,}$$
$$U_\gamma(s) \equiv U_\gamma(s,s+1)$$

$$T_\gamma[A] \equiv \tfrac{1}{2} \,\mathrm{tr}\, U_\gamma(s)$$

The trace is in the fundamental representation of SO(3). The loop functions T_γ's are s independent and gauge invariant and thus are functions on $\mathcal{A}/\mathcal{G}$. Define,

$$T^i[\gamma; s](A,E) \equiv \frac{1}{2} tr\{U_\gamma(s) E^i(\gamma(s))\}$$

For every strip $\Lambda(\sigma,\tau)$ in Σ , $\sigma \in [0,1], \tau \in (-\epsilon,\epsilon)$ define,

$$T_\Lambda(A,E) \equiv \int_\Lambda ds^{ij} \epsilon_{ijk} T^k[\gamma_\tau, \sigma]$$

These strip functions are also gauge invariant. But since these involve the E's (and A's), these are functions on Γ.

The T_γ's and T_Λ's constitute the elements of S. They form an overcomplete set i.e. there are relations among them. The Poisson bracket algebra of these is also known [7]. Thus at this stage we have got the $B^\star_{aux}$ as well. What we need now is a representation of this algebra on a suitable $\mathcal{H}_{aux}$.

To do this, concentrate on the commutative subalgebra of the T_γ's. These functions serve to separate the points of $\mathcal{A}/\mathcal{G}$ i.e. if two potentials are inequivalent then there exists atleast one loop for which the corresponding T_γ's are different. This sub-algebra is the holonomy algebra and is denoted by $\mathcal{HA}$. If γ is a trivial loop, T_γ is 1 and is the identity element of the sub-algebra. Define the norm:

$$\|T_\gamma\| \equiv \sup_{[A]\in\mathcal{A}/\mathcal{G}} |T_\gamma[A]|$$

Completion of $\mathcal{HA}$ w.r.t. this norm gives us a commutative C-$\star$ algebra with identity, $\overline{\mathcal{HA}}$.

Now one invokes the Gelfand-Naimark theorem which asserts that *Every C-$\star$ algebra with identity is isomorphic to the C-$\star$ algebra of all continuous, bounded functions on a compact, Hausdorff space.* This space is called the spectrum of the algebra. Let us use the notation,

$$\overline{\mathcal{A}/\mathcal{G}} \equiv \text{spectrum}(\overline{\mathcal{HA}})$$

This suggestive notation is used because $\mathcal{A}/\mathcal{G}$ is dense in the spectrum and thus the spectrum can be thought of as the completion of $\mathcal{A}/\mathcal{G}$. What the theorem has enabled us to do is to have a convenient model for the completed holonomy algebra. We have still to construct $\mathcal{H}_{aux}$.

The C-$\star$ algebra structure allows construction of its representations on Hilbert spaces. For every cyclic representation of $\overline{\mathcal{HA}}$ there is Borel measure μ on $\overline{\mathcal{A}/\mathcal{G}}$ using which we get a Hilbert space:

$$\mathcal{H}_{aux} \equiv L^2(\overline{\mathcal{A}/\mathcal{G}}, \mu).$$

Thus $\mathcal{H}_{aux}$ consists of square integrable functions on the "quantum configuration space" $\overline{\mathcal{A}/\mathcal{G}}$. The T_γ 's act multiplicatively and are bounded operators on $\mathcal{H}_{aux}$. The T_Λ are linear in momentum variables, the E's. So one expects T_Λ to be represented by vector fields on $\overline{\mathcal{A}/\mathcal{G}}$. This essentially turns out to be the case. Further the requirement of self-adjointness of T_Λ picks out a unique measure μ_0. At this stage the steps 1 through 3 are completed.

Now the further steps get too technical, so I will be even more brief.

To incorporate the diffeomorphism constraint, one considers its integrated form (group element as opposed to Lie algebra element). Consider the group Diff(Σ) generated by complete, analytic vector fields on Σ. One then uses the so called "group averaging" technique to construct diff-invariant states. (This requires the subsequent steps of the algebraic quantization.)

The dense set Φ is to be chosen. This is provided by the so called "Cylindrical functions", $Cyl^\infty(\overline{\mathcal{A}/\mathcal{G}})$.

The upshot is that excepting the implementation of the scalar constraint, all the steps in the quantization scheme are implemented. The scalar constraint also has bearing on the "reality conditions" necessary in the complexified case (Minkowskian case). The recent development regarding "generalised Wick transform" indicates a possibility of handling this. Let me describe it briefly.

Consider the Euclidean case. Suppose one could succeed in imposing the scalar constraint in this case and find $\mathcal{H}_{phy}$ and A_{phy}. Suppose further that one could find a transformation taking one from the Euclidean to the Minkowskian case. One can then attempt to *define* the Minkowskian solutions from the Euclidean ones. Recent works of Thiemann and Ashtekar [8, 9] propose precisely such a transformation.

The idea is to find a Poisson bracket preserving self map of the algebra of functions on the phase space which will effect the crucial change of sign in the scalar constraint. Such a self map is defined as:

$$
\begin{aligned}
f &\longrightarrow \mathrm{W(1)} \circ \mathrm{f} \qquad \text{where,} \\
W(t) \circ f &\equiv \sum_{n \geq 0} \frac{t^n}{n!} \{f, T\}_n \\
T &\equiv i\frac{\pi}{2} \int_\Sigma A_i^a E_a^i
\end{aligned}
$$

It follows that $W \circ E_a^i = iE_a^i, W \circ A_i^a = -iA_i^a$. This implies that the Gauss law and the vector constraints are invariant while $W \circ \mathcal{S}_E = -\mathcal{S}_L$.

If one can complete the Euclidean case and *if* one could construct the quantum operator corresponding to the classically defined W above, *then* one could *define*:

$$
\begin{aligned}
\hat{\mathcal{S}}_L &\equiv \hat{W} \circ \hat{\mathcal{S}}_E \circ \hat{W}^{-1} \\
\mathcal{H}_{phy}^{Mink} &\equiv \hat{W} \mathcal{H}_{phy}^{Eucl}
\end{aligned}
$$

The simplicity of the classical expression for T suggests that this step may actually be carried out. However this not yet done.

Let me emphasize:

The whole approach is conservative in the sense that one is following relatively *non radical* extensions of the usual procedure of quantization. The approach is constructive with a combination of heuristics and technical sophistication (rigour). This is necessitated in part by our ignorance about phenomenology at the Planck scale. The only guide then is one's biases and mathematical existence. Only after this is done can one hope for uniqueness and make definitive predictions to be confronted with experiments.

4 Some operators, "weaves" and semiclassical picture [10]:

Okay, so we have a state space of quantum gravity (modulo the scalar constraint.) How does our intuition regarding geometry and classical gravity fit into this? This is tied with identifying some physical operators whose expectation values could reflect some classical geometrical picture. Let me discuss this aspect now.

I will take as given that we do have a quantum theory incorporating the diffeomorphism constraint. Although I have not discussed, loop representation so far is the most convenient one [7]. Thus physical wave functions are functions of diff-equivalence classes of loops in Σ i.e. $\psi_{phy} = \psi($ knot classes of loops in $\Sigma)$. The T_γ operators act multiplicatively on these ψ's.

Consider the 3-metric q^{ij}. Classically, it is given by,

$$(detq)q^{ij} \equiv E^i_a E^{ja}(x)$$

This involves product of E's at the same point and one needs to regulate it. Obvious method is the gauge invariant point splitting. Define,

$$T^{ij}_\gamma(x,x') \equiv \tfrac{1}{2}\ \mathrm{tr}\ [\{U_\gamma(x',x)E^i(x')\}\{U_\gamma(x,x')E^j(x)\}]$$

x, x' are the split points and γ is a loop through them. The expression is gauge invariant and in the limit γ shrinks to the trivial loop $(x \to x')$, $T^{ij}_\gamma \to q^{ij}$. In the loop representation the corresponding $\hat{T}^{ij}$ can be defined and these are taken as a definition of $\hat{q}^{ij}$. Using this $\hat{T}^{ij}$ one defines two operators whose classical expressions are:

$$Q(\omega) = \int \sqrt{q^{ij}\omega_i\omega_j}$$

where ω is any smooth 1-form with compact support

$$\mathcal{A}_S = \lim_{N\to\infty}\sum_{n=1}^{N}\mathcal{A}_n^{approx} \quad , \quad where,$$

$$\mathcal{A}_n^{approx} \equiv -\frac{1}{4}\left[\int_{S_n} d^2\sigma^{jk}\epsilon_{ijk}(x)\int_{S_n} d^2\sigma^{j'k'}\epsilon_{i'j'k'}(x')T^{ii'}(x,x')\right]^{1/2}$$

Here, a two dimensional surface S is broken up into N surface elements S_n. In both these expressions we have q^{ij} and T^{ij}_γ for which we have well defined quantum definitions.

These two operators act multiplicatively:

$$\hat{Q}(\omega))|\psi(\gamma))\rangle = (l_P^2 \oint_\gamma \dot{\gamma}^i\omega_i)\ \ |\psi(\gamma)\rangle$$

$$\hat{\mathcal{A}}_S\ \ |\psi(\gamma)\rangle = (\frac{l_P^2}{2}\ \ I(S,\gamma))\ \ |\psi(\gamma)\rangle$$

Here l_P is the Planck length and I(S,γ) is the unoriented intersection number between S and γ. In other words, each loop contributes half the Planck area to any surface it intersects. Clearly the spectrum of the area operator $\hat{\mathcal{A}}_S$ is discrete. Using these operators one identifies the so called "weave states" [5].

The idea of weave states is that these are supposed to be quantum states which "semi-classically" "approximate" a given 3-geometry. Here "semiclassical" means in the limit $\hbar \to 0$ or $l_P \to 0$. "Approximation" is meant to be the following.

Both operators defined above are multiplicative. The approximation means that on weaves, their eigenvalues should equal the corresponding classical values to leading order in l_P/L with L, a length scale provided by the classical quantities, i.e.

Let h_{ij} be a given classical metric on Σ. A given surface S has an area as computed from this metric. Its square root provides a length L. In $\hat{Q}(\omega)$ the 1-form ω is supposed to vary slowly over this length scale.

The result is that there is a specific procedure to construct weave states. These states, $\psi(\gamma)$ are such that on an average only one line cuts every surface element whose area as measured by h_{ij} is one Planck area. However, such states are *not* unique.

This undoubtedly looks like an unconventional way of thinking of semiclassical approximation. This is because at the quantum level we seem to have wave functions of knot classes of loops in Σ rather that any sharp geometry. If classical picture is to emerge in an approximate way then we need one more scale other that the Planck scale. This additional length scale is supposed to be provided by an auxilliary geometry. One can then only ask if at a given coarse graining one can have expectation values of quantum operators to match with the values expected/obtained from the auxilliary geometry. This is precisely what is attempted in the construction of weaves. (This is not as outlandish as it may appear if one notes that the quantum framework it self does not need any geometrical input but its interpretation –system vs apparatus– does need at least a length scale.)

I have only touched upon the idea of weaves. For further details please consult references.

5 Summary:

1. For real A's and E's quantization steps including the diffeomorphism but excluding the scalar constraints can be carried out. The arena for quantum states at the kinematical level ($\mathcal{H}_{aux}$) is the Hilbert space $L^2(\overline{\mathcal{A}/\mathcal{G}}, \mu_0)$.
2. The scalar constraint, even for the Euclidean case, is not yet handled satisfactorily. But if it could be then the generalised Wick transform provides another strategy of incorporating the "Reality conditions".
3. A few quantum operators are well defined via the loop representation and have a purely discrete spectrum.
4. Treating the scalar constraint formally, one has a class of solutions of *all* the quantum constraints. These are functions of knot classes of loops on Σ and have support only on simple loops. This is an infinite set.
5. Semiclassical states, weaves, can be constructed which approximate a given 3-geometry with coarse graining provided by the geometry.
6. Σ is a crucial ingredient in this approach. How to handle a variety of Σ's (in the sense of "topology changing processes") is not adequately developed.

Einstein's theory of course does not stipulate the type (species etc) of matter constituting the physical world. Some suitable theory of the "matter sector" which will mesh with the above approach will clearly be needed to complete the picture.

References

[1] *Non-Perturbative Canonical Gravity*, A. Ashtekar, IUCAA Lect. Notes. No. 1, IUCAA, Pune, India, (1990);
A Lagrangian Basis for Ashtekar's reformulation of Canonical Gravity, J. Samuel, Pramana- J. Phys. **28** L429, (1987);
Covariant Action for Ashtekar's Form of Canonical Gravity, T. Jacobson and L. Smolin, Class. Quan. Grav. **5**, 583, (1988).

[2] R. Arnowitt, S. Deser and C. Misner in *Introduction to Current Research* ed. L. Witten, 227, (1962).

[3] *New Hamiltonian Formulation of General Relativity*, A. Ashtekar, Phys. Rev. **D36**, 1587, (1987).

[4] *Yang-Mills theory and General Relativity in three and four dimensions*, I. Bengtsson, Phys. Lett. **B220**, 51, (1989).

[5] *Mathematical Problem of Non-perturbative Quantum General Relativity*, A. Ashtekar, Les Houches 1992, gr-qc/9302024;

[6] *Quantization of Diffeomorphism invariant theories of Connection with local degrees of freedom*, A. Ashtekar, Lewandowski, Marolf, Mourao, Thiemann, Jour. Math. Phys. **36**, 6465,(1995), gr-qc/9504018;

[7] *Loop Representation for Quantum General Relativity*, C. Rovelli and L. Smolin, Nucl. Phys. **B331**, 80, (1990).

[8] *An Account of Transforms on* $\overline{\mathcal{A}/\mathcal{G}}$, T. Thiemann, gr-qc/9511049;
Reality Conditions Inducing Transforms for Quantum Gauge Field Theory and Quantum Gravity, T. Thiemann, gr-qc/9511057;

[9] *A Generalised Wick transform for gravity*, A. Ashtekar, gr-qc/9511083;

[10] *Gravitons and Loops*, A. Ashtekar, C. Rovelli and L. Smolin, Phys. Rev. **D44**, 1740, (1991);
Mathematical Problem of Non-perturbative Quantum General Relativity, A. Ashtekar, Les Houches 1992, gr-qc/9302024;

[11] For an extensive list of references see, *Bibliography of Publications related to Classical and Quantum Gravity in terms of Ashtekar Variables*, A. Troy, gr-qc/9409031.

Quantum Gravity on Computers

N.D. Hari Dass[1]

Institute of Mathematical Sciences,
Madras 600 113, INDIA

[1]e-mail:dass@imsc.ernet.in

Some times Thomas Hardy has been compared to Dostoevski, but Dostoevski is an incomparably finer artist. If Dostoevski's works could be compared to some of say, Rubens' paintings, then Hardy's are neatly finished diagrams drawn on graph paper with ruler and compass! Compare for instance the tragedy to Tess with that of Sonia, or the culmination of the forces in Raskolnikoff's confession with the melodramatic ending of "Return of the Native". "Crime and Punishment" is more superb in its conception than even Hugo's "Les Misérables". Great as is the tragedy of Raskolnikoff, greater still is the tragedy of Sonia. She reminds us of Fantine when poverty and starvation forces her to a prostitute's life, of Ophelia in her tragic devotion of Hamlet, of Cosette in her simplicity, but to whom could we compare her when, for instance, she reads out the resurrection of Lazarus to her lover. Dostoevski himself characterizes her most delicately in the words of Raskolnikoff as he threw himself at her feet, 'I do not bow to you personally, but to the suffering humanity in your person'.

— S. CHANDRASEKHAR

1 Introduction

At the moment quantisation of gravity, called "Quantum gravity" for short, appears to be the last frontier in Physics. This problem has proved to be formidable both conceptually and technically. At the core of these difficulties has been the fact that the dynamical degrees of freedom here are space time geometries. It is believed by many that several difficulties in the structure of physics like the singularities (space-time) of classical general Relativity theory as well as the infinities of relativistic quantum field theories would be resolved with a successful quantisation of gravity. Though some may even question the need for quantising gravity, the widespread perception is that lack of progress in quantum gravity represents a major void in our understanding of nature.

Currently many apparently diverse approaches to this problem have surfaced. For example there are the spin-2 graviton theories built around an underlying flat space where effects conventionally ascribed to the curvature of space-time are sought to be explained as dynamical manifestations of the spin-2 quanta. Classically, these theories have been shown to account for all the effects predicted on the basis of general theory of relativity as long as the gravitational fields are not too strong or are such as to change the topology of space-time as in the case of black holes, for example. Quantisation is realised only perturbatively. A more satisfactory, but essentially equivalent, formulation consists in quantising the Einstein-Hilbert action around a fixed (usually flat) background. Here too quantisation can be achieved only perturbatively. An important aspect in which perturbative quantum gravity differs from other successful quantum field theories like QED, QCD is in the lack of renormalisability. Nevertheless, a regularised version of the quantum theory exists which is identical, as a formalism, to other regularised quantum gauge field theories.

Though QG based on the Einstein-Hilbert action is non-renormalisable in $d = 4$, the renormalisability "improves", in the sense that non-renormalisable counter terms begin to appear only at higher and higher orders of perturbation theory, by the inclusion of appropriate matter with enhanced symmetries called "supersymmetry". Of all such supergravity theories, the so called $N = 8$ theory is remarkable in many respects. Apart from possible nonrenormalisability appearing only at seven loops, the matter content of these theories is uniquely fixed, making them attractive choices not only for QG but also for the unification of all fundamental forces. However, here too quantisation can only be realised perturbatively.

The fact that perturbative schemes for quantum gravity pre-select a background metric around which the fluctuations are quantised makes them rather unattractive even though they may have validity in some limited domain. The earliest approach, the canonical quantisation, avoided this pit fall. However, the technical difficulties of implementing this scheme are rather well known. In a nutshell, the theory is seen to be a constrained system with the added difficulty that the solution of the constraints needs the solution to the dynamics itself. Also, the constraints are highly nonlinear and nonpolynomial. A great ray of hope in this direction is the reformulation in terms of Ashtekar variables wherein not only do the constraints become polynomial, but the metric itself becomes a secondary object. The approach also promises to open new vistas as far as nonperturbative issues are concerned.

Finally, mention has to be made of approaches to quantum gravity based on string theory. As is well known, consistency with conformal invariance of the two dimensional world sheet theory automatically incorporates spin-2 gravitons with the added bonus that (perturbative) quantum gravity is finite in this approach. Still, a satisfactory non perturbative approach is lacking. One has to await further progress in string field theory.

The main theme of my talk today is some recent progress made in understanding some nonperturbative features of QG by using numerical simulations which I have termed "Quantum gravity on computers" in a lighter vein. The literature on even this recent development is enormous and I shall not attempt to give a full bibliographical account of this fascinating area. Instead, I shall give a few key references from which the interested reader can construct a "path of understanding".

Though recourse is made to numerics in this approach, it is derived from the finest analytical ingredients from simplicial topology, geometry, statistical mechanics etc. Before discussing the details of the numerical simulations of quantum gravity, it is worthwhile understanding how the concepts work when applied to quantum field theories in flat space-time. There are essentially two approaches to such conventional quantum field theories : a) the *Hamiltonian approach* where the dynamical variables at each space point at a given instant of time are operators acting on a huge Hilbert space a basis for which is the well known Fock space. The short distance behaviour of the product of field operators is singular

$$\phi(x)\ \phi(x+\epsilon)\ \sim \frac{1}{\epsilon^{\delta}} O\ +\ \ldots \qquad \delta > 0 \tag{1}$$

which is a reflection of the ultraviolet divergences inherent in quantum field theories owing to the infinitely many degrees of freedom.

The major technical problem of the Hamiltonian formalism is the careful treatment of such singular operator products while maintaining the basic structure of the field theory like its symmetries etc.

The other approach is the so called *Path Integral approach.* To appreciate this approach consider the quantum mechanical description of a particle. Though the trajectory of a particle has no meaning in quantum theory, in the path integral approach, one considers the space of all possible paths and assigns to each path a weight factor (in fact a phase factor) $e^{iS(X(t))}$ where $S(X(t))$ is the classical action evaluated for the path $X(t)$. Then one forms the **path integral**

$$K(x_1,t_1;x_2,t_2)\ =\ \int DX\ e^{iS(X)} \tag{2}$$

where DX is a measure (the Wiener measure) for the sum over all paths. This approach can also be called the "sum over histories" approach. Every quantity that appears in the above equation is a so called c-number and the use of mathematically subtle operators is circumvented. However, the price one has to pay for this is the care required in constructing the measure DX. As shown by Feynman, this approach is mathematically equivalent to the Hamiltonian (or Schrödinger) approach at least when the configuration space is topologically trivial. The translation to the Schrödinger approach is codified by the spectral representation for the Kernel $K(x_1,t_1;x_2,t_2)$

$$K(x_1,t_1;x_2,t_2)\ =\ \sum_n \psi_n^*(x_1,t_1)\psi_n(x_2,t_2) \tag{3}$$

A generalisation of this construction for the case of quantum field theory consists of replacing DX by $D\phi$, the measure for summing over all "histories of field configurations", and $S(X)$ by $S(\phi)$ the action-functional for the particular history of field configuration $\phi(x,t)$. The major technical problem one faces now is the construction of a meaningful measure $D\phi$ which is an infinite dimensional analog of the Wiener measure. Even if one were to succeed in finding such a measure, the technical problem of giving a meaning to the functional integral

$$Z\ =\ \int D\phi\ e^{iS(\phi)} \tag{4}$$

still remains because of the oscillatory nature of the integrand. This can be overcome by the so called **Euclideanisation**.

$$t \to iX_4^E \qquad iS(\phi) \to -iS_E(\phi) \tag{5}$$

As established by Schwinger, Wightman, Osterwalder, Schrader and others, there is an unique extrapolation from the results of euclidean field theory to those of Minkowski field theory, at least when the topology of spacetime is trivial. For most quantum field theories of interest on flat space-time it follows that

$$S_E(\phi) \geq 0 \tag{6}$$

so that the bothersome oscillatory factor has really been made into a damping factor. In fact, it appears that as long as the Hamiltonian of the theory is bounded from below, the euclidean action is positive semi-definite.

A notable exception to this is gravity! The euclideanised action here is

$$S_E \;=\; \int \sqrt{g} R_E \tag{7}$$

and clearly the scalar curvature of a euclidean manifold need not be positive semi-definite. We will have more to say on this later on.

Finally, the measure $D\phi$ is made meaningful by discretising the euclidean space-time.

$$D\phi \;=\; \Pi_x d\phi_x \tag{8}$$

Now derivatives of fields in the continuum have to be approximated by finite differences, for example,

$$\partial_\mu \phi \;\rightarrow\; \frac{\phi(\vec{x} \;-\; \phi(\vec{x} + a\vec{e}_\mu))}{a} \tag{9}$$

where a is the lattice spacing.

With these modifications the generating functional, Z, of the quantum field theory, becomes

$$Z \;=\; \int \Pi_i d\phi_i \; e^{-S(\{\phi_i\})} \tag{10}$$

and in this form is indistinguishable from the partition function of a classical statistical system in $d+1$ spatial dimensions where d is the spatial dimensionality of the quantum field theory problem. This exact mapping of quantum field theory onto a classical statistical mechanical problem is at the heart of the numerical simulations to probe the non perturbative aspects of quantum field theory.

Clearly there are many choices of discrete lattices one could consider like hypercubic lattice, triangular lattice and even a random lattice. And for each such choice, there are again many ways of approximating a derivative by a finite difference. Intuitively it appears reasonable that in the continuum limit all these differences in choice should become irrelevant. In fact, if one were to simply take the limit of the lattice spacing, a, to zero in all the mathematical expressions of the discretised theory, one would recover the formal expressions of the continuous theory. This is called the 'naive continuum' limit. It is a naive limiting procedure because the resulting continuum expressions are not mathematically well defined. But it turns out that the lattice theory affords a much more subtle and meaningful limiting procedure. To appreciate this, note that as the lattice spacing a tends to zero, physically measurable length scales like Compton wavelengths of particles etc must remain finite. That is, the ratio of the physical length scale l_{ph} to a actually diverges. By scaling all dimensionful quantities Q with appropriate powers of a to make them dimensionless Q_L (for example $m_L = m.a,\ area_L = area.a^{-2}$ etc) the partition function Z can the rewritten entirely in terms of the dimensionless quantities except for an irrelevant factor. The quantity $l_{ph}a^{-1}$ precisely corresponds to one of the many correlation lengths. Thus we see that the requirement for the existence of a continuum limit is that the classical statistical system onto which the quantum field theory has been mapped must be such that at least one of its correlation lengths diverges. But this is precisely the criterion for a statistical system to have a second or higher order (as distinguished from a first order) phase transition. Now the recipe for finding the true, as opposed to the naive, continuum limit is clear : treat the parameters of the quantum field theory like coupling constants, masses (suitably scaled to make them dimensionless) as parameters of a classical statistical system (temperature, concentration, magnetic field etc) and tune them till second order phase transition points (critical) are reached. At each of these critical points a continuum theory can be defined as follows : suppose that as the parameters are tuned to a particular set of critical values, the correlation function ξ diverges as

$$\xi \rightarrow \xi_c(\{\lambda\}) \quad \text{where} \quad \xi_c(\{\lambda_c\}) \;=\; \infty \tag{11}$$

and let O be an observable of mass dimensionality d, then O will survive the continuum limit if and only if

$$O \;\rightarrow\; \xi_c^{-d} \text{ as } \{\lambda\} \rightarrow \{\lambda_c\}\,. \tag{12}$$

This is called correlation length scaling. In particular, if there are two observables O_1 and O_2 of the same mass dimensionality surviving the continuum limit,

$$\frac{O_1}{O_2} \to c(\text{finite number}) \text{ as } \{\lambda\} \to \{\lambda_c\} \tag{13}$$

It is only the set of finite limits $\{c\}$ that represents the observable content of the continuum theory.

It is quite clear now that the "statistical continuum limit" provides much richer possibilities than the process of naive continuum limit. For one thing, the classical statistical system equivalent to the quantum field theory could have many critical points and at each of those critical points one could define a continuum theory. Even at the same critical point, in principle one could have several distinct sets of correlation lengths such that all correlation lengths belonging to a given set have the same scaling behaviour. In such a case, each of the distinct sets can define a continuum limit resulting in inequivalent continuum limits at the same critical point. This would be a statistical mechanical realisation of inequivalent quantisations. That such a possibility can indeed happen is evidenced in the case of the three dimensional compact quantum electrodynamics on the lattice where three distinct continuum limits can be defined at the same critical point corresponding to fixed string tension, fixed mass gap or the free massless photon phase.

2 Numerical Simulation of Statistical Systems

There are many techniques for numerical simulation of statistical systems chief among them being i) microcanonical simulations and ii) canonical simulations. Simulations are some times also based on the grand canonical ensemble. The essence of the numerical simulations is to generate an ensemble of configurations and perform the so called 'importance sampling' according to which a suitable selection procedure is adopted such that configurations that dominate the partition function are randomly generated.

The Monte Carlo simulation of this system consists of choosing an initial configuration and a so called move (update)- that yields another configuration for each initial configuration. Denoting a move that takes the configuration c_1 to c_2 by $W(c_1 \to c_2)$, there are various restrictions on $W(c_1 \to c_2)$ if the Monte Carlo simulation is to be reliable.

i) given a configuration c_1, *all* configurations of the system must be eventually reachable by a sequence of moves. This property is called "ergodicity". If this property is not satisfied, parts of the configuration space are never sampled.

ii) since $W(c_1 \to c_2)$ represents the probability of getting the configuration c_2 starting from the configuration c_1, the following must hold :

$$\sum_{c_2} W(c_1 \to c_2) = 1 \tag{14}$$

$$\sum_{c_1} W(c_1 \to c_2) = 1 \tag{15}$$

Given the above two properties, it follows from an application of the Frobenius-Perron theorem that W viewed as a matrix, has maximum eigenvalue of unity, and the eigenfunction corresponding to this largest eigenvalue is the equilibrium distribution $P_{eq}(C)$. That is

$$\sum_{c_1} W(c_1 \to c_2) P_{eq}(c_1) = P_{eq}(c_2) \tag{16}$$

It then follows that any initial ensemble of configurations eventually evolves into the equilibrium ensemble distribution $P_{eq}(c)$. How quickly the initial ensemble evolves into the equilibrium ensemble distribution depends inversely on the gap between the largest eigenvalue of unity and the next largest eigenvalue.

In principle any $W(c_1 \to c_2)$ that satisfies these requirements should result in a reliable Monte Carlo simulation of the system. In practice, however, an additional requirement is made on $W(c_1 \to c_2)$ called the property of "detailed balance". This requirement is

$$P_{eq}(c)W(c \to c') \; = \; P_{eq}(c') \; W(c' \to c) \tag{17}$$

There are two obvious solutions to the requirement of detailed balance : i) heat bath : $W(c \to c') = P_{eq}(c')$ irrespective of c. ii) Metropolis : if we write $P_{eq}(c)$ as $e^{-S(c)}$, the metropolis algorithm for going from a configuration c to c' is as follows : Pick the configuration c' randomly at first : if $S(c')$ is less than $S(c)$ accept the configuration c' as the final configuration. If on the other hand, if $S(c')$ is greater than $S(c)$, accept c' with the probability $P \;\; e^{-S(c')}$.

Thus a practical implementation of the Monte Carlo simulation starts with an initial configuration on which one applies a sequence of moves. Two extremes for the initial configuration are the so called *cold* and *hot* starts. In the cold start, the dynamical degree of freedom takes the same value at each lattice site while in the hot start the dynamical degree of freedom takes completely random values at the lattice sites. After applying the moves a number of times, one monitors the approach to thermal equilibrium by measuring some observable. The onset of thermal equilibrium is heralded by the constancy of the average (local) of the observable with fluctuations following the characteristics of thermal fluctuations. The constant average value should be independent of the initial configuration one started with.

Once thermal equilibrium has been reached, one again generates a sequence of configurations by making sweeps of moves throughout the lattice and computes the average values of various observables. The estimation of the statistical errors is subtle as there is no a priori guarantee that the sequence of configurations used for making measurements are statistically independent. The degree of statistical independence of subsequent configurations is estimated by the "autocorrelation time" τ. Essentially, τ is also the time scale that governs the speed with which the system reaches equilibrium (time in Monte Carlo simulations is simply the total number of sweeps carried out on the initial configuration). A practical way of estimating the autocorrelation time is by measuring the so called "integrated autocorrelation time". First, one measures the autocorrelation function for some observable O

$$A(T) \; = \; \frac{< O(t)O(t+T) > - < O(t) >^2}{< O^2(t) > - < O(t) >^2} \tag{18}$$

The integrated autocorrelation time is given by

$$\tau_{int} \; = \; \frac{1}{2} + 2 \sum_{T'=1}^{\infty} A(T') \tag{19}$$

In practice, a lot of care has to be exercised in employing this method. Because of inherent noise, $A(T)$ for large T only goes to a constant rather than vanishing exponentially. Because of this, a naive application of the method yields diverging τ_{int}. Experience has shown that cutting off $T' \sim 6\tau_{int}$ gives reliable estimates. It is fair to say that understanding autocorrelation times is more an art than a science! If the autocorrelation time is τ, then in a sequence of N sweeps only N/τ are statistically independent.

Since the continuum limit resides only at the critical point, where the correlation lengths diverge, autocorrelation times register dramatic increase. This is quantified by so called 'dynamic critical exponent' $\tau \sim L^z$,where L is the system size. Off criticality $z \sim 0$ while at criticality $z \sim 2$. Thus as we go to larger and larger systems to avoid finite size effects, the autocorrelation times at criticality become L^2 times autocorrelation times off criticality which are of order unity. Thus to improve statistical accuracy one would need unrealistically large computer resources. Special algorithms are needed that drastically reduce the dynamic critical exponent z. Examples of such algorithms are the cluster algorithm, multigrid algorithm etc. In terms of the eigenvalue spectrum associated with $W((c \to c')$, what happens as we approach criticality is that the second largest eigenvalue begins to approach the largest eigenvalue (1). This means that the thermalisation times also become very large.

In addition to the autocorrelation effects, effects of finite size often have a critical bearing on the reliability of the results.

3 Quantum gravity

The dynamical degrees of freedom for the gravitational system are the metric components with the restriction that two metrics that are related to each other by a general coordinate transformation are to be identified as the same metric. Of course, when Fermionic matter is coupled to gravity, the appropriate degrees of freedom are the Vierbeins (in $d = 4$) or tetrads, now with the restriction that two tetrads related by either general coordinate transformation or local Lorentz transformation are to be regarded as being physically equivalent. In the present discussion, we shall restrict attention to gravitational systems with utmost bosonic matter.

The relevant functional integral is

$$Z \;=\; \int Dg e^{-S(g)} \tag{20}$$

where $S(g)$ is some general coordinate invariant action and has the form

$$S(g) \;=\; \alpha \int dv + \beta \int R_E dv + \gamma \int R_{\mu v} R^{\mu v} dv + \ldots \tag{21}$$

dv being the scalar volume element $(\sqrt{g} d^n x)$. The Einstein-Hilbert action corresponds to only having the second term. Thus quantising general relativity would correspond to dealing with

$$Z \;=\; \int Dg e^{-\beta \int dv R_E} \tag{22}$$

Unlike the flat space quantum field theories discussed earlier, the euclideanised action now is not positive definite because even for euclidean manifolds the scalar curvature can have any sign. There has been a variety of responses to this situation. Some of these are motivated by the observation that this instability can essentially be attributed to the so called conformal mode. To understand this, consider some fiducial metric g_+ such that the scalar curvature is positive. Then a metric g can be found that is related to g_+ by a conformal factor, $g' = e^{\phi} g_+$, such that the scalar curvature corresponding to g' is negative. In this manner the lack of positive semi-definiteness of the Einstein-Hilbert action can be seen to be solely due to the action of the conformal mode whose kinetic term has the wrong sign.

Some of the suggestions made to overcome this situation are : i) throw away the conformal mode altogether, ii) rotate the conformal mode into purely imaginary values thereby making the kinetic term of the rotated mode of the right sign. But both these change the theory so it is not clear that it is general relativity that one is quantising. Another tantalising suggestion, made by Greensite, is to "stabilise" the action. In effect, this amounts to a non perturbative modification of the theory whose perturbative content is exactly the same as that of the original theory. But here too one is changing the theory.

Whether the unboundedness of the action is really a sickness that ought to be cured by modifying the theory, or a feature essential to the physics of quantum gravity is something that we are yet to find out. Some have argued that though the action is bounded, the entropy factor could overcome this. For this to happen, the entropy of configurations with large negative scalar curvature has to become very small. I do not see how this is possible as one can always find a conformal factor for every positive scalar curvature configuration that would map it to a negative scalar curvature configuration. In $d = 2$, the unboundedness problem does not exist as $\int \sqrt{gR}$ is the Euler characteristic of the manifold.

With these provisos, we define the problem of quantum gravity to be eqn(20). To carry out numerical simulations of this problem, we need a discretisation that still keeps the geometrical essence of the problem.

In analogy with QFT on flat spacetime, one could attempt a discretisation that would consider $g_{\mu\nu}$ at discrete lattice points as the basic variables. Already in the case of non-Abelian gauge theories on flat space-time, such a naive discretisation is incompatible with any discrete version of the non-Abelian gauge transformation. It is well known that gravity theory has many essential features of a non-Abelian gauge theory and consequently one expects the same kind of difficulties in implementing any discrete version of general covariance. In the case of flat spacetime non-Abelian gauge theories, discretisation compatible with gauge invariance is achieved by constructing the action out of elements of the holonomy of the Yang-Mills connection around closed loops of the discretised manifold. Often, the holonomy group elements are composed of elementary group elements called "Link" variables. Such a formulation of discretised quantum gravity is certainly possible. But here we shall follow a path to discretisation that is very elegant and intrinsically geometric. Further, there is a whole body of very powerful analytical work related to it that would be indispensable to both the implementation and the subsequent interpretation of numerical simulations.

3.1 Geometric discretisation - Simplicial decomposition

The spirit behind this approach is that any manifold can be approximated with "arbitrary accuracy" by a so called simplicial manifold. A d-dimensional simplicial manifold is a collection of d-simplices glued along the $d-1$ dimensional boundary simplices.

Some simplices are shown in the next figure :

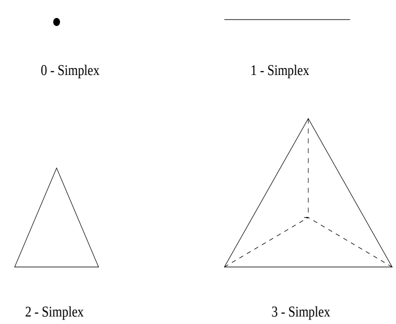


Figure 1: Some d - Simplices

A d-dimensional simplex has as its boundary $(d+1)$ sub simplices of dimensionality $d-1$. As David has pointed out there are certain subtleties about gluing simplices to get a simplicial manifold. This is illustrated by the following example :

Consider the tetrahedron UNWSED and join EW. Then make the following 2-simplicial identifications : UWS $\leftrightarrow$ UEN ,UNW $\leftrightarrow$ UES, DNW $\leftrightarrow$ DES, DWS $\leftrightarrow$ DEN. What one gets this way is a two dimensional simplicial complex where 2-simplices are glued to each other along 1-simplices (edges). Nevertheless, this simplicial complex is not a simplicial manifold because the neighbourhood of every vertex does not have the same topology. In fact, the boundary of the neighbourhood of U and D is T_2 and not S_2. Thus one has to subject the simplicial complex to a "manifold test".

An example of a two dimensional simplicial manifold is given in the next figure. The simplicial decomposition of manifolds can be used to two rather different approaches to discretised quantum gravity. These are : i) Regge calculus, and ii) dynamical triangulation (DTR).

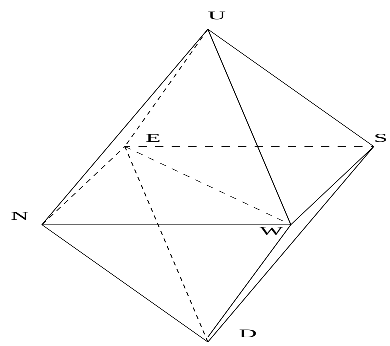


Figure 2: A simplicial complex which is *not* a simplicial manifold

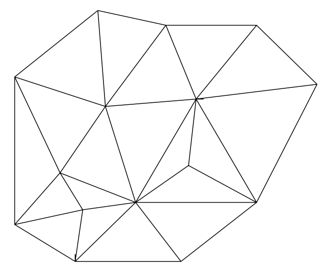


Figure 3: A simplicial complex which *is* a simplicial manifold

3.1.1 Regge calculus

In this approach one fixes the incidence number for each vertex (which for the two-dimensional example considered in figure 4 is simply the number of neighbours of each vertex). The length of each edge is considered as a dynamical variable and curvature is measured via deficit angles.

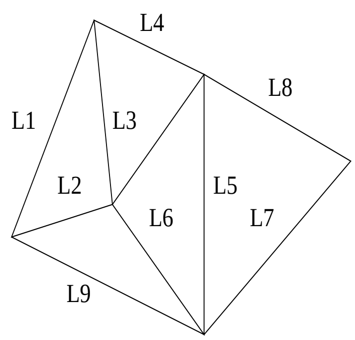


Figure 4: A Regge calculus configuration

The quantum gravity functional integral is now replaced by

$$\int Dg \to \int \left\{ \frac{dl_i}{f(l_i)} \right\} \tag{23}$$

$f(l_i)$ is the measure. As is well known, the issue of the functional measure in quantum gravity is still not very well understood. In quantum Regge calculus, this issue of the measure is further complicated by the fact that (infinitely) many different edge length configurations correspond to the same physical manifold. This is most easily demonstrated by the example of the two dimensional flat manifold for which infinitely many Regge discretisations are possible. Starting with any given edge length configuration, one can generate all the rest by simply moving the vertices around while maintaining the various triangle-like identities. In fact, this example can be generalised to the case of any maximally symmetric space.

Numerical simulations based on Regge calculus in the case of two dimensional quantum gravity has recently been criticised for its inability to reproduce known analytical results for the so called "string susceptibility" (see discussions later in the text) for quantum gravity coupled to conformal matter. On the other hand Kawai and his coworkers have shown that some of the universal features like loop length distributions are correctly reproduced by Regge calculus method. As remarked earlier, the correct measure has not been identified and this could be at the heart of the matter. Resolution of these issues is clearly an important task for the future. For the rest of the talk I shall mainly emphasise the DTR approach.

3.1.2 Dynamical Triangulation (DTR)

In this approach all simplices are taken to be equilateral i.e. the edge lengths are all the same. The interior and the boundaries of all simplices and sub simplices, as in the case of Regge calculus, are considered to be flat. The manifold one gets is so called "piecewise linear". Nontrivial curvature is produced by letting the incidence number fluctuate dynamically. The functional integral over metrics is replaced by a summation over triangulations, with some measure :

$$\int Dg \rightarrow \sum_T \frac{1}{c_T} \tag{24}$$

Again, there is no guiding principle to determine the measure, C_T ; nevertheless, results have not shown any sensitive dependence on C_T. Also, the DTR simulations in $d = 2$ case have reproduced many of the exact analytical results.

3.2 Sum over topologies

In both these approaches, the issue of the topology of the manifold has to be addressed. Here again there does not appear to be any guiding principle at present. In the case of string theory represented as two dimensional quantum gravity coupled to matter, a sum over all topologies is mandatory. If on the other hand, one wishes to investigate some statistical mechanical system on a manifold with fixed topology but fluctuating metric, then one would not sum over topologies. With this in mind, the relevant functional integral will be taken to be either with a fixed topology or summed over different topologies with appropriate weight factors as the case may be.

3.3 Alexander Moves (updates)

With the understanding that the quantum gravity functional integral is to be replaced by a sum over all possible simplicial decompositions (also called "triangulation" from now onwards), we need to specify a move, or an update, that would take one triangulation to another. The so called "Alexander moves" provide the answer. The issue of ergodicity will be discussed shortly.

The Alexander move (ij, x) is defined as follows: take any link (ij) of the complex. Insert a new site x in the interior of (ij) and divide up the simplexes which contain (ij) into twice as many simplices, half of which replace i by x and half of which replace j by x. The inverse of this Alexander move is defined accordingly. We illustrate the Alexander move in two dimensions in the next figure.

A related concept is that of $(k, l)_d$ moves which are defined as follows: consider a $(d+1)$ - simplex S_{d+1} whose boundary is the collection $\sum_d$ of d-simplices. Now partition $\sum_d$ into collections $\sum_1, \sum_2$ such that $\sum_1 \cup \sum_2 = \sum_d$ and $\sum_1 \cap \sum_2 = \phi$. Since $\sum_d$ is the boundary of $S_{d+1}(\sum_d = \partial S_{d+1})$,

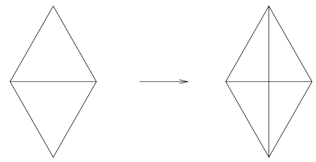


Figure 5: An Alexander move in two dimensions

the boundaries of $\sum_1$ and $\sum_2$ are the same but with opposite orientation. This follows from the deeper result that the "boundary of a boundary is zero" ($\partial \sum_d = \partial^2 S_{d+1} = 0 \Rightarrow \partial \sum_1 = -\partial \sum_2$). If $\sum_1$ has k d-simplices and $\sum_2$ has l d-simplices, $k + \ell = d + 2$. The move $(k, \ell)_d$ amounts to replacing k simplices of $\sum_1$ by l simplices of $\sum_2$. The volume of the d-dimensional simplicial manifold is the total number of d-simplices in it ; thus the $(k, l)_d$ move leads to the volume change $\Delta V = l - k$. The maximum possible change of volume is d.

Let us illustrate the $(k, l)_d$ moves by a 2-dimensional example. So we need to start with a 3-simplex which is a tetrahedron.

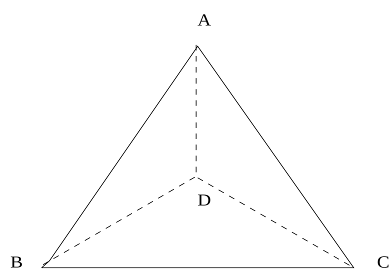


Figure 6:

The boundary of the tetrahedron ABCD is spanned by the triangles (2-simplices) ABD, BDC, DCA, ABC. thus

$$\sum_d = \{ABD, BDC, DCA, ABC\} \tag{25}$$

The inequivalent partitions of $\sum_d$ we should consider are a) into 1+3, b) into 2+2. In the latter case the move leads to no change in volume.

a) $1 \leftrightarrow 3$:

Let $\sum_1 = ABD$; $\sum_2 = \{BDC, DCA, ABC\}$

$\partial \sum_1 = loop\ (A, B, D)$; $\partial \sum_2 = loop\ (A, B, D)$ with opposite orientation.

Move $1 \rightarrow 3$ is to start with ABD, erect the 3-simplex ABCD on ABD as base and remove ABD leading to a 2-manifold finally. This move corresponds to $\Delta V = 2$ and an increase in the number of vertices by 1.

The inverse of this move is to locate a vertex with coordination number 3 (say D), remove the 3 triangles incident at D leaving a gap which is filled by adding the triangle ABD.

b) $2 \rightarrow 2$:

This can be accomplished by, say, the partition

$$\Sigma_1 = \{ABD, ABC\} \ ; \ \Sigma_2 = \{BDC, ADC\} \tag{26}$$

$$\partial\Sigma_1 = \{AD, DB, BC, CA\} \ ; \ \partial\Sigma_2 = -\partial\Sigma_1 \tag{27}$$

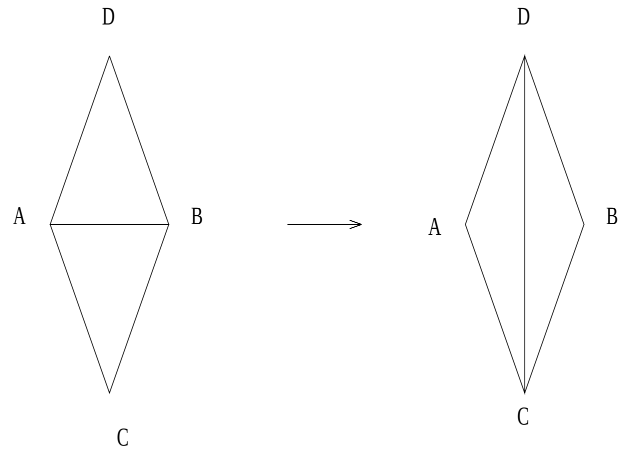


Figure 7: $2 \to 2$ move in two dimensions

4 The Dual Picture

It is often useful to introduce the dual of a simplicial manifold. Suppose we are working in D dimensions and have a simplicial manifold approximating the manifold. The simplicial manifold is obtained by gluing together D-simplices along the (D-1)-dimensional boundary simplices. Each of the (D-1)-dimensional boundary simplices has (D-2)-dimensional simplices as its boundary and so on. The 0-simplices are the vertices or sites of the discretised manifold.

The manifold dual to this is constructed as follows : to each p -simplex of the original manifold, a (D-p)-simplex of the Dual manifold is associated. If in the original manifold, two D-simplices V_D^1 and V_D^2 are connected along the (D-1)-simplex V_{D-1}^{12}, in the dual manifold the 0-simplices (Vertices) corresponding to them will be connected by a 1-simplex (edge) that is dual to V_{D-1}^{12}. This is illustrated below by a 2-dimensional example where the original edges are shown by bold lines and the dual edges in dotted lines :

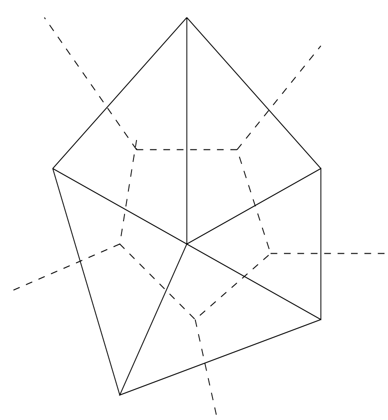


Figure 8:

It is clear that the incidence number of the dual graph is exactly (D+1).

The $(k,l)_d$ moves get increasingly more complex as we go to higher and higher dimensions. For example in $d = 3$, the possible moves are :

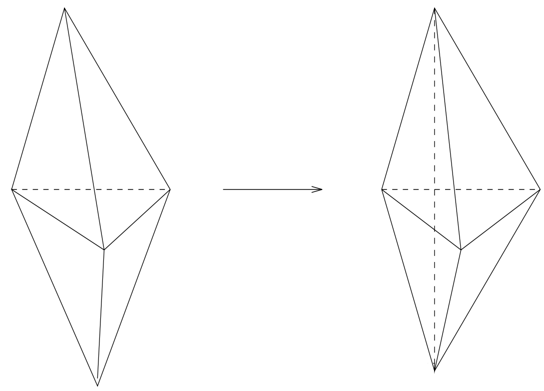


Figure 9:

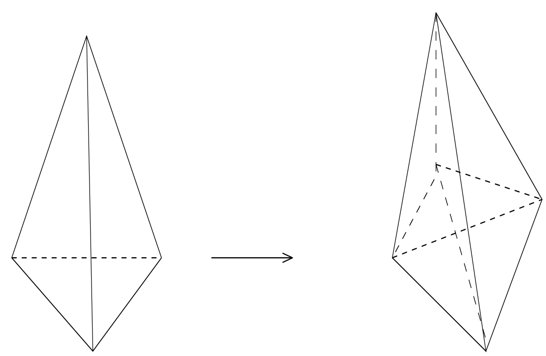


Figure 10:

These moves can be represented in a transparent manner in terms of the change in the connectivity of the dual graph. For example, the (2,2) move in 2-dimensions (the triangles are numbered)

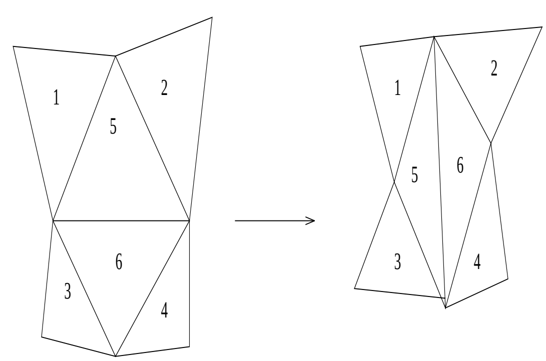


Figure 11:

has the dual representation:

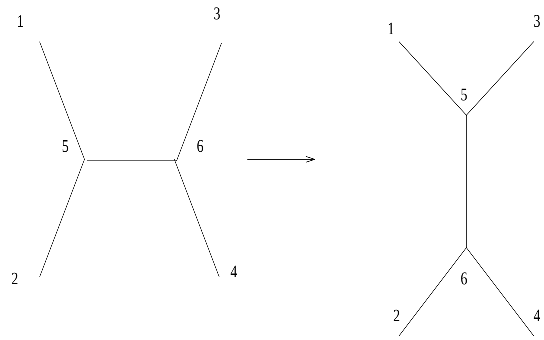


Figure 12:

Likewise, the (1,3) move

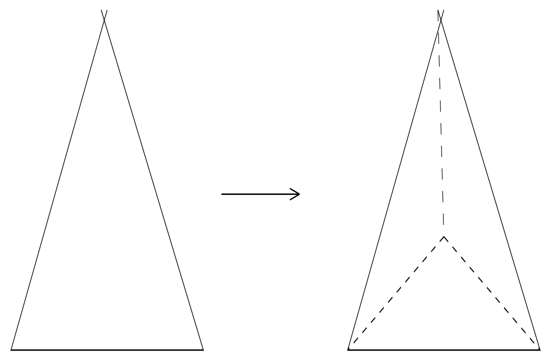


Figure 13:

has the dual representation:

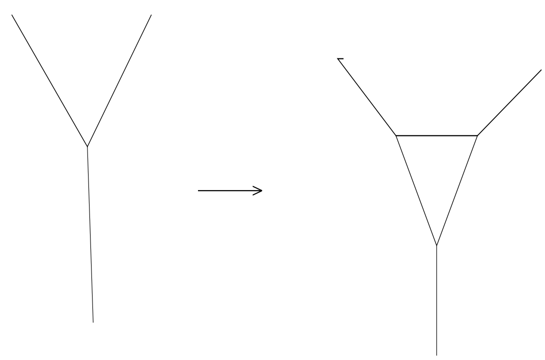


Figure 14:

The advantage of representing the $(k,l)_d$ moves by their dual representation becomes noticeable in higher dimensions. For example, in d=4, the (3,3) move which is quite clumsy to be represented in terms of the original simplices has the dual representation:

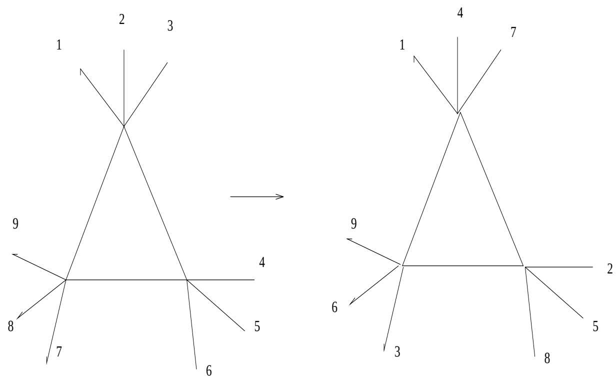


Figure 15:

5 Ergodicity

For the $(k,l)_d$ moves to be useful in the Monte Carlo simulations of Quantum Gravity through the simplicial decompositions, it is necessary that they are ergodic i.e starting with some triangulation it should be possible to reach any arbitrary triangulation through a sequence of $(k,l)_d$ moves.

For this purpose one introduces the ideas of "Alexander equivalence " and "Combinatorial equivalence" of simplicial manifolds. Two simplicial manifolds are said to be Alexander equivalent if one can go from one of them to the other through a sequence of Alexander moves. On the other hand, two simplicial decompositions are said to be combinatorially equivalent if they have a common subdivision. Now two important theorems will be stated without proof:

a) Alexander moves and $(k,l)_d$ moves are equivalent for $d \leq 4$; for $d \geq 5$ the proof requires some additional technical assumptions("$d-2$-spheres are local constructive")

b) two simplicial complexes are Alexander equivalent iff they are combinatorially equivalent.

One of the corollaries of the last statement is that any complex combinatorially equivalent to a simplicial manifold is itself a simplicial manifold.

Combining these two theorems one concludes that in $d \leq 4$ the $(k,l)_d$ moves are ergodic in the space of combinatorially equivalent simplicial manifolds.

5.1 Computational Ergodicity

There are a class of so called "computationally non-recognisable" simplicial manifolds in the sense that there is no algorithm to recognise them. For triangulations of such manifolds it can be shown that if the number of four simplices N_4 of two triangulations T_1 and T_2 are bounded by N, the number of steps in a finite algorithm to get from T_1 to T_2 cannot be bounded by a recursive definable function (examples of recursive definable functions are $N!$,$N!^{N!}$..). This for all practical purposes is a breakdown of ergodicity. But in simulations performed so far this has not caused any problems, but is potentially worrisome.

6 Action

Now that we have introduced a discretisation and a set of moves to go from one discretisation to another, we have to introduce the discrete analog of actions. Some typical actions that are general coordinate invariant are:

$$S = \int \sqrt{g}[\alpha R + \beta R^2 + \partial_\mu R \partial^\mu R + \Lambda + ...] \tag{28}$$

In two dimensions $\int \sqrt{g}R$ is a topological invariant, the Euler characteristic and its only effect is to give a weight factor depending on topology. Also, $\int \sqrt{g}$ is the volume of the manifold, and in the context of DTR is simply the number of triangles. Therefore, for microcanonical simulations(fixed area) neither of the two terms is relevant for fixed topology simulations of random surfaces with no matter couplings. This means every configuration is equally weighted. The other terms in eqn(28) can also serve as invariant observables. To construct the discrete analogs of these terms one uses, for example,

$$R = 2\pi \frac{(6 - q_i)}{q_i} \qquad \sqrt{g} = q_i/3 \tag{29}$$

In more than two dimensions all these terms are important as candidate actions. One can also think of the number of various p-simplices N_p as dynamical variables. However, not all these are independent. For example, in two dimensions $N_1 = 3N_2/2$. In four dimensions, one can choose N_2 and N_4 as independent, and a candidate action for discretised quantum gravity is

$$S = \kappa_1 N_2 + \kappa_2 N_4 \tag{30}$$

6.1 Two dimensional case

We shall give a very detailed account of the simulations in two dimensions as this is a very important testing ground by virtue of the availability of many different formulations as well as of exact analytical results. Before that it is instructive to consider some important aspects of the different analytical approaches available. First, let us consider the formulation according to Polyakov as well as some of the exact analytical results due to Polyakov, Knizhnik and Zamolodchikov(KPZ) on the one hand, and due to David, Distler and Kawai (DDK) on the other. The object of interest is

$$Z = \int \mathcal{D}_g X \mathcal{D} g e^{-\int \partial_\mu X \partial_\nu X \sqrt{g} g^{\mu\nu}} \tag{31}$$

In the above equation μ, ν take values $1....d$, where d is the dimensionality of the Euclidean space in which the two dimensional surface is embedded. d has also the interpretation of the total central charge of the matter coupled to $d = 2$ quantum gravity.

This model was solved exactly by Polyakov in the so called light-cone gauge, and these exact results were subsequently extended by KPZ to essentially derive all correlation functions and hence various critical exponents. It was found that the model was well defined only for $d \leq 1$ or $d \geq 25$.

Of particular interest is the so called "fixed area partition function" $Z(A)$ defined as

$$Z(A) \;=\; \int D_g X \; Dg \; e^{-S_g(X)} \; \delta(\int \sqrt{g} - A) \tag{32}$$

This is essentially the number of configurations with fixed Area A. In DTR, this is studied by keeping the number of triangles fixed. The large area behaviour of $Z(A)$ for $c < 1$ is expected to be

$$Z(A) \;\sim\; e^{\mu A} A^{-b} \tag{33}$$

The "string susceptibility" is defined to be

$$\gamma \;=\; 3 - b\,. \tag{34}$$

According to KPZ and DDK, for pure surfaces ($c = 0$)

$$\gamma \;=\; -2 + \frac{-5\chi}{4} \tag{35}$$

More generally, for the central charge of the matter equal to C,

$$\gamma(c) \;=\; 2 - \frac{25 - c \pm \sqrt{(25-c)(1-c)}}{24}\,\chi \tag{36}$$

where χ is the Eular character.

The analysis of DDK also showed that in the conformal gauge, the partition function (31) is the same as that of the quantum Liouville theory :

$$Z \;=\; \int D\phi e^{-\int(\partial\phi)^2 - a\int e^{v\phi}} \tag{37}$$

where a and b depend on the central charge c. For $c < 1$, all correlation functions of the quantum Liouville theory are known.

7 Matrix Models

As has already been remarked upon, the graph dual to the DTR graph in $d = 2$ has fixed incidence number of 3. Therefore, the dual graphs are Feynman graphs of ϕ^3 theory. The topology (Euler characteristic) of the dual graph is the same as that of the original DTR manifold. The important observation is that the generator of the ϕ^3 graphs with arbitrary topology can be identified with the partition function of the Hermitian $N \times N$ - Matrix Model :

$$Z_M \;=\; \int dM e^{-tr(\frac{1}{2}M^2 + \frac{g}{N}M^3)} \;\equiv\; e^{-F} \tag{38}$$

dM is the Haar measure. The asymptotic behaviour of F for large N is of the form,

$$F \sim N^2 F_0 + F_1 + N^{-2} F_2 + \ldots \tag{39}$$

All F_i are singular at some $g = g_c$ in the sense that some derivatives of F_i w.r.t. g blow up at $g = g_c$. A Taylor expansion of $F(g)$ around $g = 0$ has the form

$$F_i \;=\; \sum_{n=0} \left(\frac{g}{g_c}\right)^n a_n^{(i)} \tag{40}$$

Now the connection between this matrix model *free energy* and the fixed area partition function $Z_\chi(A)$ for manifolds with Euler characteristic χ is

$$Z_\chi(n) \;=\; a^{(i)} \;=\; a_n^{(i)} g_c^{-n} \tag{41}$$

with $\chi = 2 - i$. It is clear that F_i can also be thought of as either the free energy for a *canonical ensemble* of random surfaces with the area A playing the role of energy and $-\log g$ playing the role of $\frac{1}{kT}$, or as a *grand canonical ensemble* with the number of vertices playing the role of member of particles, and g playing the role of fugacity.

The mapping of the DTR problem onto that of the matrix model can be done for central charges $c \leq 1$. Let us illustrate this with the example of the Ising model ($c = 1/2$).

Let us consider the case where at each dual graph vertex we attach an Ising variable σ taking the values ± 1. The Hamiltonian for the Ising system can be taken to be

$$H = -J\Sigma_{<ij>}\sigma_i\sigma_j \tag{42}$$

where $< ij >$ represents the sum over all the edges of the dual-graph. It should be kept in mind that as the surfaces are updated the set of edges of the dual graph also changes dynamically. As shown by Boulatov and Kazakov, this model even in the presence of an external magnetic field is mapped onto

$$Z = \int dM_1 dM_2 e^{-tr(\frac{M_1^2}{2} + \frac{M_2^2}{2} - cM_1M_2 + g_1M_1^4 + g_2M_2^4)} \tag{43}$$

The temperature and the magnetic field of the Ising system are given respectively by $c = e^{-2J}$ and $g_1/g_2 = e^{-2H}$. Both the matrix models represented above are exactly solvable for F_i.

Before we give the results of DTR for various observables of interest it is instructive to recall a few features of two dimensional geometry: any metric can locally be brought to the form $\lambda(\xi)\cdot$ flat metric, with the help of general coordinate transformations. Naively, one would expect that if the classical action is also Weyl invariant, i.e invariant under the transformation $g \to \lambda(x) \cdot g$, the metric can everywhere be brought to the flat form. However, this is not true as it is impossible to find measures $\mathcal{D}_g X$ and $\mathcal{D}g$ which are invariant under both general coordinate transformations as well as Weyl transformations. Also, on purely geometric grounds it can be shown that there are "conformal classes" of metrics so that a metric belonging to one class can not be transformed into a metric of another class by a globally well-defined Weyl transformation. When the genus g is greater than two, there are $2g-2$ complex parameters describing the conformal classes. For genus one (Euler characteristic zero) case, the conformal class is parametrised by a single complex parameter called moduli, and the partition function takes the form

$$Z = \int d\tau f(\tau) \tag{44}$$

The function $f(\tau)$ is calculable.

8 Observables

There are several observables of interest that can be studied under DTR. Let us start with the **String Susceptibility** γ. It has already been defined in eqn(34). Both the continuum theory as well as the matrix model predict that :$\gamma = -1/2$ for the pure surface theory (c = 0), and that $\gamma = -1/3$ for the Ising model coupled to random surfaces at the critical point of the Ising system (these results are for the zero genus case). As the Ising model coupling is varied what one finds in DTR are: for $J > J_c$ and $J < J_c$, γ stays at -0.5 and at $J = J_c$ it reaches $-1/3$. Of course there is a well defined cross-over region.

It is interesting that the Ising model coupled to random surfaces is exactly solvable in the presence of an external magnetic field also, while the analogous problem in flat space remains unsolved.

In addition to the string susceptibility, one can study various critical exponents $\alpha, \beta, \delta, \nu$ for the Ising system. One can also study the magnetic susceptibility and exponents associated with it(the precise manner the magnetic susceptibility diverges with the system size at the critical point, for example). In flat space it is well known that there are scaling relations among these various exponents that are dimension dependent. It is also known that the dependence on dimensionality appears only through the combination νd. The exact results based both on the continuum theory and the Matrix models predict the exponents $\beta = 1/2$,$\gamma = 2$ and $\delta = 5$ while the Onsager values are $\beta = 1/8$,$\gamma = 7/4$ and $\delta = 15$. DTR has indeed confirmed these predictions and the validity of the scaling relations. While for the flat-space case $\nu d = 2$ with $\nu = 1$, for random surface case $\nu d = 3$. However,ν is not known for the random surface case. It appears as if the relevant dimensionality is the "Hausdorff dimensionality".

It is puzzling that simulations based on Regge calculus do not seem to be reproducing the exact results of the continuum and Matrix models.

8.1 Hausdorff Dimension

This is another quantity of interest and perhaps one of the most important lessons one has learnt from numerical simulations is that quantum fluctuations can drastically alter the scaling dimension of space-time. In flat two dimensional space through the relation $area = \pi \cdot R^2$ for a circle, one identifies the scaling dimension to be 2. For the case of fluctuating geometry one expects $A = \pi R^{d_H}$ where d_H is the Hausdorff dimension. The way to measure d_H in DTR is best illustrated with the dual graph: One starts at some site of the dual graph which is marked 'zero' to denote the 'origin'. Then, one marks all the neighbouring sites of the origin as '1', and all the distinct neighbours of the 1-sites as '2' and so on. The marking on the sites is analogous to 'radius' and the total number of sites with markings $\leq R$ is taken as a measure of the area. One has to make sure that it is a connected domain. The typical results for the measurement of fractal dimensions are shown in the next figure.

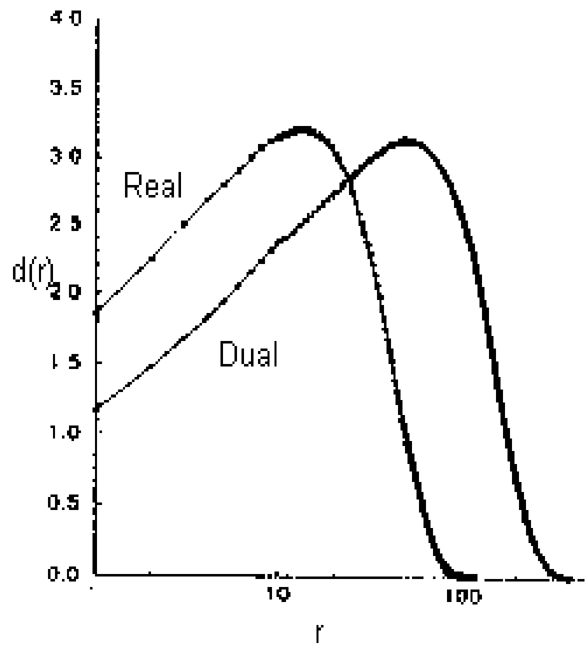


Figure 16: The numerical results of fractal dimensions measured in the real and dual spaces consisting 400,000 triangles. r represents an averaged geodesic distance.

The fractal dimension starts off at some low value, reaches a peak and starts falling off. This fall off to zero is a purely finite size effect as eventually all the triangles(dual sites) get visited. It is expected that as the system size increases, the measured fractal dimension approaches d_H. From the present simulations it appears that $d_H = 4$ for all $c \leq 1$.

8.2 Spectral Dimension

Another intrinsically geometric quantity of interest is the so called "spectral dimension". This is defined through the return time for a random walk. One essentially computes the kernel $K(\vec{X}, T = 0; \vec{X}, T = T)$ and study it in the limit $T \to 0$

$$< TrK(T) > \simeq T^{-d_s/2} \quad as \quad T \to 0 \tag{45}$$

Simulations indicate that $d_s = 2$.

8.3 Loop Length Distributions

Recall the introduction of the 'radial 'distance in connection with the definition of the Hausdorff dimension. What one finds typically at a given radial distance is not just one connected domain but many, as shown in the figure below. Kawai and coworkers have shown for the pure surface

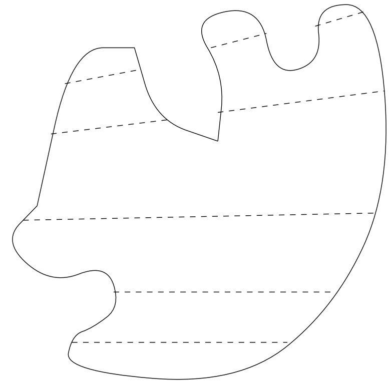


Figure 17:

case that the number of loops $N(D)$ of perimeter $L(D)$ at radial distance D have a universal distribution

$$N(D)D^2 = f(L(D)/D^2 \tag{46}$$

they were also able to compute the function f. Next figure shows the agreement between theory and simulations. Though the Regge calculus simulations have not been successful in reproducing the

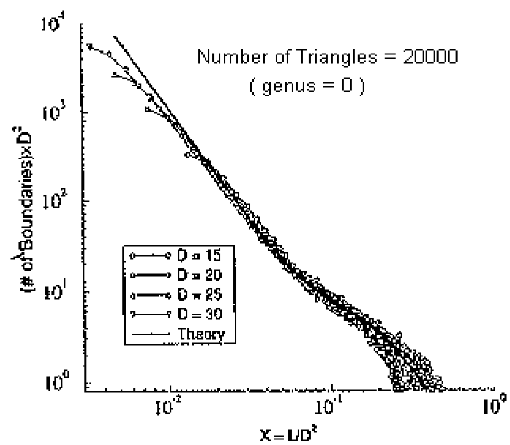


Figure 18: Loop-length distributions with Double-Log scales. The total number of triangles is 20,000 and is averaged over 500 configurations. $X = \ell(D)/D^2$ is a scaling variable, where D represents the geodesic distance (it was measured at steps D = 15,20,25 and 30) and $L(D)$ represents each loop length at step D. The small circles, triangles and quadrangles indicate the results of the numerical simulations, and the solid line indicates that of string field theory.

critical exponents, they have been reasonably successful in reproducing the loop length distributions for small loops. Thus Regge calculus is perhaps not all together on the wrong track. There is also perhaps a connection between the failure of the Regge calculus in reproducing the critical exponents and its inability to capture the correct loop length distribution for large loops.

8.4 Baby Universes

A typical configuration produced in DTR simulations looks nothing like a smooth surface; in fact, what one finds often are various surfaces connected to each other through "necks" as illustrated in the next figure. The minimum neck size in DTR is of course a triangle and it is in fact just

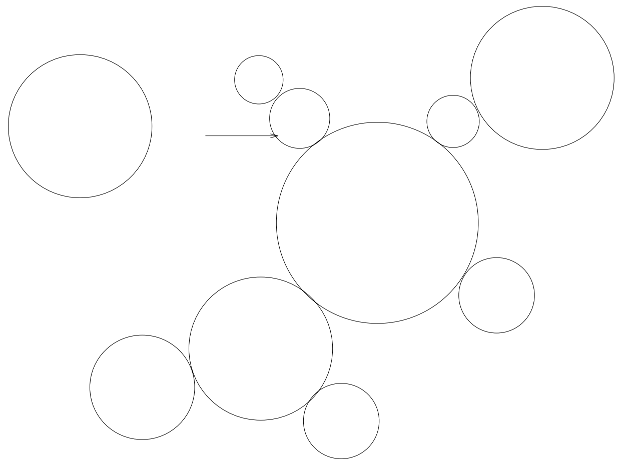


Figure 19:

a matter of counting, as shown by Jain and Mathur, to estimate the number of "minimum neck baby universes"(MINBU) once the fixed area partition sum $Z(A)$ is known

$$n(B, A) = \frac{3(A - B + 1)\ Z(A - B + 1)\ (B + 1)\ Z(B + 1)}{Z(A)} \tag{47}$$

where $n(B, A)$ is the probability of finding a MINBU of area B in a surface of total area A. This follows from the fact that a triangle can be located on a surface of area $(B + 1)$ (1 represents the neck) in Z(B+1) ways and this can be glued at any of the $(A + B - 1)$ locations of the surface with area $(A - B + 1)$ along with 3 ways of gluing a triangle onto a triangle.

It is clear that counting the MINBU distribution is a very practical way of measuring the string susceptibility γ. However, one has to be careful about the influence of finite size corrections. From the definition of γ (eqns. 33, 34) it is clear that when $\gamma \leq 0$, the average MINBU size goes to a constant as the area of the configuration becomes larger and larger, whereas for $\gamma \geq 0$ the average MINBU size grows with A.

As was mentioned earlier update algorithms should be able to effectively overcome critical slowing down. In simulations of spin systems on fixed geometry this is achieved through cluster algorithms and the like. In random surface simulations this problem of large autocorrelation lengths is quite serious. Ambjorn and coworkers have suggested "MINBU surgery" as a way of overcoming this. This amounts to cutting away the MINBU's and stitching them onto the surface in arbitrary ways.

Among other observables of interest one could think of various powers of the incidence number q_i. Of course it follows rather trivially that the average coordination number is 6, in other words the average curvature (scalar) is zero.

8.5 Resistivity

Kawai and coworkers have proposed the use of resistivity measurements as a probe of both the smoothness of a surface as well as for determining its complex structure. For the sphere topology for which the complex structure is trivial, consider the arrangement shown in the figure. Each

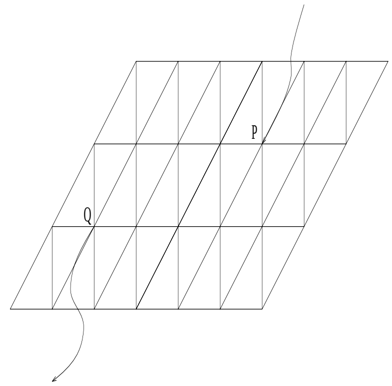


Figure 20: A resistive network

edge of the dual graph is treated as a resistor of some given resistance say 1 ohm. At the point P a current of 1 Amp flows into the surface and at the point Q a current of 1 Amp flows out. The action is modified through the addition of

$$S' = \int \sqrt{g} g^{\mu\nu} \partial_\mu V \partial_\nu V \tag{48}$$

V can be thought of as an external (non dynamical) scalar field. Now one solves the discrete version of Poisson's equation for the values of V at the sites of the dual graph. One compares this with what is expected of a continuous surface

$$V(z) = const \log \frac{(z - z_P)}{(z - z_Q)} \tag{49}$$

the constant being a measure of the resistivity. This enables one to attach complex coordinates z also. Since $\sqrt{g} g^{\mu\nu}$ is independent of the conformal factor, this method is sensitive only to the complex structure.

If the surface is a smooth surface one expects a peaking in the observed resistivity whereas for a structure like a branched polymer(to which the surface is conjectured to degenerate into when $c > 1$) the distribution in resistivity is expected to be broad. In the next figure some typical DTR data for resistivity is shown: (The vertical axis is the number of configurations with average resistivity, r.)

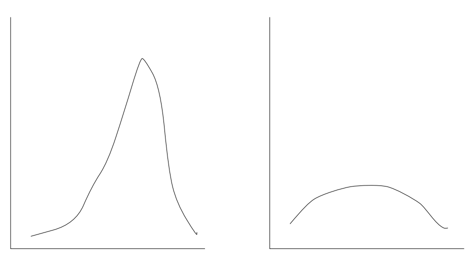


Figure 21:

8.6 Complex Structure

Kawai has also proposed a way of determining complex structure through resistivity measurements. First let us briefly discuss what complex structure means in the case of genus 1 topology(torii).

In the continuum picture the torus is equipped with a 'homology basis' of a-cycles and b-cycles as shown in the figure;

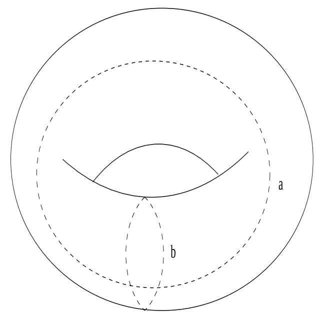


Figure 22:

Now there exists a so called Abelian differential satisfying

$$\partial_{\bar{z}}\omega = 0 \tag{50}$$

where ω is a one form. The complex moduli τ is defined as the ratio

$$\tau = \frac{\oint_b \omega}{\oint_a \omega} \tag{51}$$

The proposal for determining τ through DTR is as follows: first locate the a and b cycles on the configuration. This is done by starting at $t = 0$ with a triangle and evolving it in 'radial time' as described in the context of the Hausdorff dimension. As D increases, this elementary loop will split into other loops or stays as a single loop. If the topology is that of a sphere, all the loops will eventually shrink to their minimum size(triangle), and there will never be an instance of two loops merging into a single loop as shown in figure 23.

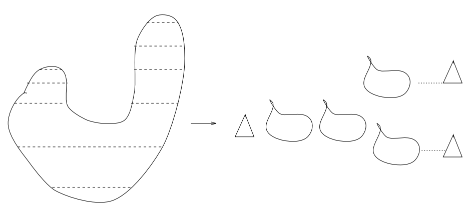


Figure 23:

On the other hand, in the case of torus topology, two loops will merge. This is shown in the figure 24:

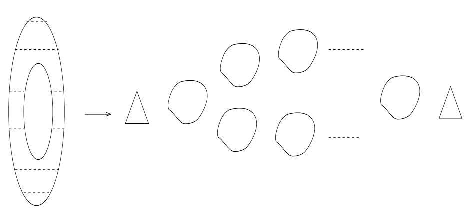


Figure 24:

The point where two loops merge will be a point on the, say, a-cycle. One of the two loops merging there can then be identified with the other(b) cycle. The entire a-cycle can now be reconstructed by following the history of the two loops that merged.

After locating the a, b-cycles, now apply a potential difference across the a-cycle and solve for the voltage distribution. On identifying

$$j_\mu = \partial_\mu V \tag{52}$$

then

$$\oint_a j = \Delta V \tag{53}$$

Now the dual of the one form j is given by

$$\tilde{j}_\mu = \epsilon_{\mu\nu} j^\nu \tag{54}$$

The holomorphic one-form $j + \tilde{j}$ is seen to be an Abelian differential(care should be taken to include some resistivity dependent factors). Now the moduli τ can be measured and the function $f(\tau)$ determined. This way, Kawai, Tsuda and Yukawa have determined the bosonic string partition function for the $c = 0$ case and it is reproducing the expected features.

8.7 Other Results

There are too many interesting results to report on here for lack of time, so I will just include a few examples. For example, Tsuda and Yukawa have studied the phase diagram for two dimensional gravity in the coupling constant space α, β where

$$S(g) = \beta \int d^2x \sqrt{g} R^2 + \alpha \int d^2x \sqrt{g} g^{\mu\nu} \partial_\mu R \partial_\nu R \tag{55}$$

The phase diagram obtained is shown in the next figure:

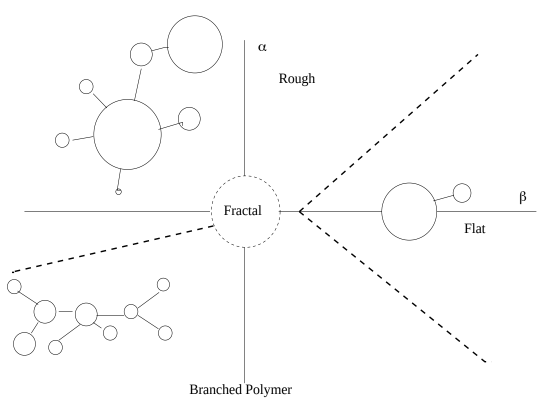


Figure 25: A phase diagram for 2d gravity

There are also some preliminary results that have been obtained by Bakker and Smit for four dimensional gravity.There is also some recent interesting work on the meaning of fixed geodesic distance in $d = 2$ quantum gravity by H. Aoki et al.

References

[1] *General Relativity Without Coordinates*, T. Regge, Il Nuovo Cimento, **XIX, No 3** (1964);

[2] *Dynamical Triangulation - A gateway to QG ?*, J. Ambjorn, hep-th/9503108;

[3] *Recent Progress in the theory of Random Surfaces and Simplicial QG*, J. Ambjorn, Lattice '94;

[4] *Ising Transition in 2-d Simplicial QG, can Regge calculus be right ?*, C. Holm and W. Janke, Lattice '94;

[5] *Regge skeletons and quantisation of 2-d gravity*, W. Bock ,Lattice '94;

[6] *Fractal Structure in 2-d Quantum Regge Calculus*, Jun Nishimura and Masaki Oshikawa, Phys. Lett. **B 338**, 187-196, (1994);

[7] *On the fractal structure of two-dimensional QG*, J. Ambjorn et al, hep-lat/9507014;

[8] *Simplicial QG and Random Lattice*, F. David, Saclay Preprint T93/028;

[9] *Simplicial Quantum Gravity*, B. V. de Bakker, Thesis submitted to University of Amsterdam, hep-lat/9508006;

[10] *Numerical Analysis of 2-d QG*, N. Tsuda, Thesis submitted to Tokyo Institute of Technology, (1994);

[11] S. Jain and S. Mathur, *Phys. Lett.* **B286**, 239,(1992);

[12] For discussion of Ergodicity see:
A. Nobutovsky and R. Ben - Av, *Comm. Math. Phys.* **157**, 93-97, (1993);
Gross M, Varsted S, *Nucl. Phys.*, **B378**, 367, (1992);

[13] H. Kawai, N. Kawamoto, T. Mogami and Y. Watabiki, *Phys. Lett.* **B306**, 19, (1993);

[14] *Complex Structure of 2-d surfaces*, H. Kawai,N. Tsuda and T. Yukawa in Lattice' 95, hep-lat/9512014;

[15] *Operator Product Expansion in 2-d QG*, H. Aoki et al, hep-th/9511117;

[16] D. Boulatov and V.A. Kazakov, *Phys. Lett.* bf B186, 379, (1987);

[17] N. D. Hari Dass, B.E. Hanton and T. Yukawa, *Phys. Lett.* **B368**, 55, (1996). 379, (1987).

List of Contributed Papers

1. Static and non static global string
   A Banerjee, N Banerjee, A Sen
2. Black hole complementarity and fermions
   Arundhati Dasgupta
3. Comments on Yilmaz's theory of gravitation
   C.S.Unnikrishnan
4. Cosmic censorship in Tolman Bondi dust collapse
   C.S.Unnikrishnan
5. Energy in a gravitational field : Can it make local sense?
   C.S.Unnikrishnan
6. Constant Decleration parameter in higher dimension
   D Panigrahi
7. The trajectory of particles around cosmic strings
   G Alagar Ramanujam, K Anandan, G Vasanthakumari
8. The trajectory of particles in the gravitational field of a tachyon
   G Alagar Ramanujam, K Meenakshi, H Sridharan
9. Certain results in the λ varying cosmology
   G Alagar Ramanujam, M Pankajavalli, S Varadharajan
10. String models in some nontrivial backgrounds
    G V Vijayagovindan
11. The orbits of charged particle in an electromagnetic fields on
    Kerr background geometry
    K N Mishra, K Chakraborty
12. Singularities, hamiltonians and infinite dimensional Lie algebras
    in general relativity
    K.H.Mariwalla
13. Metric in axially symmetric radiation zone
    M.D.Patel and R.M.Patel
14. A Modified Ozer-Taha Type cosmological model
    Moncy V John, K Babu Joseph
15. Einstein pseudotensor and total energy of the universe
    N Banerjee, Somasri Sen
16. Viscous fluid universe interacting with electromagnetic and zero-mass scalar fields
    N Ibotombi Singh
17. On the relationship between Killing-Yano tensors and electromagnetic
    fields on curved spaces
    Ng. Ibohal

18. Squeezed state representation of black hole radiation
P.K.Suresh and V.C.Kuriakose

19. Gauss map and 2+1 gravity
R.Parthasarthy

20. A cylindrically symmetric stiff fluid solution of Einstein's equations and gravitational collapse
Ramesh Tikekar, M C Sabu

21. How far singularity theorms imply the big-bang singularity?
S S Sharma

22. Distortion of GW signals from binary systems due to presense of accretion disks
Sandip Chakrabarty

23. Bianchi type-1 Vacuum cosmological model in scale invariant theory of gravitation
Sk. Md. Daud, G. Mohanty

24. Hermitian Wheeler-Dewitt Operators and the wave function of the universe
Subenoy Chakraborty Nabajit Chakravarty

25. Spherically symmetric non-static space-time and monopoles
Subenoy Chakraborty, Lalit Biswas

26. Quantum temporal logic, dynamic evolution, and symmetries in the histories approach to quantum theory
Tulsi Dass, Yogesh Joglekar

27. Can there be a theory of everything in physics?
Uma S Sharma

List of Participants

Name	Affiliation
Alagar Ramanujam G	Principal & Head, PG Dept. of Phys., NGM College, Pollachi 642 001.
Balasubramanian R.	IUCAA, Post Bag 4, Ganesh Khind, Pune, 411 007
Banerjee Narayan	Dept. of Phys., Jadavpur Univ., Calcutta 700 032.
Basu Madhumita B.	C/O. DR. S.C. BOSE, 1 Nirmal Ch. St., Calcutta, 700 012.
Bharadwaj Somnath	Raman Research Inst., C V Raman Avenue, Sadashiva Nagar, Bangalore 560 080.
Bhattacharya Pijush	Indian Inst. of Astrophysics, Koramangala, Bangalore, 560 034.
Biswas Lalit	Regional Met. Centre, Met. Dept., Weather Section, 4 Duel Ave., Alipore, Calcutta, 700 027.
Chakrabarti S.	Theoretical Astrophysics Group, T.I.F.R., Homi Bhabha Rd, Mumbai 400 005.
Chakraborty Subenoy	Dept. of Math., Jadavpur Univ., Calcutta, 700 032.
Chakravarty Nabajit	Dept of Math., Jadavpur Univ., Calcutta, 700 032.
Dadhich Naresh	IUCAA, Post Bag 4, Ganesh Khind, Pune, 411 007
Dalitz R.H.	Dept. of Theo. Phys., Oxford Univ., Oxford, UK
Das Saurya	The Inst. of Mathematical Sciences, Madras, 600 113
Dasgupta Arundhati	The Inst. of Mathematical Sciences, Madras, 600 113
Date G.	The Inst. of Mathematical Sciences, Madras, 600 113
Daud Mahammad	Hariharpur P.O.,Brahmansasan, via Balichak-Midnapur, 721 124.
Gopalkrishna A.V.	Dept. of Math., I.I.Sc., Bangalore, 560 012
Gupta Varsha	Dept. of Phys. & Astrophysics, Delhi Univ., New Delhi, 110 007.
Hari Dass N. D.	The Inst. of Mathematical Sciences, Madras, 600 113

List of Participants (Cont...)

Name	Affiliation
Ibohal Nagangbam	Dept. of Math., Univ., of Manipur, Imphal, 795 003.
Iyer B. R.	Raman Research Inst., C V Raman Avenue, Sadashiva Nagar, Bangalore 560 080.
Joglekar Yogesh N.	B 206, Hall-1, Indian Inst. of Technology, Kanpur, 208 016.
Jotania Kanti R.	Raman Research Inst., C.V. Raman Avenue, Sadashiva Nagar P.O., Bangalore, 560 080
Kar Sayan	Inst. of Physics, Sachivalaya Marg, Bhubaneshwar, 751 005
Kuriakose V. C.	Dept. of Phys., Cochin Univ. of Science and Technology, Kochi, 682 022.
Maharana J.	Inst. of Physics, Sachivalaya Marg, Bhubaneshwar, 751 005
Majumdar P.	The Inst. of Mathematical Sciences, Madras, 600 113
Mariwala K.	The Inst. of Mathematical Sciences, Madras, 600 113
Mishra Kameshwar N.	Dept. of Math.. Bhilai Inst. of Technology, B.I.T., Durg, M.P.
Mitra P.	Saha Inst. of Nuclear Physics, Block AF, Bidhannagar, Calcutta, 700 064.
Mohanty S.	Phys. Res. Laboratory, Navarangapura, Ahmedabad, 380 009.
Moncy V. John	Vilavinal Kozhencherri East P.O. Pathanamthitta, Kerala, 689 641.
Munshi Deepak	IUCAA, Post Bag 4, Ganesh Khind, Pune, 411 007
Panchapakesan N.	Dept. of Phys. Delhi Univ., New Delhi, 110 007.
Pankajavalli M.	PG Dept of Phys., NGM College, Pollachi, 642 001.
Parthasarathy R.	The Inst. of Mathematical Sciences, Madras, 600 113
Patel M. D.	Dept. of Math., Sardar Patel Univ., Vallabh Vidyanagar, Gujarat, 388 120.
Prasanna A. R.	Phys. Res. Laboratory, Navarangapura, Ahmedabad, 380 009.

List of Participants (Cont...)

Name	Affiliation
Raghunathan K.	Dept. of Theo. Phys., Univ. of Madras, Guindy Campus, Madras, 600 025.
Rajasekaran G.	The Inst. of Mathematical Sciences, Madras, 600 113
Rajesh Nayak K.	Indian Inst. of Astrophysics, Koramangala, Bangalore, 560 034.
Ramachandran R.	The Inst. of Mathematical Sciences, Madras, 600 113
Sabu M. C.	Dept. of Math., Sardar Patel Univ., Vallabh Vidyanagar, Gujarat, 388 120.
Sen Somasri	Rel. & Cosmo. Research Centre, Dept. of Phys., Jadavpur Univ., Calcutta, 700 032.
Sen Anjan Ananda	Rel. & Cosmo. Research Centre Dept. of Phys., Jadavpur Univ., Calcutta, 700 032.
Sharma Shilendra S.	118/A, Radhaswami Road Houses, Bh/Chankyapuri, Ghatalodiya, Ahmedabad, 380 061.
Sharma Uma S.	32 Devikrupa Soc., C.T.M. Four Ways Ramol Rd., Ahmedabad, 380 026.
Singh T.P.	Theoretical Astrophysics Group, T.I.F.R., Homi Bhabha Rd, Mumbai 400 005.
Srivatsava D.C.	Dept. of Phys., Gorakhpur Univ., Gorakhpur, 273 009.
Suresh P. K.	Dept. of Phys., Cochin Univ. of Science and Technology, Cochin, 682 022.
Unnikrishnan C. S.	Gravitational Expts. Group, T.I.F.R., Homi Bhabha Rd, Mumbai 400 005.
Vaidya P.C.	Dept. of Math., Gujarat Univ., Ahmedabad, 380 009.
Varadharajan S.	Dept.of Phys., Coll. of Agri. Engg., Pallavaram PO, Lalgudi T.K., Trichy, 621 712.
Vijayagovindan G. V.	School of Pure and Applied Phys., Mahatma Gandhi Univ., Kottayam, 686 560.
Vishveshwara C.V.	Indian Inst. of Astrophysics, Koramangala, Bangalore 560 034.